

Hamermesh, Daniel S.

Working Paper

Fun with matched firm-employee data: progress and road maps

IZA Discussion Papers, No. 2580

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Hamermesh, Daniel S. (2007) : Fun with matched firm-employee data: progress and road maps, IZA Discussion Papers, No. 2580, Institute for the Study of Labor (IZA), Bonn, <https://nbn-resolving.de/urn:nbn:de:101:1-20080425149>

This Version is available at:

<https://hdl.handle.net/10419/33807>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 2580

Fun with Matched Firm-Employee Data: Progress and Road Maps

Daniel S. Hamermesh

January 2007

Fun with Matched Firm-Employee Data: Progress and Road Maps

Daniel S. Hamermesh

*University of Texas at Austin,
NBER and IZA*

Discussion Paper No. 2580
January 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Fun with Matched Firm-Employee Data: Progress and Road Maps^{*}

With the beginnings of a worldwide burgeoning development of matched firm-employee data, it is worthwhile to examine the possibilities for using these data. This essay discusses a variety of areas in which some progress has been made and presents ideas for future research in a number of others, including the study of labor demand, search and unemployment, wage determination and time use. It concludes that such data could be as important for labor economics, and for generating new knowledge about labor markets, as have been longitudinal household datasets, but with existing restrictions on access this kind of success will be difficult to achieve.

JEL Classification: J2, J3

Keywords: wage determination, match quality, social norms, time use

Corresponding author:

Daniel S. Hamermesh
Department of Economics
University of Texas
Austin, TX 78712
USA
E-mail: hamermes@eco.utexas.edu

^{*} This paper was initially presented as a Keynote Address at the Conference on the Analysis of Firms and Employees in Nuremberg, Germany, September 29, 2006. I thank Julia Lane and other participants for helpful suggestions on the content and implications of the talk.

1. Introduction

One might ask what right I have to be writing on employer or employee data, as it has been eight years or so since I have done anything in this area; and in the economics publishing business eight years is about a lifetime. Trying to rationalize this paper, I am thus reminded about one of our favorite family jokes about the couple who are tired of having their tomcat going out every night “catting” around. They finally decide to have the cat neutered. The night after the cat returns from the veterinarian, lo and behold it is again out all night fooling around. The owners ask the cat, “What are you doing? We had you fixed, and you’re still out ‘catting’ around.” The cat responds, “Now I’m a consultant!”

Like the tomcat I am a consultant who has a fair amount of experience in this area. The only difference is that, unlike the cat, I do not want to talk about fooling around at night. Instead, what I do is review a number of issues about where we have been and where we are in the area of firm-employee data, suggest some on-going research projects that seem especially worth pursuing, discuss others that might be done in more interesting ways, and suggest some new data initiatives.

2. Which Came First, the Data or the Idea?

I want to frame the discussion in terms of the larger issue: The direction of causation between theoretical ideas and the availability of data—i.e., the question: What generates scientific progress (assuming that we can use the term science to talk about economics)? Does it matter whether causation runs from data to problems or from problems to data? I think it does. Stafford (1986) discusses this issue in detail, arguing for causation from data to ideas. For example, in the case of inter-temporal labor supply

we would not know so much without the creation of the PSID. The same data set has also generated a huge amount of serious thinking about the inter-generational transmission of inequality. Without the availability of micro data on establishments it is unlikely that the work of Dunne *et al* (1989) or Davis and Haltiwanger (1992) would have had their current intellectual influence. In terms of issues of firm behavior in particular, I doubt that we would even have thought about the issues in the way we now do without the availability of this type of data. No doubt the opposite is also often true, but my purpose here is to talk about the former.

To paraphrase *Matthew 5*, “Blessed are the data developers because they inspire the creation of knowledge.” Creating data is a very thankless task for which one gets very few points. Yet so many of our ideas are inspired by new data, and so much of research rests on innovations in questions and data collection that are barely, if at all, acknowledged by the more technical researchers. Indeed, large numbers of studies by people who would not be caught dead using data themselves in the end depend on novelties in the kinds of data available to empirical researchers.

3. Using Matched Firm-Employer Data to Advance Labor-Market Research

a. Wage Determination

I could define a labor economist as somebody who has written at least one paper that includes a regression with wages or earnings as the dependent variable. If you have not done that, you may not be a labor economist. Now a crucial question in all earnings regressions is: Is it firm characteristics or worker characteristics that are important in determining wages? More generally, what are the relative contributions of these characteristics? This is important for policy purposes, because one might think of

operating on the demand or supply side of the labor market to enhance workers' outcomes.

Early work by Groshen (1991) showed the importance of employer characteristics, and more recent work by Abowd *et al* (1999) and Bryan (2004) attempted to sort out employer from personal effects. The issue really only matters if we have the equivalent of a misspecification problem in econometrics. Consider the schematic in Figure 1. So long as firms' and workers' characteristics are uncorrelated, we can obtain unbiased estimates of the impact of workers' (firms') characteristics on earnings. (I assume that both types of characteristics have separate, significant impacts on earnings, that the horizontal arrows in Figure 1 matter.)

Does the vertical inequality in Figure 1 hold or not? One leading labor economist noted that one of his accounting colleagues remarked that labor economists are little better than accountants, as all we do measure the size of effects. In a sense he is more correct today than when he made this statement ten years ago, but I do hope we are doing more than that, trying to find out not merely how much, but why. We are trying to inquire about what underlies that vertical double arrow, and that is a deep question about causation in the labor market.

The difficulty with trying to infer the importance of firms' and workers' characteristics in wage determination is that the answer we obtain will depend upon how much information we have from workers and from firms and on how much variability there is in this information. If we have very little information on the variability of characteristics among workers and lots on firms, it will look like firms' characteristics matter a lot and workers' matter little. The obverse will be true if we have a lot on the

workers and very little on firms. I thus doubt that we will be able to get very good estimates of the relative importance of demand- or supply-side forces in wage determination.

We can, however, do better; and using matched firm-employee (FE) data in which there is substantial information on both workers and firms will help in this direction. The issue is not unimportant, given the failure of the vertical inequality in Figure 1. For example, getting more information will enable us to obtain better measures of the returns to various types of investment in human capital. Also, while I do not know exactly what we mean by HR practices, I would think that better matched firm-employee data would allow those people who do understand this area to estimate more carefully the productivity effects of such practices. In the end we will never be able to sort out the effects fully—one can always envision characteristics of the firms or the workers on which we lack information; but with better data we can learn more about the true effects—and we can do so without having to resort to inferences from policy and other “experiments” that are not likely to typify most labor markets.

b. Match Quality and Employment Duration

There is a very large literature on this topic going back at least to Kiefer and Neumann (1979). The entire literature builds a theoretical structure and tries to look inside the black box of wages and employment to infer things about the match of workers and firms (which we saw in the previous section is crucial to sorting out the underlying determinants of wages). The inferences are usually built on very heroic assumptions about structure, and the models are mostly tested using household data. A few papers (Baker *et al*, 1994, was the first) actually try to use new FE data; but what we really need

is to examine match quality and employment duration by looking at large numbers of firms and their workers. Such data would yield substantial additional returns if more dynamic elements were added to the theory of match quality. Going beyond the issue of who flows out of firms and which firms gain and lose workers to thinking about the causes of quitting and firing would be valuable. This is a difficult task, not merely a matter of accounting.

Advancing on this front would allow us a better handle on the impacts and returns to training and re-training programs. It would allow us to answer more generally the Baker *et al* (1994) research question about the role of early activity in a job. It would shed light on the role of policies to encourage search, e.g., on the question of how unemployment insurance and the search requirements that it usually entails affect search behavior and job mobility. While we have a huge number of studies of UI, there is very little on the search that actually goes on and on the mobility that results from offering different levels of UI benefits with different qualifying requirements.

Match and mismatch in the labor market do not just arise out of one encounter, even a successful one: They are affected by each party's prior history of searching and matching, a point that relates to work on patterns of female labor-force participation by Heckman (1981) and more recently by Hyslop (1999). The more you know about the long history of spells of work, the better you can predict next period's behavior. In other words, individual idiosyncrasies in participation affect future participation. The same persistence probably exists in spells of employment, in the sense that staying with a particular company alters the nature of one's future matches, and having workers stay a long time alters a company's behavior in its future attempts at matching. The obverse is

also true (see Pfann and Hamermesh, 2001). In order to analyze this kind of persistence one needs long-duration information on matched firms and workers. The Danish IDA, for example, would suffice for this purpose, but I do not know of any other sufficiently long and comprehensive dataset.

While we know that duration and incidence in unemployment matter, the lack of long longitudinal datasets has prevented us from going much further in assessing the welfare effects on workers of unemployment and on firms of having different histories of job-matching/hiring/firing. Even more important, especially in a time of increased protectionism, having more detailed histories of job matches would allow insights into the likely success of different alternative policies that might help trade-displaced workers find jobs. More generally, such data would be helpful in teaching us more about the losses to worker displacement: The current literature typically relies on firm (Jacobsen *et al*, 1994) or household (Hamermesh, 1987) data; and in the few cases where matched data are used, characteristics on at least one side of the market are very thin.

A further, innovative step would be to obtain interactive life histories of firms and workers. There is a large literature on firm survival (e.g., Jovanovic, 1982), building a theory of survival and asking what are its correlates. There is also a long literature on worker mobility, going back to Ph.D. theses in the late 1960's. These two strands of research are not, however, interactive, in the sense that they do not examine who is going from where to where and which firms are hiring whom. They do not show what kinds of firms generate workers who are attractive to other firms. For example, in our own industry certain institutions are known as training grounds where assistant professors invest more in themselves than elsewhere and become attractive candidates for hiring by

other institutions. More generally, are there firms that become training grounds for other firms? What are the characteristics of such firms? What is it that they actually do? On the other side, what kinds of workers make firms more productive, and what are the job histories of such workers? Of course these questions relate to issues of training, tenure and mobility; but without what are essentially matched life histories of firms and the workers who move among them, we cannot get much below the surface of these relationships.

More specifically, what are the characteristics of workers as firms are born, age and die? We know that new firms generally tend to have younger workers—an unsurprising result of the well-known fact of greater voluntary and involuntary mobility among younger labor-force participants. Beyond that knowledge, however, not much has been done. The absence of progress is, I believe, due to the absence of the kind of interactive longitudinal dataset that is needed, partly too because of the lack of deep thought that facts derivable from those data might generate. If nothing else, this approach might allow us to go beyond identifying firm-specific training as the difference between the wage effects of tenure and pre-firm experience.

Data like these might even allow us to take the notion of entrepreneurship more seriously than is done in the courses in our business schools that purport to teach this subject. I do not know what entrepreneurship is, but maybe there is something there. Perhaps one could use these data to discover what it is, in terms of the kinds of labor (and other) resources that entrepreneurs collect and organize that makes some successful—allows some firms to grow while others die. If nothing else, this approach is surely a good

way to get grants, since business-supported foundations love to fund entrepreneurship research (although that maybe an oxymoron).

In the 1950s and early 1960s there was substantial interest in the idea of the optimal amount of turnover (Ross, 1958). Since then almost nothing has been written on this topic, neither theory nor empirical work using the kind of employer and employee data that are now becoming increasingly available. Yet businesspeople talk about this problem all the time (although their main interest is in minimizing voluntary turnover and having freedom to determine the rate of involuntary turnover). There still is a question of what the right amount of turnover is and whether and to what extent one should be encouraging or discouraging mobility. If, as I think is true, we should take this topic seriously, interactive life histories of firms and employees are the empirical bases that would underlie the positive research that might inform answers to this normative question. Here is a whole line of research that is well worth following out.

c. Social Norms and Peer Effects

Going still further afield from the usual set of interests of labor economists, and going closer to topics currently near my research heart (Burda *et al*, 2006) is the idea of social norms and peer effects. These are currently very hot topics in both empirical and theoretical economics (e.g., Durlauf, 2002; Clark, 2003). Some of this work deals specifically with the economics of the workplace; and recent efforts in this area include Bandiera *et al* (2006) and Jones *et al* (2006). The study of the economics of conformity (Bernheim, 1994) is also related to this literature. Using FE data seems most appropriate in this context: Without knowing the details of worker characteristics and the nature of their interactions with their peers and their supervisors, it would seem difficult to sort out

peer effects and the roles of social norms from the usual substitution/complementarity relationships that occupy standard theory. With such data the researcher knows who the workers are, with whom they are working and when they are working, and s/he obtains a lot of variation in these relationships.

The difficulty with this area of research is that it is not easy to generate empirically refutable implications of these theories in the context of the workplace (or the marketplace generally, for that matter). Distinguishing peer effects from substitution relations in factor demand is no easy task. I am reminded of the scene in *Through the Looking Glass*, where Alice clearly sees the bottles standing on a shelf in the closet. As she approaches the closet the bottles seem blurred, and by the time she gets there they are a complete haze. The best expression of this difficulty is in a review of Frank (1984) that Bob Lucas wrote in the *Wall Street Journal*.

More serious thought is needed in this area to develop a testable theory of norms and peer effects; but I doubt that the absence of such thought will limit the burgeoning literature on these topics. So the question is how FE data can contribute to the literature. Knowing the composition of work groups, including the wage rates of their members, their productivities, and how they interact, could allow us to distinguish between social effects and price responses. This would enable us to go beyond such initial serious efforts that are based on case studies (e.g., Mas and Moretti, 2006). I am not sure how to construct these tests, but having spent five years as a manager I do believe such hard-to-measure interactions are important, and I know that they are understudied.

Aside from a general intellectual interest, taking the idea of peer effects and social norms seriously in labor economics could be very useful. One might argue that the

creation of beneficial social norms might be the central focus of human resource practices. More broadly, maybe there are certain public policies that one could introduce that might engender more positive peer effects in the workplace. Studying these issues using large FE data sets is a way that we labor economists can both ground our research more closely in workplace behavior while still basing it on serious theory (that we need to develop) and sound econometric analysis.

d. Macroeconomic Topics

Much of macroeconomic research these days is plain silly: Calibrating time series of divorce rates, for example, does not speak to the major macroeconomic issues of our day.¹ Indeed, one might view the successes of macroeconomics that occurred between the late 1960s and the early 1980s, particularly grounding it in microeconomic theory, as causing the demise of macroeconomics as a central field and as having made it into an applied field, no different from labor economics or any other. Despite the silliness of much of the research and the jihads among the various macroeconomic sects, macroeconomics is probably still the area in which economists can have the most impact and in which good applied research is very widely cited. Perhaps the best indication of this is Stigler's (1976) apologia for economics, in which that leading micro-economist's apoloigia for our existence was based on what economists have done to smooth out business cycles.

I already discussed how longitudinal household and establishment data have enabled macroeconomists to explore the empirical micro-foundations of their subject.

¹At the Conference I made this point somewhat differently: "Gestern Nachmittag bin ich durchs Park gelaufen, und ich habe eine Menge Gänse entgegnet. Es fiel mir ein, dass die Gänse sehr ähnlich die heutigen Makroökonomien sind. Erstens reden sie nur miteinander und wollen nicht mit den anderen Tieren zu tun. Zweitens machen sie immer ein grosser Lärm, und drittens, nachdem sie weggefliegen sind, bleibt nur ein grosser Misthaufen."

Additional nice work using search theory to examine the macro-dynamics of vacancies and unemployment has been produced recently by Shimer (2005) and others. How much more, however, could be learned if the data were not just employee-based, but also allowed following matches of workers and firms as vacancies arose and workers became unemployed? Here is one area where the availability and use (which would follow from the availability) of longitudinal FE data would have a very big payoff. Indeed, applying formal theory to such data would allow us to answer a crucial question that has essentially not been studied for three decades: What do we mean by a job vacancy? How does the existence of job vacancies interact with both measured and unmeasured characteristics of workers, those currently in the firm and those potentially supplying themselves for jobs? Such an analysis would allow measuring job flows **and** worker flows in a way that goes beyond accounting for their cyclical changes.

FE data also allow us to examine other heretofore unstudied (or at least unanswered) questions about macroeconomic labor-market effects. For example, what are the financial effects on the cycle—how do changes in firms' liquidity alter their hiring/firing behavior, and how does this interact with labor-market flows? Related to this, how do cyclical changes in government spending affect cyclical activity by altering consumption, wealth and firms' balance sheets? What are the margins of adjustment of effort over the cycle, in terms of hours of work, intensity of work and employment? We know a lot about this question from establishment data and household data separately, but having matched data would speak toward the vertical double-arrow in Figure 1 and solve potential mis-specification problems.

Much closer to my own research interests, current high-tech research on adjustment costs still relies on inferring those costs from observed net changes in employment. This is simply wrong (Hamermesh, 1995): Hiring and firing costs are associated with hiring and firing, not with net changes in employment, since quits are ubiquitous. With data on both job flows and matched worker flows, this distinction can be made and embedded in an appropriate econometric model.

e. Program Evaluation

Card's (1990) pioneering examination of a particular "natural experiment" used double-differences in outcomes to evaluate policy impacts and led to an explosion of searches for such experiments and to double-differencing in labor economics. Many of these searches seem only very loosely based in any underlying economic problem (but see, e.g., Costa, 2001); and much of this line of research, which now is slowing down, is also close to devoid of any grounding in economic theory. FE data could well contribute to this activity and bring it a little bit closer to economics.

We could, for example, learn something from such data about the impact of changes in the minimum wage on wage, employment and hours changes among different groups of workers, both those directly affected and others. Similarly, while studying overtime restrictions in the U.S. is not too interesting because of the paucity of temporal and spatial variations in the law, in Europe those variations are large. With FE data one could be able to do the standard evaluation research on the impact of those changes (e.g., Crépon and Kramarz, 2002), but more important, could also account for how match effects alter the outcomes in ways correlated with changes in the provisions governing overtime. Similarly, while interesting evaluation work has used standard data sets and

natural experiments on laws governing compulsory schooling (e.g., Oreopoulos, 2006) to examine the returns to schooling, here again there is the strong possibility that the vertical double-arrow in Figure 1 matters. With longitudinal FE data the labor-market sorting that occurs when legislated changes alter the supplies of workers of different skill (education) levels can be properly accounted for, and the impacts of those laws can be assessed more accurately.

f. Trade Unions

Coming from a country that essentially has no unions except in the public sector, I am somewhat hesitant about discussing research on the labor-market effects of unions. The American industry of economic research on unions that flourished in the 1960s until the mid-1980s has essentially died off. That is unsurprising, as unions increasingly approximate Henry Simon's Cheshire Cat analogy; but they may come back, and they are certainly still important in the private sectors of many European economies. It may just be my nostalgia, but there still are interesting questions here. We do not know a lot about the things studied in that now hoary literature. The impact of unions on productivity, much studied by Richard Freeman and his students (e.g., Freeman and Medoff, 1984), depends on changes in both firm and worker behavior and could be re-examined profitably using matched data. Similarly, while gross union wage effects have been measured *ad nauseam*, the impacts of unobservable characteristics of firms and workers on wages (and other outcomes) has not been studied; and considering them would very likely alter our inferences about the sizes of these union impacts.

Even though trade unions do not matter much in the U.S. labor market, their importance in Europe more than justifies extending existing research to take advantage of

FE datasets. Some interesting work has already been done using such data (Card and de la Rica, 2005); but there are many more topics that could usefully be studied even on fairly narrow issues of labor-market impacts. More important, such data may even be used to inform the study of questions involving the political-economic role of unions and their indirect effects on labor-market outcomes.

4. Expanding Data Frontiers

a. Fruitful Additional Links Among Existing Data Sets

While tremendous steps have been taken in linking datasets, even in such relatively backward countries as the U.S., there are still many fruitful routes that could be taken at relatively low cost to link existing data sets. One of the most cost-effective surveys used by labor economics is the old EOPP data, a survey of employers in metropolitan areas that obtained information on them and their most recent hire (see, e.g., Barron and Black 1985). A similar recent dataset is the four-city survey by Holzer and Danziger (2001). These are both firm-employee data, without any matching, with information only on a few workers, but with vacancy data.

Since 2000 the U.S. has had vacancy data from large numbers of firms; and other countries have much longer series of micro data on vacancies. There would be a tremendous payoff to linking these data to available data from other sources, such as Social Security records, to learn about the characteristics of those who fill vacancies, those whose departures create them, and other aspects of the creation or death of job matches. Still more useful would be a link between individual data and the mobility data that (at least in the U.S.) accompany vacancy data. That would allow us to follow individuals and learn how the characteristics of firms and workers interact to generate

mobility of different types, e.g., layoffs versus quits versus fires. It would be a big step beyond the research that examines gross job flows and worker separations and accessions jointly (Anderson and Meyer, 1994; Hamermesh *et al.*, 1996). With such linked information we could obtain a complete picture of the dynamics of the labor market. We would be able to get a serious handle on the adjustment costs facing firms, and on the choices workers make when confronting matches offering different observed and unobserved (by the researcher) characteristics. Such data would allow us to go much further in the areas of job-matching and macroeconomic effects on the labor market.

b. New Matched Firm-Employee Data

Thus far I have lauded the progress that has been made in linking firm and employee data, have indicated a number of additional research topics to which existing linked data could be applied, and have outlined a few valuable new links of existing data sets. There are other areas where the recent creation of new household data should be calling forth new firm data, with the possibility of linking the two. My own current hobby horse would have us aim toward answering the question: What do people do at work, and how do a firm's characteristics alter what its employees are engaged in doing?

My concern about this issue stems partly from increased attention to time-use data—always collected from individuals and always treating time on the job as one aggregate (except for time spent in breaks). While we spend much of our time analyzing work and recognize that it proceeds at different paces, we usually say that work is just what we are paid for. But that definition only makes good sense if we are explicitly piece-rate workers, which very few employees are. Moreover, viewing work-hours as homogeneous—ignoring variations in effort and failing to distinguish time spent learning

from time spent producing—means that we have to infer these distinctions from the output variations that they may generate. This puts us in the same position as biology before the invention of the electron microscope, much less the ability to sequence genes: Results had to be inferred from outcomes rather than from understanding the underlying process.

We need to know what happens during work—what people are actually doing on the job. Are they training, loafing, working on new tasks or working on their regular tasks? We need a survey of workers' time use on the job, with the time diaries based at the level of particular firms. Moreover, these diaries should be kept by people who have kept diaries describing a typical entire day. Half of the requirements for this new matched dataset already exist in the form of the household-based time use surveys (such as the U.S. American Time Use Survey). With the creation of a job-based complement to those data we will be able to study and to provide better answers to such questions as: 1) How fast is hourly labor productivity growing, and what are the firm and worker correlates of this growth? 2) What is the relationship between productivity at work and productivity at home? Specifically, how does production in the market interact with production at home in a world in which time prices and incomes are rising? 3) How are our inferences about the distribution of worker well-being altered when we consider variations in the distribution and burden of effort in the workplace? 4) How are workplace injuries affected by the timing of different activities on the job, and how do worker-firm interactions alter patterns of injury?² 5) How does what people actually do

²Galizzi *et al* (2006) provides an example of a simple form of the data-collection effort I am proposing, albeit restricted to several establishments.

on the job relate to promotion policies, to paths of wages and to worker mobility, and how do these relations differ by gender/race/ethnicity?

5. Conclusions

The worldwide growth of matched firm-employee datasets is a wonderful development, not only for the grist they provide to labor and other economists, but because they enable society to obtain information about itself and answers to questions of public importance. Much has already been accomplished with existing FE datasets, much more could be accomplished with them, and still more could be learned were we to create additional such data. Despite these accomplishments and the obvious potential for still more, however, I am somewhat hesitant about recommending devoting many more resources to this sort of activity.

Think about the number of studies that are written in one year using, for examples, the PSID, NLS surveys or the German SOEP. Then think about the number being done with FE datasets, which are, as I have tried to indicate here, much richer along several dimensions. Why the difference? I believe the answer is simple: These household datasets are publicly available, readily downloadable and can be freely used in the privacy of one's home (or at least at one's place of employment or study). Many FE datasets require one to relocate, at least temporarily, to a secure site where the data are housed; and in the U.S., unless you reside in one of a few places, the secure sites are very far away. (In my own case the nearest one is 1500 kilometers distant.) Failing direct access, one must send one's computer code to one of these sites where the programs can be run. If you only want to measure some averages, or obtain a few moments for calibration, that is no problem. I am absolutely convinced from my own empirical work,

however, that it is essential for the researcher to wallow in the data—to get down and dirty with the data—to understand what they are showing about the phenomenon of interest.

Appropriate access to FE data does not seem possible with the current arrangements in most countries. I am very sympathetic to employers' need for privacy (but why the need is implicitly so much greater for firm data than for household data escapes me). We economists have not been forceful enough in demonstrating the importance of better access to matched data for research that will benefit the public. We have acquiesced too readily to the slightest expression of concern about privacy. Incentives can be created to prevent any potential misuse of the data, and we must work them out with governmental agencies. Should we fail to do so, the research and policy potential of FE data will regrettably remain mostly untapped.

REFERENCES

- John Abowd, Francis Kramarz and David Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 67 (March 1999): 251-333.
- Patricia Anderson and Bruce Meyer, "The Extent and Consequences of Job Turnover," *Brookings Papers on Economic Activity, Microeconomics* (1994): 177-236.
- George Baker, Michael Gibbs and Bengt Holmstrom "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, 109 (Nov. 1994): 881-919.
- Oriana Bandiera, Iwan Barankay and Imran Rasul, "Social Connections and Incentives in the Workplace: Evidence from Personnel Data," Unpublished paper, London School of Economics, 2006.
- John Barron and John Bishop, Extensive Search, Intensive Search and Hiring Costs: New Evidence on Employer Hiring Activity," *Economic Inquiry*, 23 (July 1985): 363-82.
- B. Douglas Bernheim, "A Theory of Conformity," *Journal of Political Economy*, 102 (Oct. 1994): 841-77.
- Michael Burda, Daniel Hamermesh and Philippe Weil, "Total Work, Gender and Social Norms," Unpublished paper, Department of Economics, University of Texas at Austin, Dec. 2006.
- Mark Bryan, "Workers, Workplaces and Working Hours," ISER Working Paper No. 2004-25, University of Essex, 2004.
- David Card, "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, 43 (Jan. 1990): 245-57.
- and Sara de la Rica, "The Effect of Firm-Level Contracts on the Structure of Wages; Evidence from Matched Employer-Employee Data," National Bureau of Economic Research, Working Paper No. 11,829, 2005.
- Andrew Clark, "Unemployment as a Social Norm: Psychological Evidence from Panel Data," *Journal of Labor Economics*, 21 (April 2003): 323-51.
- Dora Costa, "Hours of Work and the Fair Labor Standards Act: A Study of Retail and Wholesale Trade, 1938-1950," *Industrial and Labor Relations Review*, 53 (July 2000): 648-64.

- Bruno Crépon and Francis Kramarz, “Employed 40 Hours or Not Employed 39: Lessons from the 1982 Mandatory Reduction of the Workweek,” *Journal of Political Economy*, 110 (Dec. 2002): 1355-89.
- Steve Davis and John Haltiwanger, “Gross Job Creation, Gross Job Destruction and Employment Reallocation,” *Quarterly Journal of Economics*, 107 (Aug. 1992): 819-63.
- Timothy Dunne, Mark Roberts and Larry Samuelson, “The Growth and Failure of U.S. Manufacturing Plants,” *Quarterly Journal of Economics*, 104 (Nov. 1989): 671-98.
- Steven Durlauf, “On the Empirics of Social Capital,” *Economic Journal*, 112 (Nov. 2002): F459-79.
- Robert Frank, *Choosing the Right Pond: Human Behavior and the Quest for Status*. New York: Oxford University Press, 1985.
- Richard Freeman and James Medoff, *What Do Unions Do?* New York: Basic Books, 1984.
- Monica Galizzi, Petra Miesmaa and Craig Slatin, “Occupational Injuries, Workers’ Reporting and Firms’ Policies in the Healthcare Industry,” Unpublished paper, University of Massachusetts-Lowell, 2006.
- Erica Groshen, “Sources of Intra-Industry Wage Dispersion: How Much Do Employers Matter?” *Quarterly Journal of Economics*, 106 (Aug. 1991): 869-84.
- Daniel Hamermesh, “The Costs of Worker Displacement,” *Quarterly Journal of Economics*, 102 (Feb. 1987): 51-75.
- , “The Source of Adjustment Costs,” *Economic Journal*, 105 (May 1995): 620-34.
- , Wolter Hassink and Jan van Ours, “Job Turnover and Labor Turnover: A Taxonomy of Employment Dynamics,” *Annales d’Économie et de Statistique*, 41-42 (January 1996): 21-40.
- James Heckman, “Heterogeneity and state Dependence,” in Sherwin Rosen, ed., *Studies in Labor Markets*. Chicago: University of Chicago Press, 1981.
- Harry Holzer and Sheldon Danzier, *Urban Inequality: Evidence from Four Cities*. New York: Russell Sage Foundation, 2001.

- Dean Hyslop, “State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women,” *Econometrica*, 67 (Nov. 1999): 1255-94.
- Louis Jacobson, Robert LaLonde and Daniel Sullivan, “Earnings Losses of Displaced Workers,” *American Economic Review*, 83 (Sept. 1993): 685-709.
- Derek Jones, Panu Kalmi and Antti Kauhanen, „Productive Efficiency: Evidence from a Food Processing Plant,“ Unpublished paper, Helsinki School of Economics, 2006.
- Boyan Jovanovic, “Selection and the Evolution of Industry,” *Econometrica*, 50 (May 1982): 649-70.
- Nicholas Kiefer and George Neumann, “An Empirical Job-Search Model, with a Test of the Constant Reservation-Wage Hypothesis,” *Journal of Political Economy*, 87 (Feb. 1979): 89-107.
- Alexandre Mas and Enrico Moretti, “Peers at Work,” Unpublished Paper, Department of Economics, University of California—Berkeley, 2006.
- Philip Oreopoulos, “Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter,” *American Economic Review*, 96 (March 2006): 152-75.
- Gerard Pfann and Daniel Hamermesh, “Two-Sided Learning, Labor Turnover and Displacement,” National Bureau of Economic Research, Working Paper No. 8273, 2001.
- Arthur Ross, “Do We Have a New Industrial Feudalism?” *American Economic Review*, 48 (December 1958): 902-20.
- Robert Shimer, “The Cyclical Behavior of Equilibrium Unemployment and Vacancies,” *American Economic Review*, 95 (March 2005): 25-49.
- Frank Stafford, “Forestalling the Demise of Empirical Economics: The Role of Microdata in Labor Economics Research,” in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics, Volume 1*. Amsterdam: North-Holland Press, 1986, 387-423..
- George Stigler, “Do Economists Matter?” *Southern Economic Journal*, 42 (Jan. 1976): 347-54.

Figure 1. Conditions Requiring Matched Firm-Employee Data

