R&D SPILLOVERS AND SCIENTIST AND ENGINEER LABOR MOBILITY

Erling Barth
Institutt for Samfunnsforskning, University of Oslo and IZA

James C. Davis
US Census

Richard Freeman
Harvard and NBER

Gerald Marschke
SUNY Albany and NBER

Andrew Wang
Harvard and NBER

Abstract:

Using Census data that link U.S. workers to their workplaces we estimate how much past employment in R&D-performing firms raises scientists' and engineers' (S&E) wages and thus their marginal product at new employers. We find that S&E workers who arrive at a new firm that is R&D-active having been exposed to R&D in previous employment are more productive. We then use Census data to estimate establishment-level production functions showing that both the establishment's own R&D and R&D activities by other firms that are geographically close and in the same industry have positive effects on establishment output. This evidence is consistent with previous studies. We then add to the production function a measure of external R&D brought to the establishment via newly hired high-wage workers. The coefficient estimate on incoming human capital-embodied R&D is both statistically and economically significant, but does not substantially reduce the coefficient estimates on external R&D. While this evidence is highly preliminary, it suggests that while worker mobility plays a role, it is not the dominant mechanism by which R&D spillovers among firms occur.

Thanks to Sifan Zhou for her invaluable research assistance.

I. Introduction:

Our focus is to understand the role of labor markets in promoting technological change and diffusion. To what extent does a firm's mixing of scientists, engineers, and R&D managers with diverse employment histories generate new ideas? How does inter-firm mobility of technical workers contribute to technology transfer and diffusion of new ideas? Answers to these questions will help us to understand the mechanisms behind economic growth. A better understanding of the role of scientist and engineer (S&E) labor markets in developing and disseminating new technologies has policy ramifications. If growth due to movement of ideas via worker mobility is important than policies that influence mobility in S&E labor markets and the movement of S&E workers among regions, states, and countries--e.g. job lock, intellectual property enforcement, immigration--should be considered in this light. For example, Gilson (1999) has argued that the rise of Silicon Valley is due in large part to the California courts' refusal to enforce noncompete clauses in employment contracts. Gilson suggests that this refusal coupled with the natural mobility of scientists and managers result in the rapid diffusion of technological innovation and in Silicon Valley's natural ability to continuously re-invent itself (Saxenian, 1994).

We use the U.S. Census Longitudinal Employer Household Database (LEHD), a matched worker-employer data set that allows one to follow individual workers over their career. We merge R&D data from the Survey of Industrial R&D (SIRD) to the LEHD to obtain for each employer a measure of R&D expenditures. From the employer's R&D spending, we derive at the worker level her R&D human capital stock. We find that workers who arrive at a new firm having been exposed to R&D in previous employment have higher earnings, possibly reflecting higher productivity.

We then estimate establishment-level production functions showing that both the establishment's own R&D and R&D activities by other firms within the same industry have positive effects on establishment output. This evidence is consistent with previous studies. We then add to the production function a measure of external R&D brought to the establishment via newly hired high-wage workers. The coefficient estimate on incoming human capital-embodied R&D is both statistically and economically significant, but does not substantially reduce the coefficient estimates on external R&D. While this evidence is highly preliminary, it suggests that while worker mobility plays a role, it is not the dominant mechanism by which R&D spillovers among firms occur.

The paper is organized as follows. The next section presents a literature review of the R&D productivity, spillover, and researcher mobility literature. Section III details our worker-level analysis of the impact of past R&D exposure on current productivity. Section IV describes our production function analysis, where we estimate the magnitude of R&D spillovers among firms and the role played by S&E mobility and facilitating these spillovers. Section V concludes.

II. Literature Review:

Econometric evidence suggests the size of R&D spillovers is large (Bloom, Schankerman, and Van Reenen, 2013) but the mechanism by which these spillovers occur is not well understood. Job-hopping researchers who carry the technology of their previous employers with them may be an important source of spillovers. Economists have long suspected that the firm-to-firm mobility of scientists transmits technological know-how across firms (Arrow, 1962; Stephan, 1996), but evidence is often anecdotal and econometric evidence is scarce. Levin et al. (1987) present survey evidence that firms count the hiring of R&D employees from innovating firms as a means of learning about new technologies. Stories in the press abound of firms in the technology sector poaching researchers (often times the whole team or laboratory) to access a rival's ideas. That spillovers among firms are generated by firm to firm worker mobility is consistent with evidence that spillovers among firm are influenced by geographic proximity (Jaffe, Trajtenberg, and Henderson, 1993).

A few studies examine the relationship between R&D and the mobility of higher educated or S&E workers (Almeida and Kogut,1999; Andersson et al. 2008; Kim and Marschke, 2005). Møen (2005) finds that technical workers in R&D-intensive firms accept lower wages earlier in their tenure with the employer for higher wages later. He argues that this finding implies that potential externalities associated with labor mobility are at least in part internalized in the labor market. He also estimates a wage equation that includes current R&D exposure, past R&D exposure with the present employer, and past R&D exposure with other employers. This sub-analysis is similar in spirit to the first part of our study.

III. R&D Human Capital

A. Empirical Specification:

We construct an unbalanced panel of S&E workers that contains quarterly data on employment, their earnings, and past R&D exposure. We analyze S&E workers from the quarter of their move and address the question: do incoming S&E workers with greater past R&D exposure earn more?

We estimate

$$w_{ift} = \delta_i + \alpha_1 x_{ift} + g(r_{ift}, c_{if't}, Ten) + \omega_{ift} \quad (1)$$

where w_{ift} is worker *i*'s wage in establishment *f* in quarter *t*; x_{ift} represents establishment, firm and worker controls; r_{ift} measures firm-level R&D expenditures at *f*; $c_{if't}$ measures knowledge capital levels *i* has brought from previous employer *f'*; *Ten* is tenure with establishment *f*; δ_i is a worker fixed effect, and ω_{ift} is the error term.

Table 1 describes the variables used. For w_{ift} we use quarterly earnings across all employers (in constant 2002 dollars). We use earnings because data do not include a direct measure of hourly wage, nor do they include hours from which we might infer wage. x_{ift} includes number of years of experience in the labor market, *Exp*, a measure of the general human capital the worker embodies, and tenure in quarters at the current firm, *Ten*, a measure of employer-specific human capital. We also include the squares of *Exp* and *Ten*. Because of the well-established finding that large employers pay higher wages, we include employment size both at the establishment (*Emp*) and the firm level. All specifications include industry (for the current employer) and year indicators.

Our proxy for the innovations that the worker is exposed to and may appropriate at her next employer is the amount of R&D expenditures at the previous job. ***g*** is that part of our regression specification that is concerned with R&D experience, where ***g*** is

$$\boldsymbol{g(r_{ift}, c_{if't}, Ten)} = \alpha_2 r_{ift} + \alpha_3 Ten \cdot r_{ift} + \alpha_4 c_{if't} + \alpha_5 Ten \cdot c_{if't} + \alpha_6 c_{if't} \cdot r_{ift} + \alpha_7 Ten \cdot c_{if't} \cdot r_{ift}$$

We use two measures of R&D human capital, c_{ift}. R&D human capital (*RDHC*) is the previous employer's average annual R&D expenditures over the period 1985-2007 in 2005 dollars.[1] This specification of R&D

[1] The previous employer is taken to be the employer of the worker in the quarter just prior to the quarter her primary employer changed. If in that prior quarter she had multiple employers, R&D expenditures is computed as the weighted average of R&D expenditures across all employers, where the weight given to a particular employer is the based on the earnings from that employer compared to overall earnings for the quarter.

human capital assumes that long and short exposures to R&D expenditures have equivalent effects on worker productivity. Thus in this formulation, R&D human capital does not accumulate in the same way that years of schooling does, and does not follow the worker around from employer to employer. (Future work may use formulations of R&D human capital that are cumulative.) We average the expenditures over a 22 year period because of the SIRD sampling strategy (smaller firms are surveyed only intermittently) and we think the annual R&D expenditure is plagued by measurement error. A dollar's worth of R&D exposure may have greater impact on the worker's human capital in a smaller firm than a larger one. Previous employer size takes care of any dilution in RDHC's impact by firm size. The second measure of c_{ift} is an indicator, *RDHCdum*, that is equal to one if *RDHC* >0 and equal to zero otherwise. We use the indicator variable because we think it is less sensitive to measurement error in R&D.

R&D expenditures at the current employer, r_{ift}, may also influence the worker's productivity. For r_{ift} we use the current employer's average annual R&D expenditures over the period 1985-2007 in 2005 dollars. We also use an indicator variable *RDCdum* that is equal to one if *RDC* >0 and equal to zero otherwise. Our prediction is that current employer's R&D expenditures have a positive effect on worker's wage ($\alpha_2 > 0$) because R&D exposes the S&E worker to new technologies and skills that the worker may exploit outside the firm. R&D expenditures within the firm may vary little from year to year. If R&D expenditures at the current employer have a cumulative effect on productivity then $\alpha_3 > 0$.

We are testing whether knowledge capital spills over from *f'* to *f* facilitated by *f*'s employment of an worker formerly employed by *f'*, that is, we are testing whether $\alpha_4 > 0$. If that knowledge capital depreciates or grows stale over time then $\alpha_5 < 0$, and if the productivity advantage accumulates with tenure at the new employer, $\alpha_5 > 0$. We expect that human capital R&D is more productive in R&D active firms and thus $\alpha_6 > 0$, and if that complementarity fades with tenure, $\alpha_7 < 0$.

B. Data:

The backbone of our study is the Longitudinal Employer-Household Dynamics (LEHD) database, a Census Bureau-administered database that contains quarterly employment and earnings data on individual workers matched to employer establishment. The LEHD does not provide occupational classifications but can be linked to both the Current Population Survey and the decennial census. We first use March CPS observations from 1990 to 2005 and the long form samples of the 1990 and 2000 decennial censuses to identify S&E workers. We then identify these workers in the LEHD and add in additional education and demographic information from the CPS, decennial census, as well as from the LEHD Individual Characteristics File (ICF). We are able to track every matched CPS and Census S&E worker across different employers observed in the LEHD Employment History Files quarterly data from 1997 through 2007, from establishment to establishment and firm to firm.

For most of the states included in the LEHD, 1990 was the first year of observation. As of 2004 31 states participated. 1997 was the first year the Census used the North American Industry Classification System or NAICS. Prior to that the Census used the Standard Industry Classification system.. To maintain a consistent industry classification system within the period of our analysis we therefore start with 1997.

The LEHD employer-level data are at an establishment-level and can be linked to Census Bureau establishment and firm level micro data via the Business Register Bridge. We use this bridge to link to the Census's Longitudinal Business Database (LBD) to obtain detailed annual data on each worker's employer at the establishment and firm level. The LBD is a census of all business establishments and firms with paid employees in all industries and in all states and is comprised of survey and administrative records. We use the LBD to obtain the employer's industry and number of workers at the firm and establishment level.

The LBD does not contain R&D information. To obtain employer R&D expenditures we link the LEHD to the Survey of Industrial Research and Development, another Census database. The SIRD samples approximately 25,000 firms each year. Firms with $5 million or more in R&D expenditure and firms that have reported positive R&D expenditures in the previous five years are included in the sample with certainty. Firms that have reported no R&D or for which R&D expenditures are uncertain are included in the R&D survey with low probability. We treat the absence of R&D expenditure as zero expenditure, though it is possible that the employer's R&D expenditure is positive but the employer was not sampled. We think that the missing R&D is most often zero,[2] but to check how safe we are with that assumption we will perform the following analysis. We will pool all of the firms in the SIRD panel 1985-2010. Then we will match the SIRD panel firms to year 2000 LBD, and examine the size distribution of the matched firms in year 2000. We will compare this to the size distribution of firms in the 2000 SIRD sample (using the sample weights in 2000 SIRD). This will assesses how well the full SIRD panel covers the non-certainty SIRD firms in the year 2000. We will include the results of this exercise in the next draft.

SIRD provides R&D figures only at the firm level. Finally we include in our analytical sample only those S&E workers from the LEHD that experience at least one change of primary employer between 1997 and 2007.

Turnover rates: According to IPUMS-CPS data, between 1997 and 2007, annual turnover rates for S&E workers averaged between 9% and 15%, compared to 12% and 16% for workers with college degrees, and 12% and 15% for workers with some college over the same period. In the sample we use for analysis, annual turnover rates average 35%, much higher than the IPUMS-CPS sample turnover rates because we are conditioning on one or more moves between 1997-2007.

Interestingly, most of the S&E workers in our sample work for employers with zero R&D expenditures. Also, when S&E workers move they tend to move within the same R&D category. Table 2 shows that about 79% of S&E workers who leave non-R&D firms in the first quarter of 2000 end up at another non-R&D firm. 70% of S&E workers who leave R&D-performing firms end up at another R&D-performing employer. By contrast, Table 3 shows that whether they start from a large or small employer (by employment size) moving S&E workers tend to move to smaller employers.

Descriptive statistics: Table 4 displays means and standard deviations of the principle variables used in our analysis. Our samples are constructed from S&E workers we have identified as having moved at least once between the first quarter of 1994 and the third quarter of 2007. Our analysis is at the worker-quarter level and for each worker in our analysis we include only the quarters within the 1997 to 2007 period after the worker's first observed move. The first column of Table 4 corresponds to worker-quarters for all employers, regardless of whether the employer conducts R&D. This is the largest sample we work with and it contains over 9 million worker-quarters. We will perform separate earnings analyses for S&E workers with R&D employers and with non-R&D employers. We therefore show descriptive statistics for the key variables of the analysis for non-R&D-performing employers in the second column and for R&D-performing employers in the third column. We identify an employer as R&D-performing if it reports positive R&D expenditures in the SIRD any time between 1985 and 2007, and as non-R&D performing, otherwise. An earnings analysis using only those worker-quarters in which both the employer reports positive R&D expenditures and the worker brings positive human capital R&D to their current employer from the previous employer is also conducted. The final column reports descriptive statistics for that subsample.

[2] The number of sampled companies that conduct R&D generally ranges from 3,500 to 4,000. The largest 1,000 companies with R&D activity account for 90% of total R&D expenditures in the U.S.

Table 4 shows that earnings of S&E workers in R&D-performing firms are much higher than in non-R&D-performing firms. Earnings also tend to be higher for workers who have R&D in their background (column 4 vs. column 3). S&E workers tend to be male, white and well-educated (average education is at least 15 years) with S&E workers with R&D-performing employers and with R&D backgrounds having the most years of education. Means and standard deviations of experience and tenure, and current employer's establishment and firm size (by number of workers) are also shown. Table 4 shows that firm and establishment size are both larger in the firms that conduct R&D.

The average annual R&D expenditure at the current employer is approximately $300 million (in 2005 dollars) for the whole sample, and over twice than that when we throw out the worker-quarters in which the current employer is not R&D-active (columns three and four). Workers' R&D human capital stock is reported in each quarter based on the R&D expenditures of the most recent previous employer. Table 4 shows that the average R&D human capital stock is higher for workers working in R&D performing firms ($500 million vs $300 million). The last two rows show the means of RDC and RDHC divided by firm employment (current employer's firm size for RDC and previous employer's firm size for RDHC).

<u>C. Results:</u>

Determinants of quarterly log earnings: Tables 5-7 show our estimates of (1). In addition to worker fixed effects, all specifications include industry and year indicators. Table 5 includes all S&E workers and Tables 6 and 7 include only workers with positive R&D human capital from the previous employer. The specifications in Table 6 use (natural) logged levels of R&D whereas in Table 7 the specifications include non-log-transformed levels of R&D. Each of these tables show separate analyses for R&D and non-R&D performing employers.

In Table 5, both the current employer's R&D expenditures and the worker's R&D human capital are in indicator form (*RDCdum*, *RDHCdum*). Consider first model I in Table 5. Model I includes all worker-quarter combinations, whether the employer is an R&D-performing firm and whether the worker has positive R&D human capital. The estimated coefficients on *Exp* and Exp^2 show that quarterly earnings rise with general experience in the labor market—about 8 percent with each year of work—but at a declining rate. The findings for experience are consistently repeated across specifications: the coefficients on *Exp* range between about .07 and .08 in Table 5 and also in Tables 6 and 8.

Holding experience constant, tenure with the current employer also increases earnings but by a smaller amount than experience: a one year increase in experience increases earnings by $0.0826-0.0022 \cdot Exp$, using the result from model I, i.e. by about eight percent for workers who have recently entered the workforce and by four percent for workers with 20 years of experience. Whereas a one year (four quarter) increase in tenure increases earnings by only 1.6 percent ($4 \cdot .004$) for newly hired workers; by approximately 12.5 years of tenure the positive tenure effect has disappeared (($.004 - .00008 \cdot 50 = 0$)). The coefficient estimate for *Ten* is much smaller for R&D performing employers compared to non-R&D employers (compare II and III). Tables 6 and 7 the effects of tenure are obscured due to multiple interactions between tenure and R&D-based measures, but perhaps for plausible values of RDC and RDHC, they follow the same pattern as in Table 5.

The size of the establishment in which the worker is currently working has a positive effect on log quarterly earnings: according to the estimates presented for Table 5 doubling establishment size increases earnings by about three percent for R&D-performing firms and 0.5 percent for non-R&D performing firms, and for Model I which includes both types of firms the implied change lies in the middle. The results for models IV-XI in Tables 6 and 7 show estimates of similar magnitude: the implied effect on S&E earnings of doubling establishment size clusters around four percent for non-R&D employers and one half percent for R&D employers. The establishment size effect for S&E workers in R&D employers is therefore negligible.

Holding establishment size constant, firm size increases earnings but by only a small amount: the point estimates are about .004 in Models I and III. In Model II, the estimate is positive but insignificant. In all other models, the coefficient estimate never exceeds .001. In Model IV, where R&D is measured as the log of the level and the earnings equation is estimated only over observations with employers with positive R&D expenditures and workers with positive R&D human capital, the firm size effect is negative and significant, though small in magnitude (-0.008).

We turn now to our measured effects of current employer R&D on S&E log earnings. Model I in Table 5 suggests a five percent earnings premium associated with working in an R&D firm. This is consistent with R&D spending producing an increase in *general* productivity. Model I also suggests that the wage-tenure profile of S&E workers in R&D firms is no different from profiles in non-R&D firms (the coefficient estimate on *Ten·RDCdum* is statistically indistinguishable from zero). By contrast, Moen found that wages were lower for technical workers but the wage-tenure profiles were steeper in R&D firms. In the split sample analyses (models II and III), the wage-tenure profiles are upward sloping but steeper in the non-R&D employers, again contrasting with Moen's results.

The coefficient estimates on *RDHCdum* are statistically significant for the whole sample (I) and the sample limited to R&D-performing employers (II). For non-R&D performing employers (III), the coefficient estimate is small in magnitude and insignificant. Model I's result suggests that S&E workers with non-R&D employers who begin employment with R&D human capital stock receive 1.5% higher initial earnings than S&E workers without R&D human capital, and this premium depreciates at a rate of .03 percentage points per quarter. For S&E workers with R&D-performing employers, the earnings premium associated with R&D human capital stock is 2.7% higher and depreciating at a rate of 0.11 (-.0011=-.0008-.0003) percentage points per quarter, taking about 10 years to vanish.

Models II and III imply new S&E workers receive no premium for R&D human capital in non-R&D firms, and a 3.9% premium in R&D firms. According to the estimated coefficients for *Ten·RDHCdum*, the RDHC premium declines with tenure, both in R&D and non-R&D firms; the 3.9% premium found in R&D firms takes 12 years to vanish (.0008·48.375=.0387).

In Table 6, we limit the analyses to workers with positive R&D human capital (*RDHCdum*=1). The first two columns, Models IV and V, estimate earnings regressions on the subsample of R&D employers; the last two columns, Models VI and VII, estimate the earnings regressions on the subsample of non-R&D employers. In Models IV and VI, the RDC and RDHC measures are based on the natural log of R&D expenditures. In Models V and VII, they are based on the log of per capita R&D. Because we include levels of R&D expenditures and estimate our models on workers or employers with positive R&D only, we think of our coefficient estimates for RDC and RDHC are estimates of the magnitude of the intensive margin of the effects of R&D on S&E earnings.

The coefficient estimates on *lnRDClev* in IV is about .011. Note that the coefficient on the interaction of RDC with RDHC is positive and significant and of course the sample upon which IV is estimated includes only S&E workers with positive R&D human capital. It makes sense then to evaluate the impact implied by this regression of RDC on starting quarterly earnings for a worker with *positive* RDHC. For a worker with mean RDHC for example the implied elasticity of earnings with respect to RDC is .014 (0.0115 + .0002*ln(748,980)) implying that doubling the employer's annual R&D expenditures increases the worker's starting quarterly earnings by about 1.4 percent.

Model IV can tell us how accumulating R&D exposure affects earnings. Counter-intuitively, the elasticity of earnings with respect to RDC falls for the worker with mean RDHC by .00007 ((.0003-.0000271·ln(748,980)= -.00007) with each quarter of R&D exposure.

The coefficient estimate on *lnRDHClev* (Model IV) is statistically significant by conventional standards and implies that a doubling of human capital R&D is associated with a 0.4 percent increase in initial earnings. Evaluated at the average employer R&D, the estimated elasticity of earnings with respect to RDHC is .0066 (.0039 + .0002·ln(669,174)) implying that a doubling of human capital R&D is associated with a 0.7 percent in quarterly earnings for a worker in an average R&D-performing firm. Consistent with the evidence presented on the extensive margin (Table 5), the intensive margin analysis suggests greater earnings premia for R&D human capital with employers who themselves spend more on R&D.

Model IV suggests that the wage premium due to R&D human capital diminishes with tenure at the new employer. The elasticity of earnings with respect to HCRD falls for the worker with mean RDC by.00016 with each quarter of tenure is (.0002-.0000271·ln(669,174)= - .00016). Takes 10 years for the starting elasticity of .0066 to be zeroed out.

The estimates generated from the sample of non-R&D firms (Models VI) show a weaker elasticity estimate for *lnHCRDlev* though statistically significant.

As a robustness check, we estimated the models of Table 6 with the R&D-based variables (*RDClev*, *RDHClev*) un-logged. These results are presented in Table 7. The coefficient estimates on *RDClev* is positive (Model VIII). The coefficient estimate in Model VIII implies that a $1increase in firm level R&D expenditures increases the worker's initial wages if we evaluate RDHC at the mean by over 400,000 percent (.0071 -.0054·748,980=-4044) which suggests this is a poor specification. Model VIII suggests similarly large in magnitude negative effects of raising RDHC.

D. Discussion:

We have run a set of alternative specifications that included the employment size of the firm that previously employed the worker, partly to pick up any differences in human capital investments correlated with size. This did not qualitatively alter the coefficient estimates for the variables of primary interest.

We also estimated separate sets of wage regressions for managerial, production and clerical workers. Generally the magnitude of the R&D human capital effects are smaller, though positive, for production and clerical workers. This gives us some confidence that R&D human capital as we have constructed it is picking up research-related skills and/or tacit technical knowledge. For managerial workers, the effects of R&D human capital on wages are also positive and in some specifications of greater magnitude than for S&E workers. Including a measure of the establishment's share of high wage workers with R&D human capital knocks down the coefficient estimate on the worker's R&D human capital (although the coefficient remains statistically significant). Approximately these results suggest that increasing the percentage of high wage workers in the firm by ten percent increases each worker's wage by three percent. We intend to explore these results and analyses of other, non-S&E occupational types in the next draft.

The evidence presented here indicates that S&E workers who begin employment with positive R&D human capital stock receive 4% higher initial earnings than new S&E workers without this human capital in R&D firms. This premium fades over the worker's tenure, consistent with obsolescence in whatever the new worker is conveying from her previous R&D active employer. In the following analysis, we attempt to assess the contribution of incoming, R&D-experienced workers to the establishment-level R&D spill-ins that have been found in the literature (e.g., Bloom et al.).

IV. Production functions, R&D spillovers, and worker mobility

A. Empirical Specification:

The R&D-productivity connection has been extensively explored in prior literature (see Griliches, 1998), including early studies at the U.S. Census Center for Economic Studies. The general model for measuring returns to R&D augments the production function to include "knowledge capital"

$$Y = AL^{\theta 1}M^{\theta 2}C^{\theta 3}K^{\theta 4}$$

where Y is output, L is labor, M are materials, C is ordinary (tangible) capital, and K is knowledge (intangible) capital. For estimation using a panel of establishments or firms, we write the model as

$$y_{it} = \beta_1 l_{it} + \beta_2 m_{it} + \beta_3 c_{it} + \beta_4 k_{it} + \eta_j + \lambda_t + \varepsilon_{it} \qquad (2)$$

where lowercase letters denote logarithms of the variables in the model, i denotes establishments, t denotes time, η_j are industry effects, λ_t is a time-specific effect, and ε_{it} is the error term.

This draft presents estimates of (2) at the establishment level, with R&D measured at the firm level.[3] The establishments used to estimate (2) are restricted to the manufacturing industry only. Table 8 describes the variables used in the production function analysis. Table 9 displays their means and standard deviations for the sample used in the production function estimations.

B. Results:

Table 10 presents the results of our production function estimations. The first column shows the estimation results of (2). All specifications include industry dummies and time dummies, in addition to the variables described in Table 10. The R&D measure is measured as R&D expenditures at the establishment's firm, averaged over 1985-2007 in constant dollars, divided by average firm employment. The R&D variable takes the form of a set of dummies: the first indicator is 1 if average R&D intensity is either zero or missing, the second through sixth indicators are for the five quintiles of the R&D intensity distribution (Table 8). The estimates for each of the variables in (2) are sensible and indeed the coefficient estimates on own-R&D intensity suggest that greater R&D expenditures correspond to greater establishment output.

Due to spillovers and diffusion to other firms, the private returns may not equal the social returns to R&D, generally measured in this literature with estimates using industry level observations. To estimate social returns to R&D, we add a spillover term to the production function to produce the second set of results (column 2 in Table 10). The spillover term is constructed in a similar manner as the R&D expenditures for the firm, but is based on the R&D expenditures of the firms of the establishments that are close to the establishment geographically and in the same 3-digit industry. Note that the coefficient estimates on *INDGEO_RDTOTDNE_FX*p (see Table 8) are positive and rise in magnitude with the quintile of external R&D intensity: that is, the greater the R&D expenditures of nearby firms in the same industry, the higher the establishment's output. At the higher quintiles, the spillover coefficient estimates are statistically significant by conventional significance standards.

[3] A focus of this project will be an investigation of methods to allocate R&D measures collected at the firm level to their establishments in order to construct establishment based R&D expenditure estimates, thus future drafts will include estimates of R&D at the establishment and not the firm level in our estimation of the establishment's production function.

The third specification includes the share of high-wage workers whose previous employer was R&D active (had positive R&D expenditures over the 1985-2007 period). A worker is deemed a high-wage worker if her earnings are greater than or equal to the earnings at the 25th percentile of the earnings of scientist and engineer workers. Because this is at the upper end of the wage distribution of all workers we are picking up in addition to highly educated scientists and engineers, managers, directors, and executives. Finding a positive coefficient on aggregated embodied R&D in (2) and that adding aggregated embodied R&D to (2) knocks down most or all of the estimated coefficient on spillover R&D would indicate that embodied R&D spillover is an economically important mechanism for spillovers. We find that indeed the coefficient estimate on HIGH_RDDM is positive, large and significant, but also its inclusion only has a small attenuating effect on the coefficient estimates of *INDGEO_RDTOTDNE_FX*p. This is consistent with a small role for labor mobility. This result stands up to a Levinsohn-Petrin-type control for unobserved heterogeneity (see Wooldridge, 2009)--these results are not reported.

C. Discussion:

In future work, we will consider other measures of both spillover R&D and R&D human capital to test this result further. One possible reason we have not found an important role for mobility in our results, is that the mobility we are measuring is not connected to the source of the spillover flows. We plan to address this for example by looking at the effect of labor mobility from specific firms on those particular firm's spillover effects. This is possible with the LEHD.

V. Conclusions

In our wage regressions, we find a five percent earnings premium associated with working in a firm that conducts R&D. This finding is consistent with R&D activities adding to S&E workers' *general* skill-sets and with spillovers. In R&D firms, Møen finds lower wages early in technical workers' tenure, but steeper wage-tenure profiles. Contrary to Møen , when we split the sample into R&D-active and R&D-inactive firms, we find R&D firms show flatter wage-tenure profiles than non-R&D firms (at least early the worker's in tenure).

The main focus of our wage equations is estimating the R&D human capital wage premium. New S&E workers who begin employment with positive R&D human capital receive a 1.5 percent higher initial earnings than new S&E workers without this human capital in non-R&D firms, and 4.2 percent higher initial earnings in R&D firms (*extensive margin*). When we split the sample, we find new S&E workers receive no premium for R&D human capital in non-R&D firms, and a 4 percent premium in R&D firms. The value of R&D human capital appears to obsolesce with tenure in both types of firms. For example, the four percent premium found in R&D firms takes about twelve years to vanish. Our analysis of the *intensive margin* shows that a higher level of R&D human capital is associated with higher earnings for S&E workers (the elasticity for R&D human capital is about .01). Also S&E workers' R&D human capital is complementary to the current employer's R&D spending, a finding consistent across specifications. Our analysis, however, does not tell us the source of the R&D human capital premium: is it training (e.g., how to conduct an assay), transfer of new ideas, or something else?

Having established at the worker level that R&D backgrounds are productivity-enhancing, we seek to evaluate the importance of inter-firm mobility in S&E workforces in explaining R&D spillovers found in the production function literature (e.g., Bloom et al.). We are in the preliminary stages of this inquiry. We estimate our own Griliches-type production functions that along with capital, labor, etc include measures of knowledge capital, both internally and externally produced. We proxy these two forms of knowledge capital with own-firm R&D expenditures and R&D expenditures of firms that are proximate, both in terms of industry and geography.   We show that these R&D expenditures by other firms have a positive influence on establishment output, though this influence is smaller than for the establishment's own firm R&D. We then add to the production function the share of high-wage workers (this includes S&E workers but also upper management) who were hired from other R&D active employers as a measure of human capital-embodied spillover R&D. We find that a firm with a greater share of high-wage workers hired away from other R&D-active firms have higher output, ceteris paribus. Adding the this measure of human capital-embodied spillover R&D however does not greatly attenuate the effect of external R&D expenditures. We thus do not (yet?) see that worker mobility is important in facilitating R&D spillovers.   It is possible that the mobility we are measuring is not connected to the source of the spillover flows. We plan to address this for example by looking at the effect of labor mobility from specific firms on those particular firm's spillover effects. This is possible within our data framework.

References

Almeida, P., and Kogut, B. "Localization of Knowledge and the Mobility of Engineers in Regional Networks," *Management Science*, Vol. 45 (1999), pp. 905-918.

Andersson, Frederick; Brown, Clair; Campbell, Benjamin; and Park, Yooki. "The Effect of HRM Practices and R&D Investment on Worker Productivity." in S. Bender, J. Lane, K.L. Shaw, F. Andersson, and T. von Wachter (eds.) *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*, University of Chicago Press.

Arrow, K, .J. "Economic Welfare and the Allocation of Resources for Invention." I n R.R. Nelson, ed., *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton, N.J.: Princeton University Press, 1962.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, Vol. 81, No. 4 (2013), pp. 1347-1393.

Foster, Lucia; Grim, Cheryl; Haltiwanger, and John Haltiwanger, 2013. "Reallocation In The Great Recession: Cleansing Or Not?," Working Papers 13-42, Center for Economic Studies, U.S. Census Bureau.

Gilson, R. J. "The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not To Compete." *New York University Law Review*, Vol. 74 (1999), pp. 575-629.

Griliches, Z. (1998) *R&D and Productivity: The Econometric Evidence*, NBER, University of Chicago Press.

Hall, Bronwyn H. (2007), "Measuring the Returns to R&D: The Depreciation Problem," NBER Working Paper 13473, October 2007.

Jaffe, Adam, Manuel Trajtenberg and Rebecca Henderson. (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics* 108 (3), pp. 577-598.

Kim, Jinyoung and Gerald Marschke (2005), "Labor Mobility of Scientists, Technological Diffusion, and the Firm's Patenting Decision," *Rand Journal of Economics*, 36(2):298-317.

Levin, R.C., Klevorick, A.K., Nelson, R.R., and Winter, S.G. "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity* 3, 1987.

Wooldridge, Jeffrey M., 2009. "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, Elsevier, vol. 104(3), pages 112-114, September.

Møen, Jarle. "Is mobility of technical personnel a source of R&D spillovers?." *Journal of Labor Economics* 23.1 (2005): 81-114.

Saxenian, A. L. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128.* Cambridge, Mass.: Harvard University Press, 1994.

Stephan, P. "The Economics of Science." *Journal of Economic Literature*, Vol. 34 (1996), pp. 1199-1235.

Table 1
Variables for Wage Regressions

Earnq	Quarterly earnings across all employers (in constant 2002 dollars)
Exp	Experience in labor market (years, Age-Education-6))
Ten	Tenure at employer (quarters)
Emp	Number of workers at employer (establishment-level, firm)
RDClev	Current employer's R&D expenditures: in 2005 dollars, GDP deflator; average annual R&D expenditures over 1985-2007
RDCdum	Dummy indicating whether current employer's R&D expenditures >0
RDCemp	Current employer's R&D expenditures/Current firm employment
RDHClev	Human capital R&D: in 2005 dollars, GDP deflator; previous employer's average annual R&D expenditures over 1985-2007
RDHCdum	Dummy indicating whether R&D human capital >0
RDHCemp	R&D human capital /Previous firm employment

Table 2
Mobility and R&D Activity

		Destination Firm	
		R&D=0	R&D>0
Origin Firm	R&D=0	8,362 .787	2,268 .213
	R&D>0	2,507 .296	5,955 .704

Year 2000, first quarter. Large: 2000+ employees

Table 3
Mobility and Firm Size

		Destination Firm	
		Large	Small
Origin Firm	Large	3,103 .384	4,987 .616
	Small	2,832 .291	6,898 .709

Year 2000, first quarter. Large: 2000+ employees

Table 4 Variables for Wage Regressions				
	Means (Standard Deviations)			
	All	RDCdum=0	RDCdum=1	RDCdum=1, RDHCdum=1
Earnq (2002$)	18,159.38 (32,890.26)	15,659.14 (22,830.61)	20,841.65 (40,855.98)	21,611.49 (42,072.86)
Age	42.68 (10.11)	42.39 (10.47)	42.99 (9.70)	43.54 (9.53)
Male	0.79 (0.41)	0.77 (0.42)	0.80 (0.40)	0.81 (0.39)
Education (years)	15.19 (2.31)	14.99 (2.34)	15.40 (2.27)	15.48 (2.25)
White	0.77 (0.42)	0.78 (0.41)	0.76 (0.43)	0.75 (0.43)
Exp (Age-Educ-6)	27.49 (10.19)	27.41 (10.51)	27.59 (9.84)	28.06 (9.70)
Ten (quarters)	11.26 (10.82)	10.12 (10.37)	12.48 (11.16)	12.63 (11.32)
Year	2,002.67 (2.97)	2,002.69 (2.98)	2,002.65 (2.96)	2,002.64 (2.97)
Emp (Establishment) (No. of workers)	1,228.80 (3,190.51)	953.77 (2,863.80)	1,523.86 (3,483.26)	1,648.09 (3,593.83)
Emp (Firm) (No. of workers)	24,545.13 (61,380.59)	8,919.08 (56,212.57)	41,308.88 (62,262.12)	43,153.02 (63,391.74)
RDClev (2005$)	297,539.20 (854,088.50)		616,741.60 (1,146,813.00)	669,173.70 (1,158,720.00)
RDHClev (2005$)	298,339.70 (950,249.70)	115,511.80 (547,168.60)	494,478.90 (1,214,978.00)	748,979.90 (1,430,146.00)
RDCemp (2005$/Workers, Current Firm)	12.10 (29.22)		25.07 (38.00)	26.70 (38.90)
RDHCemp (2005$/Workers, Previous Firm)	8.65 (22.59)	3.81 (15.41)	13.85 (27.41)	20.98 (31.44)
Observations	9,448,056	4,889,960	4,558,096	3,009,269

Table 5
Determinants of ln(Earnq)

	I	II	III
VARIABLES	All	*RDCdum*==1	*RDCdum*==0
Constant	7.7912*** (0.0086)	8.3917*** (0.0175)	7.6588*** (0.0114)
Exp	0.0826*** (0.0002)	0.0749*** (0.0002)	0.0814*** (0.0003)
Exp^2	-0.0011*** (0.0000)	-0.0010*** (0.0000)	-0.0011*** (0.0000)
Ten	0.0040*** (0.0000)	0.0017*** (0.0001)	0.0049*** (0.0001)
Ten^2	-0.00004*** (0.0000)	0.000005*** (0.0000)	-0.0001*** (0.0000)
Ln*Emp (estab)*	0.0173*** (0.0001)	0.0064*** (0.0002)	0.0287*** (0.0002)
Ln*Emp (firm)*	0.0040*** (0.0001)	-0.0001 (0.0002)	0.0035*** (0.0002)
RDCdum	0.0519*** (0.0008)		
Ten·RDCdum	0.0000 (0.0000)		
RDHCdum	0.0148*** (0.0008)	0.0387*** (0.0009)	0.0004 (0.0010)
Ten·RDHCdum	-0.0003*** (0.0001)	-0.0008*** (0.0000)	-0.0006*** (0.0001)
RDHCdum·RDCdum	0.0274*** (0.0011)		
Ten·RDHCdum·RDCdum	-0.0008*** (0.0001)		
Observations	9448056	4558096	4889960
Number of S&E workers	464953	278063	341827
R^2 (*within*)	0.0593	0.0392	0.0484

Standard errors in parentheses. Regressions include industry and year indicators and worker-specific fixed effects.

*** p<0.01, ** p<0.05, * p<0.1

Table 6
Determinants of ln(Earnq)
Workers with Human Capital R&D only

	RDCdum==1		*RDCdum*==0	
VARIABLES	IV *RDHC=lnRDHClev, RDC=lnRDClev*	V *RDHC=lnRDHCemp, RDC=lnRDCemp*	VI *RDHC=lnRDHClev*	VII *RDHC=lnRDHCemp*
Constant	8.3331*** (0.0245)	8.4449*** (0.0238)	8.0844*** (0.0497)	8.1106*** (0.0493)
Exp	0.0726*** (0.0003)	0.0713*** (0.0003)	0.0664*** (0.0007)	0.0661*** (0.0007)
Exp^2	-0.0009*** (0.0000)	-0.0009*** (0.0000)	-0.0010*** (0.0000)	-0.0010*** (0.0000)
Ten	-0.0036*** (0.0003)	-0.0009*** (0.0001)	0.0052*** (0.0002)	0.0050*** (0.0001)
Ten^2	0.0000*** (0.0000)	0.0000*** (0.0000)	-0.0000*** (0.0000)	-0.0000*** (0.0000)
Ln*Emp (estab)*	0.0047*** (0.0003)	0.0046*** (0.0003)	0.0368*** (0.0006)	0.0367*** (0.0006)
Ln*Emp (firm)*	-0.0075*** (0.0003)	0.0007*** (0.0002)	0.0001 (0.0004)	0.0002 (0.0004)
RDC	0.0115*** (0.0007)	0.0118*** (0.0003)		
Ten·RDC	0.0003*** (0.0000)	0.0002*** (0.0000)		
RDHC	0.0039*** (0.0006)	0.0109*** (0.0003)	0.0029*** (0.0005)	0.0071*** (0.0007)
Ten·RDHC	0.0002*** (0.0000)	-0.0001*** (0.0000)	0.0000 (0.0000)	-0.0001** (0.0000)
RDHC·RDC	0.0002*** (0.0001)	0.0004*** (0.0001)		
Ten·RDHC·RDC	-0.0000271*** (0.0000)	0.0000*** (0.0000)		
Observations	3009269	3009269	1250443	1250443
Number of S&E workers	203424	203424	158419	158419
R^2 (*within*)	0.0341	0.0338	0.0288	0.0289

Standard errors in parentheses. Regressions include industry and year indicators and worker-specific fixed effects.

*** p<0.01, ** p<0.05, * p<0.1

Table 7
Determinants of ln(Earnq)
Workers with Human Capital R&D only

	RDCdum==1		*RDCdum*==0	
VARIABLES	VIII *RDHC=RDHClev, RDC=RDClev*	IX *RDHC=RDHCemp, RDC=RDCemp*	X *RDHC=RDHClev*	XI *RDHC=RDHCemp*
Constant	8.4493*** (0.0238)	8.4677*** (0.0238)	8.1178*** (0.0493)	8.1173*** (0.0493)
Exp	0.0725*** (0.0003)	0.0711*** (0.0003)	0.0663*** (0.0007)	0.0663*** (0.0007)
Exp^2	-0.0009*** (0.0000)	-0.0009*** (0.0000)	-0.0010*** (0.0000)	-0.0010*** (0.0000)
Ten	-0.0003*** (0.0001)	-0.0010*** (0.0001)	0.0050*** (0.0001)	0.0051*** (0.0001)
Ten^2	0.0000*** (0.0000)	0.0000*** (0.0000)	-0.0000*** (0.0000)	-0.0000*** (0.0000)
Ln*Emp (estab)*	0.0050*** (0.0003)	0.0048*** (0.0003)	0.0368*** (0.0006)	0.0369*** (0.0006)
Ln*Emp (firm)*	0.0009*** (0.0002)	0.0013*** (0.0002)	0.0001 (0.0004)	0.0001 (0.0004)
RDC	0.0071*** (0.0007)	0.0002*** (0.0000)		
Ten·RDC	-0.0002*** (0.0000)	0.0000*** (0.0000)		
RDHC	0.0167*** (0.0007)	0.0004*** (0.0000)	0.0016 (0.0017)	0.0001** (0.0001)
Ten·RDHC	-0.0002*** (0.0000)	-0.0000*** (0.0000)	-0.0002*** (0.0001)	-0.0000*** (0.0000)
RDHC·RDC	-0.0054*** (0.0003)	-0.0000*** (0.0000)		
Ten·RDHC·RDC	0.0001*** (0.0000)	0.0000*** (0.0000)		
Observations	3009269	3009269	1250443	1250443
Number of S&E workers	203424	203424	158419	158419
R^2 (*within*)	0.0328	0.0331	0.0288	0.0288

Standard errors in parentheses. Regressions include industry and year indicators and worker-specific fixed effects.

*** p<0.01, ** p<0.05, * p<0.1

Table 8
Variables used in production function analysis

Name	*Description*	*Source*
KEQ	Capital stock, equipment	Chiang-Haltiwanger TFP files*
KST	Capital stock, structures	Chiang-Haltiwanger TFP files*
TE	Establishment total employment	Census of Manufacturers (CMF)
M	Materials expenditure	Census of Manufacturers (CMF)
E	Energy expenditure	Census of Manufacturers (CMF)
MEDUC	Mean education of workforce (years)	Decennial Census match
MU	Indicator for multi-establishment firm	Longitudinal Business Database (LBD)
ESTAB_AGE	Establishment age (years)	Longitudinal Business Database (LBD)
PW_SH	Production worker share of workforce	Longitudinal Business Database (LBD)
HIGHWAGE_SH	High-wage share of workforce (cutoff is the 25th percentile of the S&E workforce)	Longitudinal Establishment Household Database (LEHD)
RDTOTNE_FXp	Average annual firm R&D expenditures per worker: *RDTOTNE_FX0* = 1 if average R&D expenditures is missing or zero, =0 otherwise; *RDTOTNE_FXp* = 1 if R&D expenditures is positive and in the p^{th} quintile, =0 otherwise, for (p=1,…5).	Survey of Industrial R&D (SIRD)
INDGEO_RDTOTNE_FXp	Average annual R&D expenditures per worker within other firms in same industry and nearby: *INDGEO_RDTOTNE_FX0* = 1 if average R&D expenditures per worker is missing or zero, =0 otherwise; *INDGEO_RDTOTNE_FXp* = 1 if average R&D expenditures per worker is positive and in the p^{th} quintile, =0 otherwise, for (p=1,…5).	Survey of Industrial R&D (SIRD)
HIGH_RDDM	Share of high-wage workers whose previous employer was R&D active	LEHD, LBD, SIRD
*See Foster, Grim, and Haltiwanger (2013).		

Table 9 **Variables used in production function analysis**		
Name	*Mean*	*Standard Deviation*
Ln(Output)	10.10818	1.287095
ln*KEQ*	8.822684	1.410306
ln*KST*	7.416199	1.919485
ln*TE*	4.904034	.9273277
ln*M*	9.220876	1.500308
ln*E*	5.653331	1.537521
MEDUC	12.24069	.988208
MU	.6714485	.4696909
ESTAB_AGE	19.65393	9.00056
PW_SH	.7239675	.1861581
HIGHWAGE_SH	.824961	.0677311
RDTOTNE_FX2	.1268262	.3327814
RDTOTNE_FX3	.1118389	. 3151711
RDTOTNE_FX4	.0886972	.2843089
RDTOTNE_FX5	.0782967	.2686404
RDTOTNE_FX6	.065027	.2465758
INDGEO_RDTOTNE_FX2	.1239571	.3295357
INDGEO_RDTOTNE_FX3	.1707124	.376261
INDGEO_RDTOTNE_FX4	.2102948	.4075218
INDGEO_RDTOTNE_FX5	.2024425	.4018241
INDGEO_RDTOTNE_FX6	.2036506	.4027159
HIGH_RDDM	.160265	.2295784

Table 10 Determinants of ln(Output)			
	I	II	III
VARIABLES	Base	I+Spillover R&D	II+Mobility R&D
ln*KEQ*	0.0576*** (0.0023)	0.0576*** (0.0023)	0.0572*** (0.0023)
ln*KST*	0.0162*** (0.0014)	0.0162*** (0.0014)	0.0163*** (0.0014)
ln*TE*	0.3484*** (0.0033)	0.3483*** (0.0033)	0.3485*** (0.0033)
ln*M*	0.4561*** (0.0020)	0.4560*** (0.0020)	0.4553*** (-0.0020)
ln*E*	0.1177*** (0.0023)	0.1174*** (0.0023)	0.1172*** (-0.0023)
MEDUC	0.0344*** (0.0024)	0.0343*** (0.0024)	0.0324*** (0.0024)
HIGHWAGE_SH	0.8813*** (0.0332)	0.8781*** (0.0332)	0.8786*** (0.0332)
RDTOTDNE_FX2	0.0494*** (0.0057)	0.0485*** (0.0058)	0.0455*** (0.0058)
RDTOTDNE_FX3	0.0645*** (0.0060)	0.0629*** (0.0061)	0.0594*** (0.0061)
RDTOTDNE_FX4	0.1014*** (0.0066)	0.0987*** (0.0067)	0.0948*** (0.0067)
RDTOTDNE_FX5	0.1230*** (0.0072)	0.1193*** (0.0073)	0.1152*** (0.0073)
RDTOTDNE_FX6	0.1790*** (0.0084)	0.1749*** (0.0085)	0.1711*** (0.0086)
INDGEO_RDTOTDNE_FX2		0.0139 (0.0079)	0.0126 (0.0079)
INDGEO_RDTOTDNE_FX3		0.0152* (0.0076)	0.0142* (0.0076)
INDGEO_RDTOTDNE_FX4		0.0185* (0.0076)	0.0176* (0.0076)
INDGEO_RDTOTDNE_FX5		0.0260** (0.0080)	0.0250** (0.0080)
INDGEO_RDTOTDNE_FX6		0.0293*** (0.0088)	0.0285*** (0.0088)
HIGH_RDDM			0.2455*** (0.0300)
HIGH_RDDM2			-0.2573*** (0.0336)
Observations	53000	53000	53000
R^2	0.9150	0.9150	0.9150
Standard errors in parentheses. Establishment-level, includes years 1997, 2002, and 2007. Regressions include industry and year indicators, as well as establishment age and age squared, MU, and PW_SH. Observation counts rounded to insure confidentiality. *** p<0.01, ** p<0.05, * p<0.1			