Working Long Hours and Career Wage Growth*

Dora Gicheva†

September 2009

Abstract

This study establishes empirically a positive relationship between hours worked per week and growth in hourly wages. A four-wave panel survey of men and women who registered to take the Graduate Management Admission Test between June 1990 and March 1991 is used to show that this relationship is especially strong for young professional workers, it is also present in the more broadly representative NLSY79. I find the relationship to be nonlinear: in the GMAT Registrant Survey, for workers who put in 48 or more hours per week annual wage growth increases by 2 percent per 10 extra hours worked per week. The average effect is zero when hours are less than 47. The positive effect of hours on wage growth can be accounted for both by a learning-by-doing model with heterogeneous preferences for leisure, and by a model of promotions that combines higher per hour productivity in upper levels of the job ladder with worker heterogeneity. Using data on promotions and training, I provide evidence in support of the job-ladder model.

1 Introduction

The relationship between working hours and career advancement has received attention from various fields within the social sciences; Blair-Loy (2004) notes, "scholars have

*I am grateful to my advisors, Joseph Altonji and Fabian Lange, for their guidance and support. I would like to thank the participants in the Labor and Public Finance Seminar and the Graduate Summer Workshop at Yale for their valuable comments. I also wish to thank Steven Bednar, Juan Eberhard, Melissa Tartari, and Pat Kline for numerous helpful discussions. All errors are my own.

†Department of Economics, Yale University. E-mail: dora.gicheva@yale.edu

lamented how firms reserve the best jobs and opportunities for 'ideal workers', who can give long hours to their employer without being encumbered by family responsibilities." Reich (2001) points out that increasing earnings inequality means that taking a second-grade job that allows for more free time requires one to give up more current and future income. Akerlof (1976) is among the first economists to propose the idea that the labor market equilibrium can entail inefficiently long hours.

This paper links long hours to future career outcomes, in particular promotions and wage growth. It is the first of its sort to document empirically the nonlinearities in the relationship between working hours and career wage growth and to propose a theoretical framework that can be used to explain the observed relationship. I focus on percentage wage growth to avoid the endogeneity associated with the direct labor supply relationship between hours and current period wages. The empirical evidence from a four-wave panel survey of registrants for the Graduate Management Admission Test (GMAT) suggests that for young professional workers who usually put in overtime, working ten extra hours per week increases annual wage growth by around 2 percent: an effect similar in magnitude to obtaining an MBA degree.

I propose a theoretical model to explain these findings that is built upon the idea that hourly output is higher in upper-level positions within the firm and therefore employers prefer to promote workers who are consistently willing to put in long hours. The theory is based on the model of promotions with human capital accumulation in Gibbons and Waldman (1999) but adds disutility of hours that differs across workers. Willingness to work extra hours and ability are strategic complements in this model. Employees differ in their disutility of hours but their preferences are constant over time. Workers who put in very long hours or who have relatively high ability to learn on the job are the first to be promoted and experience faster wage growth.

When facing lifetime utility maximization, workers might find it optimal to choose hours that are higher than what current period utility maximization would imply. Decreasing returns to learning-by-doing combined with increasing observed wages make this framework similar to the intertemporal labor supply theory in Imai and Keane (2004).

In this study I focus on the high-end labor market; all workers in the GMAT Registrant Survey sample are college-educated (many of them holding degrees from top institutions) and tend to earn substantially more than the national average. White-collar jobs are prevalent in the dataset. The paper adds to the literature spurred by the recent increase in interest in studying this segment of the labor force, particularly with the upward trend of participation of women in such professions.[1]

I focus specifically on the future career implications of working longer or shorter workweeks; thus, my study can cast further light on the problem of the gender gap in earnings in white-collar occupations. Studies show that there remains a considerable gap in male and female earnings, even when controlling for occupation and undergraduate major (e.g. Goldin and Katz 2008, Montgomery and Powell 2003, Graddy and Pistaferri 2000).[2] Most studies of the gender gap in earnings, with the exception of Goldin and

[1]Goldin and Katz (2008) use data from the Harvard and Beyond study to examine the employment and family composition trends of Harvard alumni who graduated between 1969 and 1992; they find that the proportion of women who obtain an MBA degree increased from 5 to 14 percent while for men the change is from 11 to 19 percent.

[2]Montgomery and Powell (2003) use the first three waves of the GMAT Registrant Survey to compare GMAT takers who obtained an MBA to those who did not and find lower gender wage gap among business degree holders. Arcidiacono, Cooley and Hussey (2008) use all four waves and include fixed effects and additional program type and quality controls to show that the returns to an MBA degree are actually lower for women. These seemingly contradictory findings can be reconciled by an observation in Bertrand, Goldin and Katz (2008). Their study follows the careers of MBA graduates of the University of Chicago Graduate School of Business and finds that male and female graduates start off with very similar earnings but quickly diverge and women start earning considerably less. The paper focuses on three factors contributing to this trend: differences in training prior to obtaining an MBA, different patterns of time out of work and differences in hours worked per week. The latter two are largely

Katz (2008) and Bertrand et al. (2008), do not control for differences in the labor supply of men and women, thus ignoring the disparity in weeks and hours worked and non-education spells out of the labor force. I find that hours account for some but not all of the gender difference in wage growth.

While the notion that working hours are tied to career outcomes is not new, there are no other studies that link the two in a dynamic context and provide a rigid theoretical justification for the observed relationship. Bell and Freeman (2001) describe qualitatively the idea that preferences for hours are linked to some unobserved productivity parameter, so employers base promotion decisions on revealed preferences. The survey argues in favor of this explanation for the fact that U.S. employees work longer hours than their German peers. A similar concept is expressed in Rebitzer and Taylor (1995). In their efficiency wage model hours are directly related to productivity through workers' propensity for shirking but work incentives are created through dismissal threats instead of promotion opportunities. In Sousa-Poza and Ziegler (2003) firms also use long hours to sort productive workers but no inference is made about the effect of long hours on the wage profile. Landers, Rebitzer and Taylor (1996) consider a setting in which firms establish long-hour norms that should discourage workers with preferences for short hours from hiding their true type. Their study also links long hours to promotions but its scope is limited to the specific hierarchy within law firm partnerships; the theory that I develop and my empirical findings can be applied to a more general labor market setting.

The idea that future wages increase with current period hours is also inherent in learning-by-doing models, but I show that a learning-by-doing model alone cannot ac-

determined by having children.

count for the nonlinearity in the relationship between current hours and future wage growth. This well-known class of models, as first developed by Arrow (1962) and Rosen (1972), suggests that time spent at work directly leads to human capital accumulation, and the more time a worker spends on the job, the faster her productivity grows. Similarly, the problem can be considered in an on-the-job training framework (as discussed in Becker (1964), Mincer (1962) and many others); the difference is that there exists a tradeoff between investment and current earnings but both are increasing functions of working hours. For a learning-by-doing model to yield the prediction that the effect of hours on wage growth is more pronounced at high levels of hours, it is necessary to assume a human capital production function that is convex in time spent at work, which is counterintuitive and in disaccord with previous studies on the subject.

The rest of the paper proceeds as follows. A theoretical model of promotions which predicts the observed empirical regularities is developed in Section 2. Section 3 describes the data and presents evidence that wage growth is an increasing function of hours when hours are high. Finally, data on promotions is used in Section 4 to test some other predictions of the job ladder model, and Section 5 concludes.

2 A Model of Promotions with Heterogeneous Preferences for Leisure

As the empirical evidence summarized above and presented in detail in Section 3 implies, the relationship between hours and wage growth is highly nonlinear. Hours worked have little or no effect on the change in log wages when hours are less than average, but the effect is positive and strong for higher values of hours. In this section I present a theoretical model that can be used to explain the observed relationship. In particular, I focus on wage growth in the context of a promotions model similar to Gibbons and Waldman (1999). I add heterogeneity in the utility function, which makes workers different in their valuations of leisure. Similarly to Landers et al. (1996), it is efficient to promote workers who in lower levels of the job ladder reveal low disutility of labor.

The basic idea behind the model is that there are two types of jobs: "career" jobs, which offer higher returns to skill accumulation, and "non-career" jobs, which do not offer promotion opportunities. Only workers who have low disutility of hours or high ability to learn self-select into the first type of jobs, which induces a kink in the resulting relationship between hours and the change in wages.

I simplify the Gibbons and Waldman (1999) model to two job levels and two time periods. Workers start off with effective ability η_1, which in the second period evolves according to

$$\eta_2 = \eta_1(1 + \theta_i h_1),$$

where θ_i measures inherent ability and h_1 denotes first period hours. Hourly output in position $j = \{1, 2\}$ equals $Y_{ijt} = d_j + c_j(\eta_{it} + \epsilon_{ijt})$, where $c_2 > c_1 > 0$ and $0 < d_2 < d_1$ and ϵ_{ijt} is an i.i.d. shock. I assume that productivity with no labor market experience is lower in job 2, and all workers start in the first level of the job ladder: $d_1 + c_1\eta_1 > d_2 + c_2\eta_1$. Two additional assumptions imposed on the parameters are that $c_1/c_2 < 0.5$ and $(d_2 + c_2\eta_1)/(d_1 + c_1\eta_1) > 0.65$, from which it follows that

$$c_1(d_1 + c_1\eta_1) < c_2(d_2 + c_2\eta_1). \tag{1}$$

Utility is separable in consumption and leisure, and disutility from hours is determined by a parameter b_i in the utility function. The higher b_i is, the stronger a worker's preference for leisure. Let

$$U(w_1, w_2, h_1, h_2, b) = \left(w_1 h_1 - b h_1^2\right) + \left(w_2 h_2 - b h_2^2\right).$$

I assume that b_i and θ_i are independent; a correlation in these variables would not add much to the current analysis. In addition, it is assumed that

$$2b_i > c_2 \eta_1 \theta_i \tag{2}$$

for all b_i and θ_i.

Firms are assumed to be price takers, and effective ability η_{it} and the disutility parameter b_i are observable, so wages equal $w_1 = E(Y_{i1t}) = d_1 + c_1 \eta_{it}$ and $w_2 = E(Y_{i2t}) = d_2 + c_2 \eta_{it}$.

There are two possible paths that wages can take. The first case occurs when a worker is not promoted in the second period. Assuming perfect foresight, she solves the optimization problem

$$\max_{h_1, h_2} \left[(d_1 + c_1 \eta_1) h_1 - b_i h_1^2\right] + \left[(d_1 + c_1 \eta_1 (1 + \theta_i h_1)) h_2 - b_i h_2^2\right].$$

Maximization yields for optimal hours

$$h^*_{1_{np}} = \frac{(d_1 + c_1 \eta_1)}{2b_i - c_1 \eta_1 \theta_i} = h^*_{2_{np}}.$$

Hours are positive under assumption (2).

Alternatively, some workers will be promoted; having this information ex ante they solve

$$\max_{h_1,h_2} \left[(d_1 + c_1\eta_1)h_1 - b_i h_1^2\right] + \left[\left(d_2 + c_2\eta_1\left(1 + \theta_i h_1\right)\right)h_2 - b_i h_2^2\right].$$

This yields

$$h_{1_p}^* = \frac{(d_1 + c_1\eta_1)2b_i + (d_2 + c_2\eta_1)c_2\eta_1\theta_i}{4b_i^2 - c_2^2\eta_1^2\theta_i^2}$$

and

$$h_{2_p}^* = \frac{(d_1 + c_1\eta_1)c_2\eta_1\theta_i + (d_2 + c_2\eta_1)2b_i}{4b_i^2 - c_2^2\eta_1^2\theta_i^2}.$$

Again, assumption (2) implies that hours are positive.[3]

The following lemma implies that for a given agent, choosing to be on the promotion path means that first period hours are higher compared to hours without promotion.

Lemma 1 For given b_i and θ_i, optimal period 1 hours with promotion are higher than optimal hours without promotion:

$$\frac{(d_1 + c_1\eta_1)2b_i + (d_2 + c_2\eta_1)c_2\eta_1\theta_i}{4b_i^2 - c_2^2\eta_1^2\theta_i^2} > \frac{(d_1 + c_1\eta_1)}{2b_i - c_1\eta_1\theta_i}.$$

Proof See Appendix.

Workers evaluate their utility and choose whether to be on a promotion path or not.

[3]Note that hours without promotion do not change over time, while hours with promotion decrease from period 1 to period 2. While this feature of the model is not realistic, it is fairly straightforward to change it without altering any of the other main predictions by adding a depreciation parameter in the human capital accumulation function and making agents impatient. These additional parameters are omitted here to simplify the exposition.

Utility without promotion equals

$$U_{np} = \frac{(d_1 + c_1\eta_1)^2}{2b_i - c_1\eta_1\theta_i},$$

which can also be written as

$$U_{np} = \frac{1}{2}\left(h^*_{1_{np}}(d_1 + c_1\eta_1) + h^*_{2_{np}}(d_1 + c_1\eta_1)\right).$$

With promotion utility equals

$$U_p = \frac{(d_1 + c_1\eta_1)^2 b_i + (d_2 + c_2\eta_1)^2 b_i + (d_1 + c_1\eta_1)(d_2 + c_2\eta_1)c_2\eta_1\theta_i}{4b_i^2 - c_2^2\eta_1^2\theta_i^2},$$

or

$$U_p = \frac{1}{2}\left(h^*_{1_p}(d_1 + c_1\eta_1) + h^*_{2_p}(d_2 + c_2\eta_1)\right).$$

Then for each θ_i there exists a cutoff $\bar{b}$ such that

$$\frac{(d_1 + c_1\eta_1)^2\bar{b} + (d_2 + c_2\eta_1)^2\bar{b} + (d_1 + c_1\eta_1)(d_2 + c_2\eta_1)c_2\eta_1\theta_i}{4\bar{b}^2 - c_2^2\eta_1^2\theta_i^2} - \frac{(d_1 + c_1\eta_1)^2}{2\bar{b} - c_1\eta_1\theta_i} = 0. \quad (3)$$

Equation (3) is a quadratic equation in $\bar{b}$, which has exactly one positive root:

$$\bar{b} = \frac{2(d_1 + c_1\eta_1)(d_2 + c_2\eta_1)c_2\eta_1\theta_i - (d_1 + c_1\eta_1)^2 c_1\eta_1\theta_i - (d_2 + c_2\eta_1)^2 c_1\eta_1\theta_i + \sqrt{D}}{4((d_1 + c_1\eta_1)^2 - (d_2 + c_2\eta_1)^2)}, \quad (4)$$

where

$$
\begin{aligned}
D =&(2(d_1+c_1\eta_1)(d_2+c_2\eta_1)c_2\eta_1\theta_i-(d_1+c_1\eta_1)^2c_1\eta_1\theta_i-(d_2+c_2\eta_1)^2c_1\eta_1\theta_i)^2+\\
&+4(2((d_1+c_1\eta_1)^2-(d_2+c_2\eta_1)^2)((d_1+c_1\eta_1)^2c_2^2\eta_1^2\theta_i^2-\\
&-(d_1+c_1\eta_1)(d_2+c_2\eta_1)c_1c_2\eta_1^2\theta_i^2))>0.
\end{aligned}
$$

Lemma 2 The cutoff $\bar{b}$ satisfies assumption (2).

Proof See Appendix.

Since the cutoff $\bar{b}$ is unique, in the second period employers promote all workers with $b_i<\bar{b}$. In addition, $\bar{b}$ is a linearly increasing function of θ_i, which implies that for workers with higher "ability to learn" θ the cutoff $\bar{b}$ is higher, so they can compensate for higher disutility of hours with faster skill accumulation on the job. If the wage differential between level 1 and level 2 jobs increases (d_2 or c_2 goes up), the cutoff $\bar{b}$ will be higher, but also for workers who receive a promotion hours increase in both periods. This is consistent with the idea described in Bell and Freeman (2001) that higher wage dispersion over the job ladder translates to a longer workweek.

In this model the wages of workers who do not get a promotion grow at a rate of

$$\Delta w=w_2-w_1=c_1\eta_1(\theta_i h_1),$$

so that

$$\frac{\partial\Delta w}{\partial h_1}=c_1\eta_1\theta_i.$$

With a promotion wage growth is

$$\Delta w = w_2 - w_1 = d_2 + c_2\eta_1(1+\theta_i h_1) - d_1 - c_1\eta_1$$

and

$$\frac{\partial \Delta w}{\partial h_1} = c_2\eta_1\theta_i > c_1\eta_1\theta_i.^4$$

For two workers with the same θ but $b_i < \bar{b}$ and $b_j > \bar{b}$ it holds that $h_{1i} > h_{1j}$ (using Lemma 1 and the fact that hours are decreasing in b), thus longer hours in the first period result in faster wage growth.

Without adding multiple job levels the predictions of this model about the effect of hours on wage growth would be similar to the predictions of a pure learning-by-doing model. I assume that the human capital accumulation function is linear in hours, which would predict a linear relationship between hours in period 1 and the change in hourly wages. If instead human capital accumulation was a nonlinear increasing function of hours, this function would have to be convex in order for the findings in Section 3 to hold. However, such functions are generally assumed to be linear or concave in hours, in which case hours will not have a more pronounced effect at higher levels.

Figure 1 shows simulation results for the promotions model described above.[5] The slope of the depicted relationship is close to zero for lower values of hours and becomes positive when hours are high. With the parameter values chosen for this simulation 22

[4] Alternatively, to be consistent with the empirical specifications in this paper, one can compare $\frac{\partial(w_2/w_1)}{\partial h_1}$ for workers who are promoted and those who are not. Using the fact that $\frac{c_1\eta_1\theta_i}{d_1+c_1\eta_1} < \frac{c_2\eta_1\theta_i}{d_1+c_1\eta_1}$, it is straightforward to show that (logarithmic) wage growth is faster with promotion.

[5] I simulate an economy with 1,000 workers. The parameter values are: $\eta_1 = 0.1; c_1 = 0.075; c_2 = 0.52; d_1 = 0.616; d_2 = 0.545$. The disutility parameter b and inherent ability θ are chosen such that $300b$ and 200θ have χ^2 distributions with 50 degrees of freedom.

percent of workers are promoted in the second period. Both θ and first-period hours have a positive effect on the probability of promotion: the coefficients of a probit regression are 0.0462 and 0.0063 respectively with corresponding standard errors of 0.0016 and 0.0002.[6]

3 Empirical Evidence of the Relationship between Long Hours and Wage Growth

3.1 Data

The main dataset that I use is a Graduate Management Admission Test (GMAT) Registrant Survey conducted by the Batelle Memorial Institute on behalf of the Graduate Management Admission Council. While the GMAT Registrant Survey is smaller in size than the 1979 cohort of the NLSY for example, there are aspects of the survey design that make the data more suitable for studying the effects of long hours on career wage growth. The sampled group of workers is much more homogeneous in terms of background (for example parents' education and income), education and occupation. It is possible to make more accurate inferences about unobserved productivity than in other data sets by using standardized test scores (GMAT), undergraduate and graduate grade point average and major, and quality of undergraduate and graduate schools attended. Using the fact that about 60 percent of those present in the main sample completed an MBA program by the last installment of the survey (and it is known which program they attended), I am also able to identify the effects of a graduate management degree.

[6] The average of h_1 is 2.0044.

In addition, respondents are college-educated and for the most part hold white-collar jobs, a high proportion of which are managerial positions, on which promotion decisions are important for firm performance. Many respondents tend to work overtime, which is the focus of this paper. Women in the data have high labor market participation rates, including full-time and overtime work. I am able to focus on wage growth over roughly five to six years: between the second and fourth waves.

The GMAT Registrant Survey is a longitudinal survey that was sent out to a sample of 7,006 people from the universe of all individuals who registered to take the GMAT between June 1990 and March 1991 and were living in the U.S. at the time of registration.[7] The survey consists of four waves; the first one was conducted shortly after registration for the test and had a response rate of 84 percent. The second wave was sent out about 15 months after test registration and had 4,833 respondents. Wave 3 took place 3.5 to 4 years after registration and received 4,533 responses. The final wave was conducted about seven years after respondents registered for the GMAT (3,769 responses). Almost all responses from waves 2 to 4 were from people who responded previously; only a few of those interviewed dropped out temporarily but then returned to the survey. Appendix A provides a comparison of the survey respondents who were still in the study and mailed a response to the last interview to those who dropped out; the group that remained in the survey is on average slightly more likely to be employed and to have a few more months of experience but is otherwise similar to their peers.

I use data on reported earnings and hours on the current job to construct hourly wages. All wages are measured in 1991 dollars; I drop observations for which the

[7] Examples of other studies using this dataset are Arcidiacono et al. (2008) and Montgomery and Powell (2003).

inflation-adjusted wage is less than 2 or more than 200. The dependent variable in all wage change regressions is annualized using information on the length of time between interviews. The initial installment of the survey asks about previous labor market experience; for subsequent interviews I use the detailed employment history that is available in the data to construct an actual experience variable. I drop observations that have missing data on hours in the first, second or fourth survey, wages in the second or fourth interview, as well as for workers who are self-employed in period 2. For workers with missing observations for age or experience in the second wave, mother's education, time between interviews or undergraduate major I impute these values.[8] The biggest restriction on the sample, given the high rates of enrollment in postgraduate education, comes from the fact that I select only respondents who report working at least 15 hours per week and have valid wages at the time of the second and last interviews. Of the 3,769 individuals who responded to the fourth survey, 3,232 have valid wage observations. For 800 of those there is no valid second wave wage observation. Most workers who are unemployed are attending school full time. When I further drop workers who report fewer than 15 hours per week, the final sample consists of 1911 workers: 1103 men and 808 women.

Table 2 contains descriptive statistics for the GMAT Registrant Survey and a NLSY79 sample. The GMAT Registrant Survey focuses on young professional workers; the average age at the beginning of the study is under 28 years. Hourly wages are higher than they are in NLSY79 and increase almost twofold between the first and last interviews.

[8]The number of imputed observations is 32 for mother's education, 27 for age, 36 for experience, 7 for undergraduate major and 164 for time between interviews. Results that exclude the imputed observations are available from the author upon request; there are no noteworthy differences between the two sets of estimates.

Family background also differs from NLSY79, with 31% of GMAT Survey respondents' mothers having a college degree (compared to about 8% in NLSY79). Half of the people in the sample completed an undergraduate major in business. A large portion of workers in the GMAT Survey sample hold a managerial position and this number increases over time. Minority respondents are oversampled, similarly to NLSY79.

The GMAT Registrant Survey asks respondents to report usual hours worked per week for each job in their employment history. I drop observations for which hours in wave 2 are less than 15, and if reported hours exceed 90 I set them equal to 90 in order to eliminate any bias caused by outliers.[9] Figure 2 shows the distribution of reported hours by wave. Hours increase slightly over time; the mean is 42.3 for the first survey and increases to 47.9 by the time of the last interview. It is not surprising that there is bunching of hours at 40 but the number of people who report working 40 hours per week decreases with time at the expense of more workers reporting higher hours[10].

The data suggest that while hours increase slightly with time, there is relatively little mobility across quantiles. The correlation in the sample between period 1 and period 2 hours is 0.471 and this number increases to 0.517 when I exclude respondents who were attending school at the time of either the first or second interview. Table 1 shows a more detailed breakup of the transitions across quantiles. Of those who started off in the highest third of the distribution, 66% remained in the highest quantile at the time of the second interview. If someone initially reported 40 hours or less, the probability that she is working 40 hours or less by the last installment of the survey is 58%.

The fact that a worker's hours are strongly correlated in the short run (between

[9]This change affects a total of 17 observations.

[10]The proportion of workers who report 40 hours for survey waves 1, 2, 3 and 4 is, respectively, 33.91%, 30.93%, 28.94% and 24.18%.

the first and second waves in the case of the GMAT Registrant Survey) is particularly useful in the estimation procedures. Both in the GMAT Registrant Survey and NLSY79 respondents are given the option of reporting earnings over their preferred time period. To construct hourly wages I divide reported earnings by hours worked per week and the relevant time measure. The dependent variable in the GMAT Registrant Survey specifications is the annualized change in hourly wages between waves 4 and 2 (a period of five to six years). NLSY79 specifications use the annualized change in hourly wages between years $t+4$ and t. As Borjas (1980) and Deaton (1988) among other point out, the division method is likely to make hours in period t endogenous and lead to biased estimates. Assuming that measurement error in hours is serially uncorrelated, using previous period hours in the regression will solve this problem.[11] Thus, I use as an explanatory variable hours from the first interview for the GMAT Survey specifications and hours in $t-2$ for the NLSY79 specifications.[12]

I use the data from the 1979 cohort of the National Longitudinal Survey of Youth to complement my analysis and to show that the results I derive are not unique to the universe of GMAT registrants. For some regression specifications it is useful to follow workers over a longer period of time. Appendix B describes how I construct the NLSY79 sample. I focus on male workers; my sample consists of 16,970 observations for 3,458 workers who did not change employers between t and $t+4$. For some specifications I use a bigger sample with 23,353 observations for 4,490 males that includes job movers. Table 2 shows summary statistics for the sample of nonmovers. Because of the small number of

[11] Alternatively, one can use hours in the current period but derive separate estimates for the changes in hourly and annual earnings, which will provide a rough bound for the true coefficient on hours. See Section 3.2.4 for more details.

[12] The latter allows me to use data for the years after 1994 when NLSY79 becomes biennial.

self-employed workers among the U.S. labor force, when I look at self-employed workers in Section 4.1 the larger size of NLSY79 is an important advantage.

3.2 Estimation Results

3.2.1 Partial Linear Model

Since the relationship between hours and wage growth is likely to be nonlinear, it is useful to start by estimating a nonparametric model. In particular, I follow the procedure developed in Robinson (1988) to estimate a partial linear model of the form

$$(\ln W_{i4} - \ln W_{i2})/t = Z_i\beta + f(Hours_{i1}) + \epsilon_i. \tag{5}$$

The variables included in Z, other than demographics, are dummies for whether the respondent completed an MBA program (differentiated by type: full-time, part-time or executive), undergraduate major and whether in school at the time of the second interview. All 1911 men and women in the sample are included in the estimation.

As Robinson (1988) points out, the Z variables are likely to be correlated with hours, which will bias the results if the nonlinear function f was estimated using $(\ln W_{i4} - \ln W_{i2})/t - Z_i\beta$. I follow the approach outlined in his paper and find $E[(\ln W_4 - \ln W_2)/t|Hours]$ and $E[Z_k|Hours]$ using a Nadaraya-Watson kernel estimator. The unexplained parts $(\ln W_4 - \ln W_2)/t - E[(\ln W_4 - \ln W_2)/t|Hours]$ and $Z - E[Z|Hours]$ are then used to estimate $\hat{\beta}$ by OLS and the nonlinear portion of equation (5) is estimated also by kernel regression using the residuals

$$\hat{\eta} = (\ln W_{i4} - \ln W_{i2})/t - Z_i\hat{\beta}$$

as the dependent variable. Figure 3 shows the results for $\hat{f}(Hours_1)$.

The graph suggests that the relationship between hours and the change in log wages is nonlinear; there appears to be little or no effect at low levels of hours and a positive slope after about 50 hours. Based on the shape of the estimated function from Figure 3, I model the relationship between working hours and wage growth using an equation with change in regime at some level of hours, which is to be estimated as a separate parameter in the model.[13]

3.2.2 Change in Regime Estimation with an Unknown Breakpoint: Nonlinear Least Squares

Table 3 shows the results from a nonlinear least squares estimation of an equation of the form

$$(\ln W_4 - \ln W_2)/t = \beta_0 + \beta_1 \min(Hours_1, k) + \beta_2 \max(Hours_1 - k, 0) + f(Z),$$

where k represents the point at which there is change of slope and $f(Z)$ is a function of individual characteristics other than hours worked. In this and other specifications Z includes gender (if both men and women are included in the regression), MBA degree (full-time, part-time or executive), as well as controls for age and age squared, race, marital status, number of children, whether mother has college degree, undergraduate major and whether in school at the time of the second interview. The dependent variable

[13] The relationship between hours and the change in log wages can also be estimated using a second, third and fourth degree polynomials. The estimates suggest that the third and fourth degree terms are not significant, but if a second degree polynomial is used, the coefficient on hours is negative and significant, while the coefficient on hours squared is positive and also significant. These estimates are available from the author upon request.

is $(\ln W_4 - \ln W_2)/t$, where t is the number of days between the second and fourth interviews divided by 365. This accounts for the fact that time between interviews is not the same for all respondents. This regression model imposes continuity; in subsequent specifications I relax this constraint but find no evidence of a jump at k.

I show results for men and women together because this yields better accuracy of the estimated coefficients due to the relatively small sample size. The wages of women grow at about 1.3 percent slower rate per year than men's wages. Table 3 implies that up to about 47 or 48 hours per week, hours worked do not have an effect on wage growth; the parameter β_1 is not statistically different from zero in any of the specifications. On the other hand, β_2 is positive and significant for the full sample and for men and positive but smaller and noisier for women. The results imply that when hours exceed 47, ten extra hours (about one standard deviation) increase annual wage growth by 2 percent. A test of the restriction that β_1 equals β_2 yields a p-value of 0.0007, which implies that the relationship is indeed nonlinear.

I find that a part-time graduate business degree has no effect on wage growth, while a full-time degree increases wage growth by 1.9 percent per year. Only 89 people in the sample obtained an executive MBA, so the estimated coefficients for this variable are consistently positive but noisy. The effect of a full-time MBA degree is much stronger for women than it is for men. Obtaining the degree on average increases female wage growth by 2.57 percent, while the effect on men's wages is about half of that.

3.2.3 Maximum Likelihood Estimation of Change in Regime with an Unknown Breakpoint

This section addresses three issues that arise with the nonlinear least squares estimation method presented in the previous section: the restriction that there is no jump at the breakpoint, the potential bias associated with the heteroscedasticity of the data before and after the breakpoint, and the possibility that the optimization process finds a local maximum at 47 hours. I show that none of these is a viable concern.

The approach in this section is based on the estimation technique that Quandt (1958) proposes, which is applicable when the independent variable that determines the nonlinearity is discrete. I sort the data so that hours are smallest for the first observations and increase as i goes up. If the breakpoint is at k hours, there is an m such that $h_i \leq k$ when $1 \leq i \leq m$ and $h_i > k$ for $m < i \leq N$. Then for a given k, the log likelihood function takes the form

$$\log L = -\frac{N}{2}\log(2\pi) - m\log\sigma_1 - (N-m)\log\sigma_2 \\ - \frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(y_i - f(Z_i) - a_1 - b_1 Hours_i)^2 - \frac{1}{2\sigma_2^2}\sum_{i=m+1}^{N}(y_i - f(Z_i) - a_2 - b_2 Hours_i)^2. \tag{6}$$

Here N stands for the number of observations, and σ_1 and σ_2 are the standard deviations of the errors for observations respectively before and after the breakpoint k. I maximize the objective function in (6) for every possible k (excluding the few points at each end). The set of parameter estimates is chosen based on the k for which the log likelihood function attains its highest level. Figure 4 shows the distribution of the maximized log likelihood values over all possible values of k for the combined sample of men and women. The function has several local maxima but attains its global maximum at 48 hours (the value for 49 hours is almost identical).

The results in Table 4 show the maximum likelihood estimates of (6) when $k = 48$. The estimates for b_1, b_2 and the coefficients on *Female* and the MBA dummies are essentially the same as those estimated by the nonlinear least squares method and reported in Table 3. The coefficient b_1 is very close to zero and actually negative, while b_2 is positive and significant and higher for men than for women: 0.0020 compared to 0.0013. The effect of a full-time MBA is again much stronger for women than it is for men: for women the coefficient on this variable is 0.0237 and is highly significant. The negative effect of hours on wage growth implied for lower levels of hours can be explained with the fact that at the time of the first interview 16 percent of the workers in the sample are working (full-time or part-time) and attending school (full-time or part-time) at the same time. This can cause their hours in period 1 to be low but increases future wages (assuming a positive effect of schooling on earnings), thus the slight downward slope at small values of hours worked.

The results in Table 4 also suggest that the data are indeed heteroscedastic: the error variance is smaller at lower levels of hours and the difference is statistically significant. However, as the similarity in the estimates presented in Tables 3 and 4 suggests, the heteroscedasticity of the errors does not affect the results.

The estimates also suggest that there is no discontinuity at k. For example, in column 1 of Table 4 the point estimate for a_1 is 0.3043 and the estimate for b_1 is -0.0005. Given point estimates for a_2 and b_2 equal to 0.1968 and 0.0018 respectively and a break point at 48 hours, I have $0.1968 + (0.0018)(48) = 0.2832$. This number is very close to the point estimate for $a_1 + kb_1 = 0.2803$.

Quandt (1958) also proposes a likelihood ratio test of the hypothesis that no switch occurred against the alternative that there is one change in slope. He defines the likelihood ratio λ as

$$\lambda = \frac{L(\bar{\Omega})}{L(\Omega)},$$

where $L(\Omega)$ is the unrestricted maximum of the likelihood function under the alternative hypothesis and $L(\bar{\Omega})$ is the restricted maximum under the null. The likelihood ratio test rejects the hypothesis of equal slopes with a p-value of 0.0002.

3.2.4 The Extent of Division Bias

To give a sense of the extent of measurement error bias, Table 5 shows the results of regressions similar to the ones in Table 3, but the hours variable in these specifications corresponds to hours reported in wave 2 instead of wave 1. The coefficient β_1 is biased downwards: it is equal to -0.0012 for the whole sample and is significant at the 0.1 percent level, compared to -0.0004 and not statistically different from zero in the corresponding specification in Table 3. Additionally, β_2 appears to have a considerable upward bias: the estimated coefficient is 0.0032 compared to 0.0018. In a similar regression with dependent variable equal to the annualized change in yearly earnings (not shown here but available upon request) the corresponding coefficient of β_2 is 0.0006, so there is reason to believe that the true value is between 0.0006 and 0.0032. The preferred point estimate of 0.002 falls in this interval. The cutoff point k also has a downward bias. It is estimated at 45.2 for the whole sample but is as low as 31.7 for the subsample of women. Part of the difference can be attributed to the possibility that current period hours affect wage growth more than hours in $t-1$ but this does not explain why β_1 is much smaller. There is strong reason to believe that measurement error is important.

3.2.5 Robustness Checks

This section presents several robustness checks as evidence that the empirical results from Section 3.2.2 are not driven by the specific functional form that I use, the way I select the sample or a peculiarity of the GMAT Survey data.

Table 6 shows the results from nonlinear least squares estimation when the dependent variable is $(\ln W_3 - \ln W_2)/t$. Note that the sample sizes for these regressions are smaller because some respondents are missing wage data for the third interview. Although noisier, the results are similar to the ones from Table 3. The cutoff k is somewhat lower for these specifications: the point estimates are between 42 and 43. The parameter β_2 is of similar magnitude (around 0.002) but its standard error is larger. On the other hand β_1 is zero or negative. The variables indicating whether the respondent has an MBA degree refer to the third wave. The effect of a part-time MBA degree is not statistically different from zero, while the coefficients on full-time MBA are larger than they are in Table 3. It is still true that women benefit from a graduate business degree more than men.

Table 7 tests the hypothesis that low hours can appear to have a negative effect on wage growth because of those workers who are also attending school at the time of their first interview: in column 2 I show nonlinear least squares results for the sample that excludes respondents who are attending school at the time of the first, second or fourth interview because one can argue that school attendance will affect the wages or hours that I use in my regression specifications. This restriction reduces the sample size considerably. As expected, the coefficient on β_1 is now positive but still insignificant.

The value for β_2 increases slightly to 0.0022. The coefficient on full-time MBA stays positive but loses significance; however, in the restricted sample only 11.33 percent of the 918 respondents hold a full-time MBA degree, which explains the noisier estimates. The estimate for the break point is similar for these specifications: the break is at around 48 hours.

Column 1 of Table 7 restricts the sample to workers who did not change employers between the second and fourth interviews. There are 485 such observations. One economic reason for this restriction is that changing jobs can be positively correlated with wage growth due to a better worker-employer match (Jovanovic 1979), and job movers might have different hours than stayers. Table 7 suggests that this is not the case: the estimated coefficients (with the exception of the coefficient on full-time MBA) are very similar to what was found using the full sample.

Columns 3 and 4 of Table 7 show the results from nonlinear least squares estimation when the sample is split into MBA holders and workers who did not have an MBA degree at the time of the last interview. The estimate for β_2 for workers with a graduate business degree is 0.0021 and significant at the 0.1 percent level; it equals 0.0016 and is not significant for the other subsample. The break point k is somewhat different for the two groups (47.63 versus 45.35). It is interesting that the difference in wage growth between men and women is half as big for MBA holders than for those without an MBA.

Finally, Table 8 replicates the results from the nonlinear least squares estimation and shows polynomial regressions using NLSY79 data, both for the main sample and for the subsample of college graduates. The magnitude of the coefficient β_2 for college graduates in NLSY79 is similar to β_2 in the GMAT survey regressions: 0.002. Similarly, β_1 is not statistically different from zero. The NLSY79 estimates for college educated men are noisier. The sample contains 3,474 observations from 761 workers; the number of workers is less than half of the number of workers available in the full GMAT Survey sample. The point estimate for k is slightly higher in the NLSY79 college sample: $k = 51.22$. The estimates for k and β_2 are lower for the full NLSY79 sample. The results from the polynomial specification in column 4 suggest that the effect of hours is positive to the right of 38. NLSY79 can be used as evidence that the trends in the GMAT Registrant Survey are not sporadic; there is a persistent relationship between hours and wage growth when hours are high.

4 Empirical Evidence on Promotions

This section explores further the predictions of the theory in Section 2 and provides further empirical tests that attest its validity. The model of promotions in Section 2 is similar to the model in Gibbons and Waldman (1999), who in turn base their theory on the empirical findings in Baker, Gibbs and Holmstrom (1994a, 1994b). An essential component of the framework developed by these studies is the serial correlation of promotions over time: workers with high learning ability (or lower disutility of hours) receive promotions faster in any job level within the firm.

The dataset used in Baker, Gibbs and Holmstrom (1994a, 1994b) is highly appropriate for testing this idea; their study is based on the detailed employment records of a medium-sized U.S. firm. The GMAT Registrant Survey, while very different in design, contains questions that allow me to carry out some similar tests. Waves 1 and 4 of the survey have data on the number of promotions received since starting on the current job. While there is reason to believe that these data are measured with error, partly because the number of promotions depends largely on the specific hierarchy within each firm, they show some interesting trends. In the main sample the coefficient of correlation between the number of promotions reported in the first and last interviews is 0.2692.[14] When the sample is limited to workers who put in 48 hours or more in the last period this coefficient increases to 0.4071. This is consistent with the idea of a fast track which consistently allows certain workers to advance faster than their peers at any point along their career path.

I also find that, not including any controls, workers who report 48 hours or more in wave 1 are more likely to be promoted between waves 1 and 2: 52.36% of them report at least one promotion, compared to 40.92% of lower-hour workers (standard errors are .4999 and .4919, and the null of no difference can be rejected at the 0.001% level. Similarly, 59.52% of workers who report usual hours higher than 47 in wave 2 receive a promotion, while only 49.77% of those whose hours are lower do (standard errors are .4913 and .5002).

Tables 9 and 10 relate directly to the promotions theory. In particular, they show that both inherent ability and long hours have a positive effect on the probability of receiving a promotion. It is important to note that inherent ability is not straightforward to measure, and it is usually difficult to distinguish empirically between the ability to learn (θ_i in the model) and the parameter that represents initial level of human capital (η_1).

Table 9 shows results for the GMAT Registrant Survey. The sample is limited to respondents who report not being on a managerial position at the time of the initial interview. The dependent variable in the probit regressions from Table 9 equals one if

[14]Only 507 of the 1911 respondents report number of promotions at both the first and last interviews.

the respondent reported being a manager at the last interview and to zero otherwise. Working 48 hours or more per week has a strong effect on the probability of advancement to managerial position: the increase is on average 12 percentage points. This effect is large: for comparison, the effect of obtaining an MBA degree is only slightly higher (15 percentage points). The coefficient on GMAT score is not significant in any of the specifications. However, coming from a family with a college educated mother increases the probability of moving up the job ladder by over 10 percentage points. The results support the model of promotions in Section 2, with the caveat that GMAT scores are not a good measure of the ability to learn on the job but family background, as measured by mother's educational attainment, fits the role of θ well.

The NLSY79 results in Table 10 support these findings. In these specifications the dependent variable equals one if the respondent reported receiving a promotion since the previous interview, which is likely to be a more noisy measure of career advancement than the GMAT Survey variable I use because the question is asked in a less specific way. In NLSY79 the positive marginal effect of standardized AFQT score fits the inherent ability framework. However, the effect of working long hours is not as pronounced although it is consistently positive and significant at the 5 percent level. Workers who put in 50 hours or more per week are only about 2 percentage points more likely to be promoted two years later. The coefficient on library card is positive and significant. Interestingly, it is larger in magnitude than the effect of long hours. An extra year of education increases the probability of promotion by one percentage point.

The conclusion that can be drawn from Tables 9 and 10 is that, along with obtaining more education, working long hours has a positive effect on the probability of promotion. Assuming that family background variables like mother's education in the GMAT Survey and AFQT and the library card dummy in NLSY79 are good proxies of the ability to learn on the job, the findings also suggest that, controlling for education and long hours, inherent ability also plays a major role in employers' promotion decisions. These results support the Gibbons and Waldman (1999)-type model in which employers have full information about workers' ability and preferences for leisure.

4.1 Empirical Evidence on the Human Capital Accumulation Model

To test empirically whether a human capital accumulation model alone can account for the observed relationship between working hours and the change in log wages I follow an approach similar to Lazear and Moore (1984), who point out that learning-by-doing and OJT models should apply equally well to self-employed workers, while job ladder models are not relevant to this group. Table 11 compares NLSY79 estimates for the main sample and for self-employed workers. The specifications in columns 1 and 3 are based on a linear spline in hours; I assume that the break point is known and occurs at 48 hours. For comparison, columns 2 and 4 estimate a regression with a dummy variable that is equal to one when hours exceed 50. The results suggest that the relationship between long hours and wage growth is not present for the self-employed group. The self-employed sample is much smaller and it is not surprising that the estimates are noisier. However, apart from the hours variables in the two specifications, most other variables that are significant for the sample that is not self-employed are also significant for the self-employed group and the point estimates have the same sign and are of similar magnitude. In column 1 of Table 11 the coefficient on hours to the right of 48 equals

0.001 and has a standard error of 0.0002. The corresponding coefficient in column 3 equals -0.0011 with a standard error of 0.0008. The null hypothesis of equal coefficients can be rejected at a very low significance level. Similarly, the coefficient on the high hours indicator in column 2 equals 0.0064 (its standard error is 0.0022), while in column 4 it is -0.0012 (0.0112). The self-employed workers data serves as evidence against on-the-job training as the sole explanation of the observed nonlinear relationship.

The data in NLSY79 offer some information on training which I incorporate in order to investigate the role of human capital accumulation in the relationship of interest between hours and wages. Training information is available for 11,593 of the 16,970 observations in the sample. I construct a training variable that equals 1 if a respondent reported receiving some sort of training on the job in year t and is set to 0 otherwise. There are 2,335 worker-year observations for which this variable is equal to one. Table 12 shows the results when I incorporate NLSY79 training data in a wage growth regression.

The specification in column 1 is identical to the one in column 1 of Table 11 but uses the restricted sample for which training data are available. The estimates from the two regressions are similar, which suggests that restricting the sample in this way does not bias the results. In column 2 of Table 12 I include the indicator for training. The coefficient for this variable is positive and highly significant; obtaining more training enhances wage growth. However, even when I control for training the coefficient on hours when hours are higher than 48 remains unchanged at 0.001. This suggests that training is not the only reason why high hours imply faster wage growth. A similar conclusion can be drawn from columns 3 and 4. I divide the sample into workers who received training in t and those who did not. I find that high hours had a stronger effect for the group without training (0.0018 versus 0.0008) but β_2 is positive and significant in both cases.

Finally, columns 5 and 6 do not control for training but divide workers into two groups: those with actual labor market experience that is less than 15 years and workers with 15 or more years of experience. It is interesting that the relationship between hours and wage growth exists only for the less experienced group. This observation can be put in the context of a model of investment in human capital, in which after some point workers find it optimal to choose zero investment, but it also complies with a model similar to the one described in Section 2 when workers have reached the top level of the job ladder, from which point on their wages grow solely due to the learning-by-doing component.

Overall, NLSY79 data suggest that there is a relationship between learning on the job and career wage growth but a richer model is necessary to explain some features of the data, like the increase in the wage profile slope at high hours or the difference in trends in the self-employed sample.

5 Conclusion

In this paper I develop a model that combines learning on the job and promotions within a firm when workers differ in their preferences for leisure and use this theory to explain an empirical trend that has not been analyzed previously: of increasing wage growth as a function of working hours when hours are high. The main dataset I use in the empirical portion of the study is a panel survey of Graduate Management Admission Test registrants; most respondents in the sample considered enrollment in an MBA program.

Using the same data Arcidiacono et al. (2008) study the wages of recent MBA graduates and their peers who did not complete a graduate business degree. Their paper demonstrates that the two groups differ on observable characteristics. Workers who choose to further their education tend to be stronger in some respects: they have lower nonmonetary costs of obtaining schooling, as measured for example by undergraduate grade point average or the ability to score high on a standardized test like the GMAT. On the other hand, when asked whether they agree with the statement "A graduate management degree will not be that important because I already have the credentials I need to do well in my career," workers who did not enroll in an MBA program (particularly not in a top-25 program) consistently report their workplace skills to be stronger. Arcidiacono et al. (2008) conclude that there exists a self-selection process that causes workers with stronger career advancement skills to opt out of schooling. Thus, one needs to be cautious when extrapolating the GMAT Survey findings to all professional workers. In this case it is comforting that data from NLSY79 fully support the empirical results.

I show that working over 47 hours per week has a considerable positive effect on wage growth. The theory that this study builds upon assumes that hours depend on workers' preferences but takes these preferences to be exogenous. In addition, I assume that employees are unconstrained in their choice of hours. An appropriate next step in the line of work of Bertrand et al. (2008) would be to explore in more detail the characteristics of workers who put in overtime: if workers from certain demographic groups are likely to choose long workweeks or if there are jobs characteristics or occupations associated with long hours. Gender differences in the disutility of hours can be used to explain part of the male-female wage gap.

References

Akerlof, George A., "The Economics of Caste and of the Rat Race and Other Woeful Tales," *The Quarterly Journal of Economics*, November 1976, *90* (4), 599–617.

Arcidiacono, Peter, Jane Cooley, and Andrew C. Hussey, "The Economic Returns to an MBA," *International Economic Review*, August 2008, *49* (3), 873–899.

Arrow, Kenneth J., "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, June 1962, *29* (3), 155–173.

Baker, George, Michael Gibbs, and Bengt Holmstrom, "The Internal Economics of the Firm: Evidence from Personnel Data," *The Quarterly Journal of Economics*, November 1994, *109* (4), 881–919.

– , – , and – , "The Wage Policy of a Firm," *The Quarterly Journal of Economics*, November 1994, *109* (4), 921–55.

Barron's profiles of American Colleges, Hauppauge: Barrons Educational Series, 1994.

Becker, Gary S., *Human Capital*, Columbia University Press for the National Bureau of Economic Research, 1964.

Bell, Linda A. and Richard B. Freeman, "The incentive for working hard: explaining hours worked differences in the US and Germany," *Labour Economics*, May 2001, *8* (2), 181–202.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz, "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," 2008.

Blair-Loy, "Work Devotion and Work Time," in Cynthia Fuchs Epstein and Arne

Kalleberg, eds., *Fighting for Time: Shifting Boundaries of Work and Family*, Russell Sage Foundation, 2004, pp. 392–316.

Borjas, George J., "The Relationship between Wages and Weekly Hours of Work: The Role of Division Bias," *The Journal of Human Resources*, Summer 1980, *15* (3), 409–423.

Deaton, Angus, "Quality, Quantity, and Spatial Variation of Price," *The American Economic Review*, June 1988, *78* (3), 418–430.

Gibbons, Robert and Michael Waldman, "A Theory Of Wage And Promotion Dynamics Inside Firms," *The Quarterly Journal of Economics*, November 1999, *114* (4), 1321–1358.

Goldin, Claudia and Lawrence F. Katz, "Transitions: Career and Family Life Cycles of the Educational Elite," *American Economic Review: Papers and Proceedings*, May 2008, *98* (2), 363–369.

Graddy, Kathryn and Luigi Pistaferri, "Wage Differences by Gender: Evidence from Recently Graduated MBAs," *Oxford Bulletin of Economics and Statistics*, December 2000, *62* (Supplement 1), 837–854.

Imai, Susumu and Michael P. Keane, "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, May 2004, *45* (2), 601–641.

Jovanovic, Boyan, "Job Matching and the Theory of Turnover," *Journal of Political Economy*, October 1979, *87* (5), 972–90.

Landers, Renee M., James B. Rebitzer, and Lowell J. Taylor, "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms," *American Economic Review*, June 1996, *86* (3), 329–48.

Lazear, Edward P. and Robert L. Moore, "Incentives, Productivity, and Labor

Contracts," *The Quarterly Journal of Economics*, May 1984, *99* (2), 275–96.

Mincer, Jacob, "On-the-job training: Costs, Returns and some implications," *The Journal of Political Economy*, October 1962, *70* (5), 50–79.

Montgomery, Mark and Irene Powell, "Does an Advanced Degree Reduce the Gender Wage Gap? Evidence from MBAs," *Industrial Relations*, July 2003, *42* (3), 396–418.

Quandt, Richard E., "The estimation of the parameters of a linear regression system obeying two separate regimes," *Journal of the American Statistical Association*, December 1958, *53* (284), 873–880.

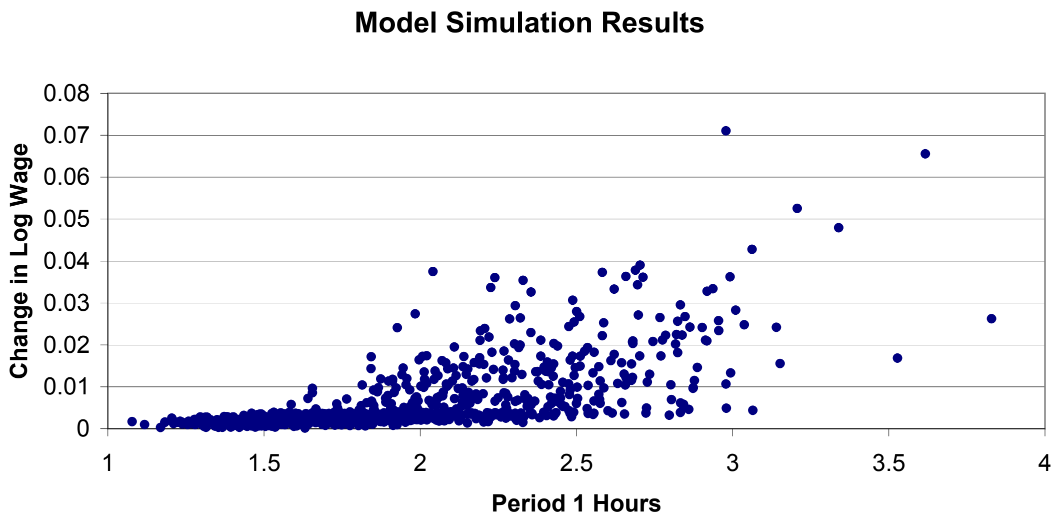
Rebitzer, James B. and Lowell J. Taylor, "Do Labor Markets Provide Enough Short-Hour Jobs? An Analysis of Work Hours and Work Incentives," *Economic Inquiry*, April 1995, *33* (2), 257–273.

Reich, Robert B., *The Future of Success*, Knopf, 2001.

Robinson, Peter M., "Root-N-Consistent Semiparametric Regression," *Econometrica*, July 1988, *56* (4), 931–954.

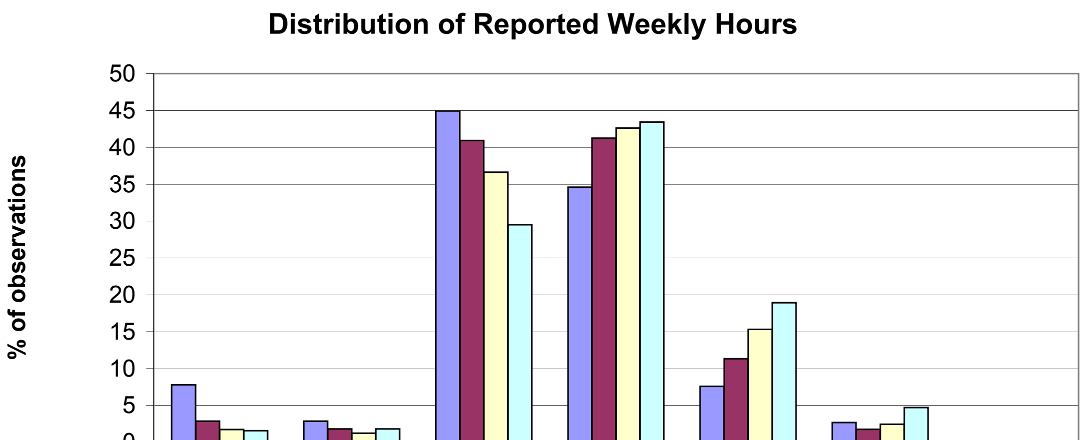
Rosen, Sherwin, "Learning by Experience as Joint Production," *The Quarterly Journal of Economics*, August 1972, *86* (3), 366–382.

Sousa-Poza, Alfonso and Alexandre Ziegler, "Asymmetric information about workers' productivity as a cause for inefficient long working hours," *Labour Economics*, December 2003, *10* (6), 727–747.



Simulation results from model of promotions. The disutility parameter b and inherent ability θ are chosen such that $300b$ and 200θ have χ^2 distributions with 50 degrees of freedom. Other parameter values are as follows: $\eta_1 = 0.1$; $c_1 = 0.075$; $c_2 = 0.52$; $d_1 = 0.616$; $d_2 = 0.545$; $N = 1000$.

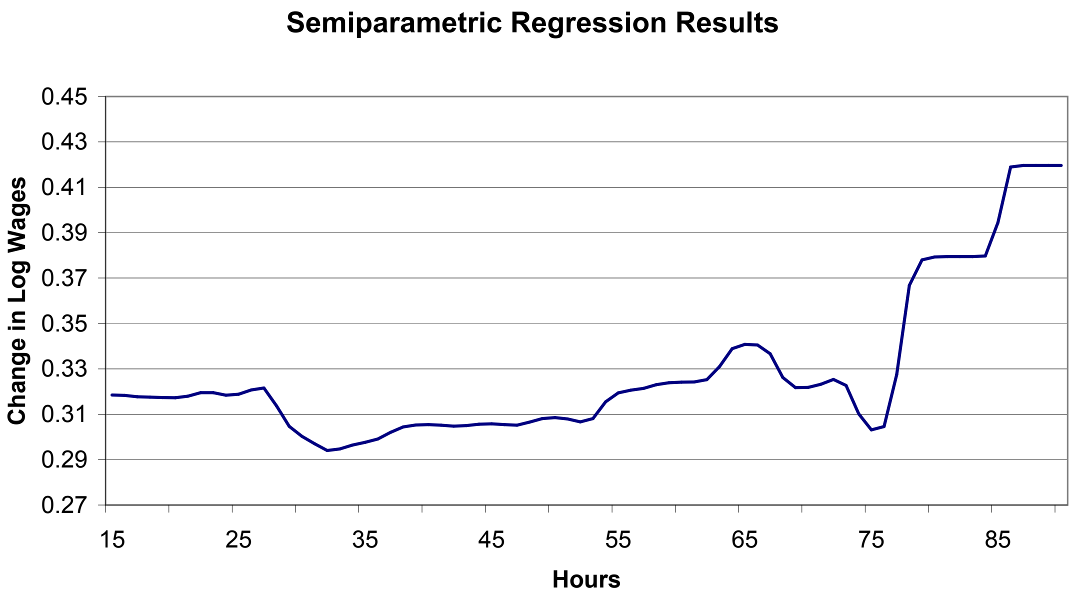
Figure 1: Hours and Change in Log Wages from Simulations



	<25	25-34	35-44	45-54	55-64	65+	Mean
Wave 1	7.8	2.88	44.94	34.59	7.59	2.68	42.3
Wave 2	2.88	1.83	40.92	41.24	11.35	1.78	44.8
Wave 3	1.75	1.25	36.63	42.62	15.31	2.44	46.3
Wave 4	1.57	1.83	29.51	43.44	18.94	4.71	47.9

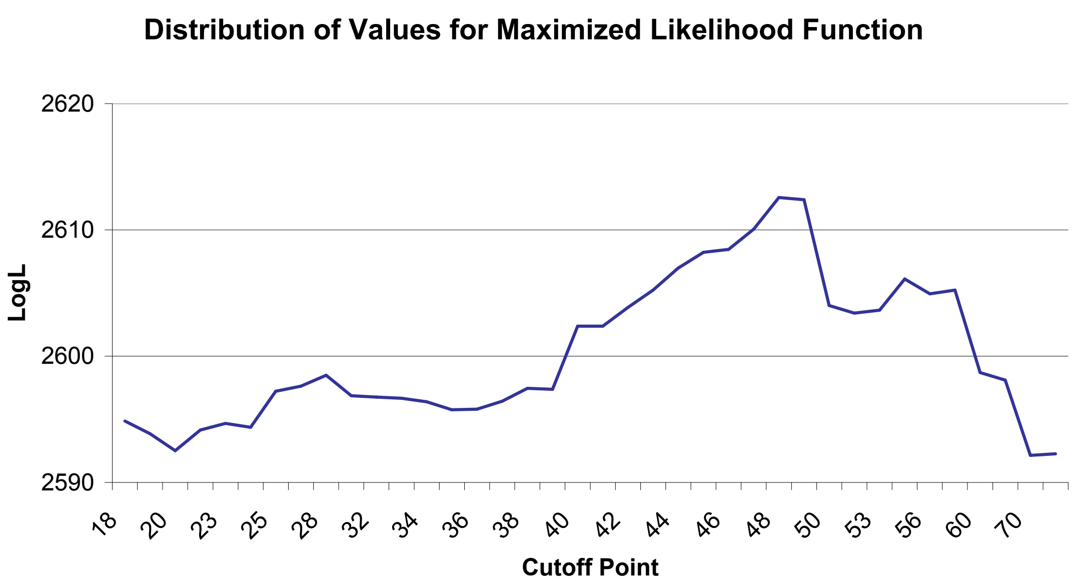
Calculations include all workers in the main sample; N = 1911 for waves 1, 2 and 4, and N = 1600 for wave 3.

Figure 2: Comparison of the Distribution of Reported Hours across Waves



Results from partially linear model $\ln W_4 - \ln W_2 = Z\beta + f(Hours_1) + \varepsilon$ following Robinson (1988). Z includes gender, completion of MBA program, age, race, marital status, number of children, maternal education, undergraduate major, and whether in school. $N = 1911$.

Figure 3: Partial Linear Model Results for the Effect of Hours on Wage Growth



Maximized values of log likelihood function over the distribution of *k* for the specification in Equation (4) with annual wage growth as the dependent variable. The full sample is used in the estimation.

Figure 4: Distribution of Log Likelihood Values for Different Levels of k

Table 1: Hours Quantile Transition Probabilities

Wave 1 Quantile	Wave 2 Quantile 1 (49-90)	2 (41-48)	3 (15-40)	Total
1 (46-90)	367 65.77%	122 21.86%	67 12.37%	558 100.00%
2 (41-45)	102 28.81%	172 48.59%	80 22.60%	354 100.00%
3 (3-40)	154 15.42%	220 22.02%	625 62.56%	999 100.00%
Total	623 32.60%	514 26.90%	774 40.50%	1,911 100.00%

Wave 1 Quantile	Wave 3 Quantile 1 (51-90)	2 (41-50)	3 (5-40)	Total
1 (46-90)	178 37.95%	209 44.56%	82 17.48%	469 100.00%
2 (41-45)	51 16.89%	181 59.93%	70 23.18%	302 100.00%
3 (3-40)	69 8.32%	338 40.77%	422 50.90%	829 100.00%
Total	298 18.63%	728 45.50%	574 35.88%	1,600 100.00%

Wave 1 Quantile	Wave 4 Quantile 1 (51-90)	2 (46-50)	3 (15-45)	Total
1 (46-90)	230 41.22%	179 32.08%	149 26.70%	558 100.00%
2 (41-45)	73 20.62%	110 31.07%	171 48.31%	354 100.00%
3 (3-40)	163 16.32%	257 25.73%	579 57.96%	999 100.00%
Total	466 24.39%	546 28.57%	899 47.04%	1,911 100.00%

Table 2: Summary Statistics

	GMAT Sample[a]		NLSY79 Sample[bc]	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Female	0.423		0	
Asian	0.127			
Black	0.120			
Hispanic	0.154			
Nonwhite			.413	
Age (Wave 1)	27.669	5.958	29.700	5.300
Took GMAT	0.840			
GMAT verbal	28.925	7.751		
GMAT math	29.189	8.526		
Majored in business	0.504			
Majored in science	0.126			
Majored in social sciences	0.150			
Majored in humanities	0.075			
Majored in engineering	0.146			
Married (Wave 1)	0.349		0.575	
Married (Wave 4)	0.654			
Number of children (Wave 4)	0.805	1.038	0.884	1.124
Self-employed (Wave 1)	0.024		0	
Manager (Wave 1)	0.417			
Manager (Wave 4)	0.590			
Actual experience in months (Wave 1)	70.528	69.102	89.581	82.337
Mom has college degree	0.311		0.076	
Attended competitive college[d]	0.206			
Has college degree (Wave 4)	1		.205	
Hourly wage (Wave 1)	18.444	8.308	12.445	7.932
Hourly wage (Wave 4)	35.370	22.325		
Hours (Wave 1)	42.303	10.573	43.528	7.913
Hours (Wave 4)	47.885	9.468		
MBA (PT)	0.438			
MBA (FT)	0.133			
MBA (Executive)	0.047			
Attended top 10 MBA program	0.028			

[a]N = 1,911

[b]N = 16,970

[c]All variables refer to current interview

[d]Undergraduate institution is among top 10% most competitive according to *Barron's profiles of American Colleges* (1994)

Table 3: Nonlinear Least Squares Estimation Results, GMAT Survey

Dependent Variable: $(lnW_4 - lnW_2)/years$			
	All	Men	Women
Constant	0.3094***	0.3619***	0.2791***
	(0.0601)	(0.0836)	(0.0824)
$\min(Hours, k)$	-0.0004	-0.0004	-0.0004
	(0.0002)	(0.0003)	(0.0003)
$\max(Hours - k, 0)$	0.0018**	0.0020***	0.0013
	(0.0006)	(0.0006)	(0.0015)
Female	-0.0126***		
	(0.0032)		
MBA (FT)	0.0190***	0.0131	0.0257***
	(0.0052)	(0.0071)	(0.0077)
MBA (PT)	0.0012	-0.0006	0.0026
	(0.0036)	(0.0051)	(0.0051)
MBA (exec)	0.0095	0.0060	0.0142
	(0.0070)	(0.0093)	(0.0094)
k	47.14***	46.50***	48.00***
	(2.34)	(2.52)	(5.66)
R^2	0.1572	0.168	0.1495
N	1911	1103	808

* p<0.05, ** p<0.01, *** p<0.001

Included are controls for age, race, marital status, number of children, maternal education, undergraduate major and whether in school. Hours refer to hours reported in Wave 1 to avoid measurement error bias. The reported errors are heteroscedasticity robust.

Table 4: ML Estimation Results with Hours as a Discrete Variable

Dependent Variable: $(lnW_4 - lnW_2)/years$			
	All	Men	Women
$Hours < k$:			
Constant	0.3043***	0.345***	0.2843**
	(0.0582)	(0.0914)	(0.0921)
Hours	-0.0005*	-0.0004	-0.0005
	(0.0003)	(0.0004)	(0.0004)
$Hours \geq k$:			
Constant	0.1968***	0.2310**	0.1961*
	(0.0633)	(0.0968)	(0.1020)
Hours	0.0018***	0.0020***	0.0013*
	(0.0004)	(0.0006)	(0.0007)
Female	-0.0133***		
	(0.0033)		
MBA (FT)	0.0150**	0.0083	0.0237***
	(0.0049)	(0.0068)	(0.0072)
MBA (PT)	-0.0012	-0.0029	0.0009
	(0.0039)	(0.0052)	(0.0061)
MBA (exec)	0.0119	0.0083	0.0195
	(0.0082)	(0.0096)	(0.0170)
σ_1	0.0571***	0.0585***	0.0547***
	(0.0003)	(0.0005)	(0.0005)
σ_2	0.0750***	0.0751***	0.0751***
	(0.0012)	(0.0015)	(0.0027)
k	48	48	48

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Included are controls for age, race, marital status, number of children, maternal education, undergraduate major and whether in school. Hours refer to hours reported in Wave 1 to avoid measurement error bias.

Table 5: The Extent of Division Bias: Using Hours in Wave 2

Dependent Variable: $(lnW_4 - lnW_2)/years$			
	All	Men	Women
Constant	0.3543***	0.4115***	0.4549***
	(0.0618)	(0.0834)	(0.0887)
$\min(Hours, k)$	-0.0012***	-0.0015***	-0.0057***
	(0.0003)	(0.0005)	(0.0017)
$\max(Hours - k, 0)$	0.0032***	0.0039***	0.0019***
	(0.0006)	(0.0007)	(0.0005)
Female	-0.0107***		
	(0.0031)		
MBA (FT)	0.0153**	0.0101	0.0184*
	(0.0051)	(0.0068)	(0.0076)
MBA (PT)	0.0018	0.0017	0.0012
	(0.0035)	(0.0049)	(0.0050)
MBA (exec)	0.0087	0.0079	0.0109
	(0.0069)	(0.0091)	(0.0092)
k	45.21***	46.04***	31.69***
	(1.20)	(1.18)	(2.26)
R^2	0.1930	0.2171	0.1899
N	1911	1103	808

* p<0.05, ** p<0.01, *** p<0.001

Included are controls for age, race, marital status, number of children, maternal education, undergraduate major and whether in school. Hours refer to hours reported in Wave 2. The reported errors are heteroscedasticity robust.

Table 6: Using Wage Growth between Second and Third Waves as Dependent Variable

Dependent Variable: $(lnW_3 - lnW_2)/years$			
	All	Men	Women
Constant	0.3842*	0.3922*	0.3806
	(0.1521)	(0.1846)	(0.2498)
$\min(Hours, k)$	-0.0007	-0.0005	-0.0009
	(0.0006)	(0.0008)	(0.0008)
$\max(Hours - k, 0)$	0.0021	0.0020	0.0023
	(0.0011)	(0.0014)	(0.0016)
Female	-0.0259***		
	(0.0071)		
MBA (FT)	0.0604***	0.0474*	0.0727**
	(0.0155)	(0.0199)	(0.0238)
MBA (PT)	0.0038	-0.0006	0.0114
	(0.0097)	(0.0141)	(0.0115)
MBA (exec)	0.0216	0.0358	-0.0048
	(0.0204)	(0.0268)	(0.0261)
k	42.00***	41.76***	43.25***
	(3.76)	(4.91)	(4.96)
R^2	0.0745	0.0797	0.0787
N	1580	920	660

* p<0.05, ** p<0.01, *** p<0.001

Nonlinear least squares estimation results. Included are controls for age, race, marital status, number of children, maternal education, undergraduate major and whether in school. Hours refer to hours reported in Wave 1 to avoid measurement error bias. The reported errors are heteroscedasticity robust.

Table 7: Robustness Checks: Excluding Job Movers and Workers Attending School and Estimation Results for MBA and Non-MBA Holders

Sample	Same job in waves 2 and 4	Not in school in waves 1,2 or 4	MBA	No MBA
Dependent Variable: $(lnW_4 - lnW_2)/years$				
Constant	0.1839**	0.4013***	0.3806***	0.2237*
	(0.0655)	(0.0925)	(0.0831)	(0.0887)
$\min(Hours, k)$	-0.0002	0.0002	-0.0007*	-0.0001
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
$\max(Hours - k, 0)$	0.0020**	0.0022**	0.0021***	0.0016
	(0.0007)	(0.0008)	(0.0006)	(0.0011)
Female	-0.0109*	-0.0149**	-0.0100*	-0.0184***
	(0.0054)	(0.0049)	(0.0040)	(0.0054)
MBA (FT)	0.0004	0.0116		
	(0.0108)	(0.0071)		
MBA (PT)	0.0028	0.0062		
	(0.0054)	(0.0050)		
MBA (exec)	0.0000	0.0343**		
	(0.0138)	(0.0129)		
k	47.91***	48.27***	47.63***	45.35***
	(3.27)	(3.92)	(2.00)	(6.54)
R^2	0.1521	0.1773	0.1651	0.1393
N	485	918	1180	731

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

Nonlinear least squares estimation results. Included are controls for age, race, marital status, number of children, maternal education, undergraduate major and whether in school (except for Column 2). Hours refer to hours reported in Wave 1 to avoid measurement error bias. The reported errors are heteroscedasticity robust.

Table 8: NLSY79 Results: Nonlinear Least Squares and Polynomial Estimation

	NLS		Polynomial	
Sample	All	College grads	All	College grads
Constant	0.06580***	0.09165**	0.08555***	0.11250**
	(0.01143)	(0.03143)	(0.01563)	(0.03740)
$\min(Hours, k)$	-0.00065***	-0.00010		
	(0.00018)	(0.00033)		
$\max(Hours - k, 0)$	0.00093**	0.00218		
	(0.00030)	(0.00113)		
Hours in $t-2$			-0.00201	-0.00228
			(0.00055)	(0.00125)
Hours squared			0.00002	0.00003
			(0.00000)	(0.00003)
Years of education	0.00393***	0.00505***	0.00391***	0.00489***
	(0.00035)	(0.00148)	(0.00035)	(0.00145)
Experience	-0.00038	-0.00043	-0.00044	-0.00042
	(0.00050)	(0.00093)	(0.00050)	(0.00093)
Experience squared	0.00003	0.00003	0.00002	0.00002
	(0.00003)	(0.00005)	(0.00003)	(0.00005)
Nonwhite	-0.00600***	-0.01325***	-0.00596***	-0.01279***
	(0.00150)	(0.00365)	(0.00148)	(0.00365)
Married	0.00005	0.00073	-0.00007	0.00058
	(0.00185)	(0.00400)	(0.00185)	(0.00400)
Number of children	0.00055	0.00170	0.00053	0.00163
	(0.00083)	(0.00178)	(0.00083)	(0.00180)
Age	-0.00185***	-0.00383***	-0.00230***	-0.00388***
	(0.00025)	(0.00078)	(0.00038)	(0.00090)
k	45.89***	51.22***		
	(2.09)	(4.76)		
R^2	0.0259	0.0629	0.0247	0.0578
N	16970	3474	16970	3474

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

The dependent variable is $(\ln W_{t+4} - \ln W_t)/4$. Included are year dummies. Reported errors are clustered on individual level. Hours refer to hours reported in $t-2$.

Table 9: The Role of Long Hours and Inherent Ability in Promotions: GMAT Survey

Dependent variable: Whether advanced to managerial position				
Worked 48+ hours (wave 1)	0.121**	0.126**	0.117*	0.116*
	(0.045)	(0.045)	(0.046)	(0.046)
GMAT verbal score	0.001	-0.000	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.003)
GMAT quantitative score	0.002	0.001	-0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
Female	-0.040	-0.034	-0.038	-0.037
	(0.039)	(0.039)	(0.039)	(0.039)
Asian	-0.175***	-0.170**	-0.166**	-0.159**
	(0.053)	(0.053)	(0.054)	(0.054)
Black	-0.109	-0.116	-0.117	-0.119
	(0.061)	(0.062)	(0.062)	(0.062)
Hispanic	-0.054	-0.067	-0.051	-0.044
	(0.054)	(0.054)	(0.055)	(0.055)
Age (wave 1)	-0.029	-0.025	-0.025	-0.044
	(0.027)	(0.027)	(0.027)	(0.038)
Age squared	0.000	0.000	0.000	0.001
	(0.000)	(0.000)	(0.000)	(0.001)
MBA (FT)		0.148*	0.144*	0.149*
		(0.061)	(0.062)	(0.062)
MBA (PT)		0.107**	0.121**	0.114**
		(0.041)	(0.041)	(0.042)
MBA (exec)		0.191*	0.201*	0.188*
		(0.083)	(0.082)	(0.085)
Mom has college degree			0.106**	0.109**
			(0.041)	(0.041)
Experience (wave 4)				0.001
				(0.001)
Experience squared				-0.000
				(0.000)
N	777	777	777	777

* p<0.05, ** p<0.01, *** p<0.001
Sample is limited to workers who were not in a managerial position at the beginning of the survey. The reported coefficients are marginal effects from a probit regression.

Table 10: The Role of Long Hours and Inherent Ability in Promotions: NLSY79

Dependent variable: Whether promoted since last interview				
Worked 50+ hours in $t-2$	0.016*	0.016*	0.015*	0.015*
	(0.007)	(0.007)	(0.007)	(0.007)
Standardized AFQT	0.031***	0.016***	0.014***	0.014**
	(0.003)	(0.004)	(0.004)	(0.004)
Nonwhite	0.006	0.000	0.001	0.000
	(0.007)	(0.007)	(0.007)	(0.007)
Age (Wave 1)	-0.017*	-0.020**	-0.020**	-0.020**
	(0.007)	(0.007)	(0.007)	(0.007)
Age squared	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Years of education		0.009***	0.009***	0.009***
		(0.002)	(0.002)	(0.002)
Library card			0.023***	0.024***
			(0.007)	(0.007)
Actual experience				-0.0005
				(0.0015)
Experience squared				0.0001
				(0.0001)
N	23353	23353	23353	23353

* p<0.05, ** p<0.01, *** p<0.001

Sample is limited to male workers. Included are year dummies (1988, 1989, 1990, 1996, 1998, 2000, 2002, 2004). Self-employed workers are excluded. The reported coefficients are marginal effects from a probit regression. The errors are clustered at the individual level

Table 11: Long Hours and Wage Growth for Self-Employed Workers: NLSY79

Dependent variable: $(\ln Wage_{t+4} - \ln Wage_t)/4$				
	Not self-employed		Self-Employed	
Min($Hours_{t-2}$,48)	-0.0005**		0.0015	
	(0.0002)		(0.0009)	
Max($Hours_{t-2} - 48$,0)	0.0010***		-0.0011	
	(0.0002)		(0.0008)	
$Hours_{t-2} \geq 50$		0.0064**		-0.0012
		(0.0022)		(0.0112)
Education	0.0038***	0.0038***	0.0080**	0.0078**
	(0.0003)	(0.0003)	(0.0029)	(0.0029)
Actual experience	-0.0004	-0.0005	0.0026	0.0027
	(0.0005)	(0.0005)	(0.0031)	(0.0031)
Experience squared	0.0000	0.0000	-0.0001	-0.0001
	(0.0000)	(0.0000)	(0.0001)	(0.0001)
Nonwhite	-0.0058***	-0.0054***	-0.0332*	-0.0347*
	(0.0015)	(0.0015)	(0.0148)	(0.0147)
Married	0.0002	-0.0006	0.0279	0.0293*
	(0.0019)	(0.0019)	(0.0147)	(0.0148)
Number of children	0.0005	0.0004	-0.0065	-0.0060
	(0.0008)	(0.0008)	(0.0057)	(0.0057)
Age	-0.0023***	-0.0024***	-0.0077**	-0.0075**
	(0.0004)	(0.0004)	(0.0028)	(0.0028)
Library card	0.0036*	0.0037*	0.0049	0.0067
	(0.0015)	(0.0015)	(0.0125)	(0.0125)
Constant	0.0630***	0.0434***	0.1337	0.1860
	(0.0118)	(0.0097)	(0.1124)	(0.1097)
R^2	0.025	0.023	0.049	0.046
N	16970	16970	900	900

p<0.05, ** p<0.01, *** p<0.001

Sample is limited to male workers who did not change jobs between t and $t+4$. Included are year dummies. The errors are clustered at the individual level.

Table 12: Training, Hours and Wage Growth: NLSY79

Dependent variable: $(\ln Wage_{t+4} - \ln Wage_t)/4$						
	All Males		Training in t	No training in t	Experience<15	Experience≥15
Min($Hours_{t-2}$, 48)	-0.0002	-0.0003	-0.0010	-0.0001	-0.0006**	0.0011
	(0.0002)	(0.0002)	(0.0005)	(0.0003)	(0.0002)	(0.0008)
Max($Hours_{t-2} - 48, 0$)	0.0010**	0.0010**	0.0018*	0.0008**	0.0016***	-0.0005
	(0.0003)	(0.0003)	(0.0007)	(0.0003)	(0.0004)	(0.0006)
Years of education	0.0034***	0.0032***	0.0035***	0.0031***	0.0038***	0.0020*
	(0.0004)	(0.0004)	(0.0008)	(0.0004)	(0.0004)	(0.0009)
Actual experience	-0.0005	-0.0005	-0.0020	-0.0001	-0.0001	-0.0004
	(0.0007)	(0.0007)	(0.0015)	(0.0008)	(0.0010)	(0.0077)
Experience squared	0.0000	0.0000	0.0001	-0.0000	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0001)	(0.0000)	(0.0001)	(0.0002)
Nonwhite	-0.0039*	-0.0038*	-0.0057	-0.0032	-0.0042*	-0.0026
	(0.0017)	(0.0017)	(0.0043)	(0.0019)	(0.0020)	(0.0035)
Married	0.0016	0.0013	-0.0043	0.0029	0.0018	0.0011
	(0.0022)	(0.0022)	(0.0049)	(0.0025)	(0.0025)	(0.0042)
Number of children	0.0007	0.0007	0.0030	0.0000	0.0005	0.0013
	(0.0009)	(0.0009)	(0.0023)	(0.0009)	(0.0010)	(0.0017)
Age	-0.0019***	-0.0018***	-0.0017*	-0.0019***	-0.0018***	-0.0017*
	(0.0004)	(0.0004)	(0.0008)	(0.0005)	(0.0005)	(0.0009)
Library card	0.0005	0.0002	0.0019	-0.0003	0.0002	0.0014
	(0.0017)	(0.0017)	(0.0041)	(0.0019)	(0.0020)	(0.0034)
Training in t		0.0083***				
		(0.0021)				
Constant	0.0523**	0.0530**	0.0839*	0.0225	0.0341*	0.0136
	(0.0194)	(0.0194)	(0.0414)	(0.0172)	(0.0169)	(0.0933)
R^2	0.014	0.015	0.024	0.011	0.024	0.002
N	11593	11593	2335	9258	8243	3350

* p<0.05, ** p<0.01, *** p<0.001

The sample is limited to male workers and workers who did not change jobs. Included are year dummies. The errors are clustered at the individual level.

Appendix

A GMAT Registrant Survey: Attrition Effects

Table 13: Attrition between Waves 1 and 4

	Present in Wave 4			Not present in Wave 4		
Variable (Wave 1)	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Employed	3718	0.768		2031	0.677	
Hourly wage	2595	18.964	8.564	1272	19.302	11.597
Hours	2714	45.424	7.929	1378	45.763	8.249
Female	3771	0.445		3235	0.395	
Age	3712	26.835	5.896	2042	26.970	5.521
Asian	3771	0.135		3235	0.133	
Black	3771	0.130		3235	0.158	
Hispanic	3771	0.157		3235	0.142	
Married	3771	0.287		3235	0.174	
Number of children under 18	3771	0.150	0.557	2082	0.159	0.558
Mom has college degree	3707	0.349		2020	0.347	
In school	3771	0.258		3235	0.168	
Undergrad GPA	3617	3.024	0.421	1876	3.009	0.428
Attended competitive college	3503	0.226		2667	0.228	
Majored in business	3745	0.515		2055	0.520	
Majored in social science	3745	0.162		2055	0.172	
Majored in engineering	3745	0.124		2055	0.117	
Majored in science	3745	0.118		2055	0.109	
Experience (months)	3658	59.452	66.650	1953	56.326	61.855
Self-employed	2695	0.041		1366	0.065	
Entry level manager	3771	0.180		3235	0.109	
Higher level manager	3771	0.122		3235	0.084	
GMAT score/100	3169	4.935	1.056	2354	4.627	1.045
Took GMAT	3771	0.840		3235	0.728	

B NLSY79 Sample Construction

I drop the 1,280 respondents from the military sample of NLSY79 because they have different employment patterns and because most respondents in this sample are dropped from the survey after 1984. Years in which a respondent was not interviewed are excluded, but if individuals returned to the survey in subsequent interviews, they are included in the sample for those years. I exclude respondents who reported being in school in every interview year and observations for which an individual was not employed at a government, private or nonprofit job. Two individuals had less than 8 years of education in all years and are dropped.

In all NLSY79 specifications hours and hourly wages apply to the main (CPS) job. Information on secondary jobs for multiple jobholders is not included. NLSY79 reports hourly wages that are constructed by dividing reported earnings by reported usual hours per week and the relevant time period, since respondents are given a choice of time unit to report earnings. Wages are deflated using the 1982-84 Consumer Price Index for all urban consumers. Actual experience is computed as the number of years after leaving school in which a respondent's usual hours per week exceeded 30 and weeks worked per year were at least 26 (which corresponds to full-time work). For the years after 1994, since the survey becomes biennial, it is assumed that if a respondent worked fulltime in year t, he was also working fulltime in $t-1$. A respondent is considered to have left school either in the first year in which he reports not being in school or, in the case when data on this variable are missing, the first year in which education level does not go up. For individuals who reported not being in school in the year prior to their first interview, I use a series of variables on employment history for the period 1975-1977 to construct actual experience if possible.

I only keep observations for which respondents are at least 19 years old and have nonmissing hours and wages. I keep observations for which the real wage is between \$2 and \$100 and usual weekly hours are between 15 and 120. Self-employed workers are dropped. Lastly, 279 respondents had no valid data for actual experience because their employment history prior to 1979 could not be fully recovered.

C Proof of Lemma 1

The difference between first period hours with and without promotion is

$$\Delta_h = \frac{(d_1 + c_1\eta_1)2b_i + (d_2 + c_2\eta_1)c_2\eta_1\theta_i}{4b_i^2 - c_2^2\eta_1^2\theta_i^2} - \frac{(d_1 + c_1\eta_1)}{2b_i - c_1\eta_1\theta_i}.$$

Then

$$\Delta_h = \frac{(d_1 + c_1\eta_1)2b_i(2b_i - c_1\eta_1\theta_i) + (d_2 + c_2\eta_1)c_2\eta_1\theta_i(2b_i - c_1\eta_1\theta_i) - (d_1 + c_1\eta_1)(4b_i^2 - c_2^2\eta_1^2\theta_i^2)}{(4b_i^2 - c_2^2\eta_1^2\theta_i^2)(2b_i - c_1\eta_1\theta_i).}$$

which is equivalent to

$$\Delta_h = \frac{-(d_1 + c_1\eta_1)2b_ic_1\eta_1\theta_i + (d_2 + c_2\eta_1)c_2\eta_1\theta_i(2b_i - c_1\eta_1\theta_i) + (d_1 + c_1\eta_1)c_2^2\eta_1^2\theta_i^2}{(4b_i^2 - c_2^2\eta_1^2\theta_i^2)(2b_i - c_1\eta_1\theta_i).}$$

We have that

$$(d_1 + c_1\eta_1)c_2^2\eta_1^2\theta_i^2 > (d_2 + c_2\eta_1)c_1c_2\eta_1^2\theta_i^2$$

because $d_1 + c_1\eta_1 > d_2 + c_2\eta_1$ and $c_2 > c_1$. In addition, Equation (1) implies that

$$(d_2 + c_2\eta_1)c_2\eta_1\theta_i2b_i > (d_1 + c_1\eta_1)c_1\eta_1\theta_i2b_i,$$

so $\Delta_h > 0$.

D Proof of Lemma 2

The inequality

$$2\bar{b} > c_2\eta_1\theta_i$$

is equivalent to

$$2(d_1+c_1\eta_1)(d_2+c_2\eta_1)c_2-(d_1+c_1\eta_1)^2c_1-(d_2+c_2\eta_1)^2c_1+\sqrt{D'} > 2(d_1+c_1\eta_1)^2c_2-2(d_2+c_2\eta_1)^2c_2,$$

where

$$\begin{aligned} D' =&(2(d_1+c_1\eta_1)(d_2+c_2\eta_1)c_2-(d_1+c_1\eta_1)^2c_1-(d_2+c_2\eta_1)^2c_1)^2+ \\ &+4(2((d_1+c_1\eta_1)^2-(d_2+c_2\eta_1)^2)((d_1+c_1\eta_1)^2c_2^2-(d_1+c_1\eta_1)(d_2+c_2\eta_1)c_1c_2)). \end{aligned}$$

Using that $(d_1+c_1\eta_1) > (d_2+c_2\eta_1)$ and $c_2 > c_1$,

$$2(d_1+c_1\eta_1)(d_2+c_2\eta_1)c_2-(d_1+c_1\eta_1)^2c_1-(d_2+c_2\eta_1)^2c_1 < \sqrt{D'},$$

so it is enough to show that

$$2(d_1+c_1\eta_1)(d_2+c_2\eta_1)c_2-(d_1+c_1\eta_1)^2c_1-(d_2+c_2\eta_1)^2c_1 > (d_1+c_1\eta_1)^2c_2-(d_2+c_2\eta_1)^2c_2.$$

Dividing through by $(d_1+c_1\eta_1)^2c_2$ yields

$$2\frac{(d_2+c_2\eta_1)}{(d_1+c_1\eta_1)}-\frac{c_1}{c_2}-\frac{(d_2+c_2\eta_1)^2}{(d_1+c_1\eta_1)^2}\frac{c_1}{c_2} > 1-\frac{(d_2+c_2\eta_1)^2}{(d_1+c_1\eta_1)^2}.$$

Denote $c_1/c_2 = \beta$ and

$$\frac{(d_2 + c_2\eta_1)}{(d_1 + c_1\eta_1)} = \alpha.$$

The assumptions on the parameters imply that $\alpha > 0.65$ and $\beta < 0.5$. The desired inequality can now be written as

$$\alpha^2(1-\beta) + 2\alpha - (1+\beta) > 0.$$

It holds that

$$\alpha^2(1-\beta) + 2\alpha - (1+\beta) > 0.5\alpha^2 + 2\alpha - 1.5.$$

When $\alpha > 0$, $0.5\alpha^2 + 2\alpha - 1.5 > 0$ for $\alpha > \sqrt{7} - 2 = 0.646$, which is true by assumption.