December 6, 2002

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

John M. Abowd (Cornell University, Census Bureau, CREST, NBER, and IZA),
John Haltiwanger (University of Maryland, Census Bureau and NBER),
Ron Jarmin (Census Bureau),
Julia Lane (Urban Institute and Census Bureau),
Paul Lengermann (Federal Reserve Board),
Kristin McCue (Census Bureau),
Kevin McKinney (Census Bureau), and
Kristin Sandusky (Census Bureau)

*This paper was prepared for the NBER/CRIW conference on "Measuring Capital for the New Economy" in Washington, D.C. in April 2002. The authors wish to acknowledge the substantial contributions of the LEHD Program staff 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 Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging, 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; please contact Ron Prevost, Director, LEHD, Ronald.C.Prevost@census.gov.

Abstract

This paper investigates and evaluates the direct and indirect contribution of human capital to business productivity and shareholder value. The impact of human capital may occur in two ways: the specific knowledge of workers at businesses may directly increase business performance, or a skilled workforce may also indirectly act as a complement to improved technologies, business models or organizational practices. We use newly created firm-level measures of workforce human capital and productivity to examine links between those measures and the market value of the employing firm. The new human capital measures come from an integrated employer-employee data base under development at the US Census Bureau. We link these data to financial information from Compustat at the firm level, which provides measures of market value and tangible assets. The combination of these two sources permits examination of the link between human capital, productivity, and market value. There is a substantial positive relation between human capital and market value that is primarily related to the unmeasured personal characteristics of the employees, which are captured by the new measures.

1. Introduction

The measurement of intangibles and human capital, important for both goods-producing and service-producing industries, has always been a difficult challenge for the statistical system. The growth of the New Economy has made responding to this challenge even more urgent: understanding how such inputs affect the value chain of productivity, growth and firm value now surpasses the need to measure the impact of bricks, mortar and equipment. Yet the changes that have brought the New Economy into existence have, at the same time, highlighted the need for improvements to traditional measures of inputs and outputs (Haltiwanger and Jarmin, 2000). This is particularly true for human capital. Finding new measures of human capital, and quantifying them in such a manner that they can be introduced into a production function and produced on a scale that provides sufficient sample size for use in official economic statistics is a formidable challenge.

This paper uses micro level data 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, and quantify the value of individual specific skills, such as innate ability, visual or spatial skills, non-algorithmic reasoning, analytic or abstract decision-making, or "people skills" (Bresnahan, *et al.* 1999).

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 there are large differences across many dimensions of firm inputs and outcomes. In particular, there is little uniformity among employers in either the methods used to hire and terminate workers or the selection of types of workers to 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 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.

2. Background, Motivation, and Specifications

The literature on human capital and intangibles separately and together are quite broad ranging and impossible to summarize here. However, we 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 focusing on its potential relationships to productivity, market value, tangible and intangible assets.

a) Human capita: 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). While we clearly stand on their shoulders, our approach is different in key ways that depend critically on data availability. In particular, our conceptual and measurement approach depends not only on the availability of longitudinal matched employer-employee data but the availability of universe files of all workers and firms.

The starting point for our approach has been well documented and investigated in a series of papers by Abowd, Kramarz and Margolis (1999, hereafter AKM) and Abowd, Lengermann, and McKinney (2002, 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 Census called the Longitudinal Employer-Household Dynamics Program (LEHD). These and related papers emphasize a point that has long been known in the study of human capital – it is very difficult to measure human capital directly. 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. JGF create detailed human capital measures from person level data in the U.S., primarily using the Current Population Survey (CPS), by exploiting wage differences across gender, experience and education groups. They aggregate these measures by industry and to the total economy level and have demonstrated that there is an enormous stock of human capital in the U.S. economy and that the stock and flows of this asset are vitally important for understanding labor productivity changes.

However, JGF (and subsequent related work including the Jorgenson, Ho, and Stiroh paper in this volume) recognize that this approach is constrained. Clearly, industry and economy-wide aggregates fail to capture the firm level variation that is a driving force in productivity growth. In addition, the existing data only provide a relatively small set of observable characteristics of workers, resulting in measurement problems, as well as the omission of measures of unobservable skill and confounding firm effects. Using the educational attainment of a college degree as a human capital measure, 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 human capital measurement, as emphasized in the recent literature on rising wage inequality in the U.S. (See, *e.g.,* Juhn, Murphy, and Pierce,1992). Finally, new empirical evidence (AKM) note that because earnings measures include the returns to working with particular of firms – for example, large, highly unionized or profitable entities – and there is sorting among workers and firms, the estimates of returns to measured and unobservable components of human capital may be biased.

While we will provide a more detailed description of the econometric and measurement approach in subsequent sections, it is useful to review the basic specification used by AKM and ALM so that we can discuss the conceptual nature of our human capital measures. The core statistical model is:

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

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 is the firm effect and the fourth component is the statistical residual, orthogonal to all other effects in the model. In what follows, we use the person effect θ plus the experience component of $x\beta$ as the core measure of human capital, called *"h"* (*i.e.,* $h_{it} = \theta_{it} + x_{it}\beta$).[1] 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- 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 place all of these workers inside their employers, we can consider the full human capital distribution 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.[2] Because 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 firm personnel policies that may impact the returns to observable and unobservable dimensions of skill.

There are, of course, limitations of our approach. For one, the estimation method provides time-invariant person and firm effects. Both theory and evidence suggests that the returns to different dimensions to skill may be time varying and there may also be time varying firm effects. Permitting time variation in the person and firm effects (*e.g.,* through estimating a mixed-effects or Bayesian model) is active area of ongoing research in the LEHD program. For now, the interpretation is that we have a time average of the relevant effects. Second, the specification does not permit any interaction between the firm and person effect. The implicit assumption in this specification is that firms pay the same premium (or discount) regardless of the type of worker. Again, this is an area for future work but in this case there is some evidence that supports this assumption. For example, Groshen (1991a and 1991b) finds that establishment wage differentials exist across occupations within establishments, and has recently been updated and confirmed by Lane, Salmon and Spletzer (2002). This research, which uses recent data from BLS's Occupational Employment Statistics survey shows that firms that pay premiums to their

---

[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 out the impact of observable characteristics such as education and unobservable components. For sub-samples of our universe files, we can measure education and some of the results (see, *e.g,.* in ALM) we refer to 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.

accountants also pay premiums to their janitors. Third, in a like manner this specification does not permit any coworker effects. Lengermann (2002a) has shown that "who you work with" matters as well as who you are and for whom you work. Despite these concerns, we regard this base specification as representing 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 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 thinking about how and why human capital might matter for productivity and market value.

b) 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:

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

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.[3]

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, there are a number of theoretical underpinnings for studying more than just the mean of the within firm-distribution of human capital. The relation between skill and productivity at the *individual* level might be nonlinear. As an example, it is commonly assumed (and the empirical results support the assumption) that the earnings-experience profile for an individual worker is strictly concave reflecting an underlying concave productivity-experience profile. Alternatively, some dimensions of skill may yield a convex productivity-skill profile. For example, the role of "superstar" workers/managers (*e.g.,* Bill Gates) in contributing to firm productivity and value may be disproportionate. A related idea is that there may be a strictly convex relation between productivity and, say, innate ability at the individual level. For our purposes, any form of nonlinear relation between the productivity of a worker and the worker's skill level implies that higher moments of the within-firm human capital distribution will account for some productivity variation at the firm level. For example, if the productivity-skill relation at the individual level on some dimension (*e.g.,* experience) is strictly concave, then firms with greater dispersion on that dimension will have lower firm-level productivity other things equal. In contrast, if the productivity-skill relationship at the individual

[3] One interesting issue that we do not explore in this paper but plan on in future work is the relationship of the firm effects to productivity and market value. Firm effects capture potentially many factors – rent sharing, firm personnel policies, efficiency wages, and/or the impact of collective bargaining. Some of these effects may be positively related to productivity and/or market value and others negatively related. Identifying and exploring these different effects is a rich potential area for future work. For more discussion of the issues and related empirical work exploring these alternative effects, see Abowd (1989, 1990).

worker level is strictly convex, then firms with greater dispersion on that dimension of skill will have higher firm-level productivity.

These examples are only one of several possible 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 (1995) suggest that in some production environments that there are interaction effects across the skills of co-workers 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 distinguish between these alternative and interesting reasons for the within-firm distribution of human capital to matter for 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 productivity, market value and human capital relationships.

Beyond these interesting issues about the role of the within-firm distribution of human capital, the empirical link between human capital and productivity may be further complicated by unmeasured (*i.e.,* omitted) tangible assets and intangible assets. If there are complementarities between unmeasured assets (or poorly measured assets) and human capital, an estimated relation between productivity and human capital may reflect such complementarities. This perspective itself is too limited since what we mean by intangible assets may be very closely connected to how human capital is organized.

c) Defining and measuring intangibles

A major issue confronting the productivity literature is the problem of studying the effects of intangible assets when they are not readily measurable. Thus, while some might argue that intangibles are the "major drivers of corporate value and growth in most sectors" (Gu and Lev 2001), there is little consensus as to what those intangibles are. In fact, they have been variously defined to be knowledge and intellectual assets (Gu and Lev, 2001), human capital, intellectual property, brainpower and heart (Gore, 1987), knowledge assets and innovation (Hall, 1998), and organizational structure (Brynjolfsson, Hitt and Yang, 2001). Measures of these variables have been equally diverse, ranging from a residual approach, to inference and, yet further, to direct measurement.

For example, while Gu and Lev conceptualize intangibles as knowledge assets (new discoveries, brands or organizational designs), they derive their measure of intangibles as a residual: the driver of economic performance after accounting for the contribution of physical and financial assets. In empirical terms, they identify the core drivers 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. Brynjolfsson, Hitt and Yang (2001), 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, using 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, the highest growth rates are in consulting, machinery and electronics industries; the lowest in retail trade, restaurants and primary metals. Gu and Lev also find that intangible-driven earnings (by two different measures) are much more highly correlated with stock market returns than are other measures, notably operating cash flow growth and earnings growth. Brynjolffson *et al*. find that organizational structure has a large impact on market valuation: firms that score one standard deviation higher than the mean on this measure have approximately $500 million greater market value. Hall finds that research and development accounts for a "reasonable fraction" of the variance of market value, but that this relation is not stable, and that there is still a great deal of unexplained variation. Patents matter, according to Hall, but less than research and development.

Empirical studies also suggests that failing to include intangibles is likely to cause considerable bias in estimates of the impact of tangibles on both market value and output. Gu and Lev find that expenditures on capital, R&D, and technology acquisitions are all highly correlated with intangible capital. Similarly, Brynjolffson *et al*. find evidence for a strong correlation between organizational structure and investment in information technology.

Even if it is 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 prior section, with the entire distribution of human capital within each firm, we can quantify the relationship between outcomes like productivity and market value and the organization of human capital.

d) 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 Brynjolfsson, Hitt and Yang (2001, hereafter BHY) and can be written as:

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

The market value of a firm is assumed to be an increasing function of the assets.[4] Defining and measuring all of the terms in this relation is 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 (absent adjustment costs and market power). From this perspective, one way of measuring intangibles is a residual approach (see, *e.g.,* Hall, 2001) since the residuals will reflect the difference between the market value and observed assets.

[4]BHY specify a linear relation and emphasize the departure of coefficients from one. 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.

Alternatively, direct measures of intangibles (*e.g.,* organizational capital as in BHY) can be included in an econometric specification explaining market value. However, note that 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 IT capital in a linear specification of (3) is larger than one. BHY provide evidence that this reflects the complementarity between market value and organizational capital.

With these remarks as a background, should human capital be included in the set of variables in an econometric specification of the market value equation? A simple model in the absence of complementarities and a basic view of the role of human capital is that human capital may not be 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 there is no correlation between human capital and unmeasured tangible or intangible assets, then human capital will not be reflected in market value. However, there may be several sources of departures from this assumption. 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, there are many factors that may yield a relation between productivity and higher moments of the within-firm distribution of human capital. Under this view, it may not be the average level of human capital at the firm that matters for market value *per se*, but rather how that human capital is organized (as measured by higher moments). Finally, in a manner analogous to the arguments and findings of BHY, human capital (its average and other measures of the distribution) may be complementary to unmeasured tangible and intangible assets. As such, human capital may be positively related to market value because of omitted measures of tangible or intangible assets from the specification.

e) 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 improve on 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 a host of econometric issues arise that could 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 discretion about 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 Acemoglou, 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 – there is abundant empirical evidence that firms have quite high levels of churning of workers through jobs, and hence abundant opportunities to change their worker mix (*e.g.*, 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 vs. indirect effects.[5] We include measures of tangible assets 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 that we measure. In particular, the impact of human capital on productivity and market value may reflect both direct and indirect effects of human capital. However, by looking at the impact on both productivity and market value we hope to make some progress in understanding the role of human capital on business outcomes. If the human capital measures are mostly capturing 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 impact on productivity (as AKM found in France) and very little impact 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 human capital measures (or some components or indices of our human capital measures) are positively related with productivity and market value, then this outcome suggests that the human capital measures are either directly or indirectly capturing some form of intangible asset associated with human capital.

We clearly recognize that exploring and separating out the direct and indirect effects of human capital in this context is important. However, as noted above, our objective is to explore the relations among 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 indirect roles of these new measures of human capital in this context will be the subject of much research in the years to come.[6]

Before proceeding, one other related interpretation issue 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. It is important to emphasize that these two measures 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 there are factors that inherently lead to a negative correlation between market value and current productivity. For example, a business with a high market value "new idea" may be actively

[5] In that sense, our approach follows very much in the spirit of Brynjolfsson *et al*.

[6] Cummins (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. As such, he attempts to identify the direct impact of the measured assets. Moreover, his approach in principle avoids another related problem of endogeneity from correlations of the asset variables with unmeasured productivity or market value shocks. Note, however, the unmeasured productivity and market value idiosyncratic shocks likely reflect the idiosyncratic factors that we are seeking to understand. By pursuing an estimation strategy for instruments that are supposedly orthogonal to these shocks, the role of intangibles may be missed entirely.

expanding and investing in physical and human capital. Adjustment costs may imply that such a firm exhibits low current productivity.

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 combined data from the integrated employer-employee, Economic Censuses and related data.

Figure 3.1 provides a brief summary of the data resources used for this project. The chart vastly simplifies the number and complexity of the links involved in constructing the matched employer-employee data. The details of these linkages can be found in the Data Appendix. However, we provide summary information about the data and matching in the next two subsections.

Figure 3.1 Summary of Data Sources

Data Set:	**Demographic Data**	Links to	**Business Data**	Links to	**Financial Data**
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.

a) The integrated employer-employee data

We exploit new Census Bureau data[7], (part of the Longitudinal Employer-Household Dynamics Program, LEHD) that integrates information from state unemployment insurance data and Census Bureau economic and demographic data in a manner that permits the construction of longitudinal information on workforce composition at the firm level. The LEHD Program represents a substantial investment made by the Census Bureau in order to permit direct linking of its demographic surveys (household-based instruments) with its economic censuses and surveys (business and business unit-based surveys).

The unemployment insurance (UI) wage records are discussed elsewhere (see Burgess Lane and Stevens, 2000). Every state in the U.S., through its Employment Security Agency, 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 UI wage record data are numerous. The data are frequent, longitudinal, and potentially universal. The sample size is generous and reporting for many data items is more accurate than survey based data. The advantage of having a universe as opposed to a sample is that movements of individuals to different employers and their consequences for earnings can be tracked. It is also possible to construct longitudinal data using the employer as the unit of analysis. The LEHD Program houses data from a number of states that currently comprise 45% of total US employment. The current states are California, Florida, Illinois, Maryland, Minnesota, North Carolina, Oregon, Pennsylvania, and Texas. In this paper, we use data from seven of these states with provisional national weights to build our human capital measures. We currently have the crosswalk between the UI files and the establishment 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.[8]

Perhaps the main drawback of the UI wage record data is the lack of even the most basic demographic information on workers (Burgess, Lane and Stevens 2000). Links to Census Bureau data overcome this for two reasons: 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 staff 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 Abowd, Lengermann and McKinney (2001) and in Abowd, Creecy and Kramarz (2002).

c) The Economic Censuses and related business-level data

The Economic Censuses (conducted every five years) provide comprehensive data on basic measures like output, employment and payroll for all of the establishments in the United States. In addition, in certain sectors (*e.g.,* manufacturing) more detailed questions on other inputs (*e.g.,* capital) are asked.

---

[7] The Census Bureau, the National Science Foundation, the Sloan Foundation and the National Institute on Aging generously supported the creation of the LEHD data bases as part of a social science database infrastructure initiative.

[8] 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.

Our first goal is to create a matched dataset linking the human capital measures to the Economic Census data. For the current paper, we focus on the Economic Censuses in 1997. One issue that immediately arises is the appropriate and feasible level of aggregation of business activity. While the Economic Censuses are conducted at the establishment level, the business level identifiers on our human capital measures are at the federal EIN, SIC, and state level.[9] As such, we aggregate the Economic Census data to that level and match to the human capital files.[10] While our unit of observation is somewhere between an establishment and a firm, most of the observations in our analysis data are at the establishment level. For multi-units reporting under a single EIN in a state we aggregate the establishment data to the two-digit SIC, 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 (out of 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 out-of-scope for the Economic Censuses (*e.g.,* agricultural businesses). It is also important to note that we have the requisite firm identifiers (Federal EINs) for six of the seven states for which we have human capital estimates and thus our analysis is this paper is restricted to the data for these six states.

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). While the SSEL has limited information, it does identify the ownership structure of firms so that we can further aggregate to the enterprise/firm level. Doing the latter aggregation permits us to match enterprise level data on human capital to Compustat.

Since 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 (*e.g.,* 50 percent of employment in the company must be in these three states) to restrict attention to companies for which we can measure human capital and firm outcomes like 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 6 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/firm level. We then restrict attention to Compustat firms that restricts the sample substantially as there are approximately 13,000 Compustat firms nationally. 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.

[9] 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 attempts to refine the most disaggregated economic entity available in these data.

[10] We use the two-digit SIC as the industry measure in this work. Industry is coded at the four-digit level in the Economic Census and in the underlying establishment data on the UI side; however, we have not yet implemented and algorithm to use additional industrial or geographic detail in the definition of the establishment.

4. Human Capital Estimates

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

Table 4.1: Summary of Pooled Human Capital Wage Components

	Standard	*Correlation with:*						
Component	*Deviation*	*ln w*	$x\beta$	θ	α	$u\eta$	ψ	ε
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 (θ)	0.835	0.468	-0.553	1.000	0.961	0.275	0.080	0.000
Unobserved part of person effect (α)	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 (ψ)	0.362	0.484	0.095	0.080	0.045	0.101	1.000	0.000
Residual (ε)	0.354	0.402	0.000	0.000	0.000	0.000	0.000	1.000

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.
Sources: Authors' calculations using the LEHD Program data bases.

Some of basic features of the estimates for the seven state dataset are shown in Table 4.1. 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 regressions based simply on worker characteristics. Third, in ALM they 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 is much more important and more highly correlated with wages than the observed component of the person effect. Fourth, the different components of human capital (*i.e.,* the person effect and the experience component) exhibit different variation and covariation. Indeed, an interesting feature of the person effect and experience effect components is that they are negatively correlated. While this result is not surprising as, for example, younger generations of workers are more highly educated, it is important to note as it reminds us that there are different dimensions of skill that 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. While we do not pursue explanations of this somewhat surprising finding, we note that aggregations of the person and firm effects to various levels of aggregation yields a strong and positive relationship. For example, ALM show that at the industry level, person and firm effects are positively related. Interestingly, Abowd, Haltiwanger, Lane and Sandusky (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 *a) A comparison of new and traditional measures of human capital*

While in principle, the JGF methodology can be applied equally well to measuring both sectoral and aggregate labor quality, in practice, the LEHD approach permits more within and across industry heterogeneity. In separate work, Lengermann (2002b) 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 labor quality indices 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 indices to sectoral labor quality derived from industry averages of our human capital measure for the period 1995-1998. 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 indices over the period 1995-1998 using the LEHD-based and JGF approach. The within-industry growth rates are highly correlated – the employment-weighted average of the sectoral correlations is 0.79. However, there is much higher average growth for any given industry and more cross-industry variation in those growth rates in the LEHD measures compared to the JGF measures (the average growth rate for the LEHD measure over the 4 years is 0.04 with the cross industry standard deviation of 0.067 while the corresponding growth for the JGF is 0.014 with a cross industry standard deviation of 0.001).

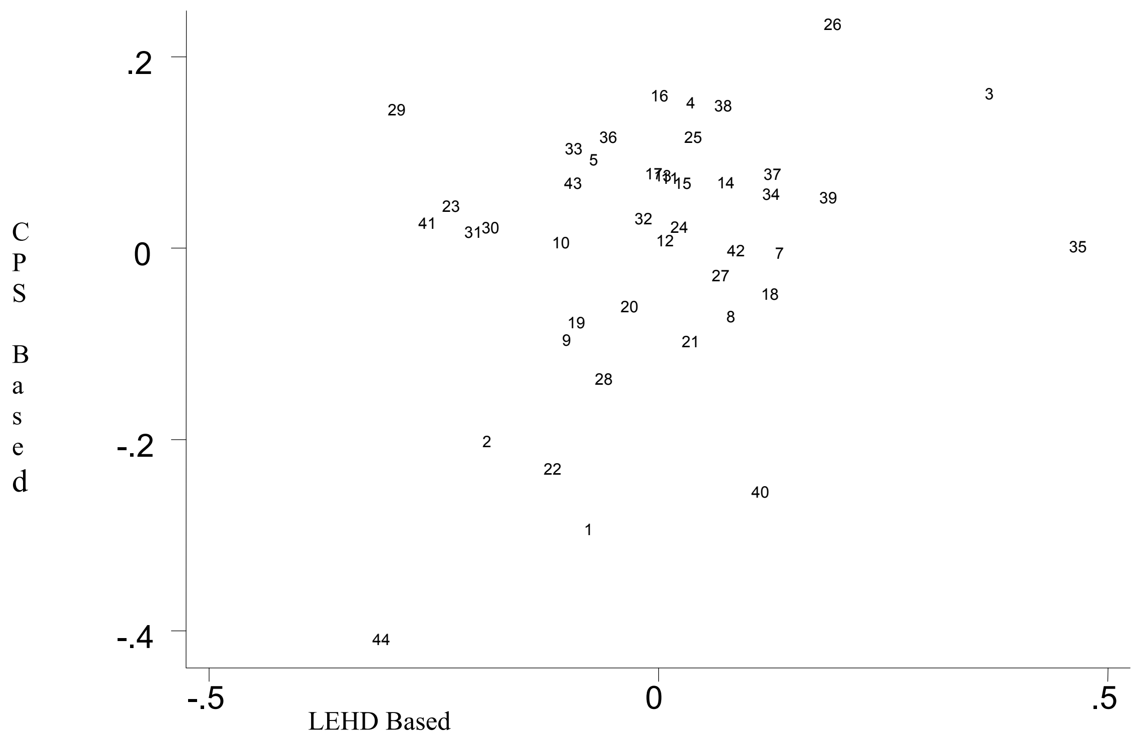
In what follows, we exploit cross sectional variation (across firms) in their human capital while the JGF procedure focuses on generating growth rates of human capital by industry. As such, the JGF measures are not well suited to examining within-year, cross-industry variation. Thus, as a second "indirect" approach we approximate the JGF labor quality indices by indices 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 the distinction of individual from firm contributions to wages, one might not expect them to yield identical results. Workers sort non-randomly 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 there exists a strong, positive correlation 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 and report them in Figure 4.1. Although the levels are normalized differently, there is clearly a great deal of correlation between the two measures – indeed, the correlation is 0.76. However, there is somewhat more cross industry variation in the LEHD-based measure than in the CPS-based measure (the standard deviation of the former is 0.15 and the standard deviation of the latter is 0.13).

In summary, the LEHD-based measures by industry are closely related to those derived by JGF or a simpler but closely related CPS-based procedure. However, LEHD-based measures generate greater average growth and more cross-sectional variation in both growth rates across industries and in levels of human capital across industries within a year.

Figure 4.1 Comparison of CPS and LEHD measures of Human Capital at the Industry Level

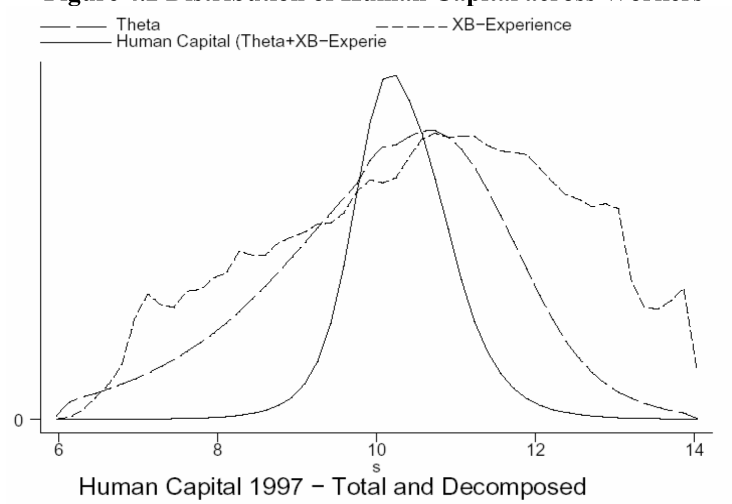


b) The construction of new human capital measures

A major contribution of this 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/firm traits to build measures of the human capital resources available to firms: the person effect (θ)), overall labor market experience of each worker captured by the experience component of $x\beta$ (denoted $x\beta$ in this section), and the sum of these two components (overall human capital, or h).

We describe the distribution of these measures in Figure 4.2. It is evident that all three components of the distribution exhibit enormous variation across workers. It is also interesting that the shapes of the distributions of the alternative measures differ: while the distribution of the person effect is bell-shaped, has thick tails and high variance, the distribution of experience is less smooth, and the distribution of human capital, the sum of θ and $x\beta$, is roughly bell-shaped, centered about zero, and has much less mass at the tails than either experience or θ. Underlying these relations is the negative correlation between experience and person effects reported in Table 4.1.

Figure 4.2 Distribution of Human Capital across Workers



c) The construction of firm level measures

While 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 in this paper 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. There are some restrictions on the sample necessitated by using this approach that are discussed in detail in the data appendix.

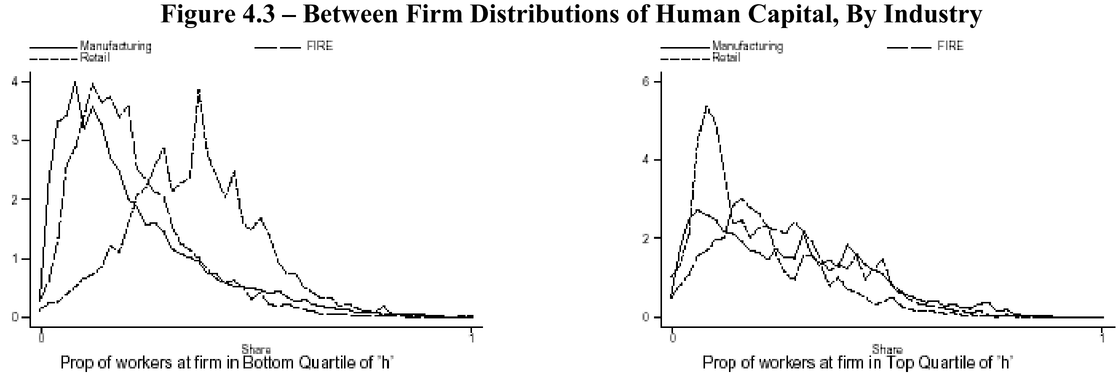
Table 4.2 shows that one key aspect of the variation across firms is driven by large variation in the human capital distributions across industries. FIRE and manufacturing are high human capital industries while retail trade is a low human capital industry. However, the components of human capital vary along these dimensions as well. The FIRE industries are high human capital industries especially because of having high median person effects while the manufacturing industries are high human capital industries more through having high experience effects.

Table 4.2: 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 θ*	*Proprtion 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 UI data from 6 states. Includes all jobs held by workers imputed to be full time at the end of the first quarter 1997.				

While there is substantial cross industry variation, the within industry variation in human capital across firms is enormous. Figure 4.3 illustrates the nature of the within industry variation in human capital and is constructed as follows. Consider first the left panel of Figure 4.3. For each firm we compute the share of workers at the firm that are in the lowest quartile of economy-wide distribution of human capital. The left panel depicts the between firm distribution of those shares. The right panel depicts the analogous exercise using the share of workers at the firm that is in the upper quartile of the distribution.

Figure 4.3 shows that there are substantial differences both within and across industries. Consistent with Table 4.2, 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, it is evident there is enormous within industry heterogeneity across firms in the shares of high and low skill workers. Apparently, firms in the same industry choose very different mixes of human capital and in the analysis that follows we will investigate whether this between-firm heterogeneity in human capital is related to between-firm heterogeneity in productivity and market value.

Figure 4.3 – Between Firm Distributions of Human Capital, By Industry



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,[11] controlling, as possible, for other relevant factors (*e.g.,* capital intensity). For this purpose, we focus on the 1997 Economic Census. Our measure of labor productivity is revenue per worker. The latter measure is the standard measure used in official BLS productivity statistics for gross output per worker.[12]

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 there is little practical guidance from either theory or prior empirical research.

Accordingly, we explore the role of the following measures: (i) the fraction of workers at the business above the economy-wide median human capital threshold; (ii) the fraction of workers at the business above the economy-wide 75th percentile human capital threshold; (iii) the fraction of workers at the business below the economy-wide 25th percentile human capital threshold; and (iv) the interaction of the latter two measures – literally the product of these two fractions. For these measures, we consider them using the overall human capital measure h and also consider these measures based upon the separate components of human capital (the person effect, θ, and the experience component, $x\beta$).[13] 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 5.1 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. First, there is tremendous heterogeneity across businesses in their mix of human capital as evidenced by the very large standard deviations in the human capital measures. Second, it is

---

[11] Recall that the level of aggregation that we use to approximate the establishment is that of an EIN/SIC2/STATE cell. While this unit is somewhere between the establishment and the firm, it remains the case that most firms operate only a single establishment.

[12]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).

[13]Note that the economy-wide thresholds are based on the universe of all workers in the seven states (not just workers employed at the businesses we match to the Economic Censuses) for which we have developed human capital estimates. Note that in an earlier version of this paper, we also generated versions of these measures based thresholds that varied by industry (so the interpretation would then be having a high fraction of highly skilled workers relative to, say, the median of the industry in the state). We found that the results are not sensitive to this distinction.

apparent that manufacturing has higher labor productivity and workers with higher human capital (on both the person effect and experience dimension).

Table 5.1: Mean Values of Variables in Log Productivity Regressions				
	All sectors		***Manufacturing***	
Variable	*Mean*	*Std Dev*	*Mean*	*Std Dev*
Log labor productivity	4.731	1.140	5.130	0.888
Log capital intensity			4.230	1.201
Overall $h = \theta + x\beta$				
Fraction of employment above 50^{th} percentile	0.480	0.215	0.538	0.225
Fraction of employment above 75^{th} percentile	0.237	0.176	0.267	0.185
Fraction of employment below 25^{th} percentile	0.266	0.172	0.208	0.169
Interaction: fraction above 75^{th} percentile with fraction below 25^{th} percentile	0.042	0.025	0.035	0.022
Person effect (θ)				
Fraction of employment above 50^{th} percentile	0.519	0.180	0.473	0.181
Fraction of employment above 75^{th} percentile	0.264	0.130	0.203	0.110
Fraction of employment below 25^{th} percentile	0.230	0.156	0.239	0.173
Interaction: fraction above 75^{th} percentile with fraction below 25^{th} percentile	0.048	0.027	0.039	0.026
Experience component ($x\beta$)				
Fraction of employment above 50^{th} percentile	0.455	0.157	0.545	0.133
Fraction of employment above 75^{th} percentile	0.220	0.116	0.285	0.115
Fraction of employment below 25^{th} percentile	0.282	0.144	0.205	0.096
Interaction: fraction above 75^{th} percentile with fraction below 25^{th} percentile	0.049	0.018	0.050	0.016
Number of observations	337,495		39,638	
Note: 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.				

Table 5.2 presents our exploratory analysis of the relation between log labor productivity and our *h* measure of the distribution of human capital. Table 5.3 presents the analogous results using the components of *h* separately. Before discussing results for alternative specifications, it is important to note some features that hold for all results. 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 uniformly high, suggesting that measures of human capital are, either directly or indirectly, important drivers of cross-sectional differences in productivity. The fact that the explanatory power for the manufacturing sector regressions is substantially less than the regressions 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.

Table 5.2: The Relation Between Labor Productivity and the Complete Human Capital Measure (Analysis level: Establisments; Dependent Variable: Log Labor Productivity)							
	All Sectors			***Manufacturing Only***			
Explanatory Variable	*(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)
Number of observations	337,495	337,495	337,495	39,638	33,926	39,638	33,926
R^2	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.

Consider first column (A) in Tables 5.2 and 5.3, it is apparent that businesses with a greater fraction of workers above the economy-wide median human capital level are much more productive. For the overall human capital measure, a one standard deviation change in this fraction is associated with a 27 log point change in labor productivity (Table 5.2). For the person effect measure (Table 5.3), a one standard deviation change in the fraction of high human capital workers is associated with a 25 log point change in labor productivity. For the experience component, a one standard deviation change in the fraction of high human capital workers is associated with a 23 log point change in labor productivity (Table 5.3). While these effects are very large, observe that they reflect only a fraction of the standard deviation in measured labor productivity across businesses (which is 114 log points).

Columns (B) and (C) of Tables 5.2 and 5.3 consider alternative measures of the distribution of human capital – focusing on the fraction of high human capital and low human capital workers. 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 and the share of workers in the upper quartile is increased by the same amount then productivity increases. This latter result holds for both the overall h measure and the components of the human capital measures. However, there are asymmetric effects 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 in the second column of Table 5.2 show that changing the share of workers in the firm that are in the lower tail of the human capital distribution has a disproportionate impact. Somewhat surprisingly the coefficient on the upper tail is negative but is small in both absolute and relative terms (*i.e.,* much smaller in magnitude than the coefficient on the lower quartile) and is not significant. The analogous column in Table 5.3 (column (B)) sheds further light on these results and shows that different components of human capital act in different ways. In particular, there is a disproportionate impact from the upper tail of the person effect and a disproportionate impact from the lower tail of the experience effect.

Table 5.3: The Relation Between Labor Productivity and Human Capital, Decomposed
(Analysis level: Establisments; Dependent Variable: Log Labor Productivity)

	All Sectors			***Manufacturing Only***			
Explanatory Variable	*(A)*	*(B)*	*(C)*	*(D)*	*(E)*	*(F)*	*(G)*
Fraction of workers above 50th percentile for θ	1.400 (0.009)			1.240 (0.024)	0.670 (0.022)		
Fraction of workers above 75th percentile for θ		1.990 (0.015)	1.700 (0.017)			2.060 (0.050)	1.480 (0.048)
Fraction of workers below 25th percentile for θ		-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 θ			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)
Number of observations	337,495	337,495	337,495	39,638	33,926	39,638	33,926
R^2	0.547	0.564	0.568	0.315	0.471	0.360	0.496

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.

Column (C) in tables 5.2 and 5.3 presents an even richer specification where we have attempted to capture the interaction between high-skill and low-skill workers. 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 quartile of the human capital distribution 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 while 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 impact from an increase in the upper tail of the distribution and the positive interaction effect reinforces this asymmetry. This 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 a one standard deviation increase in the share of workers in the highest quartile yields a 32 log point change in productivity while a one standard deviation increase in the share of workers in the lowest quartile yields a 3 log point loss in productivity.The opposite pattern holds for the experience effects. That is, the interaction effects reinforce the disproportionate impact of the lower tail of the distribution of the experience effect.

As discussed in section 2, there are a variety of possible for these asymmetries in the effects of human capital on productivity. While we cannot distinguish between competing explanations, these findings are consistent with the view that there is a concave relation between productivity and experience at the worker level and a convex relationship between productivity and the person effect at the worker level. However, the results may also reflect complementarities across co-workers that differ on different dimensions of skill.

Columns (D) – (G) of Tables 5.2 and 5.3 show results for the manufacturing sector. Column (D) replicates the column (A) but for manufacturing only. Column (E) adds capital intensity as 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 but does reduce the magnitudes of the effects substantially (although they remain very large). This aspect of the findings is important because it, suggests that human capital is complementary with physical capital. Thus, as we discussed in section 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 5.2 and 5.3 present results for manufacturing using the richer specification used in column (C) for all sectors – with and without controlling 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.

Considering the results for manufacturing as a whole, there is clear evidence of capital-skill complementarity. Interestingly, it appears that there is capital-skill complementarity for all of the dimensions of skill we are investigating. That is, including capital intensity reduces the magnitude of the impact of the person effect, the experience component and the interaction effects.

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. The results are reported in Tables 5.4 and 5.5. The qualitative results are very similar and the most parsimonious specifications yield magnitudes that are quite similar to the "establishment"-level results in tables 5.2 and 5.3. However, for the more complex specifications, especially when interaction effects are included, the magnitudes vary somewhat from the "establishment-level" results. 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. That is, there is a disproportionate impact from the lower tail of the human capital distribution for both *h* and experience at both the establishment and firm level.

Table 5.4: The Relation Between Labor Productivity and the Complete Human Capital Measure (Analysis level: Firms; Dependent Variable: Log Labor Productivity)							
	All Sectors			*Manufacturing Only*			
Explanatory Variable	*(A)*	*(B)*	*(C)*	*(D)*	*(E)*	*(F)*	*(G)*
Fraction of workers above 50^{th} percentile of human capital	1.309 (0.007)			0.690 (0.016)	0.433 (0.015)		
Fraction of workers above 75^{th} percentile of human capital		0.11 0.011	0.184 (0.013)			-0.194 (0.031)	-0.189 (0.029)
Fraction of workers below 25^{th} percentile of human capital		-1.664 (0.011)	-1.621 (0.011)			-1.133 (0.028)	-0.786 (0.027)
Interaction of above 75^{th} and below 25^{th} percentiles			-0.648 (0.058)			-0.889 (0.166)	-0.274 (0.154)
Log capital intensity					0.256 -(0.003)		0.248 (0.003)
Number of observations	303,219	303,219	303,219	34,900	34,294	34,900	34,294
R^2	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.

The differences that arise between the "establishment" and firm-level results may be due to increased measurement error of both productivity and the human capital measures at the firm level. Measuring firm level productivity is more difficult 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 as firms with many establishments may differ in their distributions of human capital across establishments. Exploring the latter is interesting in and of itself and we plan to explore this area in future work.

The results overwhelmingly make the case that understanding differences in labor productivity across businesses – particularly outside of manufacturing – involves understanding the differences in the human capital across businesses. Regardless of whether these are direct or indirect effects and regardless of endogeneity issues, it is clear that the differences in labor productivity across businesses are closely related to the differences in the human capital mix across businesses, as evidenced by the very large R^2 in the regressions. The results also clearly suggest that it is not simply a measure of central tendency of the human capital distribution that matters. The fraction of workers at the tails of the distribution and, in a related matter, 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. Our results show that the most productive firms are those that have a high fraction of workers in the top quartile of the person effect distribution and low fraction of workers in the lowest quartile of the experience effect. These findings clearly suggest that the organization and mix of the workforce matter substantially.

Table 5.5: The Relation Between Labor Productivity and Human Capital, Decomposed (Analysis level: Firms; Dependent Variable: Log Labor Productivity)							
	All Sectors			*Manufacturing Only*			
Explanatory Variable	*(A)*	*(B)*	*(C)*	*(D)*	*(E)*	*(F)*	*(G)*
Fraction of workers above 50th percentile for θ	1.562			0.923	0.614		
	0.009			(0.023)	(0.021)		
Fraction of workers above 75th percentile for θ		1.783	1.557			1.190	1.000
		(0.015)	(0.017)			(0.050)	(0.046)
Fraction of workers below 25th percentile for θ		-0.729	-1.070			-0.625	-0.385
		(0.011)	(0.016)			(0.038)	(0.036)
Interaction of above 75th and below 25th percentiles for θ			2.173			2.522	2.199
			(0.073)			(0.209)	(0.193)
Fraction of workers above 50th percentile for $x\beta$	1.483			0.604	0.224		
	(0.011)			(0.030)	(0.028)		
Fraction of workers above 75th percentile for $x\beta$		0.307	0.430			-0.113	-0.262
		(0.019)	(0.022)			(0.058)	(0.053)
Fraction of workers below 25th percentile for $x\beta$		-1.718	-1.711			-1.207	-0.871
		(0.015)	(0.017)			(0.067)	(0.062)
Interaction of above 75th and below 25th percentiles for $x\beta$			-0.569			-2.348	-1.942
			(0.084)			(0.268)	(0.247)
Log capital intensity					0.262		0.257
					(0.003)		(0.003)
Number of observations	303,219	303,219	303,219	34,900	34,294	34,900	34,294
R^2	0.530	0.547	0.549	0.288	0.410	0.310	0.427

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.

6. Investigating the Relationship 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 (See the detailed discussion of the matched Compustat sample and variable definitions are in the data appendix). We report the means and standard deviations of this subset of observations in Table 6.1 for 1997[14]. Clearly these firms are more human capital intensive than the full sample – the proportion of the workforce above the median economy-wide 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, there is still substantial heterogeneity in all measures: although the mean of each variable is different in the two samples, the standard deviations are very similar.

[14] Because we use a log specification, we eliminate firms with missing or zero values. This results in an even smaller sample than the Compustat matched sample used in the previous section. The excluded firms tend to have more skilled workers, with greater representation in the upper tail of both the person effect and experience components. They also have on average workers with greater than average levels of tenure.

Table 6.1: Mean Values of Variables in Market Value Regressions		
	All sectors	
Variable	*Mean*	*Std Dev*
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	
Overall $h = \theta + x\beta$		
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
Person effect (θ)		
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
Experience component ($x\beta$)		
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	0.664	0.126
Interaction: fraction above 75th percentile with fraction below 25th percentile	0.048	0.015
Number of observations	1,837	
Note: Sample is pooled 1995-1998 data for Business Register-based firms, defined as those with at least 50% of U.S. employment in the six analysis states matched to Economic Census and Compustat data.		

Tables 6.2 and 6.3 present the results of estimating equation (3) (the (log) market value regressions) using our two sets of human capital measures.[15] 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.[16] The value-added from our analysis is that we can also measure human capital at the firm level. In our simplest specification (column (A) of Table 6.2), a larger fraction of employees in the upper half of the human capital distribution is associated with significantly greater market value. This, in itself, is an interesting result, since, if highly skilled workers are compensated proportionately to their skill, *and* there is no correlation between unmeasured (tangible or intangible) assets and human capital, there should be no effect on market value once these other variables have been controlled. Yet the estimated effect of human capital is quite large: a one standard deviation increase in the proportion of the workforce that is above average is associated with an approximately 14 log point change in market value (set against a quite large standard deviation of market value of 200 log points).

[15] The reported results are based upon pooled data for 1995-1998.

[16] 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 6.2: The Relation Between Market Value and the Complete Human Capital Measure (Analysis level: Firm; Dependent Variable: Log Labor Productivity)			
		All Sectors	
Explanatory Variable	*(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	0.554
		(0.269)	(0.319)
Fraction of workers below 25th percentile of human capital		-0.414	-0.451
		(0.313)	(0.385)
Interaction of above 75th and below 25th percentiles			0.343
			(1.785)
Log capital	0.421	0.422	0.422
	(0.026)	(0.026)	(0.026)
Log other assets	0.539	0.538	0.538
	(0.026)	(0.026)	(0.026)
Number of observations	1,837	1,837	1,837
R^2	0.851	0.851	0.851

Notes: The human capital measure is $h = \theta + x\beta$. Data were pooled for the years 1995-1998. The analysis sample is firms in six states (defined as the firms with at least 50% 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.

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

The decomposition of these results into person and experience effects (Table 6.3) is even more striking, however. In particular, all of positive effect on market value is due to workers who have higher θ (person effects). This can be seen most clearly in column (A) of Table 6.3. It is striking that while both high-θ and highly experienced workers are more productive (Table 5.3), it is only the person effect that 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, there are asymmetries in the effects of different parts of the distribution. Column (B) of Table 6.3 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 can be seen, in column (C) of Table 6.3, 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.

Table 6.3: The Relation Between Market Value and Human Capital, Decomposed
(Analysis level: Firm; Dependent Variable: Log Labor Productivity)

	All Sectors		
Explanatory Variable	*(A)*	*(B)*	*(C)*
Fraction of workers above 50th percentile for θ	1.038 (0.242)		
Fraction of workers above 75th percentile for θ		0.867 (0.370)	0.610 (0.385)
Fraction of workers below 25th percentile for θ		-0.455 (0.392)	-1.249 (0.543)
Interaction of above 75th and below 25th percentiles for θ			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)
Number of observations	1,837	1,837	1,837
R^2	0.854	0.855	0.857

Notes: The human capital measure is $h = \theta + x\beta$. Data were pooled for the years 1995-1998. The analysis sample is firms in six states (defined as the firms with at least 50% 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.

We find it striking that it is the person effect that is important in predicting market value. Recall 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 difficult to observe dimensions of skill. We also find it striking that it is disproportionately the upper tail of the person effect that matters. 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 (*i.e.,* omitted variables).

7. Summary and Concluding Remarks

We began by noting that measurement of intangibles and human capital is an important challenge for the federal statistical system, particularly given the advent of the New Economy. We argued that it was important to find and quantify new measures of human capital that could be introduced into a firm-level production function and used in official economic statistics. This paper uses universe micro level data on both employers and employees to create new measures that begin to address this challenge.

The paper provides an overview of these new measures, and documents substantial consistency with earlier measures pioneered by Jorgenson, Gollop and Fraumeni (1987) (and subsequent closely related work). But it extends their work in ways that permit these human capital measures to vary within and between firms in the same way that other inputs and outcomes can vary. In addition, 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, the firm-level human capital distribution, in particular, to capture relevant aspects of firm-level differences in organizational capital and workplace practices.

Our preliminary results, which examine the relations among human capital, productivity, and market value, are intriguing. Not surprisingly, we find strong positive relations between human capital and productivity in the micro data that differ depending on the component of human capital used. The most skilled workers in terms of the estimated person effects have a disproportionate impact on productivity and the least skilled workers in terms of the estimated experience effects have a disproportionate (negative) effect on productivity. We find that human capital is also related to market value even after controlling for total physical assets. Interestingly, it is the component of skill that includes "unobservable" (at least to the econometrician) factors that is most closely related to market value. At this stage of our analysis, we are unable to separate out the observable and unobservable components of skill. In future work, it will be quite interesting to explore this aspect of the analysis and results.

We close by emphasizing that this work is exploratory. It is exploratory on many dimensions including the new, micro-based measures of human capital that incorporate unobservable dimensions of worker's skill and the role of such human capital in accounting for variation across firms in the U.S. economy. As we have emphasized, the strong empirical relations that we have uncovered may reflect a variety of direct and indirect effects of human capital.

References

Aaronson, Daniel and Daniel Sullivan (2001). "Growth in Worker Quality." *Economic Perspectives*, Issue Q IV, pp. 53-74.

Abowd, John M (1989) "The Effect of Wage Bargains on the Stock Market Value of the Firm" *American Economic Review*, Volume 79(4) 774-800

Abowd, John M (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." Cornell University working paper.

Abowd, John M., John Haltiwanger, Julia Lane, and Kristin Sandusky (2001). "Within and Between Firm Changes in Human Capital, Technology, and Productivity." Working paper.

Abowd, John M., Francis Kramarz, and David Margolis (1999). "High Wage Workers and High Wage Firms." *Econometrica*, pp. 251-334.

Abowd, John M., Paul Lengermann, and Kevin McKinney (2002). "The Measurement of Human Capital in the U.S. Economy." U.S. Census Bureau preliminary working paper.

Acemoglou, Daron (2001), "Technical Change, Inequality and the Labor Market." *Journal of Economic Literature*, Vol. 40, pp. 7-72.

Becker, Gary S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. 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 1.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt (1999). "Information Technology, Workplace Organization and the Demand for Skilled Labor: Firm-Level Evidence." NBER Working Paper No. 7136.

Brynjolfsson, Erik, Lorin M. Hitt, and Shinkyu Yang (2001). "Intangible Assets: How the Interaction of Computers and Organizational Structure Affects Stock Market Valuations." working paper.

Burgess, Simon, Julia Lane, and David Stevens (2000). "Job Flows, Worker Flows and Churning." *Journal of Labor Economics*, pp. 473-502.

Foster, Lucia, John Haltiwanger, and C.J. Krizan (2001). "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." *New Directions in Productivity Analysis*, (eds. Edward Dean, Michael Harper, and Charles Hulten), University of Chicago Press, (forthcoming).

Gore, Al (1997). Speech at the Microsoft CEO Summit.

Groshen, Erica L. (1991a). "Five Reasons Why Wages Vary Among Employers." *Industrial Relations*, pp. 350-381.

Groshen, Erica L. (1991b). "Sources of Intra-Industry Wage Dispersion: How Much do Employers Matter?" *The Quarterly Journal of Economics*, pp. 869-884.

Gu, Feng and Baruch Lev (2001). "Markets in Intangibles: Patent Licensing." Parker Center for Investment Research Working Paper.

Hall, Bronwyn (1990). "The Manufacturing Sector Master File: 1959-1987." NBER Working Paper No. 3366.

Hall, Bronwyn (1998). "Innovation and Market Value." paper presented at the NIESR Conference on Competitiveness and Productivity, London, England, February 1998. In Barrell, Ray, Geoffrey Mason, and Mary O'Mahoney (eds.), *Productivity, Innovation and Economic Performance*, Cambridge: Cambridge University Press, 2000.

Hall, Robert (2001). "Struggling to Understand the Stock Market." (Richard T. Ely Lecture) *American Economic Review Papers and Proceedings*, Vol. 91, pp. 1-11.

Haltiwanger, John and Ron Jarmin (2000), "Measuring the Digital Economy." In E. Byrnjolfsson and B. Kahin (eds.) *Understanding the Digital Economy*, 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." Working Paper.

Jorgenson, Dale W., F. Gollop and B.M. Fraumeni (1987). *Productivity and US Economic Growth*. Harvard University Press, Cambridge, MA.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1992). "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy*, pp. 410-442.

Kremer, Michael and Eric Maskin (1995), "Wage Inequality and Segregation by Skill," NBER Working Paper No. 5718.

Lane, Julia, Philip Moss, Harold Salzman, and Chris Tilly (forthcoming) "Too Many Cooks? Tracking Internal Labor Market Dynamics in Food Service with Case Studies and Quantitative Data," in *The Future of Low Wage Work*, New York: Russell Sage, Lane, Julia, James Spletzer and Laurie Salmon (2002), "Establishment Wage Differentials", mimeo, The Urban Institute

Longitudinal Employer-Household Dynamics Program (2002) "Longitudinal Employer-Household Dynamics Program: Employment Dynamics Estimates, Versions 2.2 and 2.3," Technical Paper No. TP-2002-05.

Lengermann, Paul, (2002a)"Is It Who You Are, Where You Work, or With Whom You Work? Reassessing the Relationship Between Skill Segregation and Wage Inequality," University of Maryland Ph.D. Thesis.

Lengermann, Paul, (2002b) "Comparing New and Traditional Measures of Human Capital," working paper.

Data Appendix

This section describes the construction of the data and key variables used for analysis. There are many steps involved in the formation of the data, and extensive detail is required to account for each of these steps. Many of these steps are part of the data infrastructure development efforts at LEHD and CES (Center for Economic Studies) and, as such, there are a number of technical documents that provide the detail of these steps. In this appendix, we provide an overview of the data construction efforts and refer to technical documents as needed.

The process of data construction can be broken into two large segments. The first is the formation of an "establishment-level" file used in the productivity analysis and containing both human capital measures constructed from the LEHD database as well as business traits from Economic Census micro-data. The second is the formation of a firm-level dataset that uses an aggregated version of the establishment file matched to Compustat data for the years 1995 through 1998. Both steps are described below.

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 economy-wide set of thresholds to generate variables characterizing the distribution of human capital at each business unit; next, using data from the 1997 Economic Census micro-data, we merge in measures of productivity for businesses in all sectors as well as capital intensity data for manufacturing businesses. Below, we elaborate on each of these steps and indicate reference documents when they exist.

We begin by using the LEHD database, which links worker identifiers and demographic detail to employer identifier variables, to compute human capital measures for each worker. The construction of the LEHD database is described in detail in LEHD Technical Working Paper No. TP-2002-05. These data are used to build human capital measures for each worker using methods from AKM and applied to the LEHD database as described in ALM. Although the time span of data used to estimate the person fixed effect varies across states, this paper makes use of human capital estimates from a pooled seven state sample and matches of these estimates to establishment and firm level data for six states for the years 1995-1998 inclusive.

Using these 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 as an Employer Identification Number (EIN), two-digit SIC, state, and year unit. This level of aggregation was selected because it is the smallest that is common to both the LEHD database and the Economic Census micro-data. At this level of aggregation, most units are establishments and 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, all workers do not contribute to our characterization of the human capital distribution at the establishment in that year. Instead, we include only those workers both imputed to work full time at any job during the year and who are working at the firm at the end of the first quarter of the year.[17] Because we do not observe information on hours worked yet wish to separate human capital from labor utilization effects, we restrict the set of jobs to those held by workers imputed to work full time. 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 data (March 12 employment).

Table A.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)*
Number of establishments	1,391,839	1,078,877	429,529
Number of jobs	40,967,355	21,829,227	20,156,292
Fracttion of firms in agriculture	0.020	0.014	0.012
Fraction of firms in mining	0.008	0.009	0.009
Fraction of firms in construction	0.072	0.063	0.060
Fraction of firms in manufacturing	0.125	0.154	0.162
Fraction of firms in public administration	0.026	0.037	0.039
Fraction of firms in FIRE	0.057	0.063	0.062
Fraction of firms in retail trade	0.191	0.162	0.161
Fraction of firms in services	0.380	0.360	0.356
Fraction of firms in TCE	0.057	0.063	0.065
Fraction of firms in 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

Note: Statistics are employment weighted. Analysis samples are based on 1997 UI-based establishments for six states (defined as EIN/2-digit SIC units).

To evaluate the consequences of the restrictions on employment, Table A.1 presents summary statistics for alternative samples. Column (A) of Table A.1 shows descriptive statistics for variables created from the establishment-level dataset that includes the full set of possible jobs – that is, all jobs held by workers imputed to work full time at any job in that year. Column (B) of this table reports the same set of measures with the added point-in-time restriction. Although the total number of jobs included falls by nearly half when this restriction is imposed (from almost 41 million jobs to about 22 million jobs), the loss of employers included in the data set is not as substantial, falling from about 1.4 million to approximately 1.1 million establishments. Establishments that are both 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 there are only slight differences in the industry composition of the two datasets. Specifically, column (B) has a slightly higher share of manufacturing establishments (rising from 12 percent to 15 percent) and a slightly lower share of retail (19 percent to 16 percent) and services (38 percent to 36 percent). It is worth noting that businesses eliminated through the point-in-time restriction appear to be less productive and, among manufacturers, less capital intensive although these differences are relatively small.

[17] End of quarter one employment is characterized as working for that employer in both quarter one and two.

Many business units at this level of aggregation employ a small number of workers. For the smallest businesses (*e.g.,* under 5 employees) our ability to characterize the distribution of human capital at the business is limited as one worker can make an enormous difference. Moreover, even for small to medium size businesses, the empirical distribution of human capital at the individual firm level is measured in a quite noisy fashion. For example, if we want to measure the fraction of workers in a business in some percentile range, the actual empirical distributions will be quite noisy and potentially misleading. 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 kernel density estimate (for h and each of the components) for each firm (see ALM for details). There are, unfortunately, tradeoffs that accompany use of this method. While we are able to measure the human capital at businesses more accurately, data requirements for the estimation procedure prevent us from using the smallest of firms in our analysis. Specifically, we generate a kernel density estimate of the for each of the three human capital measures at each business that has at least five full-time workers employed at the business at the end of quarter one in a given year.

Column (C) in Table A.1 provides descriptive statistics of the set of firms that meet the size restriction of five or more workers employed at a given point in time.[18] Comparing column (C) to column (B) reveals that the size restriction eliminates less than 10 percent of jobs (about 1.6 million out of about 22 million). However, this restriction does have a substantial impact in terms of number of business units. Specifically, we lose over 60 percent of establishments through this restriction (the number falls from about 1.1 million to less than 430,000). For this restriction, however, summary statistics suggest that, although all deleted firms are small, the firms lost through this restriction are spread evenly throughout all sectors of the economy and do not differ substantially in mean level of productivity or capital intensity.

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

Table A.2 illustrates the sensitivity of the mean value of each human capital measure to the set of sample restrictions imposed as well as to the method used to compute business-level skill. Columns (A), (B) and (C) correspond to columns (A) through (C) in Table A.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)

[18] Additional restrictions are imposed on the range of h, θ, 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 θ 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 minor loss of both jobs and firms.

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

Table A.2: Summary Statistcis 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 Kernel Density Estimates (D)*
Number of "Establishments"	1,391,839	1,078,877	429,529	429,529
Number of Jobs	40,967,355	21,829,227	20,156,292	20,156,292
Fraction of workers above 50^{th} percentile for human capital	0.403	0.496	0.502	0.497
Fraction of workers above 75^{th} percentile for human capital	0.187	0.246	0.245	0.248
Fraction of workers below 25^{th} percentile for human capital	0.339	0.254	0.242	0.257
Interaction of above 75^{th} and below 25^{th} percentiles for human capital	0.039	0.038	0.038	0.043
Fraction of workers above 50^{th} percentile for θ	0.507	0.511	0.520	0.513
Fraction of workers above 75^{th} percentile for θ	0.254	0.256	0.258	0.261
Fraction of workers below 25^{th} percentile for θ	0.253	0.245	0.232	0.236
Interaction of above 75^{th} and below 25^{th} percentiles for θ	0.051	0.046	0.046	0.049
Fraction of workers above 50^{th} percentile for $x\beta$	0.421	0.490	0.481	0.473
Fraction of workers above 75^{th} percentile for $x\beta$	0.205	0.248	0.236	0.229
Fraction of workers below 25^{th} percentile for $x\beta$	0.335	0.260	0.266	0.266
Interaction of above 75^{th} and below 25^{th} percentiles for $x\beta$	0.052	0.049	0.049	0.049

Note: 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.

Comparing columns (A) and (B) in Table A.2 reveals that 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 that businesses that are cut have a higher share of these low-experience workers. Recalling that many of these establishments that are eliminated 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 and high θ workers. Thus, the impact of the restriction appears to be primarily on the experience component of human capital. Column (C) adds the employer size restriction. In spite of the large number of establishments eliminated through this restriction, it does not appear that the smaller establishments that are eliminated employ workers of systematically different skill levels. 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 A.1 and A.2 show that the between-business distribution of skill shares is only slightly sensitive to the point-in-time employment restriction and not at all sensitive to the size restriction nor to the method used to characterize the skill distribution at each business. Table A.3 presents two sets of productivity regression results for EIN two-digit SIC state units in all sectors in 1997. The left panel is identical to the results in Table 5.3, columns (A)-(C), using the human capital measures from column (D) in table A.2. The right panel shows results obtained when we use the full dataset of all jobs held by workers employed full time at any business in

1997 (the group described in column (A) in tables A.1 and A.2) and the empirical distribution of human capital to compute fractions of workers at different percentiles for each firm.

Table A3: Comparison of the Full Data Set and the KDE Subset for the Relation Between Labor Productivity and Human Capital, Decomposed (Analysis level: Establishments in All Sectors; Dependent Variable: Log Labor Productivity)

	KDE Dataset			*Full Dataset*		
Explanatory Variable	*(A)*	*(B)*	*(C)*	*(D)*	*(E)*	*(F)*
Fraction of workers above 50th percentile for θ	1.400 (0.009)			1.514 (0.005)		
Fraction of workers above 75th percentile for θ		1.990 (0.015)	1.700 (0.017)		1.723 (0.008)	1.656 (0.009)
Fraction of workers below 25th percentile for θ		-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 θ			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.

Overall, it appears that the key productivity finding, that skill is positively related to productivity, holds up across restrictions imposed on the data as well as the methods used to build the human capital measures. There is, however, one exception. Interacting the highest and lowest quartiles of either the θ or $x\beta$ distribution at a business has a much stronger impact 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 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 (SSEL) to obtain information on business structure. This information is then used to build the firm-level data used in the market value analysis. Although there are slight differences across states and years, in general we find approximately 99 percent of the EINs from the LEHD data base 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 micro-data. To form labor productivity for each EIN, two-digit SIC state unit, we do the following. First, we sum employment on March 12 across all sub-units with non-missing sales revenue and positive employment. Second, we divide sales revenue at each sub-unit by this sum. Last, we calculate an employment-weighted average of the sub-units to aggregate to the EIN, two-digit SIC (SIC2), state level. We use a similar procedure for capital intensity.

Our objective is to maximize the number of observations on 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 on 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 discard these records, we instead apply EIN-level state-wide 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, 80,506 EIN-state only 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:
median two-digit SIC of all reporting units under a state EIN (employment-weighted).

For the firm-level productivity analysis, we aggregate all of the variables (*i.e.*, labor productivity, capital intensity, human capital variables) to the firm (*i.e.*, enterprise level) using employment weights. The Economic Census files contain firm/enterprise identifiers that make this aggregation relatively straightforward. As noted in the 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 5.4 and 5.5 we include controls for multi-unit status and diversification indicators indicating 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 SSEL. Where 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 SSEL 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/firm identifier from the SSEL, and then matching that to a list of the unique EINs in the Compustat database. There are some cases where a single EIN is associated with more than one SSEL firm or Compustat stock issue. There are also a few cases in which different EINs from the Compustat database are associated with the same firm on the SSEL. For each of these types of duplication, where there was a clear reason to think that one link should be preferred,[19] we dropped the other links. Where it was not clear how to resolve the duplication issue, we eliminated all records involved. Where an inactive Compustat case linked to different SSEL firm identifiers in different years, we used the SSEL identifier in the most recent year before the case became inactive as the link.

For cases for which 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.[20]

Table A.4 gives the distribution of outcomes of the Compustat cases in our original version of these data. Of the 14,312 Compustat cases that were traded at some point after 1995, we found a unique link for 11,170 cases. However, some of these cases either were linked to SSEL firms that were inactive in the years of interest or were missing essential Compustat data. Restricting the sample to cases that link to at least one establishment on the SSEL between 1996

[19] For example, some of the non-unique matches on the SSEL 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 prior to matching, but did so afterwards if the match was not unique. Some of the non-unique Compustat matches involve cases that have a Compustat-assigned CUSIP, which often carry alternative versions of data for companies that also have a standard CUSIP. For example, if an acquisition took place in 1999, there might be two records for the acquiring company in the data we have: one with a standard CUSIP and data that reflects 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 prior to the merger. In this sort of case, we would drop the record with the Compustat-assigned CUSIP (and consolidated data) and keep the record with the standard CUSIP.

[20] 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 SSEL, 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 SSEL to the years with non-missing data from Compustat. There were a significant fraction of cases where the years did not line up which was the main reason we decided not to use the statistically linked inactive cases. There were 346 such cases.

and 1998 (the years for which we have LEHD estimates), and have a price reported in Compustat in at least one year in that range leaves 9,917 cases.

Table A. 4 - Summary Statistics on Compustat File for Alternative Samples							
Compustat Status	*Total in 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	*Name/address/SIC link* 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

The sample that could potentially be used in conjunction with our human capital estimates is further restricted to businesses that have some employment in the six states for which we have LEHD estimates. Table A.5 gives sample sizes for this six-state sub-sample as a function of what share of employment we require to be in those states.

Table A.5 Number of firms in Compustat/SSEL database with employment in sample states				
	Required employment share in LEHD states			
Year	*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

Table A.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 1995-1998 for all Compustat cases that have non-missing data on sales and positive data on employment. The second column takes the subset of those cases that were uniquely linked to an SSEL firm, and the third column further restricts the sample to those with at least 50 percent of SSEL employment in the six LEHD states.[21]

Table A- 6 Sample means for full and matched samples			
	Full Compustat	***Matched Compustat/SSEL***	***Matched, with at least 50% employment in LEHD states***
Number 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

Clearly 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

[21] The sample means in Table 6.1 are based on a sample with non-missing data for a larger set of variables, as well as requiring an additional link to the LEHD data. The final regression sample is smaller than that in Table A.3, and has slightly smaller Compustat employment on average (Table 6.1).

more likely to be 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.

Note that there are also differences across the two data sources for employment. This is probably partially accounted for by the inclusion of overseas employees in the Compustat figures (the SSEL data include only U.S. employees). However, detailed examination of the micro-data suggests that there are differences even for firms that operate only in the U.S. and the sources of those discrepancies are not clear. Possible candidates include differences in the definition of employee or differences in dating of employment (the SSEL data reflect employment as of the week containing March 12th while 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), Hall (1998), and Brynjolfsson, Hitt and Yang (2001). The primary Compustat variables that we use are:

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.22

Physical Capital:

gross book value of capital stock is deflated by the GDP implicit price deflator for fixed investment. The deflator is applied at the calculated average age of the capital stock, based on 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.

---

[22]There are some differences in measurement methodology in the literature. Our method most closely follows that of Brynjolfsson, Hitt and Yang (1990). 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 20-year maturity. It is not difficult to construct scenarios where this assumption induces substantial measurement error.