Working Long Hours and Early Career Outcomes in the High-End Labor Market

Dora Gicheva, *University of North Carolina at Greensboro*

This study establishes empirically a positive but nonlinear relationship between weekly hours and hourly wage growth. For workers who put in over 47 hours per week, 5 extra hours are associated with a 1% increase in annual wage growth. This correlation is not present when hours are lower. The relationship is especially strong for young professionals. Data on promotions provide evidence in support of a job-ladder model that combines higher skill sensitivity of output in higher-level jobs with heterogeneous preferences for leisure. The results can be used to account for part of the gender wage gap.

I. Introduction

The relationship between working hours and career advancement has received attention from various fields within the social sciences; Blair-Loy (2004, 284) notes, "scholars have 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

I am grateful to my thesis 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 University for their valuable comments. I also wish to thank Steven Bednar, Bjoern Bruegemann, Juan Eberhard, Pat Kline, Melissa Tartari, and Michael Waldman for numerous helpful discussions and advice. Data files are available as an online supplement.
Contact the author at d_gichev@uncg.edu.

[*Journal of Labor Economics*, 2013, vol. 31, no. 4]

0734-306X/2013/3104-0005$10.00

second-grade job that allows for more free time requires a worker 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 study links long hours to future career outcomes, in particular promotions and wage growth. The paper's main contribution is to present strong empirical evidence of a positive relationship between working hours and career outcomes and to link the two in a dynamic context providing theoretical justification for the observed relationship. Data from a four-wave panel survey of registrants for the Graduate Management Admission Test (GMAT) and from the 1979 cohort of the National Longitudinal Survey of Youth (NLSY79) suggest that for young highly educated workers who usually put in long hours, working 5 extra hours per week is linked to a 1% increase in annual wage growth. The magnitude of this relationship implies that for workers who are at the start of their careers, the shadow wage is as large as twice the observed hourly wage. The estimated coefficient is statistically indistinguishable from zero when weekly hours are less than 48.

Standard learning-by-doing theory combined with a model of promotions derived from the complete-information framework in Gibbons and Waldman (1999) can explain the observed relationship. I extend the Gibbons and Waldman (1999) model to include the worker's labor supply decision and introduce heterogeneity in the employees' preferences for leisure. Workers with low disutility of hours or high inherent ability sort into career-oriented jobs, which offer promotion opportunities and are associated with longer working hours and faster wage growth. Willingness to work extra hours and inherent ability are strategic complements in this model.

The correlation between longer hours and higher pay has been explored at length by labor economists, including in the experimental literature on gift exchange (e.g., Charness 2004; Charness and Kuhn 2007; Fehr and Goette 2007), but this paper focuses on the intertemporal relationship between labor supply and earnings and examines the slope of the wage profile. In maximizing lifetime utility, forward-looking workers may choose hours that are higher than the current-period optimal value. This idea is described qualitatively in Bell and Freeman (2001), who use differences in the dispersion of wages to explain why US employees work longer hours than their German peers. Decreasing returns to learning by doing combined with increasing observed wages make the framework here similar to the intertemporal labor supply theory in Imai and Keane (2004). A lifetime optimization model with an exogenously determined wage profile instead of a learning-by-doing mechanism (MaCurdy 1981; Altonji 1986) yields a negative correlation between current hours and future wage growth in that hours are predicted to increase over a worker's life cycle as wages grow at a decreasing rate.

In my study, I focus on the high-end labor market; all workers in the GMAT Registrant Survey sample are college educated (many of them hold degrees from top institutions) and tend to have relatively high earnings. White-collar jobs are prevalent in the data set. Thus the paper adds to the literature spurred by the recent increase in interest in this segment of the labor force, particularly with the upward trend of female participation in the high-end labor market.[1]

I focus specifically on the future career implications of current labor supply choices and argue that exogenous preferences for leisure can be used to account for up to half of the gender gap in wage growth in white-collar occupations. Studies show that there remains a considerable gap between male and female earnings, even when controlling for observables like occupation and undergraduate major.[2] Most studies of the gender gap in earnings, with the exception of Goldin and Katz (2008) and Bertrand, Goldin, and Katz (2010), do not control for differences in the labor supply of men and women, thus ignoring the disparity in weeks and hours worked and noneducation spells out of the labor force. I find that hours account for some but not all of the gender difference in wage growth.

The rest of the paper proceeds as follows. A theoretical model of promotions that predicts an upward-sloping and convex relationship between hours and wage growth is developed in Section II. Section III describes the data, and Section IV presents evidence that wage growth is an increasing function of hours when hours are high. Data on promotions are used in Section V to test other predictions of the job-ladder model. Section VI discusses alternative theoretical frameworks and derives implications of

[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% while for men the change is from 11% to 19%.

[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 a lower gender wage gap among business degree recipients. Arcidiacono, Cooley, and Hussey (2008) use all four waves and include fixed effects and additional controls for program characteristics to show that the returns to an MBA degree are actually lower for women. These seemingly contradictory findings can be reconciled by Bertrand et al. (2010). Their study follows the careers of MBA graduates of the University of Chicago Booth School of Business and finds that male and female graduates start off with similar earnings but quickly diverge. Their 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 determined by having children.

the empirical results for the shadow wage and the gender wage gap, and Section VII concludes.

II. Model of Promotions with Heterogeneous Preferences for Leisure

As the empirical evidence summarized above and presented in detail in Section IV implies, the relationship between hours and wage growth is nonlinear. Hours worked have little or no correlation with the change in log wages when hours are low, but the relationship is positive and strong for higher values of hours. In this section, I present a theoretical model that predicts the observed relationship. In particular, I present a promotions model with learning by doing. The wage structure is derived from the complete information framework in Gibbons and Waldman (1999). I add the worker's labor supply decision and impose heterogeneity in workers' valuation of leisure. As in Landers, Rebitzer, and Taylor (1996), workers who early on reveal low disutility of labor advance to higher levels within the firm.

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 creates a convexity in the resulting relationship between hours and the change in wages. The idea that jobs within the same hierarchical level can differ in how likely promotions are, and workers are sorted efficiently across jobs is also developed in Clemens (2012). The model in Clemens (2012) is similarly based on Gibbons and Waldman (1999), but additional heterogeneity is introduced at the job, not worker, level.

I simplify the Gibbons and Waldman (1999) model to two job levels and two time periods. Workers start off with effective productivity η_1, which in the second period evolves according to $\eta_2 = \eta_1(1 + \theta_i h_1)$. Here θ_i represents the speed of learning,[3] and h_1 denotes first-period hours. Hourly output in position $j = \{1, 2\}$ equals $Y_{ijt} = d_j + c_j\eta_{it}$, where $0 < c_1 < c_2$ and $0 < d_2 < d_1$. This wage structure implies that workers with low η_{it} produce more in lower-level positions. Output in upper-level jobs is more sensitive to a worker's productivity and hours, since η_{i2} depends on h_1. This implies that managerial jobs require high levels of productivity, while in nonmanagerial jobs there is less variation in output associated with productivity differentials. I assume that output with no labor market experience is lower in job 2, and all workers start in the lower-level job: $d_1 + c_1\eta_1 > d_2 + c_2\eta_1$.[4]

[3] Assuming that learning is mostly general, job mobility would not affect the model.

[4] Two additional assumptions for 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 $c_1(d_1 + c_1\eta_1) < c_2(d_2 +$

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

$$U_i(w_{i1}, w_{i2}, h_{i1}, h_{i2}, b_i) = (w_{i1}h_{i1} - b_i h_{i1}^2) + (w_{i2}h_{i2} - b_i h_{i2}^2).$$

I assume that b_i and the inherent ability parameter θ_i are independent; a correlation in these variables would not add much to the results of interest. Firms are assumed to be price takers, and effective ability η_{it} and the disutility parameter b_i are observable, so wages equal $w_{ij} = d_j + c_j\eta_{it}$.

There are two possible paths that wages can take. When a worker is not promoted in the second period, the objective function is

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

and optimal period 1 hours equal optimal period 2 hours.[5]

The objective function on the promotion path is

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

Hours with and without promotion are positive under the assumption that $2b_i > c_2\eta_1\theta_i$.[6] For given b_i and θ_i, optimal period 1 hours in "career" jobs are higher than optimal hours in "noncareer" jobs.

Workers maximize utility by choosing hours and whether to be on a promotion path or not. Then for each θ_i there exists exactly one cutoff $\bar{b}(\theta_i)$ such that $V_i^{\text{no promo}}(\bar{b}, \theta_i) = V_i^{\text{promo}}(\bar{b}, \theta_1)$, where $V(b, \theta)$ denotes the value function. In addition, the cutoff $\bar{b}(\theta_i)$ satisfies the assumption that $2\bar{b}(\theta_i) > c_2\eta_1\theta_i$.

Since the cutoff $\bar{b}(\theta_i)$ is unique, in the second period employers promote all workers with ability parameter θ_i whose disutility of hours $b_i < \bar{b}(\theta_i)$. In addition, $\bar{b}(\theta)$ is a linearly increasing function of θ. This implies that for workers with higher "ability to learn" θ the cutoff $\bar{b}$ is higher, so they can

$c_2\eta_1$). Intuitively, these assumptions impose that c_2, the sensitivity of output in upper-level positions to the level of productivity, is relatively high.

5 All derivations are available in Gicheva (2010).

6 The fact that hours without promotion do not change over time is a consequence of the functional form and parameter assumptions because the investment returns from higher first-period hours $((d_1 + c_1\eta_1) + c_1\eta_1\theta_i h_2)$ are equal in equilibrium to the labor supply effects of higher productivity in the second period, $(d_1 + c_1\eta_1) + c_1\eta_1\theta_i h_1$ (I am thankful to a referee for pointing this out). Adding a discount factor to second-period utility would result in hours that are increasing over time. Hours with promotion decrease from period 1 to period 2, but adding discounting to the model can reverse this trend for certain values of the parameters. In the data used for this study, hours tend to increase over time, but the discount parameter is omitted from the model to simplify the exposition.

compensate for higher disutility of hours with faster learning on the job. If there is an exogenous increase in the wage differential between the two types of jobs (d_2 or c_2 goes up), the cutoff $\bar{b}$ will be higher for all θ, and also hours in "career" jobs increase in both periods. This is consistent with the idea in Bell and Freeman (2001) and Kuhn and Lozano (2008) that more wage dispersion is associated with higher average hours.

In this model the wages of workers who do not get a promotion grow at a rate of $\Delta w = w_2 - w_1 = c_1\eta_1(\theta h_1)$, so that $(\partial \Delta w)/(\partial h_1) = c_1\eta_1\theta$. With a promotion wage growth is

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

and $(\partial \Delta w)/(\partial h_1) = c_2\eta_1\theta > c_1\eta_1\theta$.[7] The source of differences in the relationship between hours and wage growth can be both self-selection of higher-ability workers into "career" jobs and the wage structure, but the latter is responsible for the nonlinearity.

For two workers with the same θ but $b_i < \bar{b}(\theta)$ and $b_j > \bar{b}(\theta)$, it holds that $h_{1i} > h_{1j}$ (using the result that optimal period 1 hours in "career" jobs are higher than optimal hours in "noncareer" jobs and the feature of the model that hours are decreasing in b); thus longer hours in the first period result in faster wage growth. If θ is small, variations in hours worked do not change the slope of the wage profile much. Assuming that the speed of learning declines with experience, the difference in wage growth will be most pronounced for young workers, who have the highest incentive to invest in human capital accumulation.

Conditioning separately on hours or the speed of learning, the probability of promotion is nondecreasing in both θ and h_1 (see appendix, sec. A). In other words, the probability of receiving a promotion can be written as a nondecreasing function of θ; the parameters of the model, but not hours, enter this function. Similarly, when the promotion probability is expressed as a function of hours and the parameters, but not θ, the function is nondecreasing in hours.[8] At lower levels of hours most workers' wages grow at the rate of $c_1\eta_1(\theta h_1)$, while when hours are high the probability of promotion increases and a higher proportion of workers ex-

[7] This relationship still holds true when w_1 and w_2 denote the level, rather than the natural log, of observed wages because

$$\frac{c_1\eta_1\theta}{d_1 + c_1\eta_1} < \frac{c_2\eta_1\theta}{d_1 + c_1\eta_1}.$$

[8] Using both first-period hours and the speed of learning θ, promotions can be predicted perfectly because there is no uncertainty component. In the data both h_1 and θ are measured with error (θ in particular), so randomness is introduced in the estimation through measurement error. I approximate the shape of the relationship between hours or ability and the probability of promotion using a probit model.

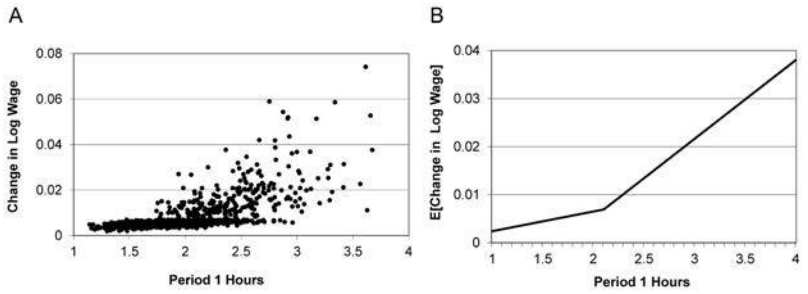


FIG. 1.—Hours and wage growth from simulations. *A*, Model simulation results, $N = 1{,}000$. *B*, Simulation results: predicted change in log wage, $N = 100{,}000$. Simulated data based on the model in Section II. The choice of the disutility parameter b and inherent ability θ is such that $300b$ and 200θ have χ^2 distributions with 50 degrees of freedom. Other parameter values: $\eta_1 = 0.1$; $c_1 = 0.075$; $c_2 = 0.52$; $d_1 = 0.616$; $d_2 = 0.545$.

perience wage growth given by $d_2 + c_2\eta_1(1 + \theta h_1) - d_1 - c_1\eta_1$. This regime change introduces the convexity in the relationship between hours and wage growth.[9]

Figure 1 shows simulation results for the promotions model.[10] To illustrate the implications of the theory, in panel A I plot the simulated data for an economy with 1,000 workers, and in panel B I show the predicted change in wages from a nonlinear least squares regression with 100,000 observations. The slope of the depicted relationship is close to zero for lower values of hours (.0028 when $h_1 < 2.155$) and increases when hours are high (the point estimate is .0115 for $h_1 \geq 2.155$). The 2.155 cutoff is estimated

[9] A different extension of the Gibbons and Waldman (1999) model that may yield some comparable predictions to the model presented here can be found in Gibbons and Waldman (2006). The authors expand on their earlier work by making experience in high-level jobs more valuable for workers in upper levels of the job ladder. Then random sorting into starting jobs, for example, due to the state of the economy at the time of job market entry (modeled through differences in the parameter d_2), can sort workers across fast-track and "noncareer" paths, keeping θ constant. In my model, for a given θ, the sorting is based on the random draw of b, not d_2. The theory here would also predict higher return to experience (θh_1) in upper-level jobs if hours are longer in higher-level positions (see n. 6). Optimal hours in my model are increasing in the rate of learning-by-doing θ, so if workers had the same disutility of leisure and it was incorporated in the Gibbons and Waldman (2006) model, a positive relationship between hours and wage growth is a likely prediction. However, it would be harder for the Gibbons and Waldman (2006) interpretation to explain differences in hours worked and career outcomes within the same cohort for a constant θ. For example, it would not be possible to make the gender gap argument in Sec. VI.B.

[10] 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.

as a parameter in the model. The relationship between hours and wage growth is positive but very small at low levels of period 1 hours because in the simulations c_1 is set to be greater than zero. With the parameter values chosen for this simulation, 22.49% of workers are promoted in the second period, and the mean of h_1 is 1.9876 with a standard deviation of 0.4493.

Extending the model to include symmetric learning about θ in a manner similar to the main analysis in Gibbons and Waldman (1999) would not change the main theoretical implications. There are two reasons specific to the extension in this paper that undermine the importance of incomplete information: the utility function implies that workers are risk neutral, and there are only two periods, which is not sufficiently long for learning to affect agents' choices. Suppose that, following Gibbons and Waldman (1999), θ is initially unobserved by workers and firms but has a known distribution and expected value $\bar{\theta}$. Note that in this setup of the model, no learning occurs between periods 1 and 2 because realized period 1 output depends only on η_1; this would be the case even if there was random noise in output so that $y_{ijt} = d_j + c_j(\eta_{it} + \varepsilon_{ijt})$. Then workers will simply use $\bar{\theta}$ instead of θ_i when solving the two-period utility maximization problem, and differences in h_1 will be based solely on the heterogeneity in b_i. Consider two cases. If second-period job assignments are made before the realization of second-period output, which parallels the setup in Gibbons and Waldman (1999), we may observe negative wage growth for some workers who receive a promotion but have values of b_i close to the cutoff $b(\bar{\theta})$. In the (probably less likely) case that period 2 productivity is revealed prior to period 2 job assignment, negative wage growth will not be observed. In both cases, the convex relationship between first-period hours and average wage growth will still be present.

III. Data

To study empirically the relationship between working hours and future wage growth, I use a panel survey of registrants for the Graduate Management Admission Test[11] and supplement the findings with results from the 1979 cohort of the NLSY. The GMAT Registrant Survey has several features that make it well suited for studying the relationship between long hours and career wage growth. The sampled group is relatively homogeneous in terms of background (e.g., parental education), schooling, and occupation. It is possible to infer ability from education variables like the quality of undergraduate institution attended or college major. I am also able to identify the effects of a graduate management degree. All respon-

[11] The survey was conducted by the Batelle Memorial Institute on behalf of the Graduate Management Admission Council. Examples of other studies using this data set are Arcidiacono et al. (2008) and Montgomery and Powell (2003).

dents obtain a college degree, and most hold white-collar jobs. A high proportion of workers report having managerial responsibilities, so promotion decisions are particularly important for this group. Long hours, which are the focus of this paper, are common. Women in the survey have high labor market participation rates, including full-time and overtime work.

The universe for the GMAT Registrant Survey consists of all individuals who registered to take the GMAT between June 1990 and March 1991 and were living in the United States at the time of registration. The survey consists of four waves; the first one was conducted shortly after registration for the test and had a response rate of 84% (5,853 responses out of 7,006 surveys sent out). The second wave was sent out about 15 months after test registration and received 4,833 responses. Wave 3 took place 3.5–4 years after registration and received 4,533 responses. The final wave was conducted about 7 years after registration for the GMAT (3,769 responses). Almost all responses came from previous survey completers; only a few of those interviewed dropped out temporarily but then returned to the study.

Respondents are given the choice of time period in which to report earnings;[12] I use reported earnings and hours for the current job to construct hourly wages. All wages are measured in 1991 dollars; I drop one observation for which the inflation-adjusted wage is less than $2 at the second interview, one for which the wage is lower than $2 at the time of the fourth interview, and four observations for which the period 4 wage exceeds $200. All other wages are between $2 and $200. The initial installment of the survey collects information on previous labor market experience, on which I base the measure of actual experience. I then use the detailed employment history provided in each installment of the survey to update actual experience accounting for periods spent out of work.

I drop observations with missing hours in the first, second, or fourth survey or wages in the second or fourth period. I also exclude workers who are self-employed in period 2. For workers with missing observations for age or experience in the second wave, mother's education, or undergraduate major, I impute these values using the predicted values from regressions on gender, race, age (if available), number of children under 18, marital status, and whether enrolled in school (with the exception of the first two variables, which are static, all others refer to the second interview). Since none of the controls are good predictors of the time between interviews, missing observations for this variable are replaced with the mean time between interviews for the rest of the sample. The number of imputed observations is 32 for mother's education, 27 for

[12] The survey asks for total pre-tax earnings, including any bonuses, commissions, and tips.

age, 36 for experience, 7 for undergraduate major, and 164 for number of days between interviews.[13] The high rates of enrollment in postgraduate education are the most common reason for dropping respondents from the sample: most workers who are unemployed are attending school full time. Of the 3,769 individuals who responded to the fourth survey, 3,232 workers reported valid wages, for 800 of which there is no valid second-wave wage observation. The final sample consists of 1,911 respondents: 1,103 men and 808 women.

Table 1 shows descriptive statistics for the GMAT Registrant Survey sample and provides a comparison with the group of respondents whom I exclude from the 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 relatively high for this age group. Around 31% of people in the sample come from a family with a college-educated mother. One unique feature of the data is that half of all respondents completed an undergraduate major in business. Over one-third of workers had managerial responsibilities at the time of their first interview. Minority respondents are oversampled, and so are women to a small degree.[14]

The sample selection procedure guarantees that all respondents are employed at the time of their first interview; this is the main difference between the subset of GMAT registrants used in the estimation and the rest of the surveyed group. As table 1 suggests, only 60% of the people I drop are employed at $t = 1$ and 31% are enrolled in school (compared to 16% in the main sample). Even though the sampled workers are on average slightly older (27.7 years compared to 26.5 years) and have an additional 1.5 years of experience, wages and hours do not differ by attrition status. The average annual growth in log wages between the second and fourth interviews is 0.062 for respondents in the sample. This is lower than the mean for respondents excluded from the sample, but 45% of observations with nonmissing wage growth in the latter group have zero years of experience at the time of their first interview, which can explain their steep wage profile. It is unclear whether one group is of higher ability: workers in the sample have slightly higher GMAT scores but fewer of them come from a family with a college-educated mother, while undergraduate grade point average is the same for both groups.

The hours variable measures usual hours worked per week in the current job. I drop observations for which hours in wave 2 are less than 15,

[13] Results that exclude the imputed observations are available from the author upon request; there are no noteworthy differences between the two sets of estimates.

[14] In the 1990–91 testing cycle, women constituted 36.8% of test takers, and 83.7% of people taking the test self-reported their race as white (non-Hispanic; Graduate Management Admission Council 1996).

Table 1
GMAT Registrant Survey Summary Statistics

	In Sample			Not in Sample			p-Value
	N	Mean	SD	N	Mean	SD	(Two-Tailed)
Female	1,911	.423		3,942	.440		.217
Asian	1,911	.127		3,942	.139		.203
Black	1,911	.120		3,942	.144		.013
Hispanic	1,911	.154		3,942	.148		.533
Married	1,911	.349		3,942	.248		.000
Number of children	1,911	.183	.613	3,942	.139	.528	.004
Age	1,884	27.668	5.959	3,870	26.501	5.631	.000
Mom has college degree	1,911	.311		3,848	.366		.000
In school	1,911	.156		3,942	.309		.000
College GPA	1,840	3.020	.420	3,653	3.018	.425	.891
Majored in business	1,911	.503		3,896	.524		.136
Majored in social science	1,911	.150		3,896	.173		.025
Majored in engineering	1,911	.146		3,896	.109		.000
Majored in science	1,911	.126		3,896	.109		.061
GMAT score/100	1,606	4.960	1.026	3,176	4.766	1.076	.000
Took GMAT	1,911	.840		3,942	.806		.002
Employed	1,911	1		3,838	.604		.000
Hours	1,911	42.295	10.581	2,318	41.901	11.471	.251
Experience	1,875	5.875	5.760	3,736	4.356	5.167	.000
Wage	1,867	13.905	6.248	2,138	13.919	7.442	.949
Annual wage growth	1,911	.062	.067	521	.086	.090	.000
Entry level manager	1,911	.219		3,942	.156		.000
Higher level manager	1,911	.144		3,942	.116		.002

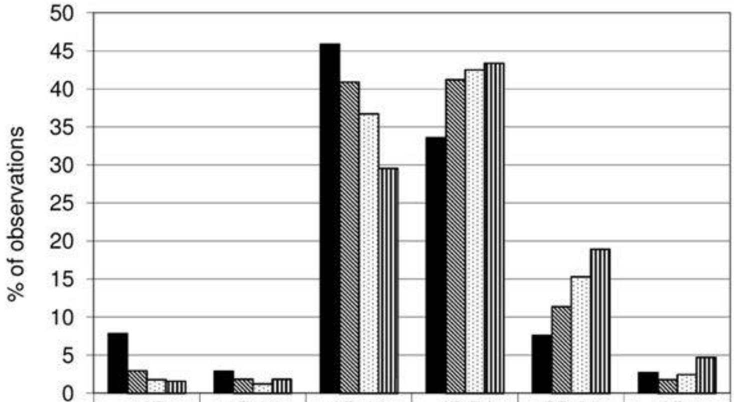
NOTE.—The omitted category for college major is humanities/other. Marital status, number of children, age, whether in school, and all employment variables refer to the first interview. Annual wage growth is the change in log wages between the second and fourth waves divided by the number of years between interviews.

and if reported hours exceed 90 I set them equal to 90 in order to eliminate any influence of outliers.[15] 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. There is some clustering of hours at 40, but the number of people who report working 40 hours per week decreases over time at the expense of more workers reporting overtime work.[16]

The data suggest that while hours increase slightly over time, there is relatively little mobility across the distribution. The sample correlation between period 1 and period 2 hours is 0.472, and this number increases to 0.517 when I exclude respondents enrolled in school at the time of either the first or second interview. Table 2 shows a more detailed transition

[15] This change affects 17 observations.

[16] The proportion of workers who report 40 hours for survey waves 1, 2, 3, and 4 is, respectively, 33.85%, 30.93%, 29.00%, and 24.23%.



	<25	25-34	35-44	45-54	55-64	65+
■ Wave 1 Mean = 42.3	7.85	2.88	45.89	33.59	7.59	2.68
▧ Wave 2 Mean = 44.8	2.93	1.83	40.92	41.19	11.35	1.78
□ Wave 3 Mean = 46.2	1.81	1.25	36.69	42.5	15.31	2.44
▥ Wave 4 Mean = 47.9	1.57	1.83	29.57	43.38	18.94	4.71

FIG. 2.—Distribution of reported weekly hours over time. The calculations include all respondents in the sample; N = 1,911 for waves 1, 2, and 4, and N = 1,600 for wave 3.

matrix. Of workers who started off in the highest third of the distribution, 66% remained in the highest quantile in the next period. If someone initially reported 40 hours or less, the probability that she is working 40 or fewer hours by the last installment of the survey is 58%.

The fact that a worker's hours are strongly correlated in the short run (1 or 2 years) is particularly useful in the estimation. To construct hourly wages I divide reported earnings in period t by hours worked per week also in period t and the relevant time measure. The dependent variable in the GMAT Registrant Survey regression specifications is the annualized change in hourly wages between the second and fourth interviews (a period of 5–6 years): $y = [\ln(E_4/h_4) - \ln(E_2/h_2)]/s$, where s is the number of years between the second and fourth surveys; s can be a fraction. As Borjas (1980) and Deaton (1988), among others, point out, the division method combined with measurement error is likely to make observed hours in period 2 endogenous and lead to biased estimates if h_2 is also used as a regressor. To solve this problem I use first-period hours h_1 instead of h_2 on the right-hand side, but it is necessary to assume that the measurement error in hours is serially uncorrelated. The NLSY79 specifications use the

Table 2
Hours Quantile Transition Probabilities

	Wave 2 Quantile			
Wave 1 Quantile	1 (46–90 Hours)	2 (41–48 Hours)	3 (15–40 Hours)	Total
1 (46–90 hours)	367	122	69	558
	65.77%	21.86%	12.37%	
2 (41–45 hours)	102	172	80	354
	28.81%	48.59%	22.60%	
3 (3–40 hours)	153	220	626	999
	15.32%	22.02%	62.66%	
Total	623	514	774	1,911
	32.55%	26.90%	40.55%	
	Wave 3 Quantile			
	1 (51–90 Hours)	2 (41–50 Hours)	3 (5–40 Hours)	Total
1 (46–90 hours)	177	209	83	469
	37.74%	44.56%	17.70%	
2 (41–45 hours)	51	181	70	302
	16.89%	59.93%	23.18%	
3 (3–40 hours)	69	337	423	829
	8.32%	40.65%	51.03%	
Total	297	727	576	1,600
	18.56%	45.44%	36.00%	
	Wave 4 Quantile			
	1 (51–90 Hours)	2 (46–50 Hours)	3 (15–45 Hours)	Total
1 (46–90 hours)	229	179	150	558
	41.04%	32.08%	26.88%	
2 (41–45 hours)	73	110	171	354
	20.62%	31.07%	48.31%	
3 (3–40 hours)	163	256	580	999
	16.32%	25.63%	58.06%	
Total	465	545	901	1,911
	24.33%	28.52%	47.15%	

annualized change in hourly wages between years $t + 4$ and t as the dependent variable and hours in $t - 2$ as an explanatory variable:

$$[\ln(E_{t+4}/h_{t+4}) - \ln(E_t/h_t)]/4 = f(h_{t-2}).$$

This allows me to include data for the years after 1994 when NLSY79 becomes biennial.

I use the 1979 cohort of the NLSY to complement my analysis and to show that the results I derive are not unique to the universe of GMAT registrants. In addition, NLSY79 is useful when I focus on self-employed workers in Section IV.D because of the larger size of the study. Appendix section B describes the construction of the NLSY79 sample using data up until 2008. The sample consists of 35,103 observations for 4,445 male workers; restricting this sample to college-educated workers leaves 6,575 observations for 915 males. Table 3 shows summary statistics for the restricted and unrestricted samples. College-educated workers in the NLSY79 sample tend to be older and to have more experience, higher wages, and lower wage growth compared to the GMAT Survey sample.

IV. Empirical Evidence of the Relationship between Long Hours and Wage Growth

A. Semiparametric Model

It is useful to start by estimating a model that does not place any parametric restrictions on the relationship between hours and wage growth. In particular, I follow the procedure developed in Robinson (1988) to estimate a partial linear model of the form

$$(\ln W_{i4} - \ln W_{i2})/t = Z_i\beta + f(\text{Hours}_{i1}) + \varepsilon_i. \quad (1)$$

The variables that comprise the vector Z_i include an experience profile, education controls (dummies for whether the respondent completed a full-time, part-time, or executive MBA program by the end of the survey period, undergraduate major, and whether enrolled in school at the time of the second interview), demographics (gender, age, age squared, and race),

Table 3
NLSY79 Summary Statistics

	Full Sample (N = 35,103)		College Grads (N = 6,575)	
Variable	Mean	SD	Mean	SD
Nonwhite	.440		.264	
Age	30.210	6.527	31.897	6.210
Married	.507		.598	
Number of children under 18	.814	1.117	.809	1.129
Actual experience	8.801	5.924	8.863	5.847
Mom has college degree	.078		.230	
Years of schooling	12.953	2.262	16.689	1.143
Hourly wage	11.629	7.529	17.199	10.867
Annualized log wage growth	.026	.100	.042	.104
Hours	43.075	8.881	44.157	9.076

NOTE.—The summary statistics reflect all survey years in which a respondent enters the sample.

and family characteristics (marital status, number of children under 18, and an indicator for college-educated mother). The denominator t of the dependent variable represents the number of days between the second and fourth interviews divided by 365. This accounts for the fact that the time between interviews is not the same for all respondents. All 1,911 men and women in the GMAT Registrant Survey sample are included in the estimation.

The Z variables are likely to be correlated with hours, which biases the results if the nonlinear function f is estimated using $(\ln W_{i4} - \ln W_{i2})/t - Z_i\beta$ (Robinson 1988). I follow the suggested approach and start by finding $E[(\ln W_4 - \ln W_2)/t|\text{Hours}_1]$ and $E[Z_k|\text{Hours}_1]$ using a Nadaraya-Watson Gaussian kernel estimator with optimal bandwidth. The unexplained parts $(\ln W_4 - \ln W_2)/t - E[(\ln W_4 - \ln W_2)/t|\text{Hours}_1]$ and $Z - E[Z|\text{Hours}_1]$ are then used to estimate $\hat{\beta}$ by ordinary least squares, and the nonlinear portion of equation (1) is estimated also by Gaussian kernel regression using the residuals $\hat{\eta}_i = (\ln W_{i4} - \ln W_{i2})/t - Z_i\hat{\beta}$ as a dependent variable. Figure 3 shows the results for $\hat{f}(\text{Hours}_1)$, including the 90% confidence bounds.

The graph suggests that the relationship between hours and the change in log wages is nonlinear; there appears to be little or no correlation at low levels of hours and a positive slope after about 50 hours. The confidence bounds in panel A are fairly wide for large values of hours, but only 26 respondents (1.4% of the sample) report working more than 65 hours per week, so the right tail of the observed hours distribution is not going to carry much weight in the parametric estimation. Panel B of figure 3 shows the same graph, but I trim the observations for which the hours measure is smaller than 35 or larger than 65. This interval includes 1,680 observations (88% of the data) and shows more clearly the nonlinearity in the relationship. Based on the shape of $f(\text{Hours})$ implied by figure 3, I model the relationship between working hours and wage growth with a two-piece linear spline; the change in regime point is estimated as a separate parameter in the model. I discuss alternative empirical specifications in appendix section C.

B. Change in Regime Estimation with Unknown Breakpoint: Nonlinear Least Squares

I estimate by nonlinear least squares the parameters of the two-piece linear spline model

$$\begin{aligned}(\ln W_{i4} - \ln W_{i2})/t = \gamma_0 &+ \gamma_1 \min(\text{Hours}_{i1}, k) \\ &+ \gamma_2 \max(\text{Hours}_{i1} - k, 0) + Z_i\beta + \varepsilon_i.\end{aligned} \tag{2}$$

I use k to denote the point at which the change of slope occurs, and the vector Z_i includes the controls listed in Section IV.A. The regression

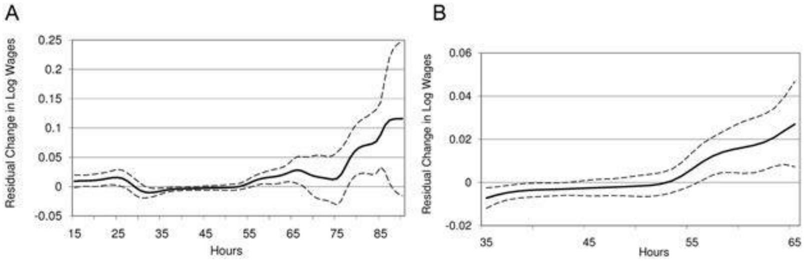


FIG. 3.—Semiparametric results for the relationship between hours and wage growth. *A*, Semiparametric regression results. *B*, Semiparametric regression results for hours between 35 and 65. Results for the partially linear model in equation (1). Z includes gender, an experience profile, age, race, marital status, number of children, mother's education, college major, MBA degree, and whether in school. The hours variable refers to first-period reported hours to avoid measurement error bias. The dashed lines represent 90% confidence intervals. $N = 1{,}911$. Gaussian kernel; bandwidth = 2.54. The graph in panel B is a close-up of panel A such that $35 \leq \text{Hours} \leq 65$ ($N = 1{,}680$).

model imposes continuity, but there is no evidence of a jump at k when this constraint is relaxed (appendix, sec. C.1). The reported errors are clustered by the occupation reported at the second interview.

The first column of table 4 shows results for the pooled sample of men and women because this yields higher precision of the estimates and allows me to estimate the unexplained gender difference in wage growth rates. The gender gap in wage growth is 1.24 log points after including all controls. In the next two columns I split the sample by gender. Table 4 implies that up to about 47 hours per week, hours worked are not correlated with wage growth; the parameter γ_1 is not statistically different from zero in the second and third columns. It is equal to -0.0004 and is significant at the 10% level for the pooled sample. At the time of the first interview, 16% of workers in the sample are employed and attending school at the same time. This is likely to decrease labor supply in period 1 but increases future wages, assuming a positive effect of schooling on earnings. The negative relationship between hours and wage growth at lower levels of hours can then be due to school enrollment. The coefficient γ_2 is positive and significant for the full sample and for the subsample of men and positive but smaller and noisy for women. It is possible that $\hat{\gamma}_2$ is lower for women because of gender differences in the propensity to select into competitive occupations (Niederle and Vesterlund 2007). The results imply that when hours exceed 47, 5 extra hours per week (about one standard deviation conditional on hours exceeding 47) are associated with 1% increase in annual wage growth. A test of the restriction that γ_1 equals γ_2

Table 4
Nonlinear Least Squares Estimation Results, GMAT Survey

	All (1)	Men (2)	Women (3)
Constant (γ_0)	.2987**	.3498**	.2713**
	(.0645)	(.0835)	(.0915)
min(Hours, k) (γ_1)	−.0004*	−.0004	−.0004
	(.0002)	(.0004)	(.0003)
max(Hours − k, 0) (γ_2)	.0019**	.0021**	.0013
	(.0004)	(.0005)	(.0011)
k	47.6906**	47.3646**	47.2595**
	(1.9960)	(2.1890)	(5.7019)
Female	−.0124**		
	(.0026)		
MBA (FT)	.0204**	.0137**	.0283**
	(.0061)	(.0059)	(.0094)
MBA (PT)	.0008	−.0007	.0018
	(.0035)	(.0055)	(.0041)
MBA (Executive)	.0098	.0064	.0149
	(.0075)	(.0125)	(.0098)
R^2	.1610	.1673	.1598
N	1,911	1,103	808

NOTE.—Estimation results for the model in eq. (2). The dependent variable is $(\ln W_4 - \ln W_2)/t$. All specifications include an experience profile and controls for age, race, marital status, number of children, mother's education, college major, and whether in school. The hours variable refers to first-period reported hours to avoid measurement error bias. The errors are clustered at the second interview occupation level.
* $p < .10$.
** $p < .05$.

rejects the null hypothesis with a p-value lower than 0.0001, which implies that the relationship is indeed nonlinear.[17]

I find that on average a part-time graduate business degree has no effect on wage growth, while a full-time degree increases wage growth by 2%

[17] I address the issue of division bias in more detail in Gicheva (2010), where I show the results when the model in eq. (2) is estimated with period 2 hours as an explanatory variable. The coefficient γ_1 is much lower: it is equal to −0.0014 for the pooled sample and is highly significant. The point estimate of γ_2 is 0.0031 and also highly significant. The estimate for the cutoff point k is 45.1 for the whole sample but is as low as 30.6 for the subsample of women. It is possible that current period hours affect wage growth more than hours in $t - 1$, which generates part of the difference, but this does not explain why γ_1 is much smaller. There is strong reason to believe that measurement error is important. I also use wage data from the third installment of the GMAT Registrant Survey to estimate the model in (2) when the dependent variable is $(\ln W_3 - \ln W_2)/t$, where t stands for time between the second and third interviews. The results are noisier but the point estimates are similar to the ones from table 3. The estimate for the cutoff k is about 5 hours lower in these specifications. The estimate of the parameter γ_2 is of similar magnitude (0.0022) and is significant at the 5% level for the pooled sample; the estimate of γ_1 is negative but not statistically distinguishable from zero. The full set of results is available upon request.

per year. The estimated effect of an executive MBA is positive in all three specifications, but only 89 people in the sample completed a degree of this type, so the estimated coefficients are imprecise. The effect of a full-time MBA degree is much stronger for women than it is for men. Obtaining the degree increases female wage growth on average by 2.8% per year, while the effect on men's wages is less than half of this number.

In table 5, I reprint the results for the pooled sample and show three other specifications that include different subsets of the control variables. All models include a gender indicator and actual experience and experience squared. The model in the second column also controls for MBA degree, college major, and school enrollment status. I add the age and race variables to the model in the third column. The estimated cutoff k changes little between specifications: it is between 46.91 and 47.69. The slope to the left of the breakpoint is also stable but has the highest absolute value when the education controls are excluded. This supports the conjecture that the estimated coefficient on hours worked is slightly negative at low levels

Table 5
Does Excluding Some Controls Affect the Estimates?

	(1)	(2)	(3)	(4)
Constant (γ_0)	.1164**	.1059**	.3052**	.2987**
	(.01195)	(.0096)	(.0649)	(.0645)
$\min(\text{Hours}, k)$ (γ_1)	−.0005*	−.0003	−.0004*	−.0004*
	(.0003)	(.0002)	(.0002)	(.0002)
$\max(\text{Hours} - k, 0)$ (γ_2)	.0021**	.0020**	.0019**	.0019**
	(.0005)	(.0005)	(.0004)	(.0004)
k	46.9076**	47.4908**	47.5330**	47.6906**
	(1.6837)	(1.9439)	(2.0031)	(1.9960)
Female	−.0107**	−.0128**	−.0128**	−.0124**
	(.0028)	(.0028)	(.0026)	(.0026)
MBA (FT)		.0215**	.0207**	.0204**
		(.0059)	(.0060)	(.0061)
MBA (PT)		.0016	.0009	.0008
		(.0035)	(.0035)	(.0035)
MBA (Exec)		.0114	.0101	.0098
		(.0073)	(.0075)	(.0075)
Education	No	Yes	Yes	Yes
Demographics	No	No	Yes	Yes
Family	No	No	No	Yes
R^2	.1191	.1479	.1589	.1610

NOTE.—Estimation results for the model in eq. (2). The dependent variable is $(\ln W_4 - \ln W_2)/t$. All specifications include an experience profile. Education controls: whether in school at the time of second interview, college major. Demographics: age, age squared, race dummies. Family controls: marital status, number of children, mother's education. The hours variable refers to first-period reported hours to avoid measurement error bias. The errors are clustered at the second interview occupation level. $N = 1{,}911$.

* $p < .10$.

** $p < .05$.

of labor supply because lower hours are likely to be observed during periods of school enrollment. The coefficient on hours worked when hours are higher than 47 increases to 0.0021 in the specification with the fewest control variables, most likely because hours are picking up the effect of variables such as age or marital status that are correlated with labor supply preferences. Overall, the results in table 5 suggest that the relationship between hours and wage growth is robust to changes in the right-hand side of the estimated equation.

The nonlinear least squares estimation method from this section has three potential problems that appendix section C.1 addresses. First, the model in equation (2) may be misspecified if there is a discrete jump at the breakpoint. Second, different variance of the error term before and after the breakpoint could bias the results in table 4. Last, the optimization process could be finding a local optimum at 47 hours. In appendix section C.1, I adjust the objective function and estimate the two-part linear spline model by maximum likelihood in order to confirm that none of these concerns are viable.

Alternatively, the relationship between hours and the change in log wages can be estimated using a higher degree polynomial. The estimated coefficients on hours and hours squared in a quadratic model suggest a relationship similar to the one implied by the two-piece linear spline parameters, as illustrated in appendix section C.2. Third and higher order terms are not statistically significant. Appendix section C.2 also illustrates the results when the coefficient of the second-period wage is not restricted to equal one. The estimated relationship between hours and wage growth remains robust.

C. Robustness of the Results

This section presents several other robustness checks as evidence that the empirical results from Section IV.B are not driven by sample selection issues or a peculiarity of the GMAT Survey data.

While job mobility has been ignored in the analysis up to now, job transitions can be positively correlated with wage growth if they are associated with improvements in the worker-employer match (Jovanovic 1979). If, in addition, job movers tend to have different hours than stayers, the estimation results will be biased. Estimating equation (2) separately for job movers and stayers can also provide evidence for or against models of careers within firms, such as tournament theory. I address these issues in the first two columns of table 6, where I show separately results for the restricted sample of 485 workers who did not change employers between the second and fourth interviews and for the remaining 1,426 respondents who reported a different employer at their last interview. The estimates for γ_2 and k are highly significant and almost identical in the two

Table 6
Some Robustness Checks

	Stayers (1)	Movers (2)	Not in School (3)	MBA (4)	No MBA (5)
Constant (γ_0)	.1838**	.3132**	.4221**	.3746**	.2078**
	(.0587)	(.0879)	(.0850)	(.0919)	(.0888)
min(Hours, k) (γ_1)	−.0002	−.0004*	−.00002	−.0008*	−.0001
	(.0004)	(.0002)	(.0005)	(.0004)	(.0003)
max(Hours − k, 0) (γ_2)	.0020**	.0019**	.0020**	.0022**	.0017*
	(.0008)	(.0006)	(.0010)	(.0005)	(.0009)
k	47.9120**	47.8081**	50.2707**	47.6695**	46.6340**
	(3.0975)	(2.2111)	(4.7024)	(1.8799)	(4.3926)
Female	−.0109*	−.0124**	−.0155**	−.0101**	−.0176**
	(.0056)	(.0029)	(.0054)	(.0034)	(.0059)
MBA (FT)	.0004	.0206**			
	(.0077)	(.0058)			
MBA (PT)	.0028	.0005	.0068		
	(.0048)	(.0055)	(.0065)		
MBA (Exec)	.0000	.0129	.0400**		
	(.0127)	(.0081)	(.0164)		
R^2	.1521	.1640	.1641	.1686	.1406
N	485	1,426	716	1,181	730

NOTE.—All specifications include an experience profile and controls for age, race, marital status, number of children, mother's education, college major, and whether in school (except for col. 3). The hours variable refers to first-period reported hours. The errors are clustered at the second interview occupation level. The dependent variable is $(\ln W_4 - \ln W_2)/t$.
* $p < .10$.
** $p < .05$.

samples: γ_2 is 0.0020 in column 1 and 0.0019 in column 2, while the corresponding cutoffs are estimated at 47.9 and 47.8 hours. The coefficient γ_1 is more negative and is significant at the 10% level in column 2, which may be due to workers who are enrolled in school in period 2 switching to a higher paid full-time job upon graduation. This is also supported by the positive and significant coefficient on full-time MBA in column 2; the same coefficient is close to zero for workers who did not change employers. The estimate for γ_0 is higher for job movers (0.3132 compared to 0.1838), which may be indicative of improved job matches, consistent with Jovanovic (1979).

To examine whether school enrollment plays a role, I estimate the model excluding respondents who are attending school at the time of any of the four interviews. I show the results in the third column of table 5. This restriction reduces the sample size to 716. The estimated coefficient on γ_1 is much closer to zero: −0.00002 with a standard error of 0.0005. The value of $\hat{\gamma}_2$ is 0.0020. The estimated value for k is slightly over 50 hours, and the unexplained gender differential increases in absolute value to −0.0155. The last two columns of table 6 show results when the sample is split into MBA completers and workers who did not have an MBA degree at $t = 4$.[18] The estimate for γ_2 for workers with a graduate business degree is 0.0022 and highly significant; it equals 0.0017 and is significant at the 10% level for respondents without an MBA. The breakpoint k is slightly different for the two groups (47.67 vs. 46.63). The unexplained gender difference in wage growth is almost 50% smaller for MBA graduates.

I next replicate the nonlinear least squares results for equation (2) using the full NLSY79 sample of males (first column of table 7)[19] and the subsample of NLSY79 college graduates (second column of table 7). The magnitude of the coefficient γ_2 for college graduates in NLSY79 is somewhat lower than the GMAT Registrant Survey value (0.0014). It is even lower but still significant if I include workers with fewer than 16 years of schooling (0.0007). The estimate for γ_1 is negative and significant in both specifications. The point estimate for k is 46.5 for the full sample and 48.9 for the subsample of college graduates. NLSY79 provides strong evidence that the relationship between hours and wage growth found in the GMAT Registrant Survey is not spurious. This relationship appears to be much stronger for highly educated workers. This supports a theoretical model representative of the high-end labor market. In the theory in Section II, promotions to managerial positions that are highly sensitive to individual productivity play a major role in the wage determination process.

D. Learning by Doing

The idea that future wages increase with current period hours is inherent in learning-by-doing models. This well-known class of models, as developed by Arrow (1962) and Rosen (1972), suggests that time spent at work leads directly to human capital accumulation, so there is a positive relationship between the length of the workweek and productivity growth. Human capital production technology in these models is usually characterized by decreasing returns to scale with respect to the inputs, time in particular (e.g., Rosen 1972). For a pure learning-by-doing model to yield the prediction that the relationship between hours and wage growth is more pronounced at high levels of hours, it is necessary to assume a human capital production function that is convex in hours worked. Such a production function cannot be ruled out, especially in some highly paid professional occupations in which there exist big nondivisible tasks. However, both the

[18] Workers who complete a graduate management degree work on average 1 extra hour per week in the first period: 42.7 compared to 41.7 hours; this difference is statistically significant at the 5% level.

[19] The corresponding results for women are shown in Gicheva (2010), but they should be interpreted with caution because females in NLSY79 tend to experience more career interruptions than females in the GMAT Registrant Survey. Nonetheless, the estimated coefficients do not differ much from the estimates for the sample of men.

Table 7
Nonlinear Least Squares, NLSY79

	Full Sample (1)	College Grads (2)
Constant (γ_0)	.0965**	.0650**
	(.0123)	(.0328)
min(Hours, k) (γ_1)	−.0007**	−.0006**
	(.0001)	(.0003)
max(Hours − k, 0) (γ_2)	.0007**	.0014**
	(.0002)	(.0004)
k	46.4915**	48.8906**
	(1.5189)	(2.2653)
R^2	.0509	.0703

Note.—Estimation results for the model in eq. (2). All specifications include number of years of completed schooling, year dummies, an experience profile, and controls for race, number of children, and marital status. The errors are clustered at the individual level. The hours variable refers to hours reported in $t-2$ to avoid measurement error bias. The dependent variable is $(\ln W_{t+4} - \ln W_t)/4$. $N = 35{,}103$ in col. 1 and 6,575 in col. 2.
* $p < .10$.
** $p < .05$.

GMAT Survey and the NLSY79 data show a convex relationship between hours and wage growth for college-educated males, which would require that a large fraction of all jobs that require a college degree is characterized by increasing returns to hours. Adding the job ladder component to the model makes it more plausible and consistent with the literature.

Lazear and Moore (1984) point out that learning-by-doing and on-the-job training models should apply equally well to self-employed workers, while job-ladder models are not relevant to this group. Their approach would be valid in the setting discussed here if the job level structure described earlier in the paper is not available to those who are self-employed; that is, if self-employed work does not consist of activities with different returns to skill that agents can switch between freely.

I estimate a two-part linear spline specification in which the breakpoint is fixed at 46 hours and compare the estimates for workers in NLSY79 who are self-employed at time t and for the main sample. The estimated coefficient on hours is very different for self-employed and not self-employed workers. The results for the latter sample are similar to the numbers in column 1 of table 7. The results for the self-employed sample fit a learning-by-doing model with a human capital production function that is concave in hours worked. The coefficient on hours equals 0.0015 and is significant at the 5% level when hours are less than 46, and is zero to the right of the breakpoint. This finding serves as some evidence against a simple learning-by-doing model as the single explanation of the observed nonlinear relationship.

Learning by doing is nonetheless an essential component of the model in Section II, and presumably the rate of human capital accumulation decreases for older workers. When the NLSY79 sample is split into workers with fewer than 10 years of experience and observations for which experience is 10 years or more and the value of the cutoff k is set at 46 hours, the estimates suggest that the slope to the right of the break is decreasing in experience, more so for workers with 16 or more years of schooling. The coefficient for the college-educated sample changes from 0.0015 to 0.0006 and loses significance when comparing the low and high experience groups. The change for the full sample is from 0.0006 to 0.0005. The coefficient on hours when hours are lower than 46 remains negative in all specifications.[20] 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.[21]

V. Empirical Evidence on Promotions

Studies of career advancement (Baker, Gibbs, and Holmstrom 1994a, 1994b; Gibbons and Waldman 1999) predict a serial correlation of promotions over time.[22] The theory in Section II, if extended to more than two time periods, would also imply that workers with high learning ability (or lower disutility of hours) receive promotions faster anywhere along the job ladder. The design of the GMAT Registrant Survey is very different from the detailed employment records of a single medium-sized US firm in Baker et al. (1994a, 1994b), but I am able to carry out some related empirical tests. The coefficient of correlation between the number of promotions reported at $t = 1$ and promotions between $t = 1$ and $t = 4$ is 0.2698.[23] When the sample is limited to workers who put in 48 hours or more in the first period, this coefficient increases to 0.3704. The correlation goes up even more (0.4721) when hours are restricted to be greater than or

[20] See Gicheva (2010) for the full set of results discussed in this section.

[21] For example, the model in Sec. II can be extended to more than two periods. Workers with lower values of the speed of learning θ would be promoted at higher levels of experience, and for them the impact of an extra hour on late-career ability would be lower.

[22] Several recent studies add to Baker et al.'s (1994a) empirical evidence in support of the Gibbons and Waldman (1999) model and extend it to different OECD countries. Lluis (2005) uses data from the German Socio-Economic Panel, which provides information on job ranks, and finds increasing returns to skill in upper levels of the job ladder. The conclusion of her study is that assignment into ranks is nonrandom with respect to ability. The investigation in da Silva and van der Klaauw (2011) is based on a large Portuguese matched employer-employee data set. They show a serial correlation in movements from one hierarchical level to the next, even after controlling for observed and unobserved heterogeneity. Kauhanen and Napari (2012) use a long matched panel survey of multiple firms and workers to find similar patterns in the Finnish labor market. In particular, the data again point to a serial correlation in promotions, including in the presence of across-firm mobility.

[23] Number of promotions data are only available for 507 of the 1,911 respondents in the sample.

equal to 55. While these data are likely measured with error and the number of promotions depends largely on the specific hierarchy within each firm, these trends are unlikely to be purely an artifact of measurement error. Workers who report 48 hours or more in wave 1 are more likely to be promoted between waves 1 and 2: 52.36% compared to 40.95% of lower-hour workers. The null hypothesis of no difference can be rejected at a very low significance level. This evidence is consistent with the idea of fast-track jobs.

A testable implication of the theory in Section II is that the probability of receiving a promotion is nondecreasing in hours and inherent ability. The GMAT Registrant Survey asks respondents whether they have managerial responsibilities in their job. I use this information to construct a promotion variable that is more uniform across different employers. I use a subsample of 940 workers who report not having managerial responsibilities at the time of their initial interview and set the promotion measure to equal 1 for those who had managerial duties at $t = 4$ and 0 otherwise. I estimate a partial linear regression model similar to the one in Section IV.A but with the promotion variable on the left-hand side and controls for demographics and schooling on the right-hand side. Figure 4 shows that the relationship between hours worked and the probability of being promoted has the shape predicted by the model: flat for very low and very high values of hours and upward sloping in the middle of the

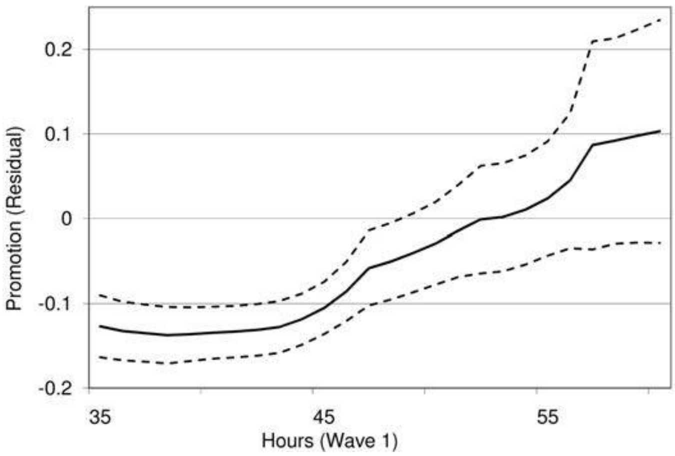


FIG. 4.—Hours and the probability of promotion. Results for a partially linear model: Promotion$_i = Z_i\beta + f(\text{Hours}_{i1}) + \varepsilon_i$. See Section V for the definition of the dependent variable; Z includes gender, an experience profile, age, race, and MBA degree. The hours variable refers to first-period reported hours. The dashed lines represent 90% confidence intervals. The sample is limited to observations for which first-period hours are between 35 and 65. $N = 914$. Epanechnikov kernel; bandwidth = 3.

distribution. In the nonlinear estimation I include only workers whose hours are between 35 and 65 to limit the influence of outliers. Thus, I use 914 of the 940 observations available. The 90% confidence bounds in the figure are fairly wide because of the small sample size. I next parameterize the relationship between hours and promotions and estimate a probit regression model using all 940 observations.

Table 8 shows marginal effects from the probit regressions. Working 5 extra hours per week is associated with a more than 2.5% increase in the probability of receiving a promotion for the average worker in the sample; the estimate is significant at the 5% level even though the sample is small. In the first column I do not include any ability controls, and the estimated relationship is the strongest. The rest of the specifications include two measures of the inherent ability parameter θ. The first measure is whether the respondent attended one of the top 10% most competitive colleges according to *Barron's Profiles of American Colleges* (1994). This variable is observed by firms, so it is possible for employers to use this information when making promotion decisions. The second measure of inherent ability in table 8 is unobserved by employers: an indicator for college-educated mother. Both variables have a positive correlation with the probability of advancing to a managerial position; the coefficients are

Table 8
The Role of Long Hours and Inherent Ability in Promotions

	(1)	(2)	(3)	(4)
Hours (wave 1)	.0068**	.0058**	.0057**	.0051**
	(.0025)	(.0026)	(.0026)	(.0026)
Female	−.0373	−.0336	−.0265	−.0277
	(.0340)	(.0342)	(.0344)	(.0345)
Attended competitive college		.0873*	.0813*	.0814*
		(.0446)	(.0455)	(.0456)
Mom has college degree		.0772**	.0828**	.0861**
		(.0367)	(.0371)	(.0372)
MBA (FT)			.1208**	.1210**
			(.0562)	(.0562)
MBA (PT)			.1148**	.1161**
			(.0367)	(.0368)
MBA (Exec)			.2242**	.2158**
			(.0697)	(.0708)
Experience (wave 1)				.0324*
				(.0185)
Experience squared				−.0009
				(.0009)

Note.—GMAT Registrant Survey sample, limited to workers who do not hold a managerial position at the beginning of the survey. The dependent variable is whether advanced to managerial position. The reported coefficients are marginal effects from probit regressions. All specifications include controls for race, age, and age squared. $N = 940$.
* $p < .10$.
** $p < .05$.

statistically significant. The estimated marginal effect of the college quality variable is over 0.08.[24] The marginal effect of the variable measuring mother's education is similar in magnitude. The results suggest that both hours and ability have a positive correlation with the probability of moving up to a managerial position.

VI. Discussion

A. Alternative Theoretical Approaches

The notion that working hours are tied to career outcomes is most common to "rat race" models, in which the disutility of hours is generally only known to the worker and strongly correlated with a component of productivity unobserved by employers. In these models employers base promotion decisions on revealed preferences. This concept is developed 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. Firms can also use maternity leave take-up as a screening device (Leslie, Manchester, and Park 2008). Landers et al. (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. A "rat race" version of my model can predict the nonlinear relationship between hours and wage growth under a separating equilibrium in which employers set high-hour norms for promotion. In the theory I present in Section II, I obtain the desired result without the need to introduce incomplete information, which is why it is the preferred model. The main difference between the complete and incomplete information models is that the latter implies that some workers may put in inefficiently long hours, while others may be inefficiently left without a promotion.[25]

[24] Seventeen percent of the sample used in the estimation graduated from a competitive college.

[25] It is hard to design a convincing test that could differentiate between the rat race and efficient sorting theories. Some suggestive evidence that most workers in the United States do not work inefficiently long hours can be found. The GMAT Registrant Survey and NLSY79 do not ask respondents about their desired number of hours, but in the PSID sample that Altonji and Paxson (1988) use for their study, only about 5.5% of respondents report being constrained in their choice of hours and working more hours than they would like (although the authors point out that the phrasing of the question may lead to an underestimate). Similarly, the International Social Survey statistics that Bell and Freeman (2001) report show that in 1997, 10.3% of sampled US workers preferred shorter workweeks, while 57.7% were content with their current hours. In contrast, in Landers et al.'s law firm survey, 65.4% of associates report that they would take a reduction in hours if

Along the lines of rat race models, one can similarly consider a career concerns model in which output is linear in ability in each job level and workers are assigned to levels based on expected ability (Holmstrom 1999). Such assignment can create a convexity in the observed returns.[26] The main difference from the separating equilibrium in Landers et al. (1996) is that in Holmstrom's (1999) theory workers determine the number of hours supplied to the labor market, and true ability is unobserved by all market participants but revealed over time. Holmstrom shows that labor supply can deviate up or down from the optimal level, so although hours may not be efficient, the outcome is not necessarily overwork. Career concerns theory is characterized by decreasing incentives over time because there is less learning by employers and less time to realize potential gains from higher perceived productivity. This feature of the model is not fully consistent with the observation that hours for young workers in the high-end labor market tend to be fairly constant or increase slightly over time.[27]

A tournament model (Lazear and Rosen 1981) can also predict a convex relationship between hours and wage growth when hours are closely linked to output, so that workers who choose to put in longer hours are more likely to win the tournament and receive a promotion. A feature of tournament theory is that it characterizes careers within firms. I show in Section IV that the relationship I find is very similar for job stayers and movers, which a tournament model alone cannot account for.

The asymmetric learning theory introduced in Waldman (1984) is an alternative to the tournament model that can be better reconciled with workers earning similar wages upon promotion regardless of whether they stay with their current firm. In this model, only current employers observe workers' ability, but other firms use promotions (or the lack of a promotion) as a signal. Unlike the tournament theory in Lazear and Rosen (1981), wages in each job level are determined by the competitive labor market; within a level, wages do not vary with ability. One way to incorporate hours into the asymmetric learning model is suggested by the extension in Zabojnik and Bernhardt (2001). In their model, ex ante identical work-

given the choice. A question that cannot be answered with the available data is which group is more similar to the population that I study. A 2006 question in NLSY79 asks about the amount of time that respondents devote to volunteer activities. College-educated men in the sample who work 46 hours per week volunteer at least as much as college graduates who work fewer hours (this is consistent with Freeman [1997]). If devoting little or no time to volunteer work is some indication of labor market work beyond the most preferred level, then this should also favor the idea that observed hours tend to be efficient.

[26] Holmstrom bases this observation on the discussion in Rosen (1982).

[27] See, e.g., fig. 2. In the college-educated NLSY79 sample, mean weekly hours conditional on working at least 15 hours are 41.2 in 1984, 44.9 in 1994, and 45.3 in 2004.

ers face different human capital investment incentives depending on their firm's technology. More productive firms hire a greater number of lower-level employees per managerial position, so promotions in these firms are less likely but are associated with higher wage premiums. Workers in these larger firms have a stronger incentive to invest in human capital accumulation, for instance, through longer hours and more learning by doing. This can result in a convex relationship between hours and wage growth. However, table 8 shows that longer hours are also associated with higher promotion probabilities, which can only be reconciled with the Zabojnik sand Bernhardt (2001) model if additional worker heterogeneity is introduced. The first, third, and fourth waves of the GMAT Registrant Survey offer some information about the size of the firms that respondents work at; this information is missing for part of the sample. Including controls for firm size, when available, does not change the estimated relationship between hours and wage growth (these results are available from the author on request), which does not follow the theory in Zabojnik and Bernhardt (2001).

B. Implication of the Results for the Gender Wage Gap

In this section, I present an example that is contingent on the assumption that male and female workers have similar distributions of the speed of learning θ, but the distribution of the disutility of hours b is shifted to the right for women. The model in Section II suggests that this exogenous difference in labor supply preferences leads to an expanding gender wage gap.[28] A contribution of my study is that I use hours at a given point in time to predict the divergence in the wages of men and women over time, while previous studies have been controlling for observed hours in each period.

I find that when controlling for hours worked, women's wages grow 1.24% slower than the wages of men (see table 4). Wave 1 hours exceed 47 for 34% of men and 21% of women. For these long-hour workers, average hours are 53 for women and 54 for men. Then, if the average wage growth for men who put in 47 or fewer hours is x (and assuming that the correlation between hours and wage growth is 0 when $h < 48$),

[28] The terms "preferences" and "disutility of hours" are used broadly here. The exercise in this section is based only on the assumption that the distribution of b is farther to the right for women than it is for men, regardless of the reasons for this shift. Preferences, in the traditional sense of the term, may play a role if female workers derive higher utility from leisure or are less willing to tolerate stressful, long-hour jobs. It is often the case that females are the primary caregivers in households with children or elderly members, either because of comparative advantage or due to gender differences in bargaining power within couples, which would also restrict the hours women are willing and able to supply to the labor market.

$$E[\ln w_1^{\text{male}} - \ln w_0^{\text{male}}] = (0.66)x + (0.34)(x + (E[h|h > 47] - 47)(0.0019)),$$

which means that $E[\ln w_1^{\text{male}} - \ln w_0^{\text{male}}] = x + (0.34)(54 - 47)(0.0019)$. For women, $E[\ln w_1^{\text{female}} - \ln w_0^{\text{female}}] = x - 0.0124 + (0.21)(53 - 47)(0.0019)$.

What these calculations suggest is that there are three main reasons why the wages of male workers tend to grow faster. First, there is the unexplained difference in wage growth, which equals 1.24% per year. Second, fewer women put in more than 47 hours per week and, last, conditional of working more than 47 hours, women average one fewer hour per week. The gender difference in wage growth after accounting for demographics and education is $0.0124 + (0.0019)((0.34)(7) - (0.21)(6)) = 0.015$. Only about 14% of the gap in wage growth is accounted for by differences in hours. These calculations use the pooled sample coefficient for the relationship between hours and wage growth (0.0019). Columns 2 and 3 of table 4 suggest that hours might account for a larger portion of the gender wage gap because, if estimated separately, the coefficients equal 0.0021 for men and 0.0013 for women (0.0018 for women with no children at the time of the second interview).

Bertrand et al. (2010) find that for graduates of the University of Chicago Booth School of Business, the gap in the wages of men and women without career interruptions increases by 20 log points in the first 9 years after graduation.[29] They attribute about half of this gap to differences in hours. This means that the growth of men's wages is on average 2.2% faster per year.[30] Women in the GMAT Registrant Survey sample experience slightly more career interruptions than men, but the way in which I select respondents imposes that they held a job in waves 1, 2, and 4. On average, women in the sample accumulate 3 fewer months of experience between the first and last installments of the survey.

I expand the findings in Bertrand et al. (2010) to include a wider segment of the labor force. Extremely long hours, higher earnings, and fast career advancement are typical for the population of their study; the GMAT Registrant Survey sample is representative of a somewhat less high-end portion of the labor market. In the GMAT Registrant Survey sample, the gender difference in hours when reported hours are greater than 47 is only about 1 hour, while in the Bertrand et al. (2010) sample this difference starts at 1.8 right after graduation and grows to 6 hours in the ninth year after graduation (for all employed women, including those with career interruptions). Given a 6-hour difference in the labor supply of men and women working

[29] Bertrand et al. (2010) run a separate wage regression for each year after graduation and find that the coefficient on the *female* indicator increases with the number of years after completing an MBA degree.

[30] The mean annual increase in the gender wage gap varies between 1.6 and 4.3 log points if estimates for the difference in wages 2 to 8 years postgraduation are used instead.

more than 47 hours, my results would predict a gender gap in annual wage growth for high-labor supply workers equal to 0.0124 + (0.0019)(6), or 2.4 log points. In this case the difference in hours worked accounts for 48% of the gap. These numbers are very close to what Bertrand et al. (2010) find.

The implications of the long-term consequences of gender differences in labor supply are dependent on the way in which one chooses to model the high-end labor market. In the rat race setup, in which there are inefficiencies associated with long hours (Akerlof 1976; Landers et al. 1996), appropriate policy interventions that favor employees who choose shorter hours due to child care or elderly care responsibilities may be Pareto improving. Landers, Rebitzer, and Taylor (1997) discuss the potential benefits and downsides of several such policies. Interventions of this sort could also be beneficial if, as discussed in note 28, the labor market is characterized by efficient sorting, but revealed labor supply preferences are based on inefficient division of household responsibilities. If, on the other hand, sorting in the high-end market is fully efficient, which is the most direct interpretation of the model in Section II, then policy interventions aimed to equalize gender differences in the labor market are not necessary.

C. Quantifying the Return to an Additional Hour

Working an extra hour per week in the present period ($t = 0$) has a twofold effect on lifetime earnings. First, there is a direct effect on current period earnings holding the wage rate constant.[31] Second, there is a marginal return to labor supply from the increase in wage growth between years 0 and 1. The relationship between hours and wage growth that I estimate in Section IV is an annualized average using wage growth over a 5–6-year period. For this example I assume that the estimated coefficient represents the relationship between hours and a 1-year increase in wages.

Consider a male worker who normally puts in 48 hours per week but for 1 year increases his workweek by 1 hour. For men who at the time of their first interview work between 48 and 53 hours, median hourly earnings equal \$15.50 (the median is \$15 with a mean of \$15.80 for all 1,103 males in the sample). The average wage growth for males who put in 47 or 48 hours in period 1 is 5.4% per year, which is the number I use below (the median annualized wage growth in the full sample of men is 5.6% with a mean of 6.4%). The present discounted value associated with the 0.19% increase in hourly wage growth between $t = 0$ and $t = 1$ is

$$PDV = \sum_{t=1}^{T} \beta^t (0.0019)(48)(\$15.50)(1.054)^{(t-1)}.$$

[31] In the GMAT Registrant Survey sample, wave 2 wages do not vary much with hours in the middle portion of the wage distribution, so I assume that an increase in hours worked does not have an immediate effect on hourly earnings.

Assuming a discount factor β of 0.9 and $T = 20$ years of experience (a fairly low number given that I focus on workers whose average age is below 30), the shadow wage is \$31.64 when the observed wage is \$15.50. This means that the marginal return from higher wage growth is as large as the observed wage.

Alternatively, one can compare the predicted earnings of two employees with the same ability θ but different labor supply preferences: one who always works 48 hours and another one who consistently puts in 49 hours per week. The first worker has a present discounted value of earnings given by

$$PDV_1 = \sum_{t=0}^{T} \beta^t (1.054)^t (48)(52)(\$15.50),$$

which is equal to \$465,478 when $\beta = 0.9$ and $T = 20$. Assuming that the relationship between hours and wage growth is present between years 0 and 10, the earnings stream with a 49-hour workweek is

$$\begin{aligned} PDV_2 = & \sum_{t=0}^{10} \beta^t (1.0559)^t (49)(52)(\$15.50) \\ & + \sum_{t=11}^{T} \beta^t (1.054)^{(t-10)} (1.0559)^{10} (49)(52)(\$15.50). \end{aligned}$$

Then $PDV_2 = \$481{,}124$ and the difference between the two numbers is \$15,646. If hours did not affect wage growth, the present discounted value of the earnings of the second worker would equal $PDV_3 = \$475{,}176$. Then $PDV_3 - PDV_1 = \$9{,}697$. This shows that 38% of the increase in the lifetime earnings of a worker who is willing to put in 1 extra hour per week is due to faster wage growth.

VII. Conclusion

In this paper, I develop a model that combines learning on the job and promotions when workers differ in their preferences for leisure. I use this theory to explain a previously undocumented trend that constitutes the focus of the empirical part of the study: wage growth is increasing in working hours when hours are high.

Using a four-wave panel survey of GMAT registrants and data from the 1979 cohort of the NLSY, I find that for employees whose workweek is 48 hours or longer, adding 5 extra hours per week is associated with a 1% increase in annual wage growth. The correlation is zero when weekly hours are less than 48. The observed relationship between hours and wage growth is similar in the two data sets when the NLSY79 sample is restricted to college graduates, which implies that the applicability of the findings is not limited to the population of GMAT registrants. The relationship is present but much weaker among workers with less than 16 years of schooling, which points to an underlying mechanism specific to the high-end labor market. The results are robust to various functional form specifications and estimation procedures.

The data confirm several other testable implications of the model. The probability of receiving a promotion is an increasing function of hours worked and proxies for the speed of learning on the job. I estimate the empirical model using a sample of self-employed workers from NLSY79 and find that for these workers the observed relationship between hours worked and the change in hourly wages is consistent with a learning-by-doing model with a concave human capital production function and no promotion-related discrete wage jumps.

The estimated relationship between hours and wage growth is fairly large in magnitude. It implies that the observed hourly wage is about half of the shadow wage for workers who are young, where the unobserved portion of the value of an additional hour is due to the increase in wage growth.[32] Under the assumption that women have inherently higher disutility of hours, my findings can be used to account for somewhere between 14% and 48% of the gender difference in wage growth, depending on the size of the labor supply gap. These numbers are important in that they provide a better understanding of the labor market choices of young highly educated workers and the career implications of these choices.

One limitation of the model and data analysis that should be addressed in future work is that the added complexity of joint household labor supply decisions is ignored. In the GMAT Registrant Survey sample, 94% of females and 86% of males have no children at the time of their first interview, which is also when the measure of hours used in the empirical analysis is collected. While household decisions may not be a driving force in the GMAT Survey data, it is likely that among other groups of the population this may be a factor that exacerbates the gender differences in wage growth.

Appendix

A. Probability of Receiving a Promotion Conditional on Inherent Ability and First-Period Hours

The promotion cutoff for the disutility of labor parameter $\bar{b}(\theta_i)$ takes the form $\bar{b}(\theta_i) = A\theta_i$, where A is a function of the parameters in the model.

[32] In his overview of some of the literature on the elasticity of labor supply, Keane (2011) also points out that the shadow value of an additional hour tends to be considerably higher than the observed wage, especially for young workers, when labor supply decisions are viewed in an intertemporal framework that allows for learning by doing.

Then Pr(Promotion$|\theta_i$) $= Pr(b_i \leq A\theta_i)$, which equals $F_b(A\theta_i)$, where F_b is the density function that characterizes the distribution of the disutility parameter. The probability of promotion is increasing in θ_i but the exact shape of the relationship depends on the distribution of b.

Suppose that for worker i observed hours equal $h_{1i} = h$. Denote

$$\underline{g}(h) = \frac{d_1 + c_1\eta_1}{(2A - c_1\eta_1)h}$$

and

$$\bar{g}(h) = \frac{2A(d_1 + c_1\eta_1) + c_2\eta_1(d_2 + c_2\eta_1)}{(4A^2 - c_2^2\eta_1^2)h}.$$

Then

$$\Pr(\text{Promotion}|h) = \frac{1 - F_\theta(\bar{g}(h))}{F_\theta(\underline{g}(h)) + 1 - F_\theta(\bar{g}(h))},$$

where F_θ is the density of θ.[33] Both $\underline{g}(h)$ and $\bar{g}(h)$ are decreasing in h when h is positive, so the probability of receiving a promotion is nondecreasing in hours. For the slope to be zero hours have to be either very high or very low. When observed hours h_L are low so that $F_\theta(\bar{g}(h_L)) = 1$ and $F_\theta(\underline{g}(h_L)) \leq 1$, the probability of receiving a promotion is zero. Even high values of θ can be insufficient to compensate for a very high disutility of hours. Alternatively, for a high value of hours, h_H, such that $F_\theta(\bar{g}(h_H)) \geq 0$ and $F_\theta(\underline{g}(h_H)) = 0$, the probability of receiving a promotion is 1.

This suggests that Pr(Promotion$|h_1$) is increasing in first-period hours, but the shape of the relationship depends on any assumptions about the distributions of the parameters.

B. Construction of NLSY79 Sample

I drop the 1,280 respondents from the military sample of NLSY79 because they have different employment patterns and most are excluded from the survey after 1984. Females are dropped from the sample. 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.

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

33 See Gicheva (2010) for more details.

period, since respondents are given a choice of time unit to report earnings. Wages are reported in 1991 dollars. 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 full-time in year t, he was also working full-time in $t-1$. When available, I use a series of variables on employment history for the period 1975–77 to construct actual experience.

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 \$200 and usual weekly hours are at least 15. I set hours equal to 90 if the reported value exceeds 90. Self-employed workers and those with not enough information to construct the actual experience measure are dropped.

C. Alternative Empirical Specifications

1. Maximum Likelihood Estimation

The approach in this section is based on the estimation technique proposed by Quandt (1958); the method is appropriate when the independent variable that determines the nonlinearity is discrete. Let N denote the number of observations in the sample. If the data are sorted on hours, for every $k \in [h_{\min};\ h_{\max}]$ there exists 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

$$\log L = -\frac{N}{2}\log(2\pi) - m\log\sigma_1 - (N-m)\log\sigma_2 \\ -\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(y_i - Z_i\beta - a_1 - b_1\text{Hours}_{i1})^2 - \frac{1}{2\sigma_2^2}\sum_{i=m+1}^{N}(y_i - Z_i\beta - a_2 - b_2\text{Hours}_{i1})^2. \tag{A1}$$

The parameters σ_1 and σ_2 denote the standard deviations of the errors for observations respectively before and after the breakpoint k, and $y_i = (\ln W_{i4} - \ln W_{i2})/t$. I maximize the objective function in (A1) over every integer value of k excluding the ends of the observed hours distribution where there are not enough observations to identify either a_1 and b_1 or a_2 and b_2. I choose the set of parameter estimates based on the k for which the log likelihood function attains its highest level. Figure A1 shows the distribution of the maximized log likelihood values over the range of k for the pooled sample of men and women. The function has several local maxima but attains its global maximum at 48 hours.

Table A1 shows the maximum likelihood estimates of the parameters in (A1) for different subsets of the controls in the vector Z_i. The optimal change in slope occurs at $k = 48$ in all specifications. The estimates for the slope before and after the cutoff are very similar to the numbers reported in table 4. The results in table A1 also imply that the data are heteroskedastic: the standard deviation of the error is smaller at lower levels of hours (0.0577 compared to 0.0708), and the difference is statistically significant. However, accounting for heteroskedasticity in the model specification does not affect the slope coefficients, as suggested by the closeness of the nonlinear least squares and maximum likelihood estimates.

It is also evident from table A1 that there is no discontinuity at k. In the specification with the full set of controls the point estimate for a_1 is 0.2870 and the estimate for b_1 is -0.0004, so the value of y approaches $a_1 + kb_1 = 0.2678$ when hours approach 48 from the left. The point estimates for a_2 and b_2 equal 0.1779 and 0.0019, respectively, so y approaches $0.1779 + (0.0019)(48) = 0.2691$ when hours approach 48 from the right.

Quandt (1958) also proposes a likelihood ratio test of the hypothesis that no switch occurred against the alternative that there is one change in regime. He defines the likelihood ratio as $\lambda = [L(\bar{\Omega})]/[L(\Omega)]$, 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 below 0.001.

2. Polynomial in Hours and Wage Level Specifications

A convex relationship between hours and wage growth can also be modeled with an mth degree polynomial function of the form

$$(\ln W_{i4} - \ln W_{i2})/t = \alpha_0 + \sum_{r=1}^{m} \alpha_r \text{Hours}_{i1}^r + Z_i\beta + \nu_i. \tag{A2}$$

For regressions with $m \geq 2$, the resulting coefficients are estimated with a high level of noise, so $m = 2$ yields the preferred specification. An even less parametrically restrictive functional form would allow the coefficient on $\ln W_{i2}/t$ to be different from one. I estimate two variations of this model: a nonlinear least squares one:

$$\begin{aligned}(\ln W_{i4} - \ln W_{i2})/t = \gamma_0 + \gamma_1 \min(\text{Hours}_{i1}, k) + \gamma_2 \max(\text{Hours}_{i1} - k, 0) \\ + \gamma_3 \ln W_{i2} + Z_i\beta + \varepsilon_i\end{aligned} \tag{A3}$$

and a quadratic in hours:

$$(\ln W_{i4} - \ln W_{i2})/t = \alpha_0 + \sum_{r=1}^{m} \alpha_r \text{Hours}_{i1}^r + \delta \ln W_{i2} + Z_i\beta + \nu_i. \tag{A4}$$

All results for the full GMAT Survey sample, along with the baseline specification in equation (2), are illustrated in figure A2, where I plot the predicted change in log wages as a function of hours: $\hat{y} = \hat{\gamma}_1 \min(\text{Hours}_{i1}, \hat{k}) + \hat{\gamma}_2 \max(\text{Hours}_{i1} - \hat{k}, 0)$ for the nonlinear least squares specifications and $\hat{y} = \hat{\alpha}_1 \text{Hours}_{i1} + \hat{\alpha}_2 \text{Hours}_{i1}^2$ for the polynomial specifications. Confidence intervals are not shown, but all estimates with the exception of $\hat{\gamma}_1$ in (2) and (A3) are significant at the 1% level. It is evident from figure A2 that the slope of the estimated relationship between hours and the change in log wages when hours exceed 47 is consistent across the different specifications.

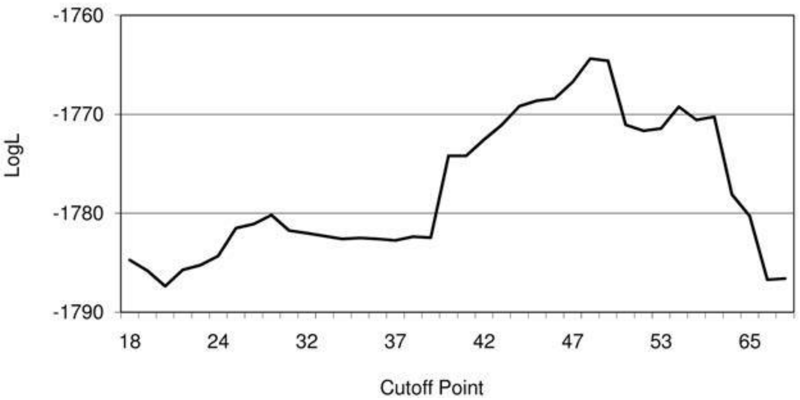


FIG. A1.—Distribution of the log likelihood function values over the cutoff k. Maximized values of the log likelihood function in equation (3) over the observed distribution of hours excluding the tails. The full GMAT Registrant Survey sample is used in the estimation. $N = 1{,}911$.

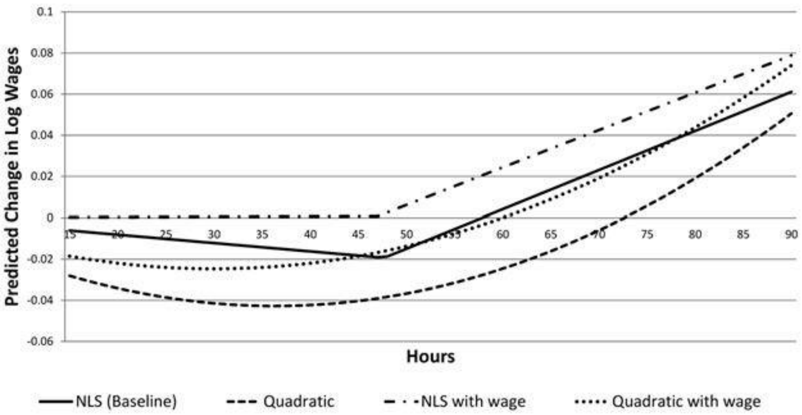


FIG. A2.—Predicted change in log wages from alternative specifications

Table A1
Maximum Likelihood Estimation Results

	(1)	(2)	(3)	(4)
Hours < k:				
Constant (a_1)	.1167**	.1038**	.2938**	.2870**
	(.0131)	(.0138)	(.0738)	(.0745)
Hours (b_1)	−.0006	−.0004	−.0004	−.0004
	(.0003)	(.0003)	(.0003)	(.0003)
Hours ≥ k:				
Constant (a_2)	−.0030	−.0050	.1863**	.1779**
	(.0358)	(.0372)	(.0821)	(.0831)
Hours (b_2)	.0020**	.0019**	.0019**	.0019**
	(.0006)	(.0006)	(.0006)	(.0006)
Female	−.0104**	−.0125**	−.0124**	−.0121**
	(.0047)	(.0045)	(.0044)	(.0045)
σ_1	.0597**	.0582**	.0578**	.0577**
	(.0009)	(.0009)	(.0008)	(.0008)
σ_2	.0712**	.0714**	.0708**	.0708**
	(.0031)	(.0029)	(.0029)	(.0029)
k	48	48	48	48
Education	No	Yes	Yes	Yes
Demographics	No	No	Yes	Yes
Family	No	No	No	Yes

NOTE.—Estimation results for the model in eq. (A1). The dependent variable is $(\ln W_4 - \ln W_2)/t$. All specifications include an experience profile. Education controls: MBA dummies, whether in school at the time of second interview, college major. Demographics: age, age squared, race dummies. Family controls: marital status, number of children, mother's education. The hours variable refers to first-period reported hours to avoid measurement error bias. $N = 1{,}911$.
* $p < .10$.
** $p < .05$.

References

Akerlof, George A. 1976. The economics of caste and of the rat race and other woeful tales. *Quarterly Journal of Economics* 90, no. 4:599–617.

Altonji, Joseph G. 1986. Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy* 94, no. 3: S176–S215.

Altonji, Joseph G., and Christina H. Paxson. 1988. Labor supply preferences, hours constraints, and hours-wage trade-offs. *Journal of Labor Economics* 6, no. 2:254–76.

Arcidiacono, Peter, Jane Cooley, and Andrew C. Hussey. 2008. The economic returns to an MBA. *International Economic Review* 49, no. 3: 873–99.

Arrow, Kenneth J. 1962. The economic implications of learning by doing. *Review of Economic Studies* 29, no. 3:155–73.

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994a. The internal economics of the firm: Evidence from personnel data. *Quarterly Journal of Economics* 109, no. 4:881–919.

———. 1994b. The wage policy of a firm. *Quarterly Journal of Economics* 109, no. 4:921–55.

Barron's Profiles of American Colleges. 1994. Hauppage, NY: Barron's Educational Series.

Bell, Linda A., and Richard B. Freeman. 2001. The incentive for working hard: Explaining hours worked differences in the US and Germany. *Labour Economics* 8, no. 2:181–202.

Bertrand, Marianne, Claudia Goldin, and Lawrence Katz. 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics* 2, no. 3: 228–55.

Blair-Loy, Mary. 2004. Work devotion and work time. In *Fighting for time: Shifting boundaries of work and family*, ed. C. F. Epstein and A. Kalleberg, 282–316. New York: Russell Sage Foundation.

Borjas, George J. 1980. The relationship between wages and weekly hours of work: The role of division bias. *Journal of Human Resources* 15, no. 3: 409–23.

Charness, Gary. 2004. Attribution and reciprocity in an experimental labor market. *Journal of Labor Economics* 22, no. 3:665–88.

Charness, Gary, and Peter Kuhn. 2007. Does pay inequality affect worker effort? Experimental evidence. *Journal of Labor Economics* 25, no. 4: 693–723.

Clemens, Adam. 2012. Position-specific promotion rates and the "fast track" effect. *Research in Labor Economics* 36:77–107.

da Silva, Antonio Dias, and Bas van der Klaauw. 2011. Wage dynamics and promotions inside and between firms. *Journal of Population Economics* 24, no. 4:1513–48.

Deaton, Angus. 1988. Quality, quantity, and spatial variation of price. *American Economic Review* 78, no. 3:418–30.

Fehr, Ernst, and Lorenz Goette. 2007. Do workers work more if wages are high? Evidence from a randomized field experiment. *American Economic Review* 97, no. 1:298–317.

Freeman, Richard B. 1997. Working for nothing: The supply of volunteer labor. *Journal of Labor Economics* 15, no. l, pt. 2: S140–S166.

Gibbons, Robert, and Michael Waldman. 1999. A theory of wage and promotion dynamics inside firms. *Quarterly Journal of Economics* 114, no. 4:1321–58.

———. 2006. Enriching a theory of wage and promotion dynamics inside firms. *Journal of Labor Economics* 24, no. 1:59–107.

Gicheva, Dora. 2010. Working long hours and early career outcomes in the high-end labor market. University of North Carolina at Greensboro, Economics Department Working Paper no. 10-3.

Goldin, Claudia, and Lawrence Katz. 2008. Transitions: Career and family life cycles of the educational elite. *American Economic Review: Papers and Proceedings* 98, no. 2:363–69.

Graduate Management Admission Council. 1996. Profile of graduate management admission test candidates, 1990–91 to 1994–95. Graduate Management Admission Council, McLean, VA.

Holmstrom, Bengt. 1999. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies* 66, no. 1:169–82.

Imai, Susumu, and Michael P. Keane. 2004. Intertemporal labor supply and human capital accumulation. *International Economic Review* 45, no. 2:601–41.

Jovanovic, Boyan. 1979. Job matching and the theory of turnover. *Journal of Political Economy* 87, no. 5:972–90.

Kauhanen, Antti, and Sami Napari. 2012. Career and wage dynamics: Evidence from linked employer-employee data. *Journal of Labor Research* 36:35–76.

Keane, Michael. 2011. Human capital, taxes and labour supply. *Economic Record* 87, no. S1:37–44.

Kuhn, Peter, and Fernando Lozano. 2008. The expanding workweek? Understanding trends in long work hours among U.S. men, 1979–2006. *Journal of Labor Economics* 26, no. 2:311–43.

Landers, Renee M., James B. Rebitzer, and Lowell J. Taylor. 1996. Rat race redux: Adverse selection in the determination of work hours in law firms. *American Economic Review* 86, no. 3:329–48.

———. 1997. Work norms and professional labor markets. In *Gender and family issues in the workplace*, ed. F. D. Blau and R. G. Ehrenberg, 166–202. New York: Russell Sage Foundation.

Lazear, Edward P., and Robert L. Moore. 1984. Incentives, productivity, and labor contracts. *Quarterly Journal of Economics* 99, no. 2:275–96.

Lazear, Edward P., and Sherwin Rosen. 1981. Rank-order tournaments as optimum labor contracts. *Journal of Political Economy* 89, no. 5:841–64.

Leslie, Lisa M., Colleen F. Manchester, and Tae-Youn Park. 2008. Screening for commitment: The effect of maternity leave usage on wages. Working paper, University of Minnesota.

Lluis, Stephanie. 2005. The role of comparative advantage and learning in wage dynamics and intrafirm mobility: Evidence from Germany. *Journal of Labor Economics* 23, no. 4:725–68.

MaCurdy, Thomas E. 1981. An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy* 89, no. 6:1059–85.

Montgomery, Mark, and Irene Powell. 2003. Does an advanced degree reduce the gender wage gap? Evidence from MBAs. *Industrial Relations* 42, no. 3:396–418.

Niederle, Muriel, and Lise Vesterlund. 2007. Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics* 122, no. 3:1067–1101.

Quandt, Richard E. 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* 53, no. 284:873–80.

Rebitzer, James B., and Lowell J. Taylor. 1995. Do labor markets provide enough short-hour jobs? An analysis of work hours and work incentives. *Economic Inquiry* 33, no. 2:257–73.

Reich, Robert B. 2001. *The future of success*. New York: Knopf.

Robinson, Peter M. 1988. Root-n-consistent semiparametric regression. *Econometrica* 56, no. 4:931–54.

Rosen, Sherwin. 1972. Learning by experience as joint production. *Quarterly Journal of Economics* 86, no. 3:366–82.

———. 1982. Authority, control, and the distribution of earnings. *Bell Journal of Economics* 13, no. 2:311–23.

Waldman, Michael. 1984. Job assignments, signalling, and efficiency. *RAND Journal of Economics* 15, no. 2:255–67.

Zabojnik, Jan, and Dan Bernhardt. 2001. Corporate tournaments, human capital acquisition, and the firm size-wage relation. *Review of Economic Studies* 68, no. 3:693–716.