

This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Measuring Capital in the New Economy

Volume Author/Editor: Carol Corrado, John Haltiwanger and Dan Sichel, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-11612-3

Volume URL: <http://www.nber.org/books/corr05-1>

Conference Date: April 26-27, 2002

Publication Date: August 2005

Title: The Relation among Human Capital, Productivity, and Market Value: Building Up from Micro Evidence

Author: John M. Abowd, John Haltiwanger, Ron Jarmin, Julia Lane, Paul Lengermann, Kristin McCue, Kevin McKinney, Kristin Sandusky

URL: <http://www.nber.org/chapters/c10621>

Chapter pages in book: (153 - 204)

---

# The Relation among Human Capital, Productivity, and Market Value

## Building Up from Micro Evidence

John M. Abowd, John Haltiwanger, Ron Jarmin,  
Julia Lane, Paul Lengermann, Kristin McCue,  
Kevin McKinney, and Kristin Sandusky

---

### 5.1 Introduction

The measurement of intangibles and human capital, important both for goods-producing and for service-producing industries, has always been a difficult challenge for the statistical system. The growth of the “new econ-

John M. Abowd is the Edmund Ezra Day Professor of Industrial and Labor Relations at Cornell University, director of the Cornell Institute for Social and Economic Research (CISER), distinguished senior research fellow at the Longitudinal Employer Household Dynamics (LEHD) program at the U.S. Census Bureau, research associate at the National Bureau of Economic Research (NBER), research affiliate at the Centre de Recherche en Économie et Statistique (CREST), research fellow at IZA (Institute for Labor Economics), and a member of the executive committee of the Conference on Research in Income and Wealth (CRIW). John Haltiwanger is professor of economics at the University of Maryland and a research associate of the NBER, a senior fellow at LEHD, and is on the executive committee of the CRIW. Ron Jarmin is Assistant Div. Chief for Research at the Center for Economic Studies (CES), U.S. Census Bureau, and is on the executive committee of CRIW. Julia Lane is program director at the National Science Foundation and senior research fellow of the LEHD program at the U.S. Census Bureau. Paul Lengermann is an economist in the Industrial Output Section of the Division of Research and Statistics at the Federal Reserve Board. Kristin McCue is an economist with CES, U.S. Census Bureau. Kevin McKinney is an economist in the LEHD program at the U.S. Census Bureau. Kristin Sandusky is an economist in the LEHD program at the U.S. Census Bureau.

The authors wish to acknowledge the substantial contributions of the staff in the Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau. We would also like to thank the participants at a pre-conference for helpful comments. This research is a part of the U.S. Census Bureau's LEHD Program, which is partially supported by National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging, the U.S. Department of Labor (ETA), and the Alfred P. Sloan Foundation. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Census Bureau or the National Science Foundation. Confidential data from the LEHD Program were used in this paper. The U.S. Census Bureau is preparing to support external researchers' use of these data under a protocol to be released in the near future.

omy” has made responding to this challenge even more urgent: the need to understand how such inputs affect the value chain of productivity, growth, and firm value now surpasses the need to measure the contribution of bricks, mortar, and equipment. Yet the changes that have brought the new economy into existence have also highlighted the need for improvements to traditional measures of inputs and outputs (Haltiwanger and Jarmin 2000), especially for human capital. Finding new measures of human capital that are both quantifiable and available for a sample large enough for use in official economic statistics is a formidable challenge.

This paper uses newly available micro-level data from the U.S. Census Bureau on both employers and employees to demonstrate a new approach to addressing this challenge. We use new measures of human capital that directly capture the market valuation of the portable component of skill, including the contribution of “observable” and “unobservable” dimensions of skill. In principle, the measures go beyond indirect proxies (such as measures of years of formal education) to quantify the value of individual specific skills, such as innate ability, visual or spatial skills, nonalgorithmic reasoning, analytic or abstract decision making, and “people skills” (Bresnahan, Brynjolfsson, and Hitt 2002). We present exploratory empirical results that relate the new human capital measures to measures of labor productivity and market value.

An additional challenge has been to document the sources of firm-level heterogeneity in productivity, growth, and value. One of the key findings of the literature using micro-level data is that large differences exist across many dimensions of firm inputs and outcomes. In particular, employers exhibit little uniformity in either the methods they use to hire and fire workers or in the selection of types of workers they employ. We therefore use measures of the dispersion of the firm-level human capital distribution to capture relevant aspects of firm-level differences in organizational capital and workplace practices.

We begin by describing the background, motivation, and underlying specifications used in this chapter. Next, we describe the newly created data sources and measures that underlie our study. The subsequent section provides an exposition of the measurement of human capital that is made possible by the new census data. We then present exploratory empirical results that relate our new human capital measures to measures of firm performance including labor productivity and market value. The final section concludes the paper.

## 5.2 Background, Motivation, and Specifications

The wide-ranging literature on human capital and intangibles is impossible to summarize here. We can, however, provide a brief background to provide some perspective on our approach. We begin with a discussion of

our methodology for measuring human capital and then consider the role of human capital at the firm level by focusing on its potential relation to productivity, market value, and tangible and intangible assets.

### 5.2.1 Human Capital: Conceptual and Measurement Issues

The importance of human capital in accounting for observed differences in wages and productivity has a very long history in economics. Becker (1964) and many others helped the profession define the components of human capital, and the contribution of human capital to productivity has been intensively and exhaustively studied (e.g., Jorgenson, Gollup, and Fraumeni 1987—hereafter JGF). Clearly, we stand on the shoulders of these researchers, but our approach differs in key ways that depend critically on the availability of data. In particular, our conceptual and measurement approach depends not only on the availability of longitudinal, matched employer-employee data but also on the availability of universe files of all workers and firms.

The starting point for our approach has been well documented and investigated in papers by Abowd, Kramarz, and Margolis (1999—hereafter AKM) and Abowd, Lengermann, and McKinney (2003—hereafter ALM). In this paper, we exploit newly developed measures of human capital that have emerged from this work and are part of a new program at the Census Bureau called the Longitudinal Employer-Household Dynamics (LEHD) program. These and related papers emphasize a point that has long been known in the study of human capital—direct measurements of human capital are very difficult. The standard approach is to take advantage of the “usual suspects”—for example, education and experience—and to build proxies for human capital using such measures. In the productivity literature referenced above, this approach has made extensive use of household data. Using primarily the Current Population Survey (CPS), JGF create detailed measures of human capital from person-level data in the United States by exploiting wage differences across groups defined by gender, experience, and education. They aggregate these measures by industry and at the total economy level. Their work demonstrates the existence of an enormous stock of human capital in the U.S. economy and shows that the stock and flows of this asset are vitally important for understanding changes in labor productivity.

However, JGF (and subsequent related work, including the Jorgenson, Ho, and Stiroh chapter in this volume) recognize that this approach has limitations. Clearly, industry- and economywide aggregates fail to capture the firm-level variation that is a driving force in productivity growth. In addition, the existing data provide only a relatively small set of observable characteristics of workers, thereby creating measurement problems, and omit measures of unobservable skill and confounding firm effects. Using a college degree as a measure of human capital, for example, fails to capture

differences in school quality and program of study (Aaronson and Sullivan 2001). The large portion of wage variation that cannot be explained by these variables highlights the important role of the unobserved component of skill in the measurement of human capital, as has been emphasized in the recent literature on rising wage inequality in the United States (see, e.g., Juhn, Murphy, and Pierce 1993). Finally, earnings measures include the returns to working with particular firms—for example, large, highly unionized, or profitable entities—and there is sorting among workers and firms. For these reasons, AKM note that the estimates of returns to measured and unobservable components of human capital may be biased.

We will describe in detail the econometric and measurement approach in subsequent sections, but we will review here the basic specification used by AKM and ALM so that we can discuss the conceptual basis of our human capital measures. The core statistical model is

$$(1) \quad w_{ijt} = \theta_i + x_{it}\beta + \psi_{J(i,t)} + \varepsilon_{ijt}$$

The dependent variable is the log wage rate of individual  $i$  working for employer  $j$  at time  $t$ , while the function  $J(i, t)$  indicates the employer of  $i$  at date  $t$ . The first component is a time-invariant person effect, the second the contribution of time-varying observable individual characteristics, the third the firm effect, and the fourth the statistical residual, orthogonal to all other effects in the model. In what follows, we use the person effect,  $\theta$ , plus the experience component of  $x\beta$  as the core measure of human capital, called “ $h$ ” (i.e.,  $h_{it} = \theta_i + x_{it}\beta$ ).<sup>1</sup> We also exploit these components separately, as they clearly represent different dimensions of human capital or skill.

For current purposes, our approach has three conceptual and measurement advantages over earlier approaches. First, because we have data on the universe of workers and of firms, we can create both firm-based and industry-based measures of human capital that include measures of dispersion as well as central tendencies. In particular, the new data permit the measurement of  $h$  and its underlying components for all workers. Further, because we can associate all of these workers with their employers, we can consider the full distribution of human capital for each firm and industry. Second, the measure of  $h$  includes a broader measure of skill—the market valuation of a number of observable and unobservable components—and, as such, encompasses various measures of skill, including education.<sup>2</sup> Be-

1. See ALM for details of the estimation procedure. Additional controls in  $x$  include year effects interacted with gender effects and full-quarter employment adjustments (not all workers work full quarters).

2. In most of the analysis that follows, we do not separate the effect of observable characteristics, such as education, from those of unobservable components. For subsamples of our universe files, we can measure education, and some of the results (see, e.g., ALM) to which we refer are based upon such analysis. There is a large ongoing effort at the LEHD program to incorporate such observable characteristics on a more comprehensive basis including the development of robust imputation procedures for our universe files.

cause it includes the person effect, which can be thought of as the portable time-invariant component of a person's wage, the measure of  $h$  also captures the influence of unobservable components of skill. Third, because the AKM approach controls for firm effects in estimating the person effects, our measure of human capital does not reflect personnel policies of the firm that may alter the returns to observable and unobservable dimensions of skill.

Our approach has some limitations. First, the estimation method provides person effects and firm effects that are time invariant. Both theory and evidence suggests that firm effects and the returns to different dimensions of skill may be time varying. Permitting time variation in the person and firm effects (e.g., through estimating a mixed-effects or Bayesian model) is an active area of research in the LEHD program. For now, our interpretation is that we have a time average of the relevant effects.

Second, the specification does not permit any interaction between the firm effects and the person effects. The implicit assumption in this specification is that firms pay the same premium (or discount) regardless of the type of worker. This, too, is an area for future work, but in this case we have some evidence to support our assumption. For example, Groshen (1991a,b) finds that establishment wage differentials exist across occupations within establishments. Groshen's result has been updated and confirmed by Lane, Salmon, and Spletzer (2002), who use recent data from the Occupational Employment Statistics survey of the Bureau of Labor Statistics (BLS); they find that firms that pay premiums to their accountants also pay premiums to their janitors.

Third, this specification does not permit any coworker effects. Yet Lengeremann (2002b) has shown that whom you work with matters, as well as who you are and for whom you work.

Despite these concerns, we regard this base specification as a significant advance over the standard approach of measuring human capital via observable education and experience, for the reasons discussed above. We explore many of these issues in what follows. For example, in section 5.3, we report the characteristics of our human capital measures and, along the way, compare our results to more traditional estimates of human capital. For now, we proceed to a discussion of how and why human capital might matter for productivity and market value.

### 5.2.2 Human Capital, Tangible Assets, Intangible Assets, and Productivity

The relation between output and inputs is summarized by the standard production function approach. Explicit recognition of human capital and intangibles augments this function in the following specification of an intensive production function:

$$(2) \quad y_{jt} = F_j(K_{jt}^T, K_{jt}^I, H_{jt}),$$

where  $y_{jt}$  is output per worker for firm  $j$  at time  $t$ ,  $K_{jt}^T$  is tangible physical capital per worker,  $K_{jt}^I$  is intangible capital per worker, and  $H_{jt}$  represents measures of the distribution of human capital of the workers at the firm.<sup>3</sup>

One of the conceptual issues that arises naturally in this setting is the possibility that various features of the within-firm distribution of human capital differentiate firms and are related to productivity. Consequently, studying the entire within-firm distribution of skill rather than only the mean is justified by a number of theoretical underpinnings. The relation between skill and productivity at the *individual* level might be nonlinear. For example, empirical results support the widely held view that, for the experience dimension of skill, the earnings-experience profile for an individual worker is strictly concave and thereby reflects an underlying productivity-experience profile that is likewise concave. Some other dimensions of skill may, however, yield a convex productivity-skill profile. For example, “superstar” managers (e.g., Bill Gates) may make a disproportionate contribution to the firm’s productivity and value. A related idea is that the relation between productivity and, say, innate ability at the individual level may be strictly convex. For our purposes, any nonlinear relation between the productivity of a worker and the worker’s skill implies that higher moments of the within-firm distribution of human capital will account for some variation in productivity at the firm level. For example, if the productivity-skill relation for, say, the dimension of experience, is strictly concave for the individual, then firms with greater dispersion on that dimension will have lower firm-level productivity, other things being equal. In contrast, if the productivity-skill relation for the individual worker is strictly convex, then firms with greater dispersion on that dimension of skill will have higher firm-level productivity.

These examples suggest only a few of the reasons that higher moments of the within-firm distribution of human capital might be related to differences in productivity across firms. The models of Kremer and Maskin (1996) suggest that, in some production environments, interaction effects arise across the skills of coworkers in the production function. For example, if the coworker interaction effect reflects complementarities across skill groups at the firm, then a business with lower dispersion will be more productive.

In the analysis that follows, we are not imposing enough structure to be able to identify which of these alternative and interesting factors may be the reason that the within-firm distribution of human capital matters for

3. One issue that we do not explore here is the relation of the firm effects to productivity and market value. Firm effects potentially capture many factors—rent sharing, firm personnel policies, efficiency wages, and the effect of collective bargaining—that may be positively or negatively related to productivity and market value. Identifying and exploring these different effects is a rich area for future work. For more discussion of this issue and related empirical work, see Abowd (1989, 1990).

differences in productivity across firms. Instead, we take an eclectic, exploratory approach and simply include various measures of the within-firm distribution of human capital in our estimation of the relation between productivity, market value, and human capital.

Besides the interesting issues about the role of the within-firm distribution of human capital, unmeasured (i.e., omitted) tangible assets and intangible assets may also complicate the empirical link between human capital and productivity. Any complementarities that may exist between unmeasured assets (or poorly measured assets) and human capital may be reflected in any estimated relation between productivity and human capital. Moreover, the nature of intangible assets may be very closely connected to how human capital is organized.

### 5.2.3 Defining and Measuring Intangibles

A major issue confronting the productivity literature is the problem of studying the effects of intangible assets when those assets are not readily measurable. Thus, although some argue that intangibles are the “major drivers of corporate value and growth in most sectors” (Gu and Lev 2001), the definition of intangibles varies widely, from knowledge and intellectual assets (Gu and Lev); to human capital, intellectual property, brainpower, and heart (Gore 1997); to knowledge assets and innovation (Hall 1998); to organizational structure (Brynjolfsson, Hitt, and Yang 2002—hereafter BHY). Measures of these variables have been equally diverse and have included a residual approach, inference, and direct measurement.

For example, although Gu and Lev (2001) conceptualize intangibles as knowledge assets (new discoveries, brands, or organizational designs), they derive their measure of intangibles as a residual; that is, as the driver of economic performance after accounting for the contribution of physical and financial assets. In empirical terms, they identify the core determinants of intangibles as research and development, advertising, information technology, and a variety of human resource practices. In a series of papers, Hall uses direct expenditures on research and development as well as patent information to proxy for knowledge assets. BHY use survey data on the “allocation of various types of decision making authority, the use of self-managing teams, and the breadth of job responsibilities” (p. 15) to construct a composite variable that acts as a proxy for organizational capital.

The results from using these measures suggest that intangibles vary considerably across firms and sectors and that they are important in accounting for fluctuations in the market. Gu and Lev (2001), who use the broadest measure of intangibles, find that the level and growth rate of intangibles vary substantially across industries. In particular, they find the highest *levels* in insurance, drugs, and telecommunications; the lowest in trucking, wholesale trade, and consulting. However, they find the highest *growth*



rates in consulting, machinery, and electronics industries; the lowest in retail trade, restaurants, and primary metals. Gu and Lev also find that intangibles-driven earnings (by two different measures) are much more highly correlated with stock market returns than are other measures, notably the growth of operating cash flow and the growth of earnings. BHY find that organizational structure has a large effect on market valuation: firms that score 1 standard deviation higher than the mean on this measure have approximately \$500 million greater market value. Hall finds that, although research and development (R&D) accounts for a “reasonable fraction” of the variance of market value, the relation is not stable, and a great deal of the variation is unexplained. Patents matter, according to Hall, but less than research and development.

Empirical studies also suggest that failing to include intangibles is likely to considerably bias estimates of the effects of tangibles on both market value and output. Gu and Lev (2001) find that expenditures on capital, R&D, and technology acquisitions are all highly correlated with intangible capital. Similarly, BHY find evidence for a strong correlation between organizational structure and investment in information technology.

Although the connections are difficult to conceptualize and measure, organizational capital is closely linked to the way workers are organized and, in turn, to the apparently different human capital mixes across firms in the same industry. As emphasized in the preceding section, knowledge of the entire distribution of human capital within each firm allows us to quantify the relation between outcomes like productivity, market value, and the organization of human capital.

#### 5.2.4 The Market Value of a Firm: Tangible Assets, Intangible Assets, and Human Capital

The general approach to describing the market value of a firm,  $V_{jt}$ , in terms of its tangible and intangible assets is well summarized, derived, and motivated in BHY and can be written as

$$(3) \quad V_{jt} = V(K_{jt}^T, K_{jt}^I, \dots)$$

The market value of a firm is assumed to be an increasing function of the assets.<sup>4</sup> Defining and measuring all of the terms in this relation are difficult, however. If the market is characterized by strong efficiency, then, as Bond and Cummins (2000) point out, the market value of a company will equal the replacement cost of its assets (in the absence of adjustment costs and market power). From this perspective, one way of measuring intangibles is a residual approach (see, e.g., Hall 2001), because the residuals will reflect

4. BHY specify a linear relation and emphasize the departure of coefficients from 1. Hall (1998) discusses an alternative log-linear relation that may be relevant. We use the log-linear specification in our analysis in part because our human capital measures are not on the same inherent scale and metric as the measures of assets and market value.

the difference between the market value and observed assets. Alternatively, direct measures of intangibles (e.g., organizational capital, as in BHY) can be included in an econometric specification explaining market value. However, such a specification potentially permits the coefficients on the various assets to reflect direct and indirect effects. One interpretation of the BHY coefficients is that they are due to complementarities with unmeasured intangibles. Thus, the coefficient on any measured asset will reflect the correlation between the measured and unmeasured assets. Thus, as BHY have found, the coefficient on information-technology capital in a linear specification of equation (3) is larger than 1. BHY provide evidence that this reflects the complementarity between market value and organizational capital.

Should the preceding considerations lead one to include human capital in the set of variables in an econometric specification of the market value equation? In the absence of complementarities, a basic view of the role of human capital might suggest that it is not relevant for the market value of the firm. That is, if all human capital is general human capital, and if it is fully compensated by the market, and if no correlation exists between human capital and unmeasured tangible or intangible assets, then human capital will not be reflected in market value.

Several factors may, however, require departures from these assumptions. First, human capital may not be fully compensated in the market. Second, the chosen mix of human capital may indeed be a key aspect of what is meant by “organizational capital.” As discussed above, many factors may yield a relation between productivity and higher moments of the within-firm distribution of human capital. Under this view, the average level of human capital at the firm may not matter for market value, but the way that human capital is organized (as measured by higher moments) may matter a great deal. Finally, human capital (its average and other measures of the distribution) may be complementary to unmeasured tangible and intangible assets in a sense analogous to the arguments and findings of BHY. As such, human capital may be positively related to market value because of measures of tangible or intangible assets that are omitted from the specification.

### 5.2.5 Econometric and Interpretation Issues

The previous subsections provided an overview of our approach. We explore newly created measures of human capital from longitudinal employer-employee data. These measures, in principle, encompass traditional measures and moderate some of the econometric difficulties. In what follows, we explore the relation of these measures to productivity and market value at various levels of aggregation. The discussion above suggests that a host of econometric issues could arise to complicate the estimation of any productivity or market value equation. At the heart of these issues is the

well-known problem that tangible assets and intangible assets, including those involving some measure of human capital, are endogenous. The observed measures for any given econometric implementation of equations (2) and (3) may also be proxies for other unobserved measures. Our specification assumes that firms have control over the level of human capital that they choose, and so an additional empirical concern is the degree to which firms are constrained in adjusting their workforce (as noted by Acemoglu 2001, among others). Work by Haltiwanger, Lane, and Spletzer (2001) notes the remarkable degree of persistence in firms' choice of workforce. However, this overall picture of firm-level persistence suggests that these firm choices are quite deliberate. Firms have ample opportunity to change the workforce—a great deal of empirical evidence shows that firms churn workers through jobs at quite high rates and hence have abundant opportunities to change their worker mix (see, for example, Burgess, Lane, and Stevens 2000).

In this paper, we focus on identifying economically and statistically significant relations rather than attempting to establish causality or to pin down direct versus indirect effects.<sup>5</sup> We include measures of tangible and intangible assets in relatively simple specifications of productivity and market value equations. We recognize that our coefficient estimates reflect both direct and indirect effects of the assets we measure. In particular, the effect of human capital on productivity and market value may include both direct and indirect components. However, by looking at the effect on both productivity and market value, we hope to make some progress in understanding the role of human capital on business outcomes. If the measures of human capital are capturing mostly general human capital, for which the worker is fully compensated, and if such human capital measures are not correlated with unmeasured intangibles (or other unmeasured assets), then human capital is likely to have a positive effect on productivity (as AKM found in France) and very little effect on market value (because firms are fully paying for the human capital and thus generate no additional value from having higher human capital). However, if measures of human capital (or some components or indices of our human capital measures) are positively related to productivity and market value, then the measures are presumably capturing, either directly or indirectly, some form of intangible asset associated with human capital.

We recognize the importance in this context of distinguishing the direct from the indirect effects of human capital. However, as noted above, our objective is to explore the links between our new measures of human capital and productivity and market value in a largely descriptive manner. We anticipate that identifying and quantifying the respective direct and indi-

5. In that sense, our approach is very much in the spirit of BHY.

rect roles of these new measures of human capital in this context will be the subject of much research in the years to come.<sup>6</sup>

One other related issue of interpretation warrants mention. We are exploring the relation between productivity and human capital measures on the one hand, and market value and human capital measures on the other. These two relations capture very different aspects of firm performance. First, as already noted, any productive input that is fully compensated in the market may be related to productivity but unrelated to market value. Second, productivity captures current activity, while market value reflects future profits and associated anticipated value. Thus, the factors that affect current activity may be very different from those affecting future profit streams. One might argue that some factors inherently lead to a negative correlation between market value and current productivity. For example, a business with a “new idea” carrying a high market value may be actively expanding and investing in physical and human capital. Adjustment costs may imply that such a firm exhibits low current productivity.

### 5.3 Data

The key measures for this project are human capital, physical capital, productivity, and market value. The integrated employer-employee data allow us to construct firm-specific measures of human capital. The data from the Economic Censuses provide measures of output, employment, and other inputs to explore the relation between (labor) productivity and human capital measures. The Compustat data on publicly traded firms provide us with measures of output, employment, physical capital, and market value at the firm level. In terms of matching, we first match our employer-employee data to the Economic Census and other business-level data at census. We then match the Compustat data to the integrated employer-employee, Economic Census, and related data.

Figure 5.1 provides a brief summary of the data resources used for this project, but of necessity it vastly reduces the number and complexity of the links involved in constructing the matched employer-employee data. The

6. Cummins (chap. 2 in this volume) takes one approach to separate out some of these effects (although not in the context of using measures of organizational or human capital). He uses instrumental variable techniques to isolate the contribution of measures of tangible assets by trying to find instruments that are correlated with the measured tangibles but uncorrelated with unmeasured intangibles. Thus, he attempts to identify the direct effect of the measured assets. Moreover, his approach in principle avoids another related problem of endogeneity arising from correlations of the asset variables with unmeasured productivity or market value shocks. The unmeasured, idiosyncratic productivity and market value shocks likely reflect, however, the idiosyncratic factors that we are seeking to understand. The pursuit of an estimation strategy for instruments that are supposedly orthogonal to these shocks may completely miss the role of intangibles.

<u>Data Set:</u>	<u>Demographic Data</u>	Links to	<u>Business Data</u>	Links to	<u>Financial Data</u>
Source:	UI files, CPS, SIPP		LBD, Economic Censuses, Annual Surveys, ACES, RD-1		Compustat
Statistical Unit(s):	Individual, EI		Establishment, EI, Firm		Firm
Key Variables:	Earnings, Age, Education, Race, etc	via EIN	Employment, Payroll, Sales, Productivity, R&D, Computer Investment, Entry, Exit, etc	via Firm Identifier	Financial Data: Market value, "Q," intangibles, R&D, advertising, Capital, Output, Employment
Availability:	By State: Currently have for several States (1990s).		Nationwide - availability of some measures varies across sectors and years. Economic Census Data available in 1992-97.		Nationwide for 1980s and 1990s.

**Fig. 5.1** Summary of data sources

details of these linkages can be found in the appendix. However, we provide summary information about the data and matching in the next two subsections.

### 5.3.1 The Integrated Employer-Employee Data

We exploit new Census Bureau data that are part of the LEHD program. These data integrate information from state unemployment insurance records and economic and demographic data from the Census Bureau in a manner that permits the construction of longitudinal information on workforce composition at the firm level. The LEHD program permits direct linking of the Census Bureau's demographic surveys (household-based instruments) with its economic censuses and surveys (business-based and business-unit-based surveys).

The unemployment insurance (UI) wage records are discussed elsewhere (see Burgess, Lane, and Stevens 2000). The Employment Security Agency in each state collects quarterly employment and earnings information to manage its unemployment compensation program. These data enable us

to construct quarterly longitudinal information on employees. The advantages of the UI wage records are numerous. The data are frequent, longitudinal, and potentially universal. The sample size is generous, and reporting for many items is more accurate than that in survey-based data. The advantage of having a universe as opposed to a sample is that the movements of individuals among employers and the consequences for earnings can be tracked. In addition, longitudinal data can be constructed with the employer as the unit of analysis. The LEHD program houses data from nine states that currently comprise 45 percent of total U.S. employment: California, Florida, Illinois, Maryland, Minnesota, North Carolina, Oregon, Pennsylvania, and Texas. In this paper, we use data from seven of these states and provisional national weights to build our human capital measures. We currently have the crosswalk between the UI files and the establishment-level and firm-level data for six of the seven states. For this reason, we restrict our analysis of the relation between productivity, market value, and human capital to data from this six-state subset.<sup>7</sup>

Perhaps the main drawback of the UI wage data is the lack of even the most basic demographic information on workers (Burgess, Lane, and Stevens 2000). Links to Census Bureau data can overcome this problem in two ways: first, the individual can be integrated with administrative data at the Census Bureau containing information such as date of birth, place of birth, and gender for almost all the workers in the data. Second, as discussed in the previous section, LEHD program staff members have exploited the longitudinal and universal nature of the dataset to estimate jointly fixed worker and firm effects using the methodology described in detail in ALM and in Abowd, Creecy, and Kramarz (2002).

### 5.3.2 The Economic Censuses and Related Business-Level Data

The Economic Censuses (conducted every five years) provide comprehensive data on basic measures such as output, employment, and payroll for all of the establishments in the United States. In addition, in certain sectors (such as manufacturing) the census asks more-detailed questions on other inputs (such as capital).

Our first goal is to create a matched data set linking the human capital measures to the Economic Census data. For the present work, we focus on the Economic Censuses in 1997. One issue that immediately arises is the appropriate and feasible level of aggregation of business activity. Although the Economic Censuses are conducted at the establishment level, the business-level identifiers on our human capital measures are at the federal Employer Identification Number (EIN), two-digit Standard Industrial Classi-

7. In order to meet the requirements of the data use agreement between census and the individual states the identity of states used in a particular analysis is normally not released.

fication (SIC) code, and state level.<sup>8</sup> Therefore, we aggregate the Economic Census data to the level of the business-level identifiers so as to match them to the human capital files.<sup>9</sup> Our unit of observation is somewhere between an establishment and a firm, but most of the observations in our analysis data are at the establishment level. For firms with multiple establishments reporting under a single EIN in a single state, we aggregate the establishment data to the two-digit SIC (SIC2), state level. In what follows, we begin our analysis of the relation between human capital and productivity using this “quasi-establishment-level” data. For this analysis, we have roughly 340,000 business units that we can match to the Economic Censuses (in a universe of roughly 430,000 business units at this level of aggregation from the UI files for the six states used in this analysis). Most of the UI businesses that we cannot match to the Economic Censuses are beyond the scope of the Economic Censuses (for example, agricultural businesses). We are also able to accomplish essentially the same thing in non-Census years using the Census Business Register, previously known as the Standard Statistical Establishment List (SSEL). Although the information in the Business Register is limited, it does identify the ownership structure of firms so that we can further aggregate to the enterprise or firm level. The latter aggregation permits us to match enterprise-level data on human capital to Compustat.

Because we are working with only six states, we are limited in our ability to examine evidence for large companies that operate in multiple states. We use a threshold rule (for example, 50 percent of employment in the company must be in these six states) to restrict attention to companies for which we can measure human capital and firm outcomes such as market value and productivity in a comparable fashion. In what follows, we first aggregate our human capital estimates up to the firm level for all firms in our six states (using the 50 percent rule as noted). The resulting sample contains roughly 300,000 firms. We use this sample to investigate the relation between human capital and productivity at the enterprise or firm level. We then restrict our attention to Compustat firms, which reduces the sample substantially because only about 13,000 Compustat firms exist nationwide. For this restricted sample, we again investigate the relations between human capital and productivity and also investigate the relation between human capital and market value.

8. The identifiers in the LEHD program’s human capital data provide additional geographic and industry information, but they are not coded down to the workplace (establishment) level. Ongoing research is attempting to refine the most disaggregated economic entity available in these data.

9. We use the two-digit SIC number as the industry measure in this work because, although industries are coded at the four-digit level in the Economic Census and in the underlying establishment data on the UI side, we have not yet implemented an algorithm to use additional industrial or geographic detail in the definition of the establishment.

## 5.4 Human Capital Estimates

The results of the human capital estimation are based on data for seven states for the years 1986–2000 and use the specification in equation (1). Although the methodology and estimates we use are discussed in detail in ALM, we provide here a brief summary of some of the features of the human capital estimates before relating these new estimates to productivity and market value.

Some basic features of the estimates for the seven-state data set are shown in table 5.1. First, the contribution of worker and firm effects to worker earnings are roughly equal. Second, the  $R^2$  of this earnings regression ( $1 - .402^2$ ) is approximately .84, a great deal higher than that for regressions based simply on worker characteristics. Third, ALM augment the analysis by decomposing the person effect into the part attributable to time-constant observable characteristics such as gender and education and the part attributable to unobservable characteristics. The fourth and fifth rows of the table illustrate the results of this decomposition (using the notation from AKM). The unobserved component of the person effect is much more important and more highly correlated with wages than the observed component.

Fourth, the different components of human capital (that is, the person effect and the experience component) exhibit different variation and co-variation. Indeed, an interesting feature of the person effect and experience component is that they are negatively correlated. This result is not surprising (because, for example, younger generations of workers are more highly

**Table 5.1** Summary of pooled human capital wage components

Component	Standard deviation	Correlation with:						
		$\ln w$	$x\beta$	$\theta$	$\alpha$	$u\eta$	$\psi$	$\varepsilon$
Log real annualized wage rate ( $\ln w$ )	0.881	1.000	0.224	0.468	0.451	0.212	0.484	0.402
Time-varying personal characteristics ( $x\beta$ )	0.691	0.224	1.000	-0.553	-0.575	-0.099	0.095	0.000
Person effect ( $\theta$ )	0.835	0.468	-0.553	1.000	0.961	0.275	0.080	0.000
Unobserved part of person effect ( $\alpha$ )	0.802	0.451	-0.575	0.961	1.000	0.000	0.045	0.000
Non-time-varying personal characteristics ( $u\eta$ )	0.229	0.212	-0.099	0.275	0.000	1.000	0.101	0.000
Firm effect ( $\psi$ )	0.362	0.484	0.095	0.080	0.045	0.101	1.000	0.000
Residual ( $\varepsilon$ )	0.354	0.402	0.000	0.000	0.000	0.000	0.000	1.000

*Sources:* Authors' calculations using the Longitudinal Employer-Household Dynamics (LEHD) Program databases.

*Notes:* Based on 287,241,891 annual observations from 1986–2000 for 68,329,212 persons and 3,662,974 firms in seven states as described in the text. No single state contributed observations for all years.



educated), but it reminds us that skill consists of several dimensions and that they need to be taken into account.

Finally, one surprising aspect of this comprehensive decomposition of the wages is that the correlation between the person effect and the firm effect is virtually zero at the observation (person-year) level. We do not seek to explain this somewhat surprising finding, but we note that various aggregations of the person and firm effects yield a strong and positive relation. For example, ALM show that at the industry level, person and firm effects are positively related. Interestingly, Abowd and others (2001) show that at the firm level, person and firm effects are positively related after controlling for output, local wage effects, and broad industry. These results by industry and at the firm level are quite relevant here since they suggest systematic sorting of workers across different firms and industries.

In the next subsections, we first provide some summary information about how these new measures compare with the JGF-like measures of human capital. We also describe the differences in the two components of the ALM measure of human capital—experience and person effects—and how they vary across workers. Finally, we examine the degree to which the human capital measures vary across firms and industries.

#### 5.4.1 A Comparison of New and Traditional Measures of Human Capital

In principle, the JGF methodology can be applied equally well to the measurement of both sectoral and aggregate labor quality, but in practice, the LEHD approach permits more heterogeneity within and across industries. Lengermann (2002a) has developed sectoral aggregates of human capital following the JGF approach and compared them to LEHD estimates. Briefly, the JGF approach incorporates data from the Censuses of Population, the Current Population Survey (CPS), and the national income and product accounts (NIPA). JGF base their labor quality indexes on totals of labor inputs cross-classified by sex, age, educational attainment, employment class, and industry. We summarize the results of two different types of comparison here.

The first “direct” approach compares the JGF indexes to sectoral labor quality derived from industry averages of our human capital measure for the period 1995–98. JGF formally define labor quality as the ratio of the total volume of labor to hours worked, where volume is measured by a constant-quality index of labor quantity. The LEHD measure of industry-average human capital follows essentially the same logic, where the measure of labor volume is also based on a constant-quality human capital measure and where total employment substitutes for total hours worked. Neither approach is completely satisfactory. The LEHD data cannot measure hours worked. The JGF constant-quality index of labor quality confounds firm heterogeneity with person heterogeneity.

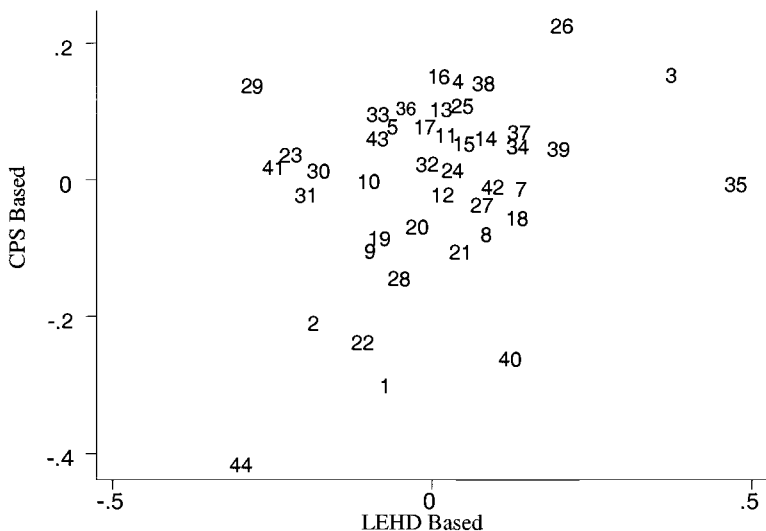
We compare the growth rates in the human capital indexes over the period 1995–98 using the LEHD-based and JGF approaches. The within-industry growth rates are highly correlated—the employment-weighted average of the sectoral correlations is 0.79. However, average growth for any given industry is much higher, and cross-industry variation in those growth rates is much greater, in the LEHD measures than in the JGF measures: the average growth rate for the LEHD measure over the four years is 0.04, with a cross-industry standard deviation of 0.067, while the corresponding values for the JGF measures are 0.014 and 0.001.

In what follows, we exploit cross-sectional variation (across firms) in their human capital, whereas the JGF procedure focuses on generating growth rates of human capital by industry. As such, the JGF measures are not well suited to examining within-year, cross-industry variation. Thus, as a second “indirect” approach we approximate the JGF labor quality indexes with indexes derived from predicted industry average wages obtained by regressing wages on age, education, and sex using the CPS. For this purpose, we use the same cells used by JGF. We show that the time series growth rates of these indirect measures are highly correlated with the actual JGF measures (the employment-weighted average correlation is 0.73). Thus, the CPS-based approach does a reasonable job of approximating the more sophisticated JGF measures.

We compare the cross-industry variation in the CPS-based measures with the same variation using the LEHD measures for the year 1998. The two measures are, in principle, comparable because both rely on regression approaches that attempt to isolate the component of wages due to individual characteristics. However, because LEHD data permit individual contributions to wages to be distinguished from firm contributions, one might not expect them to yield identical results. Workers sort nonrandomly into firms based on their own characteristics—both observable and unobservable—and the characteristics of firms. Furthermore, firm wage premiums—the firm effects in the wage regression (1)—are not distributed uniformly across industries. These two facts imply that a strong, positive correlation exists between person and firm heterogeneity at the industry level (ALM)—a correlation that the JGF cell-based analysis cannot disentangle.

We plot the industry-level aggregates for the CPS-based approach against the industry-level aggregates for the most inclusive measure of skill from the LEHD approach (figure 5.2). Although the levels are normalized differently, a great deal of correlation clearly exists between the two measures—indeed, the correlation is 0.76. However, somewhat more cross-industry variation exists in the LEHD-based measure than in the CPS-based measure (the standard deviation of the former is 0.15, and that of the latter is 0.13).

In summary, the LEHD-based measures by industry are closely related to those derived by JGF or by a simpler but closely related CPS-based pro-



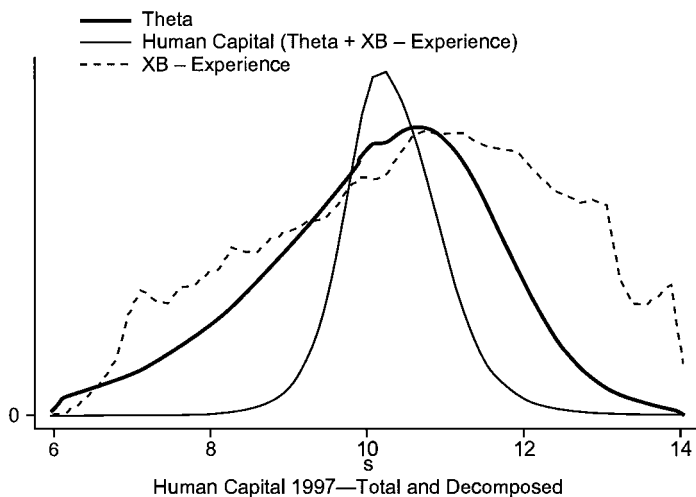
**Fig. 5.2** Comparison of CPS and LEHD measures of human capital at the industry level

cedure. However, LEHD-based measures generate greater average growth and more cross-sectional variation both in growth rates across industries and in levels of human capital across industries within a year.

#### 5.4.2 The Construction of New Human Capital Measures

A major contribution of the LEHD approach is the richness of the new measures of human capital, and these are fully discussed in ALM. Here, we explore some of the key features of the new measures, particularly aggregated to the firm level. For this purpose, we use three worker and firm traits to build measures of the human capital resources available to firms: the person effect ( $\theta$ ), the overall labor market experience of each worker captured by the experience component of  $x\beta$  (denoted  $x\beta$  in this section, though it excludes the additional controls described in note 1), and the sum of these two components (overall human capital, or  $h$ ).

We describe the distribution of these measures in figure 5.3. All three components of the distribution obviously exhibit enormous variation across workers. The shapes of the distributions of the alternative measures differ: The distribution of the person effect is bell shaped and has thick tails and high variance; the distribution of experience is less smooth; and the distribution of human capital (the sum of  $\theta$  and  $x\beta$ ) is roughly bell shaped and centered about zero and has much less mass at the tails than either experience or  $\theta$ . Underlying these relations is the negative correlation between experience and person effects reported in table 5.1.



**Fig. 5.3** Distribution of human capital across workers

### 5.4.3 The Construction of Firm-Level Measures

Although the different worker-level measures of human capital provide a useful context, the focus of this paper is on developing firm-level measures of human capital and relating them to firm outcomes. The firm-level measures that we use are those developed in ALM. They are based on kernel-density estimates of the within-firm distribution of human capital. The details of the estimation of the kernel densities are provided in ALM. Some restrictions on the sample that are necessitated by this approach are discussed in detail in the appendix.

One key aspect of the variation across firms is driven by large variation in the distribution of human capital across industries (table 5.2). Finance, insurance, and real estate (FIRE) and manufacturing are both high-human-capital industries. However, the components of human capital vary across these industries: the FIRE industries have high human capital especially because of having workers with high person effects, while the high human capital of the manufacturing industries arises more through their having workers with high experience effects.

Human capital varies substantially across industries, but the variation across firms within a given industry is enormous. We computed for each firm the share of workers that are in the lowest quartile of the economywide distribution of human capital to obtain the distribution of those shares across firms (figure 5.4, panel A). We conducted the same exercise for the share of workers at each firm that are in the upper quartile of the distribution (figure 5.4, panel B).

**Table 5.2** Summary statistics on human capital, by industry (1997)

Industry (SIC division)	Total employment	Proportion of workers above the overall median of $h$	Proportion of workers above the overall median of $\theta$	Proportion of workers above the overall median of $x\beta$
Agriculture	304,134	0.338	0.407	0.502
Construction	1,366,022	0.510	0.465	0.556
FIRE	1,382,730	0.531	0.591	0.439
Manufacturing	3,365,954	0.539	0.473	0.560
Mining	194,678	0.511	0.387	0.646
PubAdmin	811,215	0.558	0.451	0.584
Retail	3,537,787	0.383	0.542	0.388
Services	7,856,442	0.493	0.520	0.468
TCE	1,374,002	0.562	0.495	0.558
Wholesale	1,626,221	0.567	0.529	0.540

*Notes:* The sample is 1997 job-level unemployment insurance (UI) data from 6 states. Includes all jobs held by workers imputed to be full time at the end of the first quarter 1997.

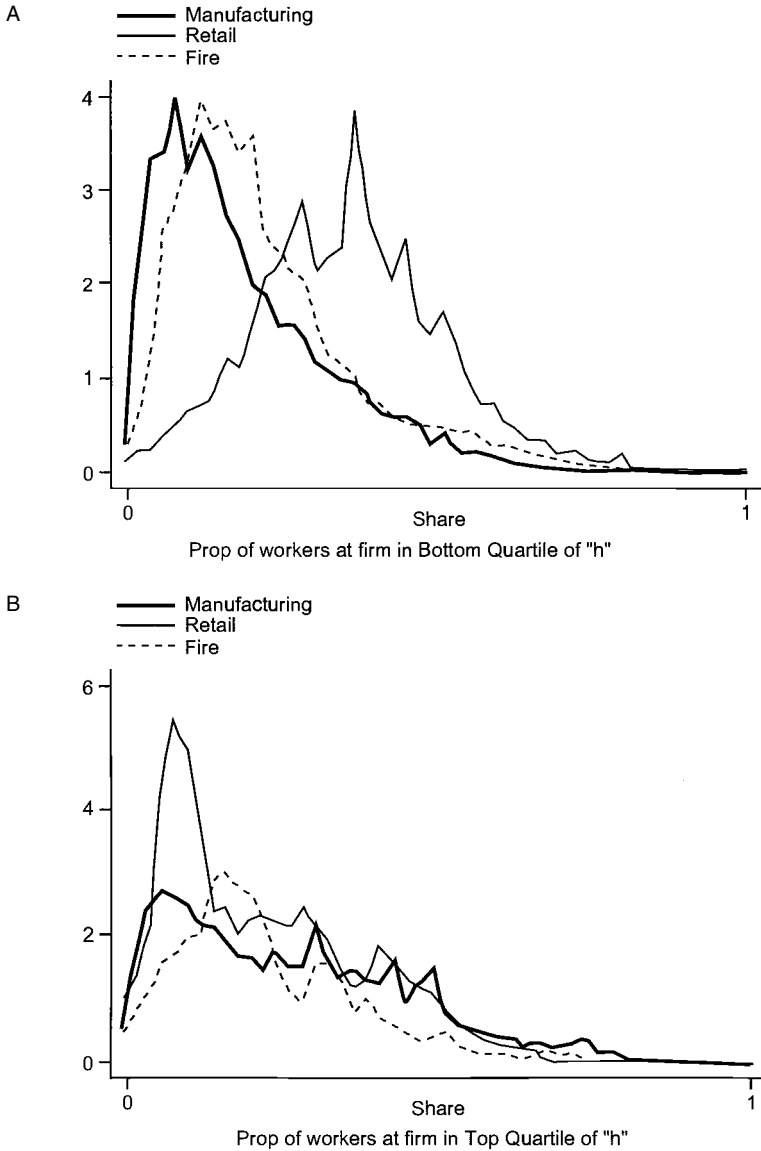
Figure 5.4 shows substantial differences both within and across industries. As is consistent with the data in table 5.2, many manufacturing and FIRE firms have low shares of low-skill workers (lowest quartile), while retail trade has many firms with high shares of low-skill workers. However, firms within an industry obviously vary enormously in their shares of high- and low-skill workers. Apparently, different firms in the same industry choose very different mixes of human capital; in the analysis that follows, we will investigate whether this heterogeneity in human capital is related to heterogeneity in productivity and market value.

## 5.5 The Relation between Productivity and Human Capital at the Micro Level

In this section, we explore the relation between our rich measures of establishment-level human capital and establishment- and firm-level productivity, controlling, as much as possible, for other relevant factors (capital intensity, for example).<sup>10</sup> For this purpose, we focus on the 1997 Economic Census. Our measure of labor productivity is revenue per worker, the standard measure used in official BLS productivity statistics for gross output per worker.<sup>11</sup>

10. Recall that the level of aggregation that we use to approximate the establishment is that of an EIN/SIC2/state cell. This unit is somewhere between the establishment and the firm, but most firms operate only a single establishment.

11. The official estimates include adjustments for changes in inventories in inventory-holding sectors. However, studies by Foster, Haltiwanger, and Krizan (2001) show that, in manufacturing, the correlation between labor productivity measured as shipments per worker and labor productivity measured as shipments adjusted for inventory changes is extremely high (almost 1).



**Fig. 5.4** Between-firm distributions of human capital, by industry: *A*, proportion of workers at firm in bottom quartile of  $h$ ; *B*, proportion of workers at firm in top quartile of  $h$

An important goal is to determine which measures of the within-firm distribution of human capital are relevant for understanding outcomes such as productivity and market value. From a traditional viewpoint, we want to control for a measure of the central tendency of the within-firm distribution of human capital. However, from the perspective of considering nonlinearities and other factors related to the organization of human capital at a business, we also want to explore additional measures of the within-firm distribution of human capital. Our approach is necessarily exploratory since neither theory nor earlier empirical research provides much practical guidance.

Accordingly, we explore the role of the following measures: (a) the fraction of workers at the business whose human capital is above the economywide median, (b) the fraction above the 75th percentile, (c) the fraction below the 25th percentile, and (d) the interaction between—more specifically, the product of—measures b and c. We consider these four measures using the overall human capital measure  $h$  and also the separate components of human capital (the person effect,  $\theta$ , and the experience component,  $x\beta$ ).<sup>12</sup> Moreover, we consider a range of specifications, some parsimonious (with only a small number of summary human capital measures) as well as richer specifications with a number of measures of the distribution included.

Table 5.3 presents the means and standard deviations of our human capital and labor productivity measures for our overall sample and for the manufacturing businesses. For the latter we can also measure capital intensity. The statistics reported in the table are based on the employment-weighted distribution. In section 4, we discussed many of the features of the human capital distribution across businesses. However, a few additional points are worth making here (table 5.3). First, businesses exhibit tremendous heterogeneity in their mix of human capital, as evidenced by the very large standard deviations in the human capital measures. Second, manufacturing apparently has higher labor productivity and workers with higher human capital (on both the person effect and experience dimensions) than other sectors.

We made an exploratory analysis of the relation between log labor productivity and our  $h$  measure of the distribution of human capital (table 5.4) and an analysis using the components of  $h$  separately (table 5.5). In all cases, the results are based upon employment-weighted regressions. All analyses included two-digit fixed industry effects, which are highly significant. Moreover, the explanatory power of each set of regressions is uni-

12. Note that the economywide thresholds are based on the universe of all workers in the six states (not just workers employed at the businesses we match to the Economic Censuses) for which we have developed human capital estimates. We also generated versions of these measures on the basis of thresholds relative to a specific industry rather than the wider economy. The results were not sensitive to this distinction.

**Table 5.3** Mean values of variables in log productivity regressions

Variable	All sectors		Manufacturing	
	Mean	SD	Mean	SD
Log labor productivity	4.731	1.140	5.130	0.888
Log capital intensity			4.230	1.201
<i>Overall <math>h = \theta + x\beta</math></i>				
Fraction of employment above 50th percentile	0.480	0.215	0.538	0.225
Fraction of employment above 75th percentile	0.237	0.176	0.267	0.185
Fraction of employment below 25th percentile	0.266	0.172	0.208	0.169
Interaction: fraction above 75th percentile with fraction below 25th percentile	0.042	0.025	0.035	0.022
<i>Person effect (<math>\theta</math>)</i>				
Fraction of employment above 50th percentile	0.519	0.180	0.473	0.181
Fraction of employment above 75th percentile	0.264	0.130	0.203	0.110
Fraction of employment below 25th percentile	0.230	0.156	0.239	0.173
Interaction: fraction above 75th percentile with fraction below 25th percentile	0.048	0.027	0.039	0.026
<i>Experience component (<math>x\beta</math>)</i>				
Fraction of employment above 50th percentile	0.455	0.157	0.545	0.133
Fraction of employment above 75th percentile	0.220	0.116	0.285	0.115
Fraction of employment below 25th percentile	0.282	0.144	0.205	0.096
Interaction: fraction above 75th percentile with fraction below 25th percentile	0.049	0.018	0.050	0.016
No. of observations	337,495		39,638	

*Notes:* The sample is 1997 data from 6 state UI-based firms (defined at the EIN 2-digit SIC level) matched to Economic Census and Annual Survey of Manufactures data. SD = standard deviation.

formly high; this result suggests that measures of human capital are, either directly or indirectly, important sources of cross-sectional differences in productivity. The fact that the explanatory power for the manufacturing sector regressions is substantially less than that for all sectors is consistent with the notion that human capital is more important for the service sector than manufacturing—and more important for the “new” economy than the “old” economy.

Businesses with a greater fraction of workers above the economywide median human capital level are much more productive (tables 5.4 and 5.5, column [A]). For the overall human capital measure, a change in this fraction of 1 standard deviation is associated with a change of 27 log points in labor productivity (table 5.4). For the person-effect measure (table 5.5), a change of 1 standard deviation in the fraction of high human capital workers is associated with a change of 25 log points in labor productivity. For the experience component, a change of 1 standard deviation in the fraction of high human capital workers is associated with a change of 23 log points



**Table 5.4**                      **The relation between labor productivity and the complete human capital measure (analysis level: establishments; dependent variable: log labor productivity)**

Explanatory variable	All sectors			Manufacturing only			
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Fraction of workers above 50th percentile of human capital	1.264 (0.007)			1.064 (0.017)	0.512 (0.017)		
Fraction of workers above 75th percentile of human capital		-0.017 (0.010)	0.268 (0.012)			0.430 (0.030)	0.173 (0.027)
Fraction of workers below 25th percentile of human capital		-1.875 (0.012)	-1.673 (0.012)			-1.143 (0.033)	-0.658 (0.031)
Interaction of above 75th and below 25th percentiles			-3.032 (0.063)			-6.563 (0.197)	-3.870 (0.187)
Log capital intensity					0.302 (0.004)		0.285 (0.003)
No. of observations	337,495	337,495	337,495	39,638	33,926	39,638	33,926
R <sup>2</sup>	0.555	0.569	0.572	0.325	0.471	0.353	0.483

*Notes:* The human capital measure is  $h = \theta + x\beta$ . The estimation sample is UI-based establishments (defined at the EIN/2-digit SIC level) for six states matched to the 1997 Economic Census and Annual Survey of Manufactures data. Standard errors in parentheses. Other controls include 2-digit industry effects. Results are based on employment-weighted regressions.

**Table 5.5**                      **The relation between labor productivity and human capital, decomposed (analysis level: establishments; dependent variable: log labor productivity)**

Explanatory variable	All sectors			Manufacturing only			
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Fraction of workers above 50th percentile for $\theta$	1.400 (0.009)			1.240 (0.024)	0.670 (0.022)		
Fraction of workers above 75th percentile for $\theta$		1.990 (0.015)	1.700 (0.017)			2.060 (0.050)	1.480 (0.048)
Fraction of workers below 25th percentile for $\theta$		-0.450 (0.010)	-0.920 (0.017)			-0.710 (0.043)	-0.299 (0.040)
Interaction of above 75th and below 25th percentiles for $\theta$			2.830 (0.078)			3.110 (0.246)	1.941 (0.230)
Fraction of workers above 50th percentile for $x\beta$	1.490 (0.120)			1.410 (0.032)	0.450 (0.031)		
Fraction of workers above 75th percentile for $x\beta$		0.200 (0.020)	0.760 (0.022)			0.540 (0.062)	0.135 (0.058)
Fraction of workers below 25th percentile for $x\beta$		-1.900 (0.017)	-1.560 (0.018)			-1.610 (0.081)	-0.713 (0.077)
Interaction of above 75th and below 25th percentiles for $x\beta$			-4.800 (0.090)			-7.020 (0.328)	-5.175 (0.308)
Log capital intensity					0.310 (0.003)		0.298 (0.003)
No. of observations	337,495	337,495	337,495	39,638	33,926	39,638	33,926
R <sup>2</sup>	0.547	0.564	0.568	0.315	0.471	0.360	0.496

*Note:* See table 5.4 notes.

in labor productivity (table 5.5). These effects are large, yet they reflect only a fraction of the standard deviation in measured labor productivity across businesses (114 log points).

We also consider alternative measures of the distribution of human capital—focusing on the fraction of workers with high human capital and those with low human capital (tables 5.4 and 5.5, columns [B] and [C]). Here the results are somewhat more complicated to interpret, but in all of the results a rightward shift in the distribution is still associated with an increase in productivity. That is, if the share of workers in the lower quartile is decreased a certain amount and the share of workers in the upper quartile is increased the same amount, then productivity increases; this result holds for the overall  $h$  measure and for each of the components of the human capital measures.

However, asymmetric effects arise from changes in the upper tail and lower tail, and the results are also sensitive to inclusion of an interaction effect. Moreover, the nature of the asymmetries differs across components of human capital. The results for the overall  $h$  measure show that changing the share of workers in the firm that are in the lower tail of the human capital distribution has a disproportionate effect (table 5.4, column [B]). Somewhat surprisingly, the coefficient on the upper tail is negative but small in absolute terms and relative to the coefficient on the lower quartile, and it is not significant. The analogous column in table 5.5 (column [B]) sheds further light on these results and shows that different components of human capital act in different ways. In particular, a disproportionate change arises from the upper tail of the person effect and from the lower tail of the experience effect.

We have also devised an even richer specification in which we attempt to capture the interaction between high-skill and low-skill workers (tables 5.4 and 5.5, column [C]). In this specification, we find that the linear terms have the expected signs: holding other things constant, including the interaction effect, businesses with more workers in the top quartile of the human capital distribution and fewer workers in the lowest such quartile are more productive. However, the interaction effects are an important part of the effects of interest. For overall  $h$  and the experience effects, we find that the interaction effect is negative, whereas for the person effects we find that the interaction effect is positive.

Putting the linear and interaction effects together reinforces the asymmetries we have already noted. That is, for the person effects, we obtain a disproportionately large change from an increase in the upper tail of the distribution, and the positive interaction effect reinforces this asymmetry. This result can be seen by noting that the combined linear and interaction effect for the person effect evaluated at the mean for the upper quartile is 2.45, and the combined linear and interaction effect for the lower quartile is  $-0.167$ . The magnitude of the implied variation in productivity is very

asymmetric as well. The combined effect implies that an increase of 1 standard deviation in the share of workers in the highest quartile yields a change of 32 log points in productivity while an increase of 1 standard deviation in the share of workers in the lowest quartile yields a loss of 3 log points in productivity. The opposite pattern holds for the experience effects. That is, the interaction effects reinforce the disproportionate change produced by the lower tail of the distribution of the experience effect.

As discussed in section 5.2, the asymmetries in the effects of human capital on productivity can be explained in a variety of ways. We cannot distinguish between competing explanations, but our findings are consistent with the view that, at the worker level, the relation between productivity and experience is concave, and the relation between productivity and the person effect is convex. However, the results may also reflect complementarities across coworkers that differ on different dimensions of skill.

Columns [D] through [G] of tables 5.4 and 5.5 show results for the manufacturing sector. Column [D] replicates column [A] but only for manufacturing. In column [E], capital intensity is an additional measure. For the most parsimonious specification, the results for manufacturing are quite similar to those for the overall economy when we do not control for capital intensity. Controlling for capital intensity does not change the qualitative nature of the results, and although it reduces the magnitudes of the effects substantially, they nonetheless remain very large. This aspect of the findings suggests that human capital is complementary to physical capital. Thus, as we discussed in section 5.2, we need to recognize that our measures of human capital are capturing both direct and indirect effects (where the latter stem in part from unobserved factors such as tangible and intangible assets).

Columns F and G of tables 5.4 and 5.5 present results for manufacturing using the richer specification used in column [C] for all sectors—with and without a control for capital intensity. Again, the results are quite similar to those for all sectors without capital intensity. Once again, adding capital intensity reduces the magnitudes of most of the effects from the human capital measures.

For manufacturing as a whole, capital-skill complementarity is clearly present. Indeed, capital-skill complementarity seems to exist for all of the dimensions of skill we are investigating. That is, the inclusion of capital intensity reduces the person effect, the interaction effects, and the effect of the experience component.

How sensitive are these results to the level of aggregation? We address this issue by aggregating establishment-level data from the 1997 Economic Censuses to the firm level and estimating a set of similar regressions (tables 5.6 and 5.7). The qualitative results are very similar, and the most parsimonious specifications yield magnitudes that are quite similar to the “establishment-level” results in tables 5.4 and 5.5. However, for the more

**Table 5.6**                    **The relation between labor productivity and the complete human capital measure (analysis level: firms; dependent variable: log labor productivity)**

Explanatory variable	All sectors			Manufacturing only			
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Fraction of workers above 50th percentile of human capital	1.309 (0.007)			0.690 (0.016)	0.433 (0.015)		
Fraction of workers above 75th percentile of human capital		0.11 (0.011)	0.184 (0.013)			-0.194 (0.031)	-0.189 (0.029)
Fraction of workers below 25th percentile of human capital		-1.664 (0.011)	-1.621 (0.011)			-1.133 (0.028)	-0.786 (0.027)
Interaction of above 75th and below 25th percentiles			-0.648 (0.058)			-0.889 (0.166)	-0.274 (0.154)
Log capital intensity					0.256 (-0.003)		0.248 (0.003)
No. of observations	303,219	303,219	303,219	34,900	34,294	34,900	34,294
R <sup>2</sup>	0.537	0.551	0.551	0.292	0.408	0.310	0.416

*Notes:* The human capital measure is  $h = \theta + x\beta$ . The estimation sample is Business Register-based firms (defined as those with at least 50% of U.S. employment in the analysis states) for six states matched to the 1997 Economic Census and Annual Survey of Manufactures data. Standard errors in parentheses. Other controls include 2-digit SIC industry effects for the firm's primary industry and indicators for multi-location status and whether the firm had establishments in 1, 2, or 3+ 2-digit SIC categories. Results are based on employment-weighted regressions.

**Table 5.7**                    **The relation between labor productivity and human capital, decomposed (analysis level: firms; dependent variable: log labor productivity)**

Explanatory variable	All sectors			Manufacturing only			
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
Fraction of workers above 50th percentile for $\theta$	1.562 (0.009)			0.923 (0.023)	0.614 (0.021)		
Fraction of workers above 75th percentile for $\theta$		1.783 (0.015)	1.557 (0.017)			1.190 (0.050)	1.000 (0.046)
Fraction of workers below 25th percentile for $\theta$		-0.729 (0.011)	-1.070 (0.016)			-0.625 (0.038)	-0.385 (0.036)
Interaction of above 75th and below 25th percentiles for $\theta$			2.173 (0.073)			2.522 (0.209)	2.199 (0.193)
Fraction of workers above 50th percentile for $x\beta$	1.483 (0.011)			0.604 (0.030)	0.224 (0.028)		
Fraction of workers above 75th percentile for $x\beta$		0.307 (0.019)	0.430 (0.022)			-0.113 (0.058)	-0.262 (0.053)
Fraction of workers below 25th percentile for $x\beta$		-1.718 (0.015)	-1.711 (0.017)			-1.207 (0.067)	-0.871 (0.062)
Interaction of above 75th and below 25th percentiles for $x\beta$			-0.569 (0.084)			-2.348 (0.268)	-1.942 (0.247)
Log capital intensity					0.262 (0.003)		0.257 (0.003)
No. of observations	303,219	303,219	303,219	34,900	34,294	34,900	34,294
R <sup>2</sup>	0.530	0.547	0.549	0.288	0.410	0.310	0.427

*Note:* See table 5.6 notes.

complex specifications, the magnitudes vary somewhat from the establishment-level results, especially when interaction effects are included. The differences in results are most apparent for the manufacturing sector when we include interaction effects for the overall  $h$  case and for the results on the experience component. Even in these cases, the overall patterns are quite similar: the lower tail of the human capital distribution for both  $h$  and experience at both the establishment level and the firm level produces disproportionate effects.

The differences that arise between the establishment-level and firm-level results may be due to increased measurement error of both the productivity and the human capital measures at the firm level. Measuring productivity is more difficult at the firm level than at the establishment level, especially for large, complex firms with many establishments that cross industry boundaries. In a like manner, the human capital measures are more complex at the firm level than at the establishment level because firms with many establishments may differ in their distributions of human capital across establishments. The latter situation is interesting in its own right, and we plan to explore it in future work.

The results demonstrate powerfully that understanding differences in labor productivity across businesses—particularly outside of manufacturing—involves understanding differences in human capital across businesses. The close relationship between labor productivity and human capital is clearly evidenced by the very large  $R^2$  in the regressions, and the result obtains regardless of whether these are direct or indirect effects and regardless of endogeneity issues. The results also clearly suggest that what matters is not simply a measure of the central tendency of the human capital distribution. The fraction of workers at the tails of the distribution matters, suggesting that the dispersion of human capital matters. Perhaps the most intriguing aspect of our results is the finding that the different components of human capital matter in different ways: the most productive firms are those that have a high fraction of workers in the top quartile of the person-effect distribution and a low fraction of workers in the lowest quartile of the experience-effect distribution. These findings strongly suggest that the organization and mix of the workforce matter substantially.

## 5.6 Investigating the Relation between Market Value and Human Capital

While we have several alternative samples and levels of aggregation at which to investigate the relation between productivity and human capital, market value is measured only at the firm level and only for publicly traded firms. Therefore, we are constrained to using the relatively small matched Compustat sample (a detailed discussion of the matched Compustat sample and variable definitions are in the appendix). We report the means and standard deviations of this subset of observations for 1996–98 (table

**Table 5.8** Mean values of variables in market value regressions

Variable	All sectors	
	Mean	SD
Employment	2,540	9,106
Log market value	4.844	2.008
Log capital	3.167	2.175
Log other assets	3.814	2.086
Multi-location indicator	0.780	
<i>Overall <math>h = \theta + x\beta</math></i>		
Fraction of employment above 50th percentile	0.545	0.189
Fraction of employment above 75th percentile	0.291	0.163
Fraction of employment below 25th percentile	0.212	0.140
Interaction: fraction above 75th percentile with fraction below 25th percentile	0.045	0.210
<i>Person effect (<math>\theta</math>)</i>		
Fraction of employment above 50th percentile	0.560	0.179
Fraction of employment above 75th percentile	0.312	0.148
Fraction of employment below 25th percentile	0.207	0.138
Interaction: fraction above 75th percentile with fraction below 25th percentile	0.049	0.023
<i>Experience component (<math>x\beta</math>)</i>		
Fraction of employment above 50th percentile	0.460	0.147
Fraction of employment above 75th percentile	0.224	0.111
Fraction of employment below 25th percentile	0.264	0.126
Interaction: fraction above 75th percentile with fraction below 25th percentile	0.048	0.015
No. of observations	1,837	

*Note:* Sample is pooled 1995–98 data for Business Register-based firms, defined as those with at least 50 percent of U.S. employment in the six analysis states matched to Economic Census and Compustat data. SD = standard deviation.

5.8).<sup>13</sup> Clearly these firms are more human capital intensive than the full sample—the proportion of the workforce above the median economywide threshold of skill (all measures) is greater, as is the proportion above the 75th percentile. The proportion below the 25th percentile, by contrast, is smaller. However, all measures still exhibit substantial heterogeneity: although the mean of each variable is different in the two samples, the standard deviations are very similar.

Tables 5.9 and 5.10 present the results of estimating equation (3), the

13. Because we use a log specification, we eliminate firms with missing or zero values. This procedure creates an even smaller sample than the Compustat matched sample described in the appendix. The excluded firms tend to have a greater fraction of skilled workers, with greater representation in the upper tail of both the person effect and experience components, and on average, the level of tenure of the workers in the excluded firms is above average.

**Table 5.9** The relation between market value and the complete human capital measure (analysis level: firm; dependent variable: log labor productivity)

Explanatory variable	All sectors		
	(A)	(B)	(C)
Fraction of workers above 50th percentile of human capital	0.732 (0.181)		
Fraction of workers above 75th percentile of human capital		0.579 (0.269)	0.554 (0.319)
Fraction of workers below 25th percentile of human capital		-0.414 (0.313)	-0.451 (0.385)
Interaction of above 75th and below 25th percentiles			0.343 (1.785)
Log capital	0.421 (0.026)	0.422 (0.026)	0.422 (0.026)
Log other assets	0.539 (0.026)	0.538 (0.026)	0.538 (0.026)
No. of observations	1,837	1,837	1,837
$R^2$	0.851	0.851	0.851

*Notes:* The human capital measure is  $h = \theta + \alpha\beta$ . Data were pooled for the years 1995–1998. The analysis sample is firms in six states (defined as the firms with at least 50 percent of U.S. employment in the analysis states) matched to Economic Census and Compustat data. All regressions include year effects, 2-digit SIC effects for the firm's primary industry, and indicators for multi-location status and whether the firm had establishments in 1, 2, or 3+ 2-digit SIC categories. Standard errors in parentheses.

(log) market value regressions, using our two sets of human capital measures.<sup>14</sup> In all specifications, we find a strong, positive relation between (log) market value and physical and other assets that is consistent with the theory and the empirical literature.<sup>15</sup> The value added by our analysis is that we can also measure human capital at the firm level. In our simplest specification (table 5.9, column [A]), a larger fraction of employees in the upper half of the human capital distribution is associated with significantly greater market value. This result is of interest because, if the compensation of highly skilled workers is proportionate to their skill, *and* no correlation exists between unmeasured (tangible or intangible) assets and human capital, market value should be unaffected once these other variables have been controlled. Yet the estimated effect of human capital is quite large: an increase of 1 standard deviation in the proportion of the workforce that is above average is associated with a change of approximately 14 log points in market value (set against a standard deviation of market value that is quite large—200 log points).

14. The reported results are based upon pooled data for 1995–98.

15. For this log-linear specification, the coefficients on a particular asset (e.g., log of physical capital) should reflect the share of that asset in the total.

**Table 5.10**      **The relation between market value and human capital, decomposed**  
**(analysis level: firm; dependent variable: log labor productivity)**

Explanatory variable	All sectors		
	(A)	(B)	(C)
Fraction of workers above 50th percentile for $\theta$	1.038 (0.242)		
Fraction of workers above 75th percentile for $\theta$		0.867 (0.370)	0.610 (0.385)
Fraction of workers below 25th percentile for $\theta$		-0.455 (0.392)	-1.249 (0.543)
Interaction of above 75th and below 25th percentiles for $\theta$		5.163	(2.061)
Fraction of workers above 50th percentile for $x\beta$	0.001 (0.292)		
Fraction of workers above 75th percentile for $x\beta$		-0.964 (0.467)	-0.327 (0.520)
Fraction of workers below 25th percentile for $x\beta$		-0.682 (0.422)	-0.251 (0.491)
Interaction of above 75th and below 25th percentiles for $x\beta$			-4.776 (2.501)
Log capital	0.424 (0.026)	0.428 (0.026)	0.428 (0.026)
Log other assets	0.529 (0.026)	0.524 (0.026)	0.521 (0.026)
No. of observations	1,837	1,837	1,837
$R^2$	0.854	0.855	0.857

Note: See table 5.9 notes.

When we estimate more-complex specifications, we find that the effects of different parts of the distribution of  $h$  on market value have some asymmetries that follow patterns similar to those found for productivity. That is, we find that the upper tail of the distribution of  $h$  has a disproportionately positive effect on market value and that the interaction effect of the lower and upper tails is positive. The positive interaction effect reinforces the asymmetry.

The decomposition of these results into person and experience effects is even more striking, however. In particular, all of the positive effect on market value is due to workers who have higher  $\theta$  (person effects; table 5.10, column [A]). A striking result is that, although both high- $\theta$  and highly experienced workers are more productive (table 5.5), only the person effect is positively related to market value.

When we examine the more detailed specifications using the upper and lower tails of the distributions of human capital, the results again show that high-person-effect firms are higher-market-value firms; but, again, asymmetries appear in the effects of different parts of the distribution. Column



(B) of table 5.10 shows that the upper quartile of the person effect has a disproportionate impact on market value. Column (C) shows that the interaction effect is positive. The disproportionate impact of the upper quartile found in column (B) is also present in column (C), which can be seen by combining the linear and interaction effects (evaluated at means). An increase in the upper quartile of the person effect yields a combined effect (linear plus interaction) of 1.374, while an increase in the lower quartile of the person effect yields a combined effect of  $-0.537$ .

We find it striking that the person effect is important in predicting market value. Recall that the person effect is the component that includes “unobservable” components of skill. Thus, one interpretation of these results is that value creation is highest for firms that do a better job of attracting and retaining workers with dimensions of skill that are difficult to observe. We also find it striking that the upper tail of the person effect matters disproportionately. Those firms that have the highest share of the “best and the brightest” workers (as measured by workers in the top quartile of the person effect) are those with the highest market value. Again, more research is necessary to determine whether these findings are due to the correlations between high-skill workers and unmeasured assets (that is, omitted variables).

## 5.7 Summary and Concluding Remarks

Finding and quantifying new measures of human capital that could be introduced into a firm-level production function is an important challenge for the federal statistical system, particularly given the advent of the new economy. This paper uses universe micro-level data on both employers and employees to create new measures that begin to address this challenge. Our new measures are closely related to those that have been developed in the existing literature, but the integration of the employer and employee data permits a much richer measurement approach to human capital.

We document the substantial consistency between our new measures and earlier measures pioneered by JGF (and subsequent closely related work). But we also extend the previous work in a variety of ways. First, our human capital measures encompass the traditional measures but also include effects not captured in traditional measures. Second, our integrated data permit these human capital measures to vary within and between firms in the same way that other inputs and outcomes can vary. Third, we examine different aspects of human capital—pure skill, experience, and a summary measure—and find marked differences in their distributions. We also use the richness of the data, especially the firm-level distribution of human capital, to capture relevant aspects of firm-level differences in organizational capital and workplace practices.

In the present analysis, we examine the relations between human capital, productivity, and market value. Not surprisingly, for measures of human capital and productivity, we find strong positive links in the microdata that differ according to the component of human capital used. The most-skilled workers (in terms of the estimated person effects) have a disproportionate positive impact on productivity, and the least-skilled workers (in terms of the estimated experience effects) have a disproportionate negative impact on productivity. We find that human capital is also related to market value even after controlling for total physical assets.

Interestingly, the component of skill with factors that are “unobservable” (at least to the econometrician) is most closely related to market value. At this stage of our analysis, we are unable to distinguish the observable from the unobservable components of skill. Future work might fruitfully explore this aspect of the analysis and results.

The new, micro-based measures of human capital incorporate unobservable dimensions of each worker’s skill. Our use of these measures to examine their role in accounting for variation across firms is clearly exploratory. The strong empirical relations that we have uncovered may reflect a variety of direct and indirect effects of human capital.

## Appendix

This section describes the construction of the data and key variables used for analysis. Many of the steps involved in the formation of the data are part of the efforts at the LEHD project and the Center for Economic Studies (CES) to develop data infrastructure; these organizations have detailed the steps in a number of technical documents. We provide here an overview of the data construction and refer to technical documents as needed.

The process of constructing the data can be broken into two large segments. The first segment is the formation of an “establishment-level” file used in the productivity analysis; the file contains human capital measures constructed from the LEHD database and business traits from Economic Census microdata. The second segment is the formation of a firm-level data set that uses an aggregated version of the establishment file matched to Compustat data for the years 1995–98.

### Building the Establishment-Level File

The process of building our approximation of an establishment-level file can be summarized as follows: we use the LEHD database to estimate human capital measures for each worker, and we use these measures along with a common economywide set of thresholds to generate variables

characterizing the distribution of human capital at each business unit. Next, using data from the 1997 Economic Census microdata, we add in measures of productivity for businesses in all sectors as well as capital intensity data for manufacturing businesses.

The LEHD database links worker identifiers and demographic detail to employer identifier variables. The construction of the LEHD database is described in detail in LEHD Program (2002). We begin by using the database to compute human capital measures for each worker using methods from AKM that are applied to the LEHD database in the manner described in ALM. The time span of the data used to estimate the person fixed effect varies across states; this paper uses human capital estimates drawn from a seven-state sample and uses matches of these estimates to establishment-level and firm-level data for six states for the years 1995–98.

Using the worker-specific human capital measures from ALM, we create variables summarizing the distribution of human capital at each business unit. We define a business unit by Employer Identification Number (EIN), two-digit Standard Industrial Classification (SIC) code, state, and year. This level of aggregation was selected because it is the smallest that is common to both the LEHD database and the Economic Census microdata. At this level of aggregation, most units are establishments, and thus we refer to this as an “establishment” file even though a multi-unit firm with multiple establishments in the same two-digit SIC state cell will have data aggregated across those establishments.

Although the LEHD data contain identifiers for all individuals employed by each business during a given year, not all workers contribute to our characterization of the human capital distribution at the establishment in that year. Rather, we identify workers who are imputed to have worked full time at any job during the year and include from that group only those who were working at the firm at the end of the first quarter of the year.<sup>16</sup> We restrict the set of jobs to those held by workers imputed to have worked full time because we do not observe information on hours worked, yet we wish to separate human capital effects from labor utilization effects. The timing restriction provides an approximation of employment at a point in time and corresponds closely to point-in-time employment as measured in the Economic Census, which collects employment for the pay period that includes March 12.

To evaluate the consequences of the restrictions on employment, we present summary statistics for alternative samples (table 5A.1). We further show descriptive statistics for variables created from the establishment-level data set that includes the full set of possible jobs—that is, all jobs held by workers imputed to have worked full time at any job in that year (col-

16. A worker employed at the end of the first quarter is characterized as having worked for that employer in both the first and second quarters.

**Table 5A.1** Summary statistics for alternative samples

	All full-time jobs (A)	All full-time jobs at end of first quarter (B)	All full-time jobs at end of first quarter with 5+ workers (C)
No. of establishments	1,391,839	1,078,877	429,529
No. of jobs	40,967,355	21,829,227	20,156,292
Fraction of firms in:			
Agriculture	0.020	0.014	0.012
Mining	0.008	0.009	0.009
Construction	0.072	0.063	0.060
Manufacturing	0.125	0.154	0.162
Public administration	0.026	0.037	0.039
FIRE	0.057	0.063	0.062
Retail trade	0.191	0.162	0.161
Services	0.380	0.360	0.356
TCE	0.057	0.063	0.065
Wholesale trade	0.064	0.074	0.073
Mean sales per employee	4.506	4.709	4.731
Mean capital intensity	4.124	4.215	4.229

*Notes:* Statistics are employment weighted. Analysis samples are based on 1997 unemployment insurance (UI)-based establishments for six states (defined as Employer Identification Number [EIN]/2-digit Standard Industrial Classification [SIC] units).

umn [A]). We also report the same set of measures with the added point-in-time restriction (column [B]). When the point-in-time restriction is imposed, the number of jobs falls nearly one-half (from almost 41 million jobs to about 22 million jobs), but the number of establishments declined at a far lower rate, from about 1.4 million to approximately 1.1 million. Establishments that are smaller and have higher worker turnover are at higher risk of being eliminated by this restriction. A comparison of the numbers across columns (A) and (B) suggests, however, that the two data sets vary only slightly from each other in terms of industry composition. Specifically, column (B) has a slightly higher share of manufacturing establishments (13 percent in column [A]; 15 percent in column [B]), a slightly lower share of retail (19 percent vs. 16 percent), and a slightly lower share of services (38 percent vs. 36 percent). Businesses eliminated through the point-in-time restriction appear to be less productive (lower sales per employee) and, among manufacturers, less capital intensive, although these differences are relatively small.

Many business units at this level of aggregation employ a small number of workers. Our ability to characterize the distribution of human capital at small businesses (here, those with fewer than five full-time employees) is limited because one worker can make an enormous difference. Moreover,

even for small to medium-sized businesses, the empirical distribution of human capital at a firm is quite noisy and potentially misleading. For example, the firm may have some positive probability of having workers in a particular percentile range but in fact have no workers in that range at a point in time. For these practical as well as conceptual reasons, we generate measures of the within-firm distributions based upon the kernel-density estimate (for  $h$  and for each of the components) for each firm (see ALM for details).

The use of this method involves trade-offs, however. The method allows us to measure the human capital at businesses more accurately, but the data requirements for the estimation procedure prevent us from using the smallest of firms in our analysis. Specifically, we generate a kernel-density estimate for each of the three human capital measures at each business that has at least five full-time workers at the end of the first quarter in a given year.

The size restriction eliminates less than 10 percent of jobs (about 1.6 million out of about 22 million; table 5A.1, columns [B] and [C]).<sup>17</sup> However, this restriction causes the number of business units to decline more than 60 percent (from about 1.1 million to less than 430,000). Summary statistics suggest, however, that although the firms lost through this restriction are small, they are spread evenly throughout all sectors of the economy and do not differ substantially in mean level of productivity or capital intensity.

We use a common set of thresholds to construct the establishment-level skill measures for the pooled distribution of all jobs held in all states currently in the LEHD database by workers imputed to have worked full time at the end of the first quarter of 1997. These thresholds are the median, the 75th percentile, and the 25th percentile value of each of the three human capital measures. We then calculate the cumulative density at each business between each of the thresholds to generate the proportions above the economywide median and economywide 75th percentile of  $h$  and the proportion below the economywide 25th percentile of  $h$ . We construct similar measures for  $\theta$  and  $x\beta$ .

The mean value of each human capital measure is sensitive to the set of sample restrictions imposed as well as to the method used to compute business-level skill (table 5A.2). Columns (A) through (C) correspond to columns (A) through (C) in table 5A.1 in terms of the restrictions imposed on the data. To generate the mean human capital measures reported in each of these three columns, we use the same set of thresholds described above along with the empirical distribution at each business unit rather than the smoothed estimate. Column (D) presents summary statistics derived from

17. Additional restrictions are imposed on the range of  $h$ ,  $\theta$ , and  $x\beta$  values included in these measures. If a worker has a value for  $h$  that is below 6 or above 14 or a value of  $\theta$  that is below -2 or above 2, that worker is excluded from the computation of the kernel-density estimate at the business. This restriction removes only the extreme outliers and results in a minor loss of both jobs and firms.

**Table 5A.2 Summary statistics on human capital distribution across alternative samples**

	All full-time jobs (A)	All full-time jobs at end of first quarter (B)	All full-time jobs at end of first quarter with 5+ workers (C)	All full-time jobs at end of first quarter with 5+ workers, using KDE (D)
No. of establishments	1,391,839	1,078,877	429,529	429,529
No. of jobs	40,967,355	21,829,227	20,156,292	20,156,292
Fraction of workers above 50th percentile for human capital	0.403	0.496	0.502	0.497
Fraction of workers above 75th percentile for human capital	0.187	0.246	0.245	0.248
Fraction of workers below 25th percentile for human capital	0.339	0.254	0.242	0.257
Interaction of above 75th and below 25th percentiles for human capital	0.039	0.038	0.038	0.043
Fraction of workers above 50th percentile for $\theta$	0.507	0.511	0.520	0.513
Fraction of workers above 75th percentile for $\theta$	0.254	0.256	0.258	0.261
Fraction of workers below 25th percentile for $\theta$	0.253	0.245	0.232	0.236
Interaction of above 75th and below 25th percentiles for $\theta$	0.051	0.046	0.046	0.049
Fraction of workers above 50th percentile for $x\beta$	0.421	0.490	0.481	0.473
Fraction of workers above 75th percentile for $x\beta$	0.205	0.248	0.236	0.229
Fraction of workers below 25th percentile for $x\beta$	0.335	0.260	0.266	0.266
Interaction of above 75th and below 25th percentiles for $x\beta$	0.052	0.049	0.049	0.049

*Notes:* Statistics are employment weighted. Analysis samples are based on 1997 UI-based establishments for six states (defined as EIN/2-digit SIC units). Columns (A), (B), and (C) are based on the empirical establishment-level distributions. Column (D) was computed using kernel-density estimates (KDE).

the smoothed kernel-density estimate of the distribution at each business. Thus, differing methods are used on the same set of workers and firms to generate the means in columns (C) and (D). In this way, we are able to isolate the effect of each sample restriction as well as the effect of using kernel-density estimates of the distribution of human capital at each business.

A disproportionate number of those workers eliminated through the point-in-time restriction are coming from the bottom quartile of the overall experience distribution, and businesses that are cut have a higher share of these low-experience workers (table 5A.2, columns [A] and [B]). Recalling that many of the eliminated establishments are retail or service-sector businesses, this fall in the share of low-experience workers is consistent with the change in industry composition. Both the full and restricted data sets appear to have a similar mean share of low- $\theta$  and high- $\theta$  workers. Thus, the effect of the restriction appears to be primarily on the experience component of human capital. In spite of the large number of establishments eliminated through the employer-size restriction, the smaller establishments that are eliminated apparently do not employ workers of systematically different skill levels (column [C]). Finally, columns (C) and (D) show mean human capital measures estimated by the two different methods using the same group of workers and firms. The similarities across columns suggest that use of the “smoothed” kernel-density skill distribution at each business does not notably change the average of the human capital distributions across businesses.

Tables 5A.1 and 5A.2 show that the between-business distribution of skill shares is only slightly sensitive to the point-in-time employment restriction; and it is not at all sensitive to the size restriction nor to the method used to characterize the skill distribution at each business. Table 5A.3 presents two sets of productivity regression results for EIN, two-digit SIC, state units in all sectors in 1997. The results in the left panel, obtained by using the human capital measures from table 5A.2, column (D) are identical to the results in table 5.5, columns (A) through (C). The right panel shows results obtained when we use the full data set of all jobs held by workers employed full time at any business in 1997 (the group described in column [A] in tables 5A.1 and 5A.2) and the empirical distribution of human capital to compute fractions of workers at different percentiles for each firm.

Overall, the key productivity finding, that skill is positively related to productivity, appears to hold up across restrictions imposed on the data as well as the methods used to build the human capital measures, with one exception. Relative to the full sample, the interaction of the highest and lowest quartiles of either the  $\theta$  or  $x\beta$  distribution at a business has a much stronger effect on productivity in the restricted sample using the smoothed distributions. This difference may arise in part because we have removed many of the smaller firms that are more likely to suffer from measurement

**Table 5A.3** Comparison of the full data set and the kernel density estimate (KDE) subset for the relation between labor productivity and human capital, decomposed (analysis level: establishments in all sectors; dependent variable: log labor productivity)

Explanatory variable	KDE data set			Full data set		
	(A)	(B)	(C)	(D)	(E)	(F)
Fraction of workers above 50th percentile for $\theta$	1.400 (0.009)			1.514 (0.005)		
Fraction of workers above 75th percentile for $\theta$		1.990 (0.015)	1.700 (0.017)		1.723 (0.008)	1.656 (0.009)
Fraction of workers below 25th percentile for $\theta$		-0.450 (0.010)	-0.920 (0.017)		-0.805 (0.006)	-0.891 (0.008)
Interaction of above 75th and below 25th percentiles for $\theta$			2.830 (0.078)			0.602 (0.037)
Fraction of workers above 50th percentile for $x\beta$	1.490 (0.120)			1.360 (0.006)		
Fraction of workers above 75th percentile for $x\beta$		0.200 (0.020)	0.760 (0.022)		0.388 (0.009)	0.488 (0.010)
Fraction of workers below 25th percentile for $x\beta$		-1.900 (0.017)	-1.560 (0.018)		-1.484 (0.007)	-1.441 (0.008)
Interaction of above 75th and below 25th percentiles for $x\beta$			-4.800 (0.090)			-0.713 (0.039)

*Notes:* The human capital measure is  $h = \theta + x\beta$ . The estimation sample is UI-based establishments (defined at the EIN/2-digit SIC level) for six states matched to the 1997 Economic Census and Annual Survey of Manufactures data. Standard errors in parentheses. Other controls include 2-digit industry effects. Results are based on employment-weighted regressions.

error (particularly for the tails of the distribution) in their human capital measures.

Two steps remain in the construction of the establishment-level files used for analysis. In each of the four years, we match the human capital measures described above to the Business Register (formerly the SSEL) to obtain information on business structure. This information is then used to build the firm-level data used in the market value analysis. In general, we find approximately 99 percent of the EINs from the LEHD database in the Business Register. The last step involved in building the establishment-level file is to aggregate 1997 Economic Census data on labor productivity and capital intensity to the EIN, two-digit SIC, state level and link these aggregates to the human capital data for 1997.

### Calculating Labor Productivity and Capital Intensity

We obtain establishment-level data from the 1997 Economic Census microdata. To form labor productivity for each EIN, two-digit SIC, state unit, we first sum employment on March 12 across all subunits with nonmissing sales revenue and positive employment. We then divide sales revenue at each subunit by this sum. Last, we calculate an employment-weighted



average of the subunits to aggregate to the EIN, SIC2, state level. We use a similar procedure for capital intensity.

Our objective is to maximize the number of observations in the human capital file for which we are able to obtain, from business data, some measure of labor productivity and, for manufacturers, capital intensity. In the majority of cases, we are able to link the two files by EIN, SIC2, and state. We are also able to incorporate business information at this same level of aggregation. However, some records in the human capital file and the business data file match by EIN and state but do not match by both EIN and SIC2. Rather than discarding these records, we instead apply EIN-level statewide measures to each of the EIN, SIC2, state observations in our matched file. We link to 354,549 units (274,043 EIN-SIC2-state matches and 80,506 EIN-state matches). Of these matches, we are able to construct a labor productivity measure for 337,495 units and a capital intensity measure for 33,926 manufacturers. The key variables constructed from the economic censuses are defined more formally below.

*Log labor productivity*: log of sales revenue per worker employed on March 12 at each EIN, two-digit SIC, state unit or each EIN state unit.

*Log capital intensity*: log of the capital stock per worker employed on March 12. The capital stock is measured as the book value of capital in the Census of Manufactures.

*Two-digit SIC*: modal two-digit SIC code of all reporting units under a state EIN (employment-weighted).

For the firm-level productivity analysis, we aggregate all of the variables (labor productivity, capital intensity, and the human capital variables) to the firm (enterprise level) using employment weights. The Economic Census files contain firm or enterprise identifiers that make this aggregation relatively straightforward. As noted in the main text, we retain only those firms who have 50 percent or more of their employment in the six states that are used for this analysis. Also, as noted in tables 5.6 and 5.7, we include controls for multi-unit status and diversification indicators to reveal whether the firm operates in more than one industry.

### **Building the Compustat-Matched File**

The Compustat database has two types of cases, those still being traded at the time the data are released and those that are no longer traded but were at some point since 1981. The Compustat documentation refers to “Active” and “Research” cases, but we refer to them as active and inactive cases here. Two types of matching procedures were used to identify links between the Compustat data and the Census Bureau’s Business Register. When possible, we used exact matching of EINs to identify the link. When that approach did not succeed (often because no EIN was available from

Compustat), we used business names, addresses, and industry codes to do probabilistic record linking.

Each establishment on the Business Register has an EIN associated with its payroll tax filings, while most stock issues in the Compustat database have an EIN from SEC filings. We carry out the EIN matching by first extracting a list of each unique combination of EIN and firm identifier from the Business Register and then matching each such combination to the unique EINs in the Compustat database. In some cases, a single EIN is associated with more than one Business Register firm or Compustat stock issue. Also, in a few cases, more than one EIN from the Compustat database is associated with the same firm on the Business Register. In some cases of duplication, we had a clear reason to think that one link should be preferred, and we dropped the other links.<sup>18</sup> Otherwise, we eliminated all the records involved. If an inactive Compustat case linked to several Business Register firm identifiers in several years, we used the Business Register identifier in the year closest to that in which the case became inactive as the link.

If we did not find an exact EIN match, we tried probabilistic record linkage using information on name, address, industry (SIC code), and EIN from the two databases. Stock issues from businesses that are based overseas account for a large portion of the cases for which we tried statistical linkage because they often do not have an EIN in the Compustat data. Overseas and inactive cases also generally do not have complete address information, so the statistical linkage is based on name, state, and industry for a large fraction of these cases. In this paper, we use the statistical links only for the active cases (plus EIN matches for both active and inactive cases) because of concern about the quality of links for the inactive cases.<sup>19</sup>

18. For example, some of the nonunique matches on the Business Register involve one business that appears to be active and one that appears to be inactive. To match as many cases as possible, we did not eliminate inactive establishments before matching but did so afterward if the match was not unique. Some of the nonunique Compustat matches involve cases that have a Compustat-assigned Committee on Uniform Security Identification Procedure (CUSIP), which often carries alternative versions of data for companies that also have a standard CUSIP. For example, if an acquisition took place in 1999, the data we have might contain two records for the acquiring company: one with a standard CUSIP that has data reflecting the company's holdings in each year, and another with a Compustat-assigned CUSIP that has consolidated data for the two businesses for some years before the merger. In this case, we would drop the record with the Compustat-assigned CUSIP (and consolidated data) and keep the record with the standard CUSIP.

19. Identifying the appropriate link was more complicated for the inactive cases both because of a lack of detailed address information and because it was not clear at the outset which years we should use for statistical linking. Because the statistical linking was quite time intensive, we did not try to match all cases to all available years of the Business Register but rather tried first with earlier years and then worked forward if a match was not found. After having identified links using that approach, we compared years in which the identified firm was active on the Business Register to the years with nonmissing data from Compustat. In a significant fraction of cases, the years did not line up, and that problem was the main reason we decided not to use the 346 statistically linked inactive cases.

**Table 5A.4** Summary statistics on Compustat file for alternative samples

Compustat status	Total (March 2001 Compustat database)	Deleted from Compustat before 1996	Deleted due to duplication	No SSEL link	Exact EIN link	Name/ address/ SIC link	
						Had EIN	No EIN
Active	9,885	0	569	1,457	6,905	532	422
Inactive	11,151	6,722	300	818	3,311	0	0
Total	21,036	6,722	869	2,275	10,216	532	422

*Note:* SSEL = Census Business Register, previously known as the Standard Statistical Establishment List.

Of the 14,312 Compustat cases that were traded at some point after 1995, we found a unique link for 11,170 cases (table 5A.4). However, some of these cases were linked to Business Register firms that were inactive in the years of interest or were missing essential Compustat data. We restricted the sample to cases that link to at least one establishment on the Business Register between 1996 and 1998 (the years for which we have LEHD estimates) and that have a price reported in Compustat in at least one year in that range; the result was 9,917 cases.

The sample for use in conjunction with our human capital estimates is limited to businesses that have some employment in the six states for which we have LEHD estimates. We calculated sample sizes for this six-state subsample as a function of the share of employment we require in those states (table 5A.5).

Table 5A.6 gives evidence on how the linking process affects the composition of the sample. The first column gives means for the full Compustat sample using pooled data from 1995–98 period for all Compustat cases that have nonmissing data on sales and positive data on employment. The second column takes the subset of those cases that were uniquely linked to a Business Register firm, and the third column further restricts the sample to those with at least 50 percent of Business Register employment in the six LEHD states.<sup>20</sup>

All three of these samples consist of firms that are very large simply because of the restriction to publicly traded firms in Compustat. The matched samples have somewhat smaller firms on average than the full Compustat sample. Very large, complex firms may be more likely to be

20. The sample means in table 5.8 are based on a sample with nonmissing data for a larger set of variables and an additional link to the LEHD data. The final regression sample is smaller than that in table 5A.3 and has slightly smaller Compustat employment on average (table 5.8).

**Table 5A.5** Number of firms in Compustat/Standard Statistical Establishment List database with employment in sample states

Year	Required employment share in LEHD states			
	Any	>50%	>90%	All
1995	3,529	1,254	828	725
1996	3,578	1,283	840	757
1997	3,851	1,397	902	800
1998	3,903	1,370	871	793
In at least 1 year	4,841	1,942	1,294	1,170

Note: LEHD = Longitudinal Employer-Household Dynamics Program.

**Table 5A.6** Sample means for full and matched samples

	Full Compustat	Matched Compustat/SSEL	Matched, with at least 50% employment in LEHD states
No. of firms	32,613	22,911	4,367
Market value	\$2,149 million	\$1,631 million	\$1,036
Compustat employment	6,281	5,068	2,925
SSEL employment		3,497	1,878
Sales, net	\$1,320 million	\$1,028 million	\$684 million
Compustat labor productivity (sales/employment)	\$363,000	\$316,000	\$358,000

dropped in linking the two databases because of problems with apparent duplication or because multinationals may not have an EIN in the database. Very large companies are also more likely to have employment spread across many states. Thus, they are less likely to be included in the third column of table 5A.6.

Differences exist across the two data sources for employment, in part, probably, because of the inclusion of overseas employees in the Compustat figures (the Business Register data include only U.S. employees). However, detailed examination of the microdata suggests that differences exist even for firms that operate only in the United States, and the sources of those discrepancies are not clear. Possible candidates include differences in the definition of employee or differences in the dating of employment (the Business Register data reflect employment as of the week containing March 12, whereas the Compustat employment numbers do not refer to a particular date).

For the Compustat variables, we follow the measurement methodology in the literature (e.g., Hall 1990, 1998; and BHY). We use the following primary Compustat variables.

*Market Value*: value of common stock at the end of the fiscal year plus preferred stock value plus total debt. In Compustat mnemonics, it is  $MKVALF + PSTK + DT$ .<sup>21</sup>

*Physical Capital*: gross book value of capital stock is deflated by the gross domestic product (GDP) implicit price deflator for fixed investment. The deflator is applied at the calculated average age of the capital stock on the basis of the three-year average of the ratio of total accumulated depreciation to current depreciation.

*Other Assets*: total assets minus the book value of physical capital. This item includes receivables, inventories, cash, and other accounting assets such as goodwill reported by companies.

## References

- Aaronson, Daniel, and Daniel Sullivan. 2001. Growth in worker quality. *Economic Perspectives* 25 (4): 53–74.
- Abowd, John M. 1989. The effect of wage bargains on the stock market value of the firm. *American Economic Review* 79 (4): 774–800.
- . 1990. Does performance-based managerial compensation affect corporate performance. *Industrial and Labor Relations Review* 43 (3): 52S–73S.
- Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. Computing person and firm effects using linked longitudinal employer-employee data. Longitudinal Employer-Household Dynamics Technical Paper no. TP-2002-06. Washington, DC: U.S. Census Bureau.
- Abowd, John M., John Haltiwanger, Julia Lane, and Kristin Sandusky. 2001. Within and between firm changes in human capital, technology, and productivity. Longitudinal Employer-Household Dynamics Technical Paper no. TP-2001-03. Washington, DC: U.S. Census Bureau.
- Abowd, John M., Francis Kramarz, and David Margolis. 1999. High wage workers and high wage firms. *Econometrica* 67 (2): 251–334.
- Abowd, John M., Paul Lengermann, and Kevin McKinney. 2003. The measurement of human capital in the U.S. economy. Longitudinal Employer-Household Dynamic Technical Paper no. TP-2002-09. Washington, DC: U.S. Census Bureau.
- Acemoglu, Daron. 2001. Technical change, inequality and the labor market. *Journal of Economic Literature* 40 (1): 7–72.
- Becker, Gary S. 1964. *Human capital: A theoretical and empirical analysis, with special reference to education*. New York: Columbia University Press.
- Bond, Stephen R., and Jason G. Cummins. 2000. The stock market and investment in the new economy: Some tangible facts and intangible fictions. *Brookings Papers on Economic Activity*, Issue no. 1:61–124.

21. Some differences in measurement methodology appear in the literature. Our method most closely follows that of BHY. One difference is that Hall (1990) suggests adjusting the value of long-term debt for differences in the age structure of the debt. However, in the absence of firm-level information on the age structure of the debt, Hall assumes all long-term debt has a twenty-year maturity. One can readily develop scenarios in which this assumption induces substantial measurement error.

- Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. 2002. Information technology, workplace organization and the demand for skilled labor: Firm-level evidence. *Quarterly Journal of Economics* 117 (1): 339–76.
- Brynjolfsson, Erik, Lorin M. Hitt, and Shinkyu Yang. 2002. Intangible assets: How the interaction of computers and organizational structure affects stock market valuations. *Brookings Papers on Economic Activity: Macroeconomics*, Issue no. 1:137–99.
- Burgess, Simon, Julia Lane, and David Stevens. 2000. Job flows, worker flows and churning. *Journal of Labor Economics* 18 (3): 473–502.
- Foster, Lucia, John Haltiwanger, and C. J. Krizan. 2001. Aggregate productivity growth: Lessons from microeconomic evidence. In *New developments in productivity analysis*, ed. Charles Hulten, Edward Dean, and Michael Harper, 303–63. Chicago: University of Chicago Press.
- Gore, Al. 1997. Remarks by vice president of the United States. Speech at the Microsoft CEO Summit. 8 May, Seattle, Washington.
- Groschen, Erica L. 1991a. Five reasons why wages vary among employers. *Industrial Relations* 30 (3): 350–81.
- . 1991b. Sources of intra-industry wage dispersion: How much do employers matter? *The Quarterly Journal of Economics* 106 (3): 869–84.
- Gu, Feng, and Baruch Lev. 2001. Markets in intangibles: Patent licensing. New York University Working Paper.
- Hall, Bronwyn. 1990. The manufacturing sector master file: 1959–1987. NBER Working Paper no. 3366. Cambridge, MA: National Bureau of Economic Research.
- Hall, Bronwyn. 1998. Innovation and market value. Paper presented at the National Institute of Economic and Social Research Conference on Competitiveness and Productivity, London, England, February 1998. In *Productivity, Innovation and Economic Performance*, ed. Ray Barrell, Geoffrey Mason, and Mary O'Mahoney, 177–98. Cambridge: Cambridge University Press, 2000.
- Hall, Robert. 2001. Struggling to understand the stock market. *American Economic Review Papers and Proceedings* 91 (2): 1–11.
- Haltiwanger, John, and Ron Jarmin. 2000. Measuring the digital economy. In *Understanding the digital economy*, ed. E. Brynjolfsson and B. Kahin, 13–33. Cambridge: MIT Press.
- Haltiwanger, John C., Julia I. Lane, and James R. Spletzer. 2001. Productivity differences across employers: The role of employer size, age, and human capital. University of Maryland, Urban Institute, and Bureau of Labor Statistics. Working Paper.
- Jorgenson, Dale W., F. Gollop, and B. M. Fraumeni. 1987. *Productivity and US economic growth*. Cambridge, MA: Harvard University Press.
- Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101 (3): 410–42.
- Kremer, Michael, and Eric Maskin. 1996. Wage inequality and segregation by skill. NBER Working Paper no. 5718. Cambridge, MA: National Bureau of Economic Research.
- Lane, Julia, Laurie Salmon, and James Spletzer. 2002. Establishment wage differentials. Washington, DC: The Urban Institute. Mimeograph.
- Lengermann, Paul. 2002a. Comparing new and traditional measures of human capital. University of Maryland, Department of Economics. Manuscript.
- . 2002b. *Is it who you are, where you work, or with whom you work? Reassessing the relationship between skill segregation and wage inequality*. Ph.D. Diss. College Park, MD: University of Maryland.

Longitudinal Employer-Household Dynamics (LEHD) Program. 2002. Longitudinal employer-household dynamics program: Employment dynamics estimates, versions 2.2 and 2.3. Longitudinal Employer-Household Dynamics Technical Paper no. TP-2002-05. Washington, DC: U.S. Census Bureau.

## Comment Katharine G. Abraham

Understanding how human capital affects output, productivity, and firm performance is key to understanding today's knowledge economy. Although observable human capital measures such as years of education and years of experience help to explain the systematic variation in earnings across individuals, they also leave a great deal unexplained. Abowd and coauthors use a rich longitudinal data set containing information on the universe of workers and establishments in six states to identify persistent differences in earnings across individuals that cannot be explained by observable characteristics or by fixed employer effects. Together with the estimated value of a worker's labor market experience, the individual-specific component of the earnings residual is treated as a measure of human capital. Both components of human capital—along with measures of the way that people with different kinds and amounts of human capital are grouped within establishments and firms—turn out to be related in interesting ways both to productivity and to the market value of enterprises.

Abowd and coauthors take pains to describe their work as an exploratory analysis, and the approach they have taken certainly strikes me as sensible. It nonetheless seems worthwhile to think a bit about the underpinnings of the measurement strategy the authors have adopted. Start with the wage equation that provides the coefficient estimates used to construct their human capital variables:

$$(1) \quad w_{ijt} = \theta_i + \mathbf{x}_{it}\beta + \psi_{J(i,t)} + \varepsilon_{ijt}$$

In this equation,  $w_{ijt}$  is the log of quarterly earnings for person  $i$  employed by firm  $j$  at time  $t$ ,  $\theta_i$  is a fixed person effect,  $\mathbf{x}_{it}$  is a vector of time-varying individual variables that capture labor market experience, and  $\psi_{J(i,t)}$  is the fixed firm effect for whatever firm  $j$  employs person  $i$  at time  $t$ . Differences in wage rates across individuals captured by  $\theta_i$  are assumed to reflect corresponding differences in productivity. Compared to the variables included in the typical human capital wage equation, the  $\theta_i$ s have the potential to capture far more of the individual variation in human capital, though with-

Katharine G. Abraham is professor of survey methodology and affiliate professor of economics with the Joint Program for Survey Methodology at the University of Maryland, and a research associate of the National Bureau of Economic Research.

out providing much insight into the specific attributes responsible for this variation.

One concern with the empirical implementation of equation (1) lies with the use of quarterly earnings information as a basis for drawing conclusions about the distribution of human capital. Available evidence suggests that those with higher earnings capacities tend to work longer hours (see, for example, Rones, Ilg, and Gardner 1997). If quarterly earnings are higher for more productive workers not only because their hourly wages are higher but also because they work more hours, the estimated  $\theta_i$  will reflect both the higher quality and the greater quantity of these workers' labor services.

Given the "black box" nature of the estimated  $\theta_i$ , one might also wonder whether factors other than differences in individuals' human capital that could affect wages have been controlled for properly. A strength of equation (1) is the inclusion of establishment effects as a determinant of observed wage rates. The existence of employer effects on wages has been well documented (see, for example, the seminal work by Blau 1977 and Groshen 1991), and standard human capital equations that relate wages to individual characteristics but ignore employer effects likely produce biased coefficient estimates. But high-wage employers may not offer equally large wage premiums to all groups of workers, and this complicates the interpretation of the wage equation coefficients the authors have estimated.

To illustrate, consider a simple hypothetical economy consisting of equal numbers of type 1 and type 2 firms, each of which employs equal numbers of low-skilled and high-skilled workers. Type 1 firms pay no wage premiums; type 2 firms pay no wage premiums to low-skilled workers, but wage premiums of 10 percent to high-skilled workers. Workers move back and forth between the two types of firms, spending half their careers with each. If data for these firms and workers were used to estimate a wage equation like equation (1), the estimated coefficients would imply firm effects for type 2 firms of 5 percent rather than the zero percent that would be estimated using data for low-skilled workers or the 10 percent that would be estimated using data for high-skilled workers. And the estimated  $\theta_i$ s would imply person effects on wages for high-skilled workers that were 5 percentage points larger than would be estimated in a model that allowed for firm effects that varied by worker skill level.

There is a limited body of empirical work that has explored employers' internal wage structures—and specifically the question of whether employer wage premiums differ across groups of workers—using data sets that contain establishment-level observations on pay by occupation. Data of this sort can be used to estimate models that allow for occupation, establishment, and the interaction between the two to affect wages:

$$(2) \quad w_{jk} = \alpha_k O_k + \omega_j E_j + \gamma_{jk} O_k E_j.$$



In this equation,  $w_{jk}$  is the log of the average wage rate for workers in occupation  $k$  at establishment  $j$ ,  $\mathbf{O}_k$  is a vector of occupational dummies, and  $\mathbf{E}_j$  is a vector of establishment dummies. If establishment wage premiums were completely uniform across occupations, all of the  $\gamma_{jk}$ s would be equal to zero; if they sometimes vary by occupation, some of the  $\gamma_{jk}$ s will be non-zero and the job cell (occupation by establishment) effects on wages will explain some portion of wage variation over and above the portion explained by the occupation and establishment dummies.

Lane, Salmon, and Spletzer (2001) use data from the Bureau of Labor Statistics Occupational Employment Survey to fit a model of this general sort. As noted by Abowd and coauthors, the Lane, Salmon, and Spletzer findings document a propensity for establishments that pay high wages to their accountants, for example, to pay high wages to their janitors too. But the findings also show variation in  $\ln(\text{wages})$  associated with differences in the size of establishment wage effects by occupation that are roughly two-thirds as large as the variation in  $\ln(\text{wages})$  attributable to the establishment main effects.<sup>1</sup> Using National Compensation Survey data on average earnings by detailed occupation within establishments, Kaplan and Pierce (2000) find a correlation of just 0.33 between  $\ln(\text{average wages})$  for pairs of detailed occupations within the same establishment. Groshen (1991) analyzes Industry Wage Survey data and finds that, in her sample of six manufacturing industries, occupation-establishment interactions explain an average of only about 10 percent of the variation in observed individual wages, as contrasted with an average of about 30 percent for the establishment main effect. But Groshen and Levine (1998), using data from an employer wage survey conducted by the Federal Reserve Bank of Cleveland that covers establishments in a broader array of industries but may be less representative in other ways, find that in recent years the standard deviation of the  $\ln(\text{wage})$  component associated with occupation-establishment interactions is roughly equal in magnitude to that associated with the establishment main effects.

The evidence just discussed does not necessarily imply that differences in internal wage structures across establishments are a source of bias in the estimated  $\theta_j$ s. Indeed, the occupation-establishment interaction effects documented in the research I have described may reflect the characteristics of the individuals who hold jobs in particular occupations within particular establishments rather than true internal wage structure effects, and the available empirical evidence does not speak to the question of whether certain groups are systematically more likely to receive wage premiums than others. But there is also reason to think that employers may pursue wage

1. For this purpose, the variance attributable to the establishment main effects is measured as the incremental variance captured by the addition of establishment dummies to a model that already contains occupation dummies.

policies that are more complex than simply paying uniformly high or uniformly low wages to all of their employees. Additional empirical research on firms' pay-setting policies and practices could help to resolve the proper interpretation of human capital estimates derived from models like equation (1).

Using their human capital measures, Abowd and coauthors estimate equations with measures of productivity and market value on the left-hand side and the human capital variables on the right-hand side. The assumption in these models is that firms choose the mix of human capital they will employ, which in turn affects their output and profitability. It seems likely, however, that the mix of human capital a firm employs is at least somewhat constrained by its history and the history of the industry to which it belongs. One clear illustration of this appears in the paper, which draws a contrast between the higher years of worker experience in the manufacturing sector as compared to the finance, insurance, and real estate sector. Given the very different histories of these two industries, I am skeptical that these differences represent a true long-run equilibrium. Manufacturing employment has been stagnant or shrinking for many years, whereas employment in the finance, insurance, and real estate sector generally has trended upward. These different employment trajectories seem very likely to have affected the age and experience distribution of the two industries' workforces. It would not be surprising if, thirty years from now, the distribution of age and experience within a more mature finance, insurance, and real estate sector looked quite different from the distribution within the industry today.

There is also a question, of course, about the direction of causality that underlies the observed relationship between the human capital a firm employs, on the one hand, and its output and profitability, on the other. The authors are interested in how human capital affects output and profitability, but there may be causality running in the opposite direction, with managers at highly productive or highly profitable firms being freer to pay the high wages that make it easier to attract higher-caliber workers. Most managers would prefer to hire more capable employees; put simply, having more capable employees makes life easier for those in charge! Managers who work for some firms may be better able than those who work for others to indulge their preferences in this regard. I am not certain how one would sort this out, but it seems worth thinking about.

One notable pattern in the empirical findings is the very strong association that employing more high- $\theta_i$  workers has with both productivity and market valuation. In table 5.5, for example, the column (B) coefficients imply that increases in the fraction of the workforce with  $\theta_i$ s above the 75th percentile are associated with increases in labor productivity four times as large as those associated with decreases in the fraction of workers with  $\theta_i$ s below the 25th percentile. And in table 5.10, the column (B) coefficients

imply that increases in the fraction of the workforce with  $\theta_i$ s above the 75th percentile are associated with increases in a firm's market valuation that are nearly twice as large as those associated with declines in the fraction of the workforce with  $\theta_i$ s below the 25th percentile. Because quarterly earnings data were used to estimate the  $\theta_i$ , the large implied effect of changing the concentration of high- $\theta_i$  workers may reflect, at least in part, differences in the quantity of the labor services provided by high- as compared to low-productivity workers. But to the extent that the observed association is robust to taking differences in work hours into account, it certainly merits further exploration.

One of the more innovative features of the authors' analysis is the use of variables designed to capture not just the average level but also the mix of different kinds of human capital to explain cross-establishment differences in productivity and cross-firm differences in market valuation. The pattern of the coefficients in column (C) of table 5.5, for example, implies that establishments with a larger fraction of workers with  $\theta_i$ s above the 75th percentile have higher labor productivity, while those with a higher fraction of workers with  $\theta_i$ s below the 25th percentile have lower labor productivity. But the positive coefficient on the interaction between these two variables suggests that a worker's contribution to establishment productivity may not be independent of who his or her coworkers are. Future research may be able to help with making the lessons to be learned about how and why the mix of human capital matters more concrete. It might be possible, for example, to identify establishments within selected industries whose employment is highly skewed toward high- $\theta_i$  workers, establishments whose employment is skewed toward low- $\theta_i$  workers, and establishments that employ a mix. One then could look for differences in the way these establishments operate that could shed light on the sources of the differences in their productivity.

This paper clearly is not the end of this line of research. It is to be hoped that future work, both by these authors and by others, will further explore the validity of the assumptions on which the analysis is based, the robustness of the findings, and the real-world underpinnings of the patterns observed.

## References

- Blau, Francine D. 1977. *Equal pay in the office*. Lexington, MA: D.C. Heath and Company.
- Groshen, Erica L. 1991. Sources of intra-industry wage dispersion: How much do employers matter? *Quarterly Journal of Economics* 106 (3): 869–84.
- Groshen, Erica L., and David I. Levine. 1998. The rise and decline (?) of U.S. internal labor markets. Research Paper no. 9819. New York: Federal Reserve Bank of New York.
- Kaplan, David S., and Brooks Pierce. 2000. Firm-wide versus establishment-specific

- pay practices. Paper presented at Labor Studies Program meeting, National Bureau of Economic Research. 31 July–2 August, Cambridge, Massachusetts. Available at <http://www.nber.org/~confer/2000/si2000/kaplan.pdf>.
- Lane, Julia, Laurie Salmon, and James R. Spletzer. 2001. Earnings determination: Contributions from the occupation and the establishment. Paper presented at the sixth annual meeting of the Society of Labor Economists. 20–21 April, Austin, Texas. Available at <http://gsbwww.uchicago.edu/research/workshops/wae/lanesole.pdf>.
- Rones, Phillip L., Randy E. Ilg, and Jennifer M. Gardner. 1997. Trends in hours of work since the mid-1970s. *Monthly Labor Review* 120 (4): 3–14.

