Cornell University ILR School

Cornell University ILR School
DigitalCommons@ILR

Working Papers ILR Collection

October 2007

The Signaling Role of Promotions: Further Theory and Empirical Evidence

Jed DeVaro
Cornell University, jed.devaro@csueastbay.edu

Michael Waldman
Cornell University, mw46@cornell.edu

Follow this and additional works at: http://digitalcommons.ilr.cornell.edu/workingpapers
Thank you for downloading an article from DigitalCommons@ILR.
Support this valuable resource today!

This Article is brought to you for free and open access by the ILR Collection at DigitalCommons@ILR. It has been accepted for inclusion in Working Papers by an authorized administrator of DigitalCommons@ILR. For more information, please contact hlmdigital@cornell.edu.

The Signaling Role of Promotions: Further Theory and Empirical Evidence

Abstract
[Excerpt] An extensive theoretical literature has developed that investigates the role of promotions as a signal of worker ability. There have been no tests, however, of the empirical validity of this idea. In this paper we develop the theory in a manner that allows us to generate testable predictions, and then investigate the validity of these predictions using a longitudinal data set that contains detailed information concerning the internal-labor-market history of a medium-sized firm in the financial-services industry. Our results support the notion that signaling is both a statistically significant and economically significant factor in promotion decisions. The paper also contributes to the extensive literature on the role of education as a labor-market signal.

Keywords
worker, ability, signal, promotion, signaling, education, labor market

Comments
DeVaro, J., & Waldman, M. (2007). *The signaling role of promotions: Further theory and empirical evidence.* Retrieved [insert date], from Cornell University, ILR School site: http://digitalcommons.ilr.cornell.edu/workingpapers/8

This article is available at DigitalCommons@ILR: http://digitalcommons.ilr.cornell.edu/workingpapers/8

THE SIGNALING ROLE OF PROMOTIONS:

FURTHER THEORY AND EMPIRICAL EVIDENCE

by

Jed DeVaro
School of Industrial and Labor Relations
Cornell University
357 Ives Hall
Ithaca, NY 14853
(607) 255-8407, devaro@cornell.edu

and

Michael Waldman
Johnson Graduate School of Management
Cornell University
Sage Hall
Ithaca, NY 14853
(607) 255-8631, mw46@cornell.edu

October 2007

* We would like to thank Dan Bernhardt, Bob Gibbons, Michael Gibbs, Heather Haveman, James Malcomson, Joshua Pinkston, Bryan Ricchetti, Eric Van den Steen, Jan Zabojnik, and seminar and conference participants at Colgate University, Columbia University, Cornell University, University of Rochester, University of Washington, St. Louis, the Harvard-MIT Organizational Economics Workshop, and the 2005 Society of Labor Economists Conference for helpful comments, Fidan Ana Kurtulus and David Rosenblum for research assistance, and George Baker and Michael Gibbs for providing us with the data used in the study.

ABSTRACT

An extensive theoretical literature has developed that investigates the role of promotions as a signal of worker ability. There have been no tests, however, of the empirical validity of this idea. In this paper we develop the theory in a manner that allows us to generate testable predictions, and then investigate the validity of these predictions using a longitudinal data set that contains detailed information concerning the internal-labor-market history of a medium-sized firm in the financial-services industry. Our results support the notion that signaling is both a statistically significant and economically significant factor in promotion decisions. The paper also contributes to the extensive literature on the role of education as a labor-market signal.

I. INTRODUCTION

An extensive theoretical literature has investigated the idea that promotions serve as a signal of worker ability.[1] By this we mean that when a worker is promoted the event is observed by other potential employers, and these other firms infer the worker is of high ability. Despite the significant theoretical attention paid to this idea, however, there has been no empirical investigation of the theory's real-world validity. In this paper we first extend the theory in a manner that allows us to develop testable predictions, and then empirically investigate the validity of these predictions. Our results support the idea that promotions serve as a signal of worker ability.

Most of the papers in this literature consider a model similar to the one originally investigated in Waldman (1984a). That analysis considers a two-period model in which a firm's job ladder consists of two jobs, all young workers are assigned to the low-level job, and after the first period each worker's first-period employer privately observes the worker's ability. Then in the second period each firm must decide whether or not to promote each old worker it employed in the previous period, where this promotion decision is publicly observed. There are four main results. First, when a worker is promoted other potential employers infer the worker is of high ability and thus increase the amount they are willing to offer the worker. Second, anticipating this behavior, the first-period employer offers a large wage increase with the promotion in order to stop the worker from being bid away. Third, because a large wage increase is necessary, firms promote fewer workers than is first-best optimal. Fourth, this distortion decreases with the importance of firm-specific human capital.

The first step of our analysis is to enrich the standard theoretical approach to this issue in a manner that allows us to develop testable implications. In the standard approach described above, all workers are observationally equivalent when they enter the labor market. In our theoretical analysis, in contrast, workers are heterogeneous in a publicly observable fashion when they enter the labor market. In particular, workers vary in terms of their publicly observed schooling levels. Workers with more education have higher expected ability than workers with less education, although as in the standard

[1] Papers in this literature include Waldman (1984a,1990), Ricart i Costa (1987), MacLeod and Malcomson (1988), Bernhardt (1995), Zabojnik and Bernhardt (2001), Owan (2004), and Golan (2005). See also the related analysis of Milgrom and Oster (1987). Other papers that allow for asymmetric information of the sort considered in this literature but do not consider the possibility of promotions serving as signals include Greenwald (1986), Lazear (1986), and Chang and Wang (1996).

approach each worker's actual ability is initially unobserved by firms.[2]

In our analysis we first show that the model exhibits the basic signaling results found in the previous literature. That is, when a worker is promoted the worker receives a large wage increase in order to prevent the worker from being bid away. In turn, because of the large wage increase, firms distort the promotion decision so that fewer workers are promoted than in the first best. We then show that these basic signaling results vary with a worker's schooling level. Because a worker with a higher level of schooling is thought of as having higher expected ability when he or she enters the labor market, the signal associated with being promoted improves beliefs about worker ability less for workers with high education levels. This, in turn, yields the following testable implications. First, because the wage associated with not being promoted is higher for those with higher levels of education, firms distort the promotion decision less for these workers. Second, the wage increase associated with being promoted decreases with worker education because the signal is smaller for more highly educated workers. Note that in terms of testing, the first prediction implies that, holding performance fixed, the probability of promotion is higher for more highly educated workers (this is explained in detail in Section III).

To more clearly understand the logic for these two predictions, consider a firm that hires two workers into the same job, where the workers are similar with the exception being that one worker has an MBA while the other has only an undergraduate degree. Because employers believe that MBAs are more productive on average, other firms learn little about the worker from a promotion. The result is that when a promotion takes place the firm does not offer a large wage increase since there is not a big effect on other firms' wage offers. Further, since a promotion is not associated with a large wage increase, the firm does not distort the promotion decision in a significant fashion.

In contrast, suppose the firm promotes the worker who has only an undergraduate degree. Because such workers are, on average, not as highly thought of as those with MBAs, in this case when the worker is promoted other firms positively update their beliefs concerning the worker's ability by a significant amount and, in turn, significantly increase the amount they are willing to pay the worker. The result is that the current employer provides a large wage increase upon promotion to stop the worker

[2] Bernhardt (1995) incorporates heterogeneous education levels into a model characterized by the promotion-as-signal hypothesis. He then derives one of the results that we focus on below which is that the incentive for a firm to distort the promotion decision is smaller for more highly educated workers. More recently, Ishida (2004a,2004b) also incorporates education into models in which promotions serve as signals of worker ability, where his focus is on the interaction between education as a signal and promotion as a signal. In contrast to our paper, these papers do not focus on testable implications and also do not provide any empirical testing.

from leaving and, because of the large wage increase associated with a promotion, the firm only promotes the worker if the worker is significantly more productive at the higher level job. In other words, consistent with our two testable predictions, the wage increase due to the promotion is larger for the less educated worker and the promotion decision is more biased against this worker.

After establishing these two testable predictions, we investigate the validity of these predictions using panel data on the personnel records for managerial workers in a single medium-sized firm in the financial-services industry over a twenty-year time period (this data set was first investigated in the classic empirical study of internal labor markets found in Baker, Gibbs, and Holmstrom (1994a,b)). Two aspects of this data set make it ideal for testing the predictions just described. First, the personnel records contain annual supervisor ratings of each worker's job performance. Second, the data set includes detailed information on the firm's job ladder constructed by Baker, Gibbs, and Holmstrom from the raw data on job titles and typical promotion paths. Together, these features of the data set allow us to test with confidence both how the education level affects the probability of promotion, and how wage increases received upon promotion are associated with the worker's education level.

The results of our analysis support the signaling theory of promotions. First, we find that from the standpoint of both statistical and economic significance increasing a worker's education level increases the probability of promotion, and other than for high school graduates we similarly find that the wage increase associated with promotion decreases with worker education. For example, after controlling for a variety of worker attributes such as initial job level, job performance, and firm tenure, decreasing the education level from masters degree to bachelors degree decreases by about twenty percent the probability a worker is promoted in the following year. In terms of the prediction concerning the relationship between education and wage growth, we find that after controlling for a number of worker attributes, decreasing the education level from masters degree to bachelors degree increases the average percentage wage increase due to promotion by over seventy percent. We also discuss a possible explanation for why the wage growth prediction is not strongly supported for high school graduates.

Second, we consider other potential explanations for our results, both theoretical and empirical, and find none that match the evidence as well as the promotion-as-signal hypothesis. For example, one alternative explanation for why there is a positive relationship between promotion probability and education level even after controlling for performance is that the skills acquired during education are more useful at higher levels of the job ladder. But our investigation finds that this idea does not explain our findings. For example, if this were the correct explanation, then including predicted post-promotion performance as an explanatory variable should significantly weaken the positive relationship between promotion probability and education. But when we include predicted post-promotion performance as an explanatory variable there is no effect on this relationship. We also consider a number of other potential explanations, including symmetric rather than asymmetric learning, the possibility that performance evaluations are a coarse measure of actual worker performance, that the performance ratings themselves are biased, and that the firm runs biased promotion tournaments. Our investigation indicates that none of these ideas matches the evidence as well as the promotion-as-signal hypothesis, although we do find some evidence consistent with one or more of the alternatives.

This paper contributes to a small but growing empirical literature on asymmetric learning in labor markets. The classic paper in this literature is Gibbons and Katz (1991) which focuses on the idea that being laid off sends a more damaging signal of worker ability than being fired in a plant closing.[3] They develop a number of theoretical predictions and then empirically test the predictions and find supporting evidence. Doiron (1995) and Grund (1999) also find supporting evidence for the Gibbons and Katz predictions using data for Canada and Germany, respectively, while Acemoglu and Pischke (1998) extend the Gibbons and Katz framework to consider incentives for firms to provide their workers with general human capital and then show supporting evidence using German data. More recently, Pinkston (2006) and Schonberg (2007) develop further tests of asymmetric learning in labor markets and also find supporting evidence. For example, Schonberg finds evidence consistent with asymmetric learning for university graduates but not for high school graduates and dropouts. Note that, as was true of the Gibbons and Katz approach, Pinkston and Schonberg focus on implications of asymmetric learning other than those associated with promotions and signaling. Hence, our paper is the first to focus on testable predictions of the promotion-as-signal hypothesis.[4]

This paper also contributes to the extensive literature on education as a labor-market signal that

[3] The Gibbons and Katz predictions can be interpreted either as a layoff serving as a negative signal, or in terms of the adverse-selection theory of labor-market turnover first put forth in Greenwald (1986).

[4] Another related paper is Belzil and Bognanno (2005). They employ an eight-year panel of promotion histories of 30,000 American executives to test between two explanations for fast tracks within firms – a symmetric learning explanation for fast tracks developed in Gibbons and Waldman (1999a) and an asymmetric learning/signaling explanation developed in Bernhardt (1995). Although their main result is that fast tracks are mostly explained by the Gibbons and Waldman symmetric learning explanation, they also find a result concerning schooling's effect on the probability of promotion that they interpret as the signaling role of promotion varying negatively with education.

grew out of the seminal work of Spence (1973) (see Riley (2001) for a survey). Much of the work on that topic focuses on the return to education as a signal in terms of the initial wage the worker receives when the worker enters the labor market, or how the higher wage dissipates over the career as firms learn true ability. A major point of our theoretical analysis, however, is that in a world of asymmetric learning, an important part of the return to education as a signal is that it improves promotion prospects possibly long after the start of a worker's career. Further, our empirical analysis shows clear support for the idea that higher education results in improved promotion prospects.

II. MODEL AND THEORETICAL ANALYSIS

Our theoretical analysis is related to that of Bernhardt (1995) in that we consider the promotion-as-signal hypothesis in a setting characterized by multiple schooling levels. Also, the specific production technology we consider is closely related to that investigated by Gibbons and Waldman (2006) in a recent study that employs symmetric learning.

A) The Model

There is free entry into production, where all firms are identical and the only input is labor. A worker's career lasts two periods, where in each period labor supply is fixed at one unit for each worker. We call workers in their first period in the labor market young and those in their second period old. Worker i enters the labor market with a schooling level, S_i, that can take on any integer value between 1 and N. We assume that there is a positive number of workers at each value of S. Note that given much of our focus will be on the information transmitted by a worker's schooling level, a simple interpretation is that a worker's schooling level represents the highest degree earned by the individual.

Let η_{it} denote worker i's "on-the-job human capital" in period t, where

(1) $$\eta_{it}=\theta_i f(x_{it}).$$

In equation (1), θ_i is the worker's ability to learn on the job and x_{it} is the worker's labor-market experience prior to period t, i.e., $x_{it}=0$ for young workers and $x_{it}=1$ for old workers. Also, $f(1)>f(0)>0$.[5] We assume that worker i with schooling level S_i has a starting value of on-the-job human capital equal to

[5] We assume $f(0)>0$ rather than $f(0)=0$. One interpretation is that this model is a simplified version of a continuous-time model where production in the first period represents production in the early part of the worker's career. Since, on average, during this early part of the career the worker has a positive amount of labor-market experience, it is natural to assume $f(0)>0$ rather than $f(0)=0$.

$[\varphi_i+B(S_i)]f(0)$, where $B(S)>B(S-1)$ for $S=2,3,\ldots,N$. φ_i is a random draw from the probability density function $g(\varphi)$, where $g(\varphi)>0$ for all $\varphi_L<\varphi<\varphi_H$ and $g(\varphi)=0$ for all φ outside of this interval. Also, let $\theta^E(S)$ denote the expected value of θ for workers with schooling level S. Note that in this specification schooling is positively correlated with a worker's ability to learn on the job. This can be because schooling enhances human capital and thus increases a worker's ability to learn on the job, or because there is a positive relationship between schooling and innate ability to learn on the job and schooling serves as a signal as in Spence (1973). In Section V we discuss these two different ways of interpreting/extending the model.

A firm consists of two different jobs, denoted 1 and 2. If worker i is assigned to job j in period t, then the worker produces

$$y_{ijt}=(1+k_{it})[d_j+c_j\eta_{it}]+G(S_i), \tag{2}$$

where d_j and c_j are constants known to all labor-market participants, $G'>0$ and $G''<0$, and k_{it} equals k, k>0, if the worker was employed at the firm in the previous period and zero otherwise (this means all young workers in any period t are characterized by $k_{it}=0$). In this specification $G(S_i)$ represents productivity due to general-purpose human capital accumulated prior to a worker entering the labor force, while k represents the importance of firm-specific human capital in this economy.

Let η' denote the amount of on-the-job human capital at which a worker is equally productive at jobs 1 and 2. That is, η' solves $d_1+c_1\eta'=d_2+c_2\eta'$. We assume $c_2>c_1>0$ and $0<d_2<d_1$, i.e., as in Rosen (1982) and Waldman (1984b) output increases more quickly with ability in the high-level job. Thus, given full information about worker abilities, the efficient assignment rule for period t is to assign worker i to job 1 if $\eta_{it}<\eta'$ and to job 2 if $\eta_{it}>\eta'$.

We assume that each worker's schooling level is known to all labor-market participants when the worker enters the labor market. In contrast, each worker's value for θ_i is not known by either the firms or the worker, although the density function g(.) and the function B(S) are common knowledge. Learning about θ_i takes place at the end of the worker's first period in the labor market when the worker's first-period employer privately observes the worker's output level. In addition to this private information, we assume that the job assignment offered to an old worker by the worker's first-period employer is public information. The result of this last assumption is that, as discussed earlier, a promotion at the beginning of a worker's second period in the labor force serves as a signal of the worker's ability.

Workers and firms are risk neutral and have a zero rate of discount, while there is no cost to workers of changing firms or to firms from hiring or firing workers. To make the model consistent with standard wage determination at most firms, we assume wages are determined by spot-market contracting. In addition, since each worker's output is privately observed rather than publicly observed and verifiable, the wage specified in such a contract consists of a wage determined prior to production rather than a wage determined by a piece-rate contract where compensation depends on the realization of output.

The wage setting process and timing of events is similar to that found in Zabojnik and Bernhardt (2001). At the beginning of each period, each firm offers each worker it employed in the previous period a job assignment or fires the worker, where this decision is publicly observed. We assume that for this decision a firm does not retain any worker it anticipates leaving with probability one during the wage determination process. This assumption is consistent with the existence of a small cost of retaining a worker who then chooses to leave. Following Greenwald (1986), Lazear (1986), and Milgrom and Oster (1987), we then assume that the wage determination process is characterized by counteroffers. That is, after the initial stage just described, all firms simultaneously offer each worker in the economy a wage for that period and then each firm makes a wage counteroffer to each old worker it employed in the previous period.[6] Each worker then chooses to work at the firm that offers the highest wage. If there are multiple firms tied at the highest wage, the worker chooses randomly among these firms unless one of these was the worker's employer in the previous period, in which case the worker remains with that firm. This tie-breaking rule is equivalent to assuming an infinitesimally small moving cost. Finally, at the end of each period each firm privately observes the output of each of its workers.

To reduce the number of cases that need to be considered, we restrict the analysis to parameterizations that satisfy the following conditions. First, $\theta^E(N)f(0)<\eta'$. This condition states that it is efficient for all young workers to be assigned to job 1. Second, $[\varphi_H+B(S)]f(1)>\eta'>[\varphi_L+B(S)]f(1)$ for all S. This condition states that for each schooling level it is efficient for old workers with high values for on-the-job human capital to be on job 2 while it is efficient for those with low on-the-job human capital to remain on job 1.

Finally, we limit attention to Perfect Bayesian Equilibria, where we also impose a Trembling-

[6] One exception is that a firm does not make initial wage offers or counteroffers to any worker it fired at the beginning of the period. Also, when a firm makes an offer to an old worker it did not employ in the previous period, we assume the firm observes whether the worker was retained or fired and the job assignment if the worker was retained. The firm does not observe how many old workers the original employer fired, retained, or assigned to each job in the current period.

Hand Perfection assumption on counteroffers (see Selten (1975) for a discussion of Trembling-Hand Perfection). Specifically, we assume that in the second period there is a small probability that the first-period employer mistakenly does not make a counteroffer when the first-period employer has the smallest cost of choosing that action. Restricting attention in this way means that our analysis is characterized by a winner's curse result similar to that found in the related analysis of Milgrom and Oster (1987), i.e., for old workers retained by their first-period employer other firms are only willing to offer the worker a wage equal to the productivity at such a firm of the lowest productivity worker who has the same labor-market signal (meaning schooling level and job assignment).[7]

B) Analysis

We begin with a benchmark analysis which is what happens when output is publicly observable so that all firms learn a worker's ability to learn on the job after the worker's first period in the labor market (but compensation is still spot-market wages where the wage is determined prior to production). Given our parameter restriction $\theta^E(N)f(0)<\eta'$, every young worker is assigned to job 1. Let $w_Y{*}(S)$ denote the wage for young workers with schooling level S. We have that $w_Y{*}(S)>d_1+c_1\theta^E(S)f(0)+G(S)$ for all S, i.e., young workers are paid more than their expected output. This occurs because old workers are paid less than their expected output – see below – and the zero-profit condition associated with competition means that young workers must be paid more than their expected output.

Now consider old workers. There are three conditions that define what happens when workers become old. First, every old worker remains with the same firm that employed the worker when he or she was young. Second, given that all firms learn each worker's ability to learn on the job and can thus infer the worker's current on-the-job human capital, old worker i is assigned to job 1 if $\eta_{it}<\eta'$ and is assigned to job 2 if $\eta_{it}\geq\eta'$.[8] Third, let $w_O{*}(\eta_{it})$ be the wage paid to old worker i as a function of the worker's current on-the-job human capital, where $w_O{*}(\eta_{it})=\max\{d_1+c_1\eta_{it}+G(S_i),d_2+c_2\eta_{it}+G(S_i)\}$.

[7] In our model this is not an immediate implication of restricting the analysis to Perfect Bayesian Equilibria both because of the presence of firm-specific human capital which Milgrom and Oster did not incorporate into their analysis, and because in our model a promoted worker is signaled to have high productivity while in theirs a worker's productivity becomes public knowledge. Note that this equilibrium refinement is similar to the notion of a Proper Equilibrium first discussed in Myerson (1978). We also assume that, if a worker's first-period employer makes an off-the-equilibrium path promotion decision at the beginning of the second period, then other firms infer the worker's ability is of the type that makes this decision least costly to the first-period employer.

[8] To simplify descriptions of behavior, throughout the paper we assume that an old worker is assigned to job 2 by the worker's current employer whenever the firm is indifferent between assignment to jobs 1 and 2.

The logic for these results is as follows. Given that output is publicly observable, there is no asymmetric information in this benchmark case, i.e., at any date all firms (and the worker) are equally informed about a worker's on-the-job human capital. Hence, at every date, given the information available, workers are assigned to jobs in the efficient fashion and switch employers in the efficient fashion. Given there is firm-specific human capital, this last condition means that each old worker remains with the same firm that employed the worker when he or she was young. Finally, each old worker is paid the wage that the market, i.e., other firms, offers the worker which is the worker's expected productivity given that he or she switches employers (note that a worker who switches employers would have zero firm-specific human capital which explains the expression for $w_O*(\eta_{it})$).

The main point of the benchmark analysis is that, if output is publicly observable, then job assignments as well as turnover decisions are efficient. As we show below, in contrast, once a worker's output is privately observed by the worker's employer, then job assignments are no longer efficient. Rather, firms assign too few old workers to the high-level job in order to avoid sending the positive signal about productivity associated with assignment to the high-level job.

Suppose that a worker's output each period is privately observed by the worker's employer. We start with some preliminary results. Equilibrium behavior when the worker is young is similar to what happened in the benchmark case. That is, as in the benchmark case, our parameter restriction $\theta^E(N)f(0) < \eta'$ yields that all young workers are assigned to the low-level job. Also similar to what was true in the benchmark case, the wage paid to young worker i, $w_Y(S_i)$, is above expected output and is such that a firm hiring a young worker earns zero expected profits from the hire. One difference is that in the benchmark case the wages paid to young workers exceeded expected output because of profits earned in the following period due to the presence of firm-specific human capital. In contrast, now the model exhibits this feature because of both future profits due to the presence of firm-specific human capital and future profits due to the presence of asymmetric information about worker productivity.

We now formally state what happens in this case. Below $w_O(S_i,\eta_{it})$ is the wage paid to an old worker as a function of the worker's schooling level and on-the-job human capital, while j_{it} is the firm that individual i works at in period t.[9] All proofs are in the Appendix.

[9] We focus on the unique equilibrium characterized by no workers being fired. This is the only equilibrium if the two jobs are sufficiently similar or k is sufficiently large. Also, in the description of equilibrium behavior, we ignore what happens when in the second period the first-period employer mistakenly fails to make a counteroffer.

<u>Proposition 1</u>: If a worker's output is privately observed by the worker's employer, then there exists a function $\eta^+(S)$, $\eta'<\eta^+(S)\leq[\varphi_H+B(S)]f(1)$ for all S, such that i) through iii) describe equilibrium behavior in each period t.[10]

. i) Each young worker i is assigned to job 1 and is paid $w_Y(S_i)>d_1+c_1(\theta^E(S_i)f(0))+G(S_i)$.

ii) If old worker i is such that $\eta_{it}\geq\eta^+(S_i)$, then the worker is assigned to job 2, remains at firm j_{it-1}, and is paid $w_O(S_i,\eta_{it})=d_2+c_2\eta^+(S_i)+G(S_i)$.

iii) If old worker i is such that $\eta_{it}<\eta^+(S_i)$, then the worker is assigned to job 1, remains at firm j_{it-1}, and is paid $w_O(S_i,\eta_{it})=d_1+c_1[\varphi_L+B(S)]f(1)+G(S_i)$.

Proposition 1 tells us that, if a worker's output is privately observed by the worker's employer, then for each schooling group there is a critical value for η_{it}, $\eta^+(S_i)$, that determines what happens when the worker becomes old. If η_{it} is below the critical value, then the worker is not promoted and stays with the initial employer. Otherwise the worker is promoted but again stays with the initial employer. Further, in each case the wage paid to a worker equals the productivity at another potential employer of the lowest productivity worker with the same labor-market signal. For example, the wage paid to a worker with schooling level 1 who is not promoted equals the productivity at an alternative employer of an old worker with on-the-job human capital equal to $[\varphi_L+B(1)]f(1)$. The logic here is that, given our Trembling-Hand assumption concerning counteroffers, there is a winner's curse problem similar to that found in Milgrom and Oster (1987) in which other potential employers are only willing to pay the lowest possible productivity of a worker with the same labor-market signal (see Greenwald (1986) and Lazear (1986) for related analyses). Hence, in order to retain a worker, this is all an old worker's previous employer needs to offer.

In addition to these results concerning the existence of a critical value for η_{it} for each schooling group and how wages are determined, another interesting aspect of the proposition is that promotion decisions are not efficient. That is, since $\eta^+(S)>\eta'$ for all S, fewer workers are assigned to the high-level job than is efficient given the initial employer's knowledge concerning the worker's on-the-job human capital. The logic here is the same as that initially explored in Waldman (1984a). Because assigning an old worker to the high-level job rather than the low-level job sends a signal that the worker has high

[10] When no old workers of schooling level S are promoted, we set $\eta^+(S)$ equal to $[\varphi_H+B(S)]f(1)$.

productivity, firms give promoted workers large wage increases in order to stop them from being bid away. In turn, because of the need to pay a promoted worker a high wage, a worker's initial employer will assign the worker to the high-level job only if his or her productivity in the high-level job significantly exceeds productivity in the low-level job.

C) Testable Implications

The model's first testable implication concerns how output produced when a worker is young translates into the firm's decision concerning whether or not to promote the worker in the following period. Below, let $y^P(S)$ denote the minimum output level required for a young worker with schooling level S to be promoted when he or she becomes old.

Corollary 1: Suppose there is a strictly positive number of promotions for workers of schooling levels S_1 and S_2, $S_2>S_1$.[11] Then $\eta^+(S_2)<\eta^+(S_1)$ and, if k is sufficiently small, $y^P(S_2)<y^P(S_1)$.

Corollary 1 captures our first testable implication which is that, if k is sufficiently small, then the performance level required to achieve promotion falls with the education level. This is closely related to the idea that the incentive to distort the promotion decision is decreasing in the schooling level, i.e., $\eta^+(S_2)<\eta^+(S_1)$ for $S_2>S_1$. There are two steps to the argument. First, as discussed earlier, a firm promotes fewer workers than is efficient because of the high wage that needs to be paid to promoted workers. But since the wage paid to workers who are not promoted, i.e., $[d_1+c_1(\varphi_L+B(S))]f(1)+G(S)$, is increasing in the schooling level, the incentive to distort the promotion decision and avoid paying the higher wage associated with promotion is smaller for workers with higher education. Hence, a firm will distort the promotion decision less for workers with more schooling or, in other words, the critical value for $\eta^+(S)$ will be closer to η' for higher values for S.

The second step is to translate this first result concerning how $\eta^+(S)$ varies with S into a statement concerning how $y^P(S)$ varies with S. By definition, $y^P(S)=c_1+d_1\eta^+(S)+G(S)$. Thus, there are two countervailing effects as S rises from S_1 to S_2. First, G(S) rises. Second, as just discussed, $\eta^+(S)$

---

[11] It is possible that the incentive for firms to distort the promotion decision results in no promotions for workers of specific schooling levels.

falls. When k is small, the difference between education levels in the incentive to distort the promotion decision is large, i.e, $\eta^+(S_1)$-$\eta^+(S_2)$ is large. The basic logic here is that the return to efficiently assigning a worker is lower the lower is k. As a result, when k falls both $\eta^+(S_1)$ and $\eta^+(S_2)$ move further from η' and, further, $\eta^+(S_1)$-$\eta^+(S_2)$ gets larger. Finally, because when k is small $\eta^+(S_1)$-$\eta^+(S_2)$ is large, we have that with small enough k the second effect concerning how increasing S affects $y^P(S)$ dominates with the result that $y^P(S_2)<y^P(S_1)$.[12]

The second testable implication concerns how the wage increase due to a promotion varies with a worker's education level. Note that what we mean here is the wage increase when a worker is promoted minus the wage increase the same worker would have received if there had been no promotion. Below let $\Delta w^P(S)$ denote the wage increase due to a promotion as a function of the education level.

<u>Corollary 2</u>: Suppose there is a strictly positive number of promotions for workers of schooling levels S_1 and S_2, $S_2>S_1$. Then $\Delta w^P(S_2)<\Delta w^P(S_1)$.

Corollary 2 captures our second testable implication which is that the wage increase due to promotion is decreasing in the schooling level. The basic logic is as follows. As discussed earlier, the wage of a promoted worker is the expected productivity at an alternative employer of the lowest productivity worker with the same schooling level who is promoted. Similarly, the wage of a worker who is not promoted is the expected productivity at an alternative employer of the lowest productivity worker with the same schooling level who is not promoted. Combining these ideas with the idea that no one is fired yields that the wage increase due to promotion equals $[[d_2+c_2\eta^+(S)]-[d_1+c_1[\varphi_L+B(S)]]]f(1)$. Given this, there are two reasons why the wage increase due to promotion is decreasing in the schooling level. First, as discussed earlier, $\eta^+(S)$ falls with the education level because the incentive to distort the promotion decision falls. Second, B(S) is increasing in the education level or, in other words, the expected productivity of the worst overall worker rises with education.

As a final point, we note that although our testable implications are derived from a specific model of the promotion-as-signal hypothesis, the two predictions are in fact robust predictions of this

---

[12] In a less realistic version of the model that omits the general human capital term, G(S), $y^P(S_2)<y^P(S_1)$ holds regardless of the level of k. In this case, however, the model would yield a prediction that is at odds with the standard finding in the empirical literature that, even after controlling for job assignment and experience, a worker's wage is positively related to the worker's education level.

hypothesis, i.e., various alternative models of the promotion-as-signal hypothesis will yield these predictions. The reason is that the basic logic of the signaling argument leads to these two results. This basic logic is that a promotion serves as a positive signal of worker ability, so firms limit the number of promotions in order to avoid the higher wage associated with a promotion which is necessitated by the positive signal. Now add to this basic logic workers who vary in terms of schooling, where higher schooling levels are correlated with higher ability. Since workers with higher schooling levels are already thought of as being of higher ability, the signal associated with promotion and thus the wage increase associated with promotion is smaller for such workers, i.e., our second testable implication. Further, since the wage increase associated with promotion is smaller for workers with more education, the incentive to distort the promotion decision is smaller for such workers. In turn, this will frequently translate into the performance level required to achieve promotion being smaller for more highly educated workers, i.e., our first testable implication.[13]

III. TESTING THE PREDICTIONS

This section first describes the data and then presents our basic testing of the theoretical predictions of the promotion-as-signal hypothesis developed in the previous section.

A) Data

Our data consist of the complete set of annual personnel records during the period 1969 to 1988 for all white-male-managerial employees of a medium-sized US firm in the financial-services industry. The data were originally constructed by George Baker, Michael Gibbs, and Bengt Holmstrom from the raw data contained in the firm's personnel records, and then analyzed in their classic empirical study of internal labor markets found in Baker, Gibbs, and Holmstrom (1994a,b) (see the first of these papers for a detailed description of the data). Their analyses used the full sample of managerial employees,

---

[13] A specific alternative model of interest is the one in which everything is the same as in the model we consider but there is a stochastic term in the production function. We have considered this alternative specification and most of our results are unchanged. First, the basic description of equilibrium captured in Proposition 1 is qualitatively unchanged, where cutoff values for on-the-job human capital are defined in terms of expected values rather than actual values. Second, our second testable implication that the wage increase due to promotion is a decreasing function of worker education is unchanged. Third, our first testable implication that, given k sufficiently small, the performance level required to achieve promotion falls with education does not hold generally, but does hold as long as the variance on the production function's stochastic term is not too large, i.e., as long as a worker's output is sufficiently informative of the worker's ability to learn on the job.

including females and nonwhite males, for a total of 68,437 employee-years of data. The sample of white males that Baker, Gibbs, and Holmstrom shared with us has 50,556 employee-years. The key variables for our analysis are promotions, salaries, education, and supervisor subjective performance ratings measured each year on a five-point scale where 1 denotes the highest performance level and 5 the lowest. As control variables we also employ demographic characteristics, firm tenure, job title, and level in the job hierarchy.

All variables are measured on December 31 for each employee in each year and pertain to that year. We do not observe the exact date of changes in job title or pay, so if an individual is promoted in, for example, 1979, we do not know whether the promotion occurred early in the year or late. Thus, the meaning of the worker's 1979 performance rating is unclear. If the worker received the promotion early in the year, then the rating likely reflects performance in the post-promotion job. However, if the promotion occurred late in the year, then the performance rating likely pertains to the pre-promotion job. To avoid this ambiguity, we define pre-promotion performance as performance in the year prior to the promotion and post-promotion performance as performance in the year after the promotion.

Most of the variables are observed for each employee for each of the sample years in which the individual worked as a managerial employee at the firm.[14] One exception is that job titles were not recorded for some new hires in the last years of the data set, though other variables were. This means that we lose some observations in our tests since we include job title dummies as controls. Another source of missing observations results from our use of subjective performance ratings, since some workers were not rated in some years. Yet another source of missing observations concerns tenure with the firm, since we do not observe the year in which workers observed in 1969 entered the firm. Since all of our tests control for tenure with the firm, all workers observed in 1969 are dropped.

To define a promotion, we begin with the job ladder constructed by Baker, Gibbs, and Holmstrom from information on job titles. There are eight job levels in the firm, where level 8 is the highest level job held by the CEO. We define a promotion as a transition to a higher level job, so, for example, a worker who moves from level 4 to level 5 in a given year is counted as receiving a promotion,

[14] Since we only observe managerial employees, we do not know if a new entrant to the sample in a given year is a new entrant to the firm or has instead been promoted from a clerical to a managerial position. As suggested in Baker, Gibbs, and Holmstrom (1994a), however, such promoted workers would likely be treated the same way as new hires to the firm, since the promotion entails a large change in job tasks and thus the retirement of most of the task-specific human capital acquired in the pre-promotion job. See the end of this section for a related discussion.

as is a worker who moves from level 4 to levels 6 or 7. In other words, we do not distinguish between one-step and multiple-step promotions, and we do not think this is a concern since roughly ninety-eight percent of the promotions in the sample are one-step promotions. Furthermore, non-promoted workers include a very small number of workers who were demoted (less than one percent of the sample).

Salary is measured as the real annual salary in 1988 dollars, deflated by the CPI. The salary data do not include bonuses, since bonus information is only available for 1981 to 1988. Baker, Gibbs, and Holmstrom (1994a) also ignore the bonus data and argue that bonuses change total compensation very little for most employees in the firm. Using the full sample (including females and nonwhite males), they found that only twenty-five percent of employees received bonuses in the 1981 through 1988 time period and that these workers were heavily concentrated in the highest levels of the job ladder. Since our sample restrictions eliminate workers in levels 4 and higher, ignoring bonus data should have little effect on our results. Also, for the few workers at lower levels who receive bonuses, these bonuses account for a relatively modest fraction of total compensation (the median bonus for workers who receive bonuses in levels 1, 2, and 3 is less than ten percent of salary, while this value is less than fifteen percent for workers in level 4). Finally, a small number of observations in the data set concern employees operating in branches outside of the United States. Since the nominal salary data were recorded in local currencies, we follow Baker, Gibbs, and Holmstrom and drop these observations from our tests that use salary data.

Education is recorded in years in the data set, though in our empirical work we aggregate this variable into a set of dummy variables designed to capture different degrees. Specifically, we construct dummy variables for high school graduate (including some college), bachelors degree, MBA or other masters degree, and Ph.D. degree. As discussed in more detail in Section V, we believe the most plausible interpretation of our model is that a higher level of schooling serves as a signal to other potential employers that the worker belongs to a higher productivity group, and thus it is the receipt of a degree that is important rather than the number of years of education. In other words, focusing on education as a signal, taking five years to complete a bachelors degree does not signal higher quality than taking four. We thus exploit only the variation in educational attainment that occurs at discrete cutoffs defined by years of typical degree completion.[15]

[15] We have also run our tests using years of education rather than degree dummy variables to test our theory's predictions. Consistent with the above discussion, in general this approach is less consistent with our theoretical predictions than the approach of employing degree dummy variables.

To be precise, a high school graduate is defined as a worker with twelve, thirteen, or fourteen years of education. A bachelors degree holder is defined as a worker with sixteen years of education. An MBA or other masters degree holder is defined as a worker with eighteen years of education. Finally, a Ph.D. is defined as a worker with twenty-one or more years of education (although there are no workers in the data set with exactly twenty-one years of education). We exclude from the sample workers for whom years of education was fifteen, seventeen, nineteen, and twenty, since these workers do not fall clearly into one of our four degree categories. This exclusion sacrifices roughly four percent of our sample, where about ninety-seven percent of these excluded workers have seventeen years of education. A worker with seventeen years of education could be someone who took five years to complete a bachelors degree or someone who took three years and completed a two-year masters degree. Given the importance in our analysis of distinguishing between workers with bachelors degrees from those with masters degrees, we exclude these workers from our sample.

In the theoretical model of the previous section, workers in any of the N education groups could be employed in either of two jobs. In the data, however, some jobs are never held or almost never held by workers with specific education levels. There are seventeen job titles in the data set, labeled A through Q.[16] As seen in Table 1, no Ph.D.s are present in job titles A, B, H, I, and J, and no high school graduates are present in titles J and M. Further, several of the other job titles have a negligible fraction of workers from one or more education groups. Since our theoretical analysis is based on career paths that are regularly traversed by workers from all education groups, in our main specifications we restrict attention to workers in titles for which the fraction of occupancy is greater than one percent for each of the four education groups. This selection criterion means that our main empirical analysis includes job titles C, D, G, K, and Q which are the job titles highlighted in italics and boldface in Table 1. We also consider alternative samples as sensitivity checks on our results. First, we estimate the model using a less stringent rule for including job titles, allowing any job title for which the occupancy rate is positive for each educational group, even if only because of a single worker. This rule includes job titles C, D, E, F, G, K, L, N, O, P, and Q. Second, we consider the full sample (see footnotes 18 and 22).

In summary, we omit observations of workers in foreign plants, of workers who were already at the firm in 1969, for which the subjective performance rating is missing, for which job title data are

[16] As in the original papers by Baker, Gibbs, and Holmstrom, the actual job titles are disguised to protect the anonymity of the firm.

missing, for which years of schooling equals fifteen, seventeen, nineteen, or twenty, and in our main specifications for which the job title is other than C, D, G, K, and Q. Table 2 displays descriptive statistics for the main variables in our analysis using the sample selection rules just described. Note that our sample restrictions eliminate workers in levels 4, 5, 6, 7, and 8.

B) The Basic Tests

We begin our empirical analysis with the testable prediction captured in Corollary 1, i.e., given k sufficiently small, the threshold level of output required to achieve promotion, $y^P(S)$, is a decreasing function of the education level. Because the BGH firm exhibits significant turnover at all job levels rather than being characterized by ports of entry, we believe the evidence supports the firm having a low level of firm-specific human capital. Hence, for the BGH firm, we interpret Corollary 1 as stating that the threshold level of output required to achieve promotion should be a decreasing function of the education level.[17]

Since the threshold level of output required to achieve promotion is unobserved by the econometrician, we develop an empirical specification that allows us to test our first prediction using the observed data on promotions, schooling levels, and performance ratings. Letting the subscript i index workers and t index years, in what follows $PROMOTION_{it}$ is a dummy variable that equals one if worker i is promoted in year t and zero otherwise, P_{it-1} is the performance rating of worker i in year t-1, while HS_{it-1}, MA_{it-1}, and PHD_{it-1} are dummy variables each of whose value equals one if at date t-1 worker i's number of years of schooling is consistent with that schooling level being the highest level of educational attainment (as defined earlier) and zero otherwise.

Similar to the theoretical model, let y_{it-1} and y^P_{it-1} denote, respectively, worker i's output in t-1 and the minimum output in t-1 required for worker i to be promoted in t, where by definition this means $PROMOTION_{it}=1$ if $y_{it-1}-y^P_{it-1}\geq 0$ and $PROMOTION_{it}=0$ if $y_{it-1}-y^P_{it-1}<0$. Since both of these variables are unobserved by the econometrician, we specify them as latent index variables.

$$y_{it-1}=h(P_{it-1})+e_{it-1} \tag{3}$$

$$y^P_{it-1}=\psi_0+\psi_1 HS_{it-1}+\psi_2 MA_{it-1}+\psi_3 PHD_{it-1}+\mathbf{X_{it-1}\tau}+v_{it-1} \tag{4}$$

---

17 In a private correspondence, Michael Gibbs has indicated to us that, based on his knowledge of the actual identity of the firm, he also believes that the BGH firm is characterized by little firm-specific human capital.

In this specification $h(\cdot)$ is a monotonically decreasing function, while e and v are stochastic disturbances. Note that $h(\cdot)$ is a decreasing function because of the way the performance rating is defined, i.e., 1 is the highest rating and 5 the lowest. The vector of controls, $\mathbf{X_{it-1}}$, includes age, age squared, tenure at the firm, tenure at the job level, and dummies for job level, job titles, and years. Note that Corollary 1 implies $\psi_3<\psi_2<0<\psi_1$ since the minimum output level required to achieve promotion is a decreasing function of schooling, where the excluded schooling group is the college educated.

Substituting (3) and (4) into the expressions for $PROMOTION_{it}$ given above yields (5a) and (5b).

(5a) $$PROMOTION_{it}=1 \text{ if } h(P_{it-1})-\psi_0-\psi_1 HS_{it-1}-\psi_2 MA_{it-1}-\psi_3 PHD_{it-1}-\mathbf{X_{it-1}}\boldsymbol{\tau}\geq v_{it-1}-e_{it-1}$$

(5b) $$PROMOTION_{it}=0 \text{ if } h(P_{it-1})-\psi_0-\psi_1 HS_{it-1}-\psi_2 MA_{it-1}-\psi_3 PHD_{it-1}-\mathbf{X_{it-1}}\boldsymbol{\tau}<v_{it-1}-e_{it-1}$$

Assuming $h(\cdot)$ is a linear function of P_{it-1}, (5a) and (5b) can be rewritten as (6a) and (6b), where $\mu_{it-1}=v_{it-1}-e_{it-1}$, $\beta_0=-\psi_0$, $\beta_1=-\psi_1$, $\beta_2=-\psi_2$, $\beta_3=-\psi_3$, $\boldsymbol{\delta}=-\boldsymbol{\tau}$, and $\xi<0$.

(6a) $$PROMOTION_{it}=1 \text{ if } \mu_{it-1}\leq\beta_0+\beta_1 HS_{it-1}+\beta_2 MA_{it-1}+\beta_3 PHD_{it-1}+\mathbf{X_{it-1}}\boldsymbol{\delta}+\xi P_{it-1}$$

(6b) $$PROMOTION_{it}=0 \text{ if } \mu_{it-1}>\beta_0+\beta_1 HS_{it-1}+\beta_2 MA_{it-1}+\beta_3 PHD_{it-1}+\mathbf{X_{it-1}}\boldsymbol{\delta}+\xi P_{it-1}$$

Assuming that μ_{it-1} has the standard normal distribution, the promotion rule is described by the following probit model.

(7) $$Prob(PROMOTION_{it}=1)=\boldsymbol{\Phi}(\beta_0+\beta_1 HS_{it-1}+\beta_2 MA_{it-1}+\beta_3 PHD_{it-1}+\mathbf{X_{it-1}}\boldsymbol{\delta}+\xi P_{it-1})$$

Since $\beta_j=-\psi_j$ for j=1,2, and 3, the prediction $\psi_3<\psi_2<0<\psi_1$ translates into $\beta_1<0<\beta_2<\beta_3$. In other words, controlling for worker performance and the additional control variables in **X**, the probability that a worker is promoted in any year t should be an increasing function of the worker's education level.

We report results from our probit estimation of equation (7) for the basic sample in the first column of Table 3, where for ease of interpretation we report marginal effects rather than the probit coefficients themselves. As can be seen the results exactly match the theoretical prediction and thus provide clear support for our first testable prediction. Other things equal, and in particular holding constant current performance, the probability of promotion is 5.9 percentage points lower for high school graduates than for bachelors degree holders, 5.2 percentage points higher for masters degree holders than for bachelors degree holders, and 15.7 percentage points higher for Ph.D.s than for bachelors degree holders. As discussed earlier, we also estimate the model using a less stringent rule for including job titles, allowing any job title for which the occupancy rate is positive for each educational group, even if only because of a single worker. As shown in the second column of Table 3, the predicted results also hold given this less stringent rule for including job titles, with all three marginal effects strongly statistically significant.[18]

A potential concern is that the performance rating in year t-1 might not fully capture worker performance in the pre-promotion job. Specifically, relevant pre-promotion performance might span multiple years rather than being the performance rating for the most recent year. To investigate this issue, we estimated a specification that includes performance in years t-1 and t-2. These results are reported for the basic sample in the third column of Table 3. As can be seen, the marginal effects found in this alternative specification are qualitatively identical and similar in magnitude to those found in the first column.

We now turn to the second testable implication derived in the previous section which is that the wage increase due to promotion is decreasing in the education level. Let ln w_{it} denote the natural log of worker i's real annual salary as measured on the last day of year t. We consider the following regression specification.[19]

$$\text{(8a)}\quad \ln w_{it} - \ln w_{it-1} = \gamma_0+\gamma_1 HS_{it-1}+\gamma_2 MA_{it-1}+\gamma_3 PHD_{it-1}+\mathbf{Y_{it-1}}\boldsymbol{\lambda}+\omega P_{it-1}+\varepsilon_{it} \qquad \text{if } PROMOTION_{it}=1$$

$$\text{(8b)}\quad = \alpha_0+\alpha_1 HS_{it-1}+\alpha_2 MA_{it-1}+\alpha_3 PHD_{it-1}+\mathbf{Z_{it-1}}\boldsymbol{\rho}+\rho P_{it-1}+\upsilon_{it} \qquad \text{if } PROMOTION_{it}=0$$

The prediction that the wage increase due to promotion is decreasing in the education level refers to the wage increase relative to what the worker would have received in the absence of a promotion. In terms of our regressions, this wage premium due to promotions is given by (8a) minus (8b). The theoretical prediction to be tested is therefore γ_3-α_3<γ_2-α_2<0<γ_1-α_1. In regression (8a), $\mathbf{Y_{it-1}}$ is the same vector of controls included in our promotion probability test given in equation (7) except for the substitution of job title transition dummy variables for the job title dummies. These job title transition dummies are of the form d_{jk} indicating a transition from job title j in year t-1 to job title k when the worker is promoted in

[18] In both tests the difference between the marginal effects for masters degree holders and Ph.D.s is statistically significant at the five percent level. Also, our basic test yields a probability of promotion of 0.206 for bachelors degree holders and 0.260 for masters degree holders. This is the source for the "about twenty percent" statement in the Introduction for the decrease in promotion probability when the education level is decreased from masters degree to bachelors degree. Note that we have also conducted this test on the full sample and there was no change in the qualitative nature of the results.

[19] One discrepancy between the theory and the empirical testing is that in the theory the promotion process is deterministic. In other words, in contrast to our empirical testing, if a worker is promoted in equilibrium, there is no similar worker with whom to compare the promoted worker with who is not promoted in equilibrium. Our empirical methodology relies on such comparisons. However, although we do not formally show it here, it is possible to introduce "slot constraints" into our theoretical framework with the result that little is changed except that equilibrium behavior would allow for such comparisons.

year t. This is important because the job title to which a worker is promoted is likely to affect the wage change from promotion. In regression (8b), $\mathbf{Z_{it-1}}$ differs from $\mathbf{Y_{it-1}}$ in two respects. First, it includes job title dummy variables instead of job title transition dummies. Second, it includes individual-specific fixed effects which we necessarily omit from equation (8a) since educational attainment is in most cases time invariant during a worker's tenure with the firm, and we are interested in the relationship between education and wage growth.[20]

The first two columns of Table 4 display the estimation results for equations (8a) and (8b) for the basic sample. As seen in the lower panel of the table, the results support the theoretical prediction. That is, we find γ_3-α_3<γ_2-α_2<0<γ_1-α_1, though our estimate of γ_1-α_1 is statistically insignificant. The point estimates suggest that, other things equal, the wage premium from promotion is 2.3 percentage points lower for masters degree holders than bachelors degree holders, 4.1 percentage points lower for Ph.D.s than for bachelors degree holders, and 0.4 percentage points higher for high school graduates than for bachelors degree holders.[21] The estimate of γ_3-α_3 is significant at the ten percent level, and the estimate of γ_2-α_2 is significant at the one percent level.

To test the robustness of the findings concerning our second theoretical prediction to changes in the sample selection criteria, we also estimated the wage-growth regression using the less stringent selection rule concerning job titles. The results, which are reported in the third and fourth columns of Table 4, are qualitatively similar to those for our main test. The estimated values for γ_1-$\alpha_{1,}$, γ_2-α_2, and γ_3-α_3, respectively, are 0.493, -1.723, and -3.363. As was true for our main test, the estimate of γ_1-α_1 is statistically insignificant, whereas the estimate of γ_2-α_2 is significant at the one percent level, and the estimate of γ_3-α_3 is significant at the ten percent level. We also estimated the model on our basic

[20] Another way to understand this test is as follows. One can construct for each observation of a promoted worker a wage increase due to the promotion using the following three-step procedure. First, estimate regression (8b) for the subsample of observations of workers who were not promoted in year t. Second, for each observation of a worker who was promoted, use the parameter estimates from the first step to derive a predicted wage increase that the worker would have received in the absence of the promotion. Third, subtract this predicted wage increase from the worker's actual wage increase. After this construction, one can use this "wage increase from promotion" variable as the dependent variable in a regression that has the same independent variables as in (8a). The coefficients on the education dummy variables in this regression would be identical to the estimates of γ_1-α_1, γ_2-α_2, and γ_3-α_3 found in Table 4. Note that including individual-specific fixed effects in (8b) is sensible because this allows us to more accurately estimate the "predicted wage increases."

[21] In this analysis the difference between masters degree holders and Ph.D.s, although it has the correct sign, is not statistically significant (Z=0.782). Also, this test yields an average percentage wage increase due to promotion of 0.053 for bachelors degree holders and 0.031 for masters degree holders. This is the source for the "over seventy percent" statement in the Introduction for the increase in the average percentage wage increase due to promotion when the education level is decreased from masters degree to bachelors degree.

subsample controlling for performance in years t-1 and t-2. As reported in the fifth and sixth columns of Table 4, these results are also qualitatively similar to those found in the first two columns. One difference worth mentioning is that the estimate of γ_1-α_1, which was positive and insignificant in our main analysis, becomes negative and insignificant in the specification that includes performance in t-1 and t-2.[22]

One possible explanation for the weak empirical support among high school graduates for our second theoretical prediction is related to the fact that our data set only contains the white collar part of the labor force (see footnote 14). Many of the high school graduates that we observe in our data set likely started their careers at the firm in a blue collar job and were promoted into a white collar job (we think this is likely much less frequent for the other education groups). Although we control as much as is feasible for the nature of the job by including controls for job level and job title transition, it is possible that workers being promoted into managerial positions out of the blue collar part of the firm are on systematically different career tracks for which promotions are associated with smaller wage increases.

IV. ALTERNATIVE EXPLANATIONS

In this section we investigate a number of potential alternative explanations for the results found in the previous section. We consider five alternatives: i) education provides higher level skills; ii) symmetric learning; iii) coarse information; iv) biased performance ratings; and v) biased promotion contests. We find none that matches the evidence as well as the promotion-as-signal hypothesis, although we do find some evidence consistent with one or more of the alternatives.

Alternative 1: Education Provides Higher Level Skills

One potential explanation for why education matters in promotion decisions even after controlling for performance is that higher levels of education are associated with skills that are more useful at higher levels of the firm's job ladder. For example, consider two workers who are equally productive at job level 1, where one worker has a bachelors degree and the other an MBA. Suppose further that having an MBA relative to a bachelors degree provides a worker with skills that are only

[22] We have also conducted our basic test using the full sample. There were two qualitative differences between this test and the test on the basic sample reported in columns 1 and 2. First, as in the fifth and sixth columns of Table 4, in this test γ_1-α_1 is negative and insignificant. Second, in this test γ_3-α_3, although it does have the correct sign, is greater than rather than less than γ_2-α_2, and further it is not statistically significant at standard significance levels.

useful on jobs at level 2 or higher. Then, even though performance on level 1 is the same, the firm will have a greater incentive to promote the MBA because his or her expected performance on level 2 is higher.

The first problem with this alternative explanation concerns our second main empirical finding. That is, although the idea that higher education levels provide higher level skills could potentially explain our promotion probability findings, it is not consistent with our finding that, except for high school graduates, the wage increase upon promotion is a decreasing function of the worker's education level. If the role of education in the promotion process is that education provides higher level skills, wage increases upon promotion would be positively, not negatively, related to the education level. The reason is that, because in that explanation more highly educated workers have skills that are only valued or more highly valued at high-level jobs, such workers should get particularly large wage increases upon promotion since promotion for these workers is associated with particularly large increases in productivity.

Furthermore, one way to address this alternative explanation is to include predicted post-promotion performance as an explanatory variable in our promotion-probability tests. That is, suppose we estimate how workers with various education levels are expected to perform after promotion and include these predictions as an explanatory variable in our probit analysis. Then, if the alternative explanation is correct, education level should become unimportant or at least less important in the promotion decision since it matters only because it translates into higher expected post-promotion performance.

To pursue this idea, we first constructed a measure of predicted performance in the post-promotion job as a function of the education level and controls. To do this we considered the subsample of promotions in each year t and estimated an ordered probit in which performance in year t+1 was the dependent variable and the independent variables were the same right hand side variables as in the probit in equation (7). Then for each observation (including both promotions and non-promotions) we used the resulting estimates to compute predicted probabilities for each of the five possible performance outcomes in year t+1. Denoting these predicted probabilities as p_1, p_2, …, p_5, our estimate of predicted post-promotion performance conditional on the information in year t-1 is $\sum_{k=1}^{5} kp_k$. We then included this measure of predicted performance as an additional control in the probit equation for promotion in year t and the results were very similar to our main results in Table 3. That is, as reported in the last column of Table 3, the marginal effect for the HS dummy variable was -0.063 (Z=4.77), for the MA dummy variable 0.049 (Z=3.51), and for the Ph.D. dummy variable 0.160 (Z=3.82). This leads us to again reject the alternative explanation for our promotion probability results that higher education gives workers skills more useful at higher level jobs.[23]

Alternative 2: Symmetric Learning

We now discuss the possibility of symmetric learning. As will be discussed in more detail in the next section, symmetric learning refers to a situation in which, rather than a worker's current employer learning more about the worker's ability than do other potential employers as in the model analyzed in Section II, learning comes from publicly available information so at any date all firms have the same information and beliefs about each worker's ability. As an example, the benchmark analysis of Section II is a symmetric learning model since in each period each worker's output realization is publicly observed rather than privately observed by the worker's current employer.

Symmetric learning can potentially explain our promotion probability findings and, in contrast to the first alternative explanation discussed above, it can also potentially explain our wage growth findings. To see the former, suppose output is stochastic and is publicly rather than privately observed. Then a worker's final performance rating will not contain all the relevant information concerning what the worker's true ability actually is, but rather both previous performance ratings and the worker's education level (to the extent a more highly educated worker is drawn from a higher ability group on average) will also contain relevant information. Hence, one possible explanation for why our probit analysis of promotion probabilities produces a positive relationship between education and probability of promotion is that, even after controlling for current performance, higher education translates into higher values for the expected underlying ability level.

The idea that symmetric learning can explain our wage growth findings follows from an argument related to one found in Gibbons and Waldman (1999a). That paper showed that symmetric

[23] Note that we also looked to see whether education level predicts post-promotion performance after including all of our standard controls including pre-promotion performance. We investigated this using both OLS and ordered probit. The answer is that, except for high school graduates, education is positively related to post-promotion performance. But as discussed above, controlling for this generally positive relationship in our promotion probit analysis does not reduce the positive relationship between education and promotion probability.

learning can explain large wage increases upon promotion because, on average, we would expect promoted workers to be those for whom there were large improvements in beliefs concerning the worker's underlying abilities. Now consider a symmetric learning world where workers vary in terms of their education levels and higher education translates on average into higher underlying ability. It would be natural in such a setting for promoted workers from lower education groups to experience larger improvements, on average, in beliefs concerning the workers' underlying abilities. In turn, these larger improvements in beliefs should translate into larger promotion wage increases for these lower education groups.

One test of this alternative explanation for our results concerns incorporating multiple performance measures into our promotion probability analysis. That is, if the reason that education level and promotion probability are positively related in our probit analysis is that a worker's education level provides valuable incremental information about a worker's true ability, then the magnitudes of the coefficients on the education variables should fall in absolute value as more performance measures are added to the probit analysis (see Altonji and Pierret (2001) for a related discussion and analysis). Note, this is true as long as the performance measures themselves are only moderately positively correlated, so that adding more performance measures adds important incremental information. In Table 5 we provide a bivariate correlation matrix for performance in period t and three lagged values for performance. Not surprisingly there is positive serial correlation in performance ratings, but the correlations are sufficiently below one that adding performance measures to our probit analysis must clearly add important incremental information.

Table 6 reports results from probit analyses that include the performance ratings in period t-1, the performance ratings in periods t-1 and t-2, the performance ratings in periods t-1, t-2, and t-3, and the performance ratings in periods t-1, t-2, t-3, and t-4 (note that the first two of these probit analyses also appear in Table 3). As reported in Table 6, the positive relationship between education and promotion probability is still strongly statistically significant even when four years of performance ratings are included in the probit analysis. More importantly, the results in Table 6 are not consisistent with the absolute magnitudes of the education coefficients falling as more performance measures are added. For example, although the absolute value of the coefficient on the high school variable is lower in column 4 than in column 1, the absolute value of the MA coefficient is basically unchanged and the Ph.D.

coefficient is actually higher in column 4 than in column 1. These findings do not support symmetric learning serving as the correct explanation for our promotion probability findings.[24]

A further test along this same line is to include a predicted post-promotion performance variable into a probit analysis that includes multiple performance ratings. The symmetric learning argument would suggest that including predicted post-promotion performance should eliminate or at least significantly reduce the effect of education on promotion probability. In the last column of Table 6 we add the predicted post-promotion performance variable from above into the probit analysis that employs four performance measures.[25] The coefficients on the education variables all continue to have the predicted sign and two of the three education coefficients continue to be strongly statistically significant. As before, the results do not support symmetric learning being the correct explanation for our findings.

Alternative 3: Coarse Information

Another potential explanation for our findings concerning the effect of education level on promotion probability is that performance ratings are coarse measures of true performance. Remember that the performance rating is an integer value between 1 and 5. Saying that the performance rating is a coarse measure of true performance simply means that five categories do not capture fine gradations of performance. To see why coarse measurement can potentially explain our promotion probability results, consider two workers who both receive a one – the highest rating – but one of the workers actually had a higher true performance so this worker's probability of promotion is in fact higher. If, as seems quite plausible, considering all such worker pairs the workers with higher true performance on average have higher education levels, then in our probit analysis education may serve as a proxy for the unmeasured higher true performance. In turn, given higher true performance positively affects the probability of promotion, this is an explanation for the positive relationship in our probit analysis between education and probability of promotion even though we control for the performance rating.

---

[24] We have also conducted this test restricting the sample for each regression to the 1796 data points for which performance in years t-1 through t-4 are all available. There was no change in the qualitative nature of the results.

[25] We attempted to construct a measure of post-promotion performance that used the ordered probit technique described above, but which employed performance measures in periods t-1, t-2, t-3, and t-4 rather than just t-1 as above. However, in this case the probit estimation failed to converge.

One problem with this argument concerns our wage growth findings. If the only reason that education is positively related to promotion probability is that higher education is proxying for unmeasured higher true performance, then we would expect that education would be positively related to the “gross” wage increase associated with promotion (by gross wage increase we mean the wage increase that does not net out what the worker would have received in the absence of promotion). But Table 4 indicates that the gross wage increase due to promotion is, except for the high school group, decreasing not increasing with the education level. Note, the prediction for the net wage increase due to promotion, which is what we focused on in Section III, is unclear. The reason is that, according to the coarse information argument, higher education (which means on average higher true performance) should increase both the wage increase upon promotion and the wage increase in the absence of promotion.

Another approach for investigating the coarse information argument involves including performance measures from multiple periods in the probit analysis. If a single period’s performance rating is a coarse measure of true performance, then we would expect that in aggregate performance ratings from multiple periods would more accurately capture true performance. In turn, this means that if the reason there is a positive relationship between education and promotion probability is that a single period’s performance rating is a coarse measure of true performance, then including measures from multiple periods should significantly reduce the positive relationship between education level and promotion probability. As already discussed, Table 6 reports promotion probability results when multiple performance ratings are included. Given the table shows that the absolute value of the coefficient on the HS variable becomes smaller as more performance measurements are added, there is some evidence consistent with coarse information mattering for the high school group. But given this relationship does not hold for the other education groups, given the coefficients of interest are still statistically significant even when four performance ratings are included, and given our point in the above paragraph that this explanation does not match our wage growth findings, we do not believe that coarse information explains our empirical results.

Alternative 4: Biased Performance Ratings

Gibbons and Waldman (1999a) discuss the idea that performance ratings may not be an unbiased measure of true performance, but rather measure performance relative to the average expected performance of the relevant group. So, for example, a worker in a new job who does not perform well from an absolute standpoint may get an average performance rating if his or her performance is equal to the average performance of all new workers at that job. They go on to argue that this idea serves as a potential explanation for the puzzling findings concerning the relationship between pay and performance ratings found in the well known studies of Medoff and Abraham (1980,1981).[26]

Biased performance ratings of the sort just described also serve as a potential explanation for our promotion probability findings. The logic here is as follows. Suppose that each worker's performance rating is measured relative to workers with the same level of education and that, on average, higher levels of education translate into higher average or expected levels of performance. Now consider two workers with the same performance rating but different levels of education. Given that each worker is evaluated relative to the average true performance for workers with the same education level, the worker with higher education must have higher true performance and thus, as we find, should also have a higher probability of promotion.

The problem with this argument is the same as one of the problems discussed above concerning the possibility of performance ratings being coarse measures of true performance. That is, if the reason that education is positively related to promotion probability even after controlling for the performance rating is that higher education captures unmeasured higher true performance, then we would expect the wage increase upon promotion not netting out the wage increase in the absence of promotion to be increasing with rising education. Since, other than for the high school group, we find a negative relationship rather than the predicted positive relationship, we believe our results are not due to biased performance ratings.

Alternative 5: Biased Promotion Contests

Meyer (1991) investigates a T-period tournament model in which there are two workers and a promotion decision at the end of the T periods. Her basic point is that, if one focuses on the performance in the last period, T, then the firm should bias its decision rule so that the worker who has been more productive in the first T-1 periods is promoted after period T even if in period T the worker's output is somewhat (but not too much) below the output of the other worker. The logic is that in a world where

[26] An alternative explanation for the Medoff and Abraham findings is put forth in Bernhardt (1995). That explanation relies on the same ideas of asymmetric learning and promotions serving as signals investigated here.

output is stochastic, each period's output is informative of worker ability so promotions are more efficient when outputs across many periods are considered rather than being determined solely by a single period's output. Note that this argument is just a variant of the symmetric-learning explanation discussed above. Hence, consistent with that discussion, we believe the results in Table 6 indicate that this argument is unlikely to be the correct explanation for our findings.

Meyer (1992) puts forth an alternative argument for biased promotion tournaments. In that analysis there is a sequence of two contests, and it is optimal for the firm to bias the second promotion contest in favor of the winner of the first. The reason is that a small bias causes a second-order decrease in effort levels in the second contest but a first-order increase in effort levels in the first contest. This could potentially explain our promotion probability findings. The idea here is that for promotions for workers starting at higher levels of the job ladder education may serve as a proxy for having been a "bigger" winner earlier on.

One problem with this explanation is that, given all the controls we include in our analysis, it is not clear how plausible this explanation is. Nevertheless, ignoring this problem with the explanation, this argument does not hold for promotions out of job level 1 since there are no earlier promotion contests for which education could be serving as a proxy. In Table 7 we reproduce our promotion probability test and show that our results hold even when the sample is restricted only to workers in job level 1. We therefore again reject biased promotion contests as an explanation for our results.

V. DISCUSSION

This section discusses two issues. In the first subsection we discuss the literature on learning in labor markets and our view concerning what this literature, including our paper, tells us about the nature of this learning. The second subsection considers the related issue of how our analysis contributes to the extensive literature concerning the role of education as a labor-market signal.

A) Symmetric Versus Asymmetric Learning

From a theoretical perspective, there are two basic approaches for thinking about learning in labor markets. One approach investigated in papers such as Harris and Holmstrom (1982) and Gibbons and Waldman (1999a) is that learning is symmetric. This means that any information revealed about a worker's ability during the worker's career is public knowledge. The other main approach first investigated in Waldman (1984a) and Greenwald (1986) is that learning is asymmetric. This means that information about a worker's ability is only directly revealed to the worker's current employer, while other firms observe the current employer's actions such as promotion decisions and firing decisions in making inferences about the worker's ability.

A substantial empirical literature investigates the nature of learning in labor markets. A number of studies such as Gibbons and Katz (1992), Farber and Gibbons (1996), and Altonji and Pierret (2001) develop testable implications of the symmetric-learning approach, while Gibbons and Katz (1991) and our own study focus on asymmetric learning. The former papers find evidence consistent with symmetric learning, while the latter find evidence consistent with asymmetric learning. Schonberg (2007) criticizes the literature because it has paid insufficient attention to developing testable implications that allow the researcher to distinguish between the two types of learning. She derives such implications in the context of a specific model of labor-market turnover and then presents empirical evidence consistent with university graduates being characterized by asymmetric learning and high school graduates and dropouts being characterized by symmetric learning.[27]

Although we find the Schonberg analysis quite interesting, we disagree with her emphasis on the idea that for any group learning in the labor market is primarily either symmetric or asymmetric. We think that when one views the evidence in its entirety the answer is that even within groups learning is somewhere between the pure symmetric and pure asymmetric cases. For example, as argued in Gibbons and Waldman (1999a), a pure asymmetric learning model cannot easily explain the empirical findings of Baker, Gibbs, and Holmstrom (1994a,b). In particular, Baker, Gibbs, and Holmstrom find that on average workers receive wage increases even in periods in which they are not promoted and these wage increases vary across workers. This is difficult to reconcile with a pure asymmetric learning model, but is easily captured in a model characterized by some symmetric learning. On the other hand, as we argued

[27] To be more precise, Schonberg develops two testable implications for asymmetric learning. The first which she finds clear evidence for among university graduates is the result originally due to Greenwald that workers who turnover should be drawn from the low end of the ability distribution. The second is that in a world of asymmetric learning the return to ability should be positively related to tenure at the firm. For this prediction she finds evidence among university graduates that is quite mixed. However, although the second testable implication is predicted by the specific model of asymmetric learning that she considers, we do not believe it is a robust prediction of asymmetric learning models. So our feeling is that Schonberg's findings overall are consistent with asymmetric learning being important for university graduates.

in previous sections, other aspects of the Baker, Gibbs, and Holmstrom data set suggest that asymmetric learning is also important.

Just to make clear what we have in mind, consider, for example, the manager of a large division of a Fortune 500 firm. To the extent that everyone can observe the overall success of the division's products, some of the learning about the manager's ability is of a symmetric or public nature. But the CEO of the firm has access to much more information than the success of the division's products in judging the ability of the division head. The CEO can look at the details of the division's accounting numbers which would typically not be available in a detailed way to individuals outside of the firm. The CEO can also judge better than other potential employers the extent to which the success or failure of the division is due to the division head and how much to other factors.

Note that, although we do not show it formally, we could incorporate an element of symmetric learning into the model analyzed in Section II, and there would be no change in terms of the two testable implications derived. For example, suppose that in addition to a worker's current employer observing the worker's output, there was a publicly observable "announcement" of the worker's ability. Further, suppose this announcement could only take on the values high and low, where the probability the announcement was high was an increasing function of the worker's true ability. Making this change would clearly change the equilibrium outcome, but qualitatively our two testable implications would be unchanged. First, holding performance fixed and given a low value for firm-specific human capital, the probability of promotion would be an increasing function of the education level. Second, the wage increase due to promotion would still be a decreasing function of worker education (and now we would get the additional result that this wage increase would also be a decreasing function of earlier values of this announcement).

As a final point concerning the general issues of symmetric and asymmetric learning, consistent with the discussion in this section we believe that an interesting direction for future research would be to consider analyses that do not focus solely on one type of learning or the other. That is, we feel it would be interesting to look at analyses characterized by a mix of symmetric and asymmetric learning and other intermediate cases between the two polar cases. Moving in this direction would be realistic and quite likely yield insights not captured by current analyses. A specific direction along this line that we feel holds particular promise is to pursue research along the lines suggested by Granovetter (1973,1995) who focuses on how hires frequently occur through personal connections. This suggests that much of the learning in the labor market is neither symmetric nor asymmetric, but rather something intermediate between the two where some information about any specific worker leaks out from the worker's current employer but only in a limited way.[28]

B) Education as a Labor-Market Signal

There is an extensive literature, both theoretical and empirical, that explores the role of education as a labor-market signal (see Riley (2001) for a survey). In this subsection we discuss how our analysis contributes to this literature. We start with the theoretical literature. From a theoretical standpoint, the seminal paper in this literature is, of course, Spence (1973). That paper looks at a world in which education does not add to productivity directly through human-capital accumulation, but rather workers with higher ability have a lower cost of acquiring education. Note that, since education does not directly contribute to productivity in that analysis, the socially optimal level of education in the economy is zero. Spence shows, however, that there are equilibria – in fact, many equilibria – characterized by positive levels of education.

Much of the theory that followed that paper focused on the definition of equilibrium and Spence's finding of multiple equilibria. Spence did not consider his model as a formal game, but rather focused on outcomes in which actions lead to self-fulfilling beliefs. Cho and Kreps (1987), for example, showed that all of the equilibria focused on by Spence are Bayesian-Nash equilibria when the model is analyzed as a formal game, and then derived conditions, i.e., Cho and Kreps' intuitive criterion, that result in a unique equilibrium. Another important contribution is Riley (1979a). That paper shows how the idea of education as a labor-market signal extends to the case in which worker types are continuous.

From our perspective the more interesting theoretical extension is to add realism by considering what happens when a worker's career lasts more than a single period. This is considered in Riley (1979b). Riley shows that, as a worker's career progresses and firms learn about true ability, then the

---

[28] See Montgomery (1991) for a theoretical analysis along this line, and more recently Pinkston (2006) conducts a theoretical and empirical exercise that allows for both symmetric and asymmetric learning. Although Pinkston finds evidence for both types of learning, his evidence more strongly supports the presence of asymmetric learning. Also, DeVaro (2007) estimates a structural model of employer recruitment choice in which hires occur either via personal connections or via formal methods such as advertising. The focus in that analysis is on the role of recruitment strategies as information-generating devices in the labor market and, although the model does not try to distinguish between the cases of symmetric and asymmetric learning, that is one direction in which the analytical framework might be extended.

importance of the initial education signal as a factor determining compensation should decrease. Note that the results in Altonji and Pierret's (2001) analysis of symmetric learning mentioned in the previous subsection are similar. They argue that, as careers progress and firms observe output realizations, education becomes less important and true ability more important as factors determining compensation (see also Lange (2007)).

Although our analysis does not formally treat education as a signal since education levels are given exogenously rather than being a choice variable for the worker, it would be easy to extend our analysis in this way. Such an extension would contribute to the theoretical literature on education as a labor-market signal in two ways.[29] First, in contrast to Riley's analysis and that of Altonji and Pierret, because of asymmetric rather than symmetric learning after workers enter the labor market, the importance of education in determining compensation would not necessarily be a monotonically decreasing function of labor-market experience. That is, because as in our analysis education would increase the probability of promotions during a worker's career, education could be an important factor in compensation even late in careers because it is only the old workers with high ability and high education who earn promotions to the top rungs of firms' job ladders.

Second, there is the related point of why individuals invest in the signal in the first place. In the standard Spence-type model the return to signaling through educational investments is the higher wage the worker receives in the labor market, while analyses that consider multiple-period careers generalize this result to the higher compensation the worker receives early in the career prior to the firms learning true ability. One of our points is that, if there is asymmetric learning after workers enter the labor market, then the return to the signal is not just higher wages early on but also higher promotion probabilities and thus higher wages later in careers. We think the idea that individuals focus on improved long-term promotion prospects in choosing educational investments, even if education is partly or even purely a signal, is a very realistic perspective.

Now let us turn to the empirical evidence. There is an extensive empirical literature on this topic, but much of it is subject to the criticism that the testable implications considered are also consistent with a world in which education simply serves to enhance human-capital accumulation. For

---

[29] Note that the discussion that follows is related to recent analyses that appear in Ishida (2004a,b) concerning the promotion-as-signal hypothesis. See also Habermalz (2006) for a related analysis that does not incorporate the promotion-as-signal hypothesis.

example, Layard and Psacharopoulos (1974), Hungerford and Solon (1987), and Heywood (1994) all investigate whether earning an educational degree enhances compensation as the signaling story would suggest, where the first paper finds no evidence for the hypothesis while the latter two find supporting evidence. But even supporting evidence here is weak evidence for the education-as-signal hypothesis. Consider a pure human-capital-accumulation world in which individuals who drop out before earning a degree do so because they realize the education is not providing valuable human capital for them. Then the attainment of a degree will be correlated with higher compensation even though the degree itself is not serving as a signal.

Papers that are not subject to this criticism include Riley (1979b), Lang and Kropp (1986), and Altonji and Pierret (2001). Riley develops predictions based on the idea that some sectors rely on education as signals and some do not, and finds supporting evidence using the *Current Population Survey*. But Riley's predictions depend on average ability not varying between the sectors and it is unclear how reasonable this assumption is. Lang and Kropp develop predictions concerning how an increase in educational levels at the bottom of the ability distribution affects educational attainment at the top through the operation of incentive compatibility constraints. They find evidence consistent with these predictions using changes in mandatory minimum education levels as the exogenous change affecting educational attainment at the bottom of the distribution. Finally, Altonji and Pierret find evidence consistent with the earlier theoretical discussion that the importance of education as a factor in compensation should decline as workers age.

Although not initially described as a test of the education-as-signal hypothesis, we believe that our empirical work does add to the existing evidence in favor of education having a signaling role. The theoretical extension we describe above would yield as a testable implication that, as we find in our empirical work, given low firm-specific human capital education improves promotion prospects even holding as fixed both the worker's current performance on the low-level job and his or her expected performance on the high-level job. Further, as we also find, the theory would predict that it is more the attainment of a degree rather than simply years of education that increases the probability of promotion (see footnote 15). Hence, as indicated, our empirical results are consistent with a world in which education has a signaling role.[30]

---

[30] One could also construct a model yielding similar predictions that is characterized by uncertainty concerning the innate abilities of workers, but in which education itself serves a purely human-capital-enhancing function. Note,

VI. CONCLUSION

An extensive theoretical literature has argued that one of the roles of promotions is that they serve as signals of worker ability. However, previously there have been no empirical tests of this idea. In this paper we first extended the theoretical literature on this topic by incorporating education into a standard model of asymmetric learning in labor markets and then derived two testable implications of the promotion-as-signal hypothesis. Our tests are based on the intuitively plausible idea that the signal associated with a promotion is more important the lower the worker's education level. After developing these predictions, we then tested for their validity using a data set covering the internal-labor-market history of a single medium-sized firm in the financial-services industry over a twenty-year period. Our empirical investigation strongly supports the two theoretical predictions. First, holding performance fixed, the probability of promotion increases with a worker's education level. Second, except for high school graduates, there is clear evidence that the wage increase due to promotion falls with educational attainment.

One interesting implication of our analysis concerns the correct way to model labor-market tournaments or, in other words, the incentives associated with promotions. The traditional approach, as first explored in the seminal analysis of Lazear and Rosen (1981), is that the firm commits to a prize for promoted workers that is independent of who is actually promoted. In contrast, we find that the wage premium associated with being promoted is not independent of the characteristics of the promoted worker, but rather is a decreasing function of worker education as suggested by the promotion-as-signal hypothesis. This suggests that the correct way to model labor-market tournaments may be by having the signal endogenously determine the size of the prize as explored in Zabojnik and Bernhardt (2001) (see also Gibbs (1995)). In their approach, firms do not commit to a wage structure in advance. Rather, workers provide effort or invest in human capital in order to increase the probability of being promoted

---

however, that for education not to serve as a signal of innate ability given there is uncertainty concerning such abilities, there would need to be no correlation between innate abilities and education levels. Although theoretically possible by, for example, having heterogeneous schooling levels arise solely from differences across individuals in their access to capital markets, we believe such a model would not be a plausible description of the real world. In the real world innate ability, along with other factors such as family wealth and access to capital markets, is clearly an important determinant of educational attainment. In other words, our results are consistent with education having a signaling role, and we believe there is no plausible alternative in which education does not have a signaling role.

and earn the higher wage associated with the resulting signal. If this is the correct approach for modeling promotion tournaments, one implication of our findings is that, restricting the analysis to workers with bachelors degrees or more, promotions may serve as more effective incentive devices for those with less education since those are the workers who receive the largest wage premia upon promotion.[31]

There are a number of directions in which the analysis in this paper could be extended. From a theoretical perspective, we could incorporate effort choice or human-capital investment as just mentioned and see whether this enrichment yields any additional testable implications. There are also a number of empirical extensions. In this paper we have focused on the validity of our theoretical predictions for the internal-labor-market operation of a single medium-sized firm in the financial-services industry. One extension would be to consider the validity of our theoretical predictions for the internal-labor-market operation of other firms for which detailed data are available. In fact, it might be worthwhile adding our empirical tests to the extensive list of standard tests of internal-labor-market operation found in Gibbons (1997). Another extension would be to investigate how our theoretical predictions hold up in a cross section of firms and industries. For example, our predictions concerning the importance of education on the probability of promotion and the wage premium associated with promotion rely on the presence of asymmetric information. Thus, in industries such as academia where asymmetric information is less important because publication records serve as publicly observable measures of performance, our prediction is that these two effects of education should be less important than is the case for our firm in the financial-services industry.[32]

A final extension that would have both theoretical and empirical components would be to enrich our model so that there is turnover and then investigate whether the additional predictions that result are consistent with the empirical evidence. Although we have not done this formally, based on the adverse-selection analyses in Greenwald (1986) and Gibbons and Katz (1991), a prediction from such an

[31] This prediction is potentially testable using the type of data found in van Herpen, Cools, and van Praag (2006).
[32] The specific prediction is that, holding fixed the publication record, the quality of the Ph.D. granting institution should have a relatively small effect on the likelihood of promotion and the wage premium associated with promotion. This should be easily testable using the type of data found in Coupe, Smeets, and Warzynski (2006). Note, although Coupe, Smeets, and Warzynski do not conduct this test, they conclude that there is little asymmetric learning in academia based on how the probability of promotion varies during careers. Another prediction along the same line is that, based on results in Waldman (1984a) and Corollary 1 of Section II, the effects we are focusing on concerning the role of education should be larger the smaller the level of firm-specific human capital. Thus, to the extent the magnitude of firm-specific human capital in an industry can be measured, there are testable implications concerning how our education effects should vary with measured firm-specific human capital at the industry level.

enrichment is likely to be that individuals who separate from the firm will on average be individuals whose productivity is below the average productivity of individuals who look observationally equivalent but do not separate. Although the results are preliminary, we have conducted tests of this prediction and found that in our data set this is indeed the case. That is, we find that after controlling for various observables such as education, job level, tenure at the firm, etc., a worker's probability of turnover is higher the worse the worker's performance rating. Although there are alternative explanations for this result, it is consistent with asymmetric learning and potentially warrants further investigation.[33]

APPENDIX

Proof of Proposition 1: As indicated in footnote 9, our focus is the unique equilibrium in which no workers are fired. We start with what happens when a worker is old. Consider wages. Because the initial employer can make counteroffers and because there is a small probability the initial employer will mistakenly not make a counteroffer when the initial employer has the smallest cost of committing that mistake, other firms are willing to offer a worker assigned to job j the worker's minimum possible output at one of these other firms which is based on who the initial employer assigns to job j in equilibrium.

Now consider job assignments. Since output on job 2 rises faster with on-the-job human capital than output on job 1, for each schooling group S there must be a value $\eta^+(S)$ such that old worker i in schooling group S is assigned to job 1 (job 2) if $\eta_{it}<(\geq)\eta^+(S)$ (see footnote 8). In turn, given the above discussion concerning wages, the wage paid to a worker in schooling group S assigned to job 1 (job 2) is given by $d_1+c_1[\varphi_L+B(S)]f(1)+G(S)$ $(\max\{d_1+c_1\eta^+(S)+G(S),d_2+c_2\eta^+(S)+G(S)\})$.

Now consider $\eta^+(S)$ for a specific value S. Suppose $[\varphi_L+B(S)]f(1)<\eta^+(S)<[\varphi_H+B(S)]f(1)$. Then $\eta^+(S)$ is the value for η_{it} such that a firm is indifferent between assigning an old worker to jobs 1 and 2. In this case $\eta^+(S)$ satisfies (A1).

(A1) $$(1+k)[d_1+c_1\eta^+(S)]-[d_1+c_1(\varphi_L+B(S))f(1)]=(1+k)[d_2+c_2\eta^+(S)]-\max\{d_1+c_1\eta^+(S),d_2+c_2\eta^+(S)\}$$

Suppose $\eta^+(S)=\eta'$. Then (A1) reduces to $d_1+c_1(\varphi_L+B(S))f(1)=d_1+c_1\eta^+(S)$, which contradicts $\eta^+(S)=\eta'$. Suppose $\eta^+(S)<\eta'$. Then (A1) reduces to (A2).

(A2) $$(1+k)[d_2+c_2\eta^+(S)]-(1+k)[d_1+c_1\eta^+(S)]=[d_1+c_1\eta^+(S)]-[d_1+c_1(\varphi_L+B(S))f(1)]$$

[33] This result could alternatively be explained, for example, by Jovanovic's (1979) job-search model in which worker-firm matches are an experience good, i.e., a worker only learns about his match with any specific employer by working at the firm.

But if $\eta^+(S)<\eta'$, then the left-hand side of this expression is strictly negative while the right-hand side is positive so we have a contradiction. Thus, if $[\varphi_L+B(S)]f(1)<\eta^+(S)\leq[\varphi_H+B(S)]f(1)$ for all S, then $\eta^+(S)>\eta'$ for all S and this, in turn, means $\max\{d_1+c_1\eta^+(S),d_2+c_2\eta^+(S)\}=d_2+c_2\eta^+(S)$ for all S.

Now suppose $\eta^+(S)=[\varphi_L+B(S)]f(1)$. Consider the return to promoting a worker whose value for $\eta_{it}=[\varphi_L+B(S)]f(1)+\gamma$, γ small. The extra productivity associated with such a promotion equals $[d_2+c_2[(\varphi_L+B(S))f(1)+\gamma]]-[d_1+c_1[(\varphi_L+B(S))f(1)+\gamma]]$ which is strictly negative for γ close to zero. Starting from a situation in which $\eta^+(S)=[\varphi_L+B(S)]f(1)$, when the off-the-equilibrium path event of a worker not being promoted is observed by the market the inference is that the worker's on-the-job human capital is $[\varphi_L+B(S)]f(1)$ (this follows from our assumption concerning off-the-equilibrium path actions – see footnote 7). The extra cost of promoting such a worker is therefore zero. Thus, since the extra productivity of promoting such a worker is less than the extra cost, the firm will not want to promote the worker so we have a contradiction. Hence, $\eta^+(S)>\eta'$ for all S.

Now consider young workers. Given that from above we know that a firm earns positive expected profits from an old worker it employed when young, competition across firms means that the wage for young workers must exceed expected productivity. We also know that, given our assumption $\theta^E(N)f(0)<\eta'$, all young workers are assigned to job 1. Combining this with young workers being paid more than expected productivity yields $w_Y(S)>d_1+c_1(\theta^E(S)f(0))+G(S)$ for all S.

<u>Proof of Corollary 1</u>: From the proof of Proposition 1, we know given there is a positive number of promotions for workers of schooling level S_1 that (A3) must be satisfied.

(A3) $$(1+k)[d_1+c_1\eta^+(S_1)]-[d_1+c_1(\varphi_L+B(S_1))f(1)]=(1+k)[d_2+c_2\eta^+(S_1)]-[d_2+c_2\eta^+(S_1)]$$

Rearranging yields (A4).

(A4) $$(1+k)[(d_2+c_2\eta^+(S_1))-(d_1+c_1\eta^+(S_1))]-[(d_2+c_2\eta^+(S_1))-(d_1+c_1(\varphi_L+B(S_1))f(1))]=0$$

Holding S fixed and taking the derivative of the left-hand side of (A4) with respect to η^+ yields $(1+k)(c_2-c_1)-c_2$. Given this, suppose $\eta^+(S_1)=\eta'$. Then the left-hand side of (A4) is strictly negative so consistent with Proposition 1 we have $\eta^+(S_1)\neq\eta'$. But we also know from Proposition 1 that $\eta^+(S_1)>\eta'$ so there must be a higher value at which (A4) is satisfied. Given the derivative of the left-hand side of (A4) with respect to η^+ equals $(1+k)(c_2-c_1)-c_2$, we now have that $(1+k)(c_2-c_1)-c_2>0$.

Now consider S_2. Given there is a positive number of promotions for workers of schooling level S_2, (A5) must be satisfied.

(A5) $(1+k)[(d_2+c_2\eta^+(S_2))-(d_1+c_1\eta^+(S_2))]-[(d_2+c_2\eta^+(S_2))-(d_1+c_1(\varphi_L+B(S_2))f(1))]=0$

Suppose $\eta^+(S_2)=\eta^+(S_1)$. Given $B(S_2)>B(S_1)$, a comparison of (A4) and (A5) yields that the left-hand side of (A5) is positive. In turn, given $(1+k)(c_2-c_1)-c_2>0$, we now have that $\eta^+(S_2)<\eta^+(S_1)$.

We now consider the relationship between $y^P(S_2)$ and $y^P(S_1)$. Subtracting (A5) from (A4) and rearranging yields (A6).

(A6) $\eta^+(S_1)-\eta^+(S_2)=c_1[B(S_2)f(1)-B(S_1)f(1)]/[(1+k)(c_2-c_1)-c_2]$

Let k* be the value for k such that $[\eta^+(S_1)-\eta^+(S_2)]f(1)/c_1f(0)=G(S_2)-G(S_1)$. (A6) tells us that for any $k<k^*$ such that $(1+k)(c_2-c_1)-c_2>0$, $[\eta^+(S_1)-\eta^+(S_2)]f(1)/c_1f(0)>G(S_2)-G(S_1)$. By definition $y^P(S)=d_1+[c_1\eta^+(S)f(0)/f(1)]+G(S)$. Given $[\eta^+(S_1)-\eta^+(S_2)]f(1)/c_1f(0)>G(S_2)-G(S_1)$ for all $k<k^*$ such that $(1+k)(c_2-c_1)-c_2>0$, we now have that $y^P(S_2)<y^P(S_1)$ if k is sufficiently small.

Proof of Corollary 2: From Proposition 1, the wage increase due to a promotion as a function of the worker's schooling level is given by (A7).

(A7) $\Delta w^P(S)=[d_2+c_2\eta^+(S)-w_Y(S)]-[d_1+c_1(\varphi_L+B(S)]f(1)-w_Y(S)]$

This can be rewritten as (A8).

(A8) $\Delta w^P(S)=(d_2-d_1)+c_2\eta^+(S)-c_1[\varphi_L+B(S)]f(1)$

We now have that $\Delta w^P(S)$ is decreasing in S for schooling groups with a strictly positive probability of promotion because by assumption B(S) is increasing in S and, from Corollary 1, $\eta^+(S)$ is decreasing in S.

REFERENCES

Acemoglu, D. and J. Pischke, "Why Do Firms Train? Theory and Evidence," *Quarterly Journal of Economics*, 113, 1998, 79-119.

Altonji, J. and C. Pierret, "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116, 2001, 313-350.

Baker, G., M. Gibbs, and B. Holmstrom, "The Wage Policy of a Firm," *Quarterly Journal of Economics*, 109, 1994a, 881-919.

Baker, G., M. Gibbs, and B. Holmstrom, "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics*, 109, 1994b, 921-955.

Belzil, C. and M. Bognanno, "The Promotion Dynamics of American Executives," mimeo, Temple University, 2005.

Bernhardt, D., "Strategic Promotion and Compensation," *Review of Economic Studies*, 62, 1995, 315-339.

Chang, C. and Y. Wang, "Human Capital Investment Under Asymmetric Information: The Pigovian Conjecture Revisited," *Journal of Labor Economics*, 14, 1996, 505-519.

Cho, I. and D. Kreps, "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 1987, 179-221.

Coupe, T., V. Smeets, and F. Warzynski, "Incentives, Sorting and Productivity Along the Career: Evidence from a Sample of Top Economists," *Journal of Law, Economics, and Organization*, 22, 2006, 137-167.

DeVaro, J., "The Labor Market Effects of Employer Recruitment Choices," mimeo, Cornell University, 2007.

Doiron, D., "Layoffs as Signals: The Canadian Evidence," *Canadian Journal of Economics*, 28, 1995, 899-913.

Farber, H. and R. Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics*, 111, 1996, 1007-1047.

Gibbons, R., "Incentives and Careers in Organizations," in D. Kreps and K. Wallis, eds., *Advances in Economic Theory and Econometrics*, Cambridge University Press: New York, 1997, 1-37.

Gibbons, R. and L. Katz, "Layoffs and Lemons," *Journal of Labor Economics*, 9, 1991, 351-380.

Gibbons, R. and L. Katz, "Does Unmeasured Ability Explain Inter-Industry Wage Differentials," *Review of Economic Studies*, 59, 1992, 515-535.

Gibbons, R. and M. Waldman, "A Theory of Wage and Promotion Dynamics Inside Firms," *Quarterly Journal of Economics*, 114, 1999a, 1321-1358.

Gibbons, R. and M. Waldman, "Careers in Organizations: Theory and Evidence," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Volume 3, North Holland, 1999b, 2373-2437.

Gibbons, R. and M. Waldman, "Enriching a Theory of Wage and Promotion Dynamics Inside Firms," *Journal of Labor Economics*, 24, 2006, 59-107.

Gibbs, M., "Incentive Compensation in a Corporate Hierarchy," *Journal of Accounting and Economics*, 19, 1995, 247-277.

Golan, L., "Counteroffers and Efficiency in Labor Markets with Asymmetric Information," *Journal of Labor Economics*, 23, 2005, 373-393.

Granovetter, M., *Getting a Job: A Study of Contacts and Careers*, Harvard University Press: Cambridge, MA, 1973.

Granovetter, M., *Getting a Job: A Study of Contacts and Careers, 2nd ed.*, University of Chicago Press:

Chicago, IL, 1995.

Greenwald, B., “Adverse Selection in the Labour Market,” *Review of Economic Studies*, 53, 1986, 325-347.

Grund, C., “Stigma Effects of Layoffs? Evidence from German Micro-Data,” *Economics Letters*, 64, 1999, 241-247.

Habermalz, S., “More Detail on the Pattern of Returns to Educational Signals,” *Southern Economic Journal*, 73, 2006, 125-135.

Harris, M. and B. Holmstrom, “A Theory of Wage Dynamics,” *Review of Economic Studies*, 49, 1982, 315-333.

Heywood, J., “How Widespread are Sheepskin Returns to Education in the U.S.?,” *Economics of Education Review*, 13, 1994, 227-234.

Hungerford, T. and G. Solon, “Sheepskin Effects in the Returns to Education,” *Review of Economics and Statistics*, 69, 1987, 175-177.

Ishida, J., “Education as Advertisement,” *Economics Bulletin*, 10, 2004a, 1-8.

Ishida, J., “Signaling and Strategically Delayed Promotion,” *Labour Economics*, 11, 2004b, 687-700.

Jovanovic, B., “Job Matching and the Theory of Turnover,” *Journal of Political Economy*, 87, 1979, 972-
990.

Lang, K. and D. Kropp, “Human Capital versus Sorting: The Effects of Compulsory Attendance Laws, *Quarterly Journal of Economics*, 101, 1986, 609-624.

Lange, F., “The Speed of Employer Learning,” *Journal of Labor Economics*, 25, 2007, 1-35.

Layard, R. and G. Psacharopoulos, “The Screening Hypothesis and the Returns to Education,” *Journal of Political Economy*, 82, 1974, 985-999.

Lazear, E., “Raids and Offer Matching,” in R. Ehrenberg, ed., *Research in Labor Economics, Vol. 8*, JAI Press: Greenwich, CT, 1986, 141-165.

Lazear, E. and S. Rosen, “Rank-Order Tournaments as Optimum Labor Contracts,” *Journal of Political Economy*, 89, 1981, 841-864.

MacLeod, W.B. and J. Malcomson, “Reputation and Hierarchy in Dynamic Models of Employment,” *Journal of Political Economy*, 96, 1988, 832-854.

Medoff, J. and K. Abraham, “Experience, Performance, and Earnings,” *Quarterly Journal of Economics*, 95, 1980, 703-736.

Medoff, J. and K. Abraham, “Are Those Paid More Really More Productive?,” *Journal of Human Resources*, 16, 1981, 186-216.

Meyer, M., "Learning from Coarse Information: Biased Contests and Career Profiles," *Review of Economic Studies*, 58, 1991, 15-41.

Meyer, M., "Biased Contests and Moral Hazard: Implications for Career Profiles," *Annales d'Economie et de Statistique*, 25/26, 1992, 165-187.

Milgrom, P. and S. Oster, "Job Discrimination, Market Forces, and the Invisibility Hypothesis," *Quarterly Journal of Economics*, 102, 1987, 453-476.

Montgomery, J., "Social Networks and Labor Market Outcomes: Toward an Economic Analysis," *American Economic Review*, 81, 1991, 1408-1418.

Myerson, R., "Refinements of the Nash Equilibrium Concept," *International Journal of Game Theory*, 7, 1978, 73-80.

Owan, H., "Promotion, Turnover, Earnings, and Firm-Sponsored Training," *Journal of Labor Economics*, 22, 2004, 955-978.

Pinkston, J., "A Model of Asymmetric Employer Learning with Testable Implications," BLS Working Paper 390, 2006.

Ricart i Costa, J., "Managerial Task Assignment and Promotion," *Econometrica*, 56, 1988, 449-466.

Riley, J., "Informational Equilibrium," *Econometrica*, 47, 1979a, 331-359.

Riley, J., "Testing the Educational Screening Hypothesis," *Journal of Political Economy*, 87, 1979b, S227-S252.

Riley, J., "Silver Signals, Twenty-Five Years of Screening and Signaling," *Journal of Economic Literature*, 39, 2001, 432-478.

Rosen, S., "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, 13, 1982, 311-323.

Schonberg, U., "Testing for Asymmetric Employer Learning," *Journal of Labor Economics*, 25, 2007, 651-691.

Selten, R., "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 1975, 25-55.

Spence, M., "Job Market Signaling," *Quarterly Journal of Economics*, 87, 1973, 355-374.

van Herpen, M., K. Cools, and M. van Praag, "Wage Structure and the Incentive Effect of Promotions," *Kyklos*, 59, 2006, 441-459.

Waldman, M., "Job Assignments, Signalling, and Efficiency," *Rand Journal of Economics*, 15, 1984a, 255-267.

Waldman, M., "Worker Allocation, Hierarchies, and the Wage Distribution," *Review of Economic Studies*, 51, 1984b, 95-109.

Waldman, M., "Up-or-Out Contracts: A Signaling Perspective," *Journal of Labor Economics*, 8, 1990, 230-250.

Zabojnik, J. and D. Bernhardt, "Corporate Tournaments, Human Capital Acquisition, and the Firm Size-Wage Relation," *Review of Economic Studies*, 68, 2001, 693-716.

TABLE 1: Distribution of Job Titles for Each Educational Group

	High School	**College**	**M.A.**	**Ph.D.**
Job Title A	9.22	4.48	1.32	0.00
Job Title B	0.35	1.02	0.45	0.00
Job Title C	***16.08***	***14.98***	***11.08***	***6.54***
Job Title D	***6.84***	***10.66***	***10.07***	***8.60***
Job Title E	15.49	6.83	4.26	0.93
Job Title F	0.84	1.03	0.25	1.12
Job Title G	***26.78***	***23.26***	***24.06***	***17.48***
Job Title H	0.18	0.80	0.67	0.00
Job Title I	5.76	2.49	1.43	0.00
Job Title J	0.00	0.11	0.00	0.00
Job Title K	***1.57***	***1.66***	***1.70***	***2.52***
Job Title L	0.22	0.64	0.75	0.47
Job Title M	0.00	0.01	0.07	0.47
Job Title N	0.69	1.71	0.59	3.83
Job Title O	0.60	3.37	3.66	8.32
Job Title P	0.81	1.16	0.47	3.36
Job Title Q	***14.57***	***25.78***	***39.17***	***46.36***
TOTAL	**100%**	**100%**	**100%**	**100%**

TABLE 2: Descriptive Statistics

	Mean	**Std. Deviation**
Promotion	0.131	0.337
HS Graduate+	0.370	0.483
College Graduate	0.376	0.484
MA	0.223	0.416
PhD	0.031	0.173
Performance	1.901	0.770
Age	42.168	9.371
Years at Company	6.123	3.793
Years at Title	3.847	2.836
Years at Level	3.855	2.837
Job Level 1	0.137	0.343
Job Level 2	0.146	0.353
Job Level 3	0.360	0.480
Job Level 4	0.357	0.479

Notes: Computed on subsample that: i) includes only job titles C, D, G, K, and Q; ii) omits workers with years of education equaling 15, 17, 19, or 20; iii) omits observations with missing performance data; iv) omits observations for which the history of job titles is incomplete over the worker's career at the firm.

TABLE 3: Probit Marginal Effects for Probability of Promotion in Year t

	(1)	(2)	(3)	(4)
	Basic Sample[1]	**Less Stringent Sample[2]**	**Basic Sample[1]**	**Basic Sample[1]**
HS Graduate+	-0.059 (5.09)***	-0.036 (4.78)***	-0.053 (3.89)***	-0.063 (4.77)***
MA	0.052 (3.96)***	0.040 (4.36)***	0.031 (1.98)**	0.049 (3.51)***
PhD	0.157 (4.05)***	0.093 (3.64)***	0.154 (3.17)***	0.160 (3.82)***
Performance (t-1)	-0.052 (7.30)***	-0.044 (9.44)***	-0.040 (4.11)***	-0.046 (3.07)***
Performance (t-2)	•	•	-0.035 (3.81)***	•
Age (t-1)	-0.038 (7.96)***	-0.017 (5.39)***	-0.032 (5.36)***	-0.038 (6.40)***
Age squared (t-1)	0.0003 (5.69)***	0.0001 (2.94)***	0.0003 (3.73)***	0.0003 (4.80)***
Years at Company (t-1)	-0.006 (1.99)**	-0.010 (4.92)***	-0.009 (2.47)**	-0.007 (2.21)**
Years at Title (t-1)	-0.013 (1.05)	-0.002 (0.36)	-0.002 (0.16)	-0.011 (0.86)
Years at Level (t-1)	0.044 (3.50)***	0.028 (4.27)***	0.031 (2.00)**	0.045 (3.54)***
Expected Performance (t+1)	•	•	•	-0.028 (0.50)
Job Level Dummies (t-1)	YES	YES	YES	YES
Job Title Dummies (t-1)	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES
Sample Size	N = 6514	N = 11,170	N = 4400	N = 6346
Pseudo R^2	0.14	0.19	0.14	0.14

[1] Includes only job titles C, D, G, K, and Q.
[2] Less stringent sample uses job titles C, D, E, F, G, K, L, N, O, P, and Q.

Notes: Z-statistics in parentheses below each estimate. Statistical significance at the 10%, 5%, and 1% levels denoted by *, **, and ***. All effects are evaluated at the means for all covariates. Marginal effects displayed for continuous covariates. For dummy covariates, cell entries are the differences in predicted probabilities when the dummy equals 1 and when it equals 0.

TABLE 4: OLS Estimates of Change in Annual Log-Wage

	(1)	(2)	(3)	(4)	(5)	(6)
	Basic Sample[1]		**Less Stringent Sample[2]**		**Basic Sample[1]**	
	Promotions	No Prom.	Promotions	No Prom.	Promotions	No Prom.
High School Graduate+	-0.404 (0.91)	-0.807 (1.06)	-0.275 (0.91)	-0.769 (1.20)	-0.339 (0.68)	0.848 (0.87)
MA	-1.210 (3.14)***	1.047 (1.63)	-0.948 (3.10)***	0.775 (1.39)	-0.515 (1.13)	1.147 (1.56)
PhD	-1.310 (1.47)	2.836 (1.31)	-0.891 (1.23)	2.472 (1.32)	-0.864 (0.81)	5.485 (2.52)**
Performance (t-1)	-0.397 (1.63)	-0.347 (3.23)***	-0.367 (2.02)**	-0.297 (3.32)***	-0.267 (0.083)	-0.408 (3.23)***
Performance (t-2)	•	•	•	•	0.594 (1.98)**	0.265 (2.36)**
Age (t-1)	-1.012 (6.10)***	•	-0.916 (7.27)***	•	-0.760 (3.67)***	•
Age squared (t-1)	0.010 (4.92)***	0.001 (0.75)	0.009 (5.87)***	0.001 (0.51)	0.007 (2.89)***	0.0003 (0.19)
Years at Company (t-1)	0.207 (1.71)*	0.008 (0.05)	0.185 (1.98)**	0.036 (0.28)	0.117 (0.90)	0.039 (0.20)
Years at Title (t-1)	-0.315 (0.84)	0.244 (0.67)	-0.131 (0.47)	0.100 (0.50)	-0.258 (0.51)	0.254 (0.60)
Years at Level (t-1)	0.656 (1.69)*	-0.664 (1.85)*	-0.672 (2.30)**	-0.473 (2.35)**	-0.655 (1.27)	-0.637 (1.53)
Constant	47.704 (11.15)***	7.216 (3.54)***	32.211 (9.95)***	9.355 (5.13)***	32.56 (6.46)***	5.661 (2.63)***
Job Level Controls (t-1)	YES	YES	YES	YES	YES	YES
Job Title Controls (t-1)	YES	YES	YES	YES	YES	YES
Year Dummies	YES	YES	YES	YES	YES	YES
Differences in Coefficients						
High School Graduate+	0.403 (0.46)		0.493 (0.70)		-1.187 (1.09)	
MA	-2.257 (3.01)***		-1.723 (2.70)***		-1.66 (1.92)*	
PhD	-4.146 (1.78)*		-3.363 (1.67)*		-6.345 (2.62)***	
Sample Size	N = 1302	N = 7442	N = 2030	N = 9153	N = 829	N = 5295

[1] Uses only job titles C, D, G, K, and Q.
[2] Less stringent sample uses job titles C, D, E, F, G, K, L, N, O, P, and Q.

Notes: All coefficients are multiplied by 100. Statistical significance at the 10%, 5%, and 1% levels denoted by *, ** and ***. Specification also includes interactions of the promotion dummy with all other covariates in addition to the education interactions. Age variable is dropped in the "no promotions" models due to collinearities in the presence of individual fixed effects.

TABLE 5: Correlation Matrix for Performance Ratings over Time

	Performance_t	Performance_{t-1}	Performance_{t-2}	Performance_{t-3}
Performance_t	1.000			
Performance_{t-1}	0.581^*	1.000		
Performance_{t-2}	0.394^*	0.590^*	1.000	
Performance_{t-3}	0.249^*	0.398^*	0.610^*	1.000

Note: * indicates correlation is statistically significantly different from zero at the 1% level. Correlations computed using “stringent sample” (i.e. job titles C,D,G,K,Q).

TABLE 6: Probit Marginal Effects for Probability of Promotion in Year t Controlling for Various Lags of Performance

	(1) Performance in Year t-1	(2) Performance in Years t-1 and t-2	(3) Performance in Years t-1, t-2, and t-3	(4) Performance in Years t-1,t-2,t-3, and t-4	(5) Performance in Years t-1,t-2,t-3, and t-4
HS Graduate+	-0.059 (5.09)***	-0.053 (3.89)***	-0.046 (2.82)***	-0.041 (2.04)**	-0.027 (1.19)
MA	0.052 (3.96)***	0.031 (1.98)**	0.055 (2.85)***	0.051 (2.11)**	0.061 (2.37)**
PhD	0.157 (4.05)***	0.154 (3.17)***	0.129 (2.08)**	0.191 (2.11)**	0.194 (2.00)**
Performance (t-1)	-0.052 (7.30)***	-0.040 (4.11)***	-0.034 (2.78)***	-0.021 (1.29)	-0.051 (1.99)**
Performance (t-2)	•	-0.035 (3.81)***	-0.040 (3.29)***	-0.029 (1.82)*	-0.033 (2.01)**
Performance (t-3)	•	•	-0.018 (1.75)*	-0.025 (1.71)*	-0.028 (1.91)*
Performance (t-4)	•	•	•	-0.007 (0.57)	-0.006 (0.48)
Expected Performance (t+1)	•	•	•	•	0.126 (1.55)
Sample Size	N = 6514	N = 4400	N = 2932	N = 1819	N = 1796
Pseudo R^2	0.14	0.14	0.15	0.13	0.13

Notes: Each specification is estimated on the subsample of job titles C, D, G, K, and Q. Apart from the the number of lagged performance measures, the specification is identical to our main specification in Column 1 of Table 3. Z-statistics in parentheses below each estimate. Statistical significance at the 10%, 5%, and 1% levels denoted by *, **, and ***. All effects are evaluated at the means for all covariates. Marginal effects displayed for continuous covariates. For dummy covariates, cell entries are the differences in predicted probabilities when the dummy equals 1 and when it equals 0.

TABLE 7:
Probit Marginal Effects, Levels 1 and 2

	Basic Sample
HS Graduate+	-0.084 (3.94)***
MA	0.084 (2.99)***
PhD	0.192 (1.84)*
Performance (t-1)	-0.023 (1.72)*
Age (t-1)	-0.055 (6.97)***
Age squared (t-1)	0.001 (5.41)***
Years at Company (t-1)	-0.069 (1.18)
Years at Title (t-1)	-0.034 (0.57)
Job Title Dummies (t-1)	YES
Year Dummies	YES
Sample Size	N = 2085
Pseudo R^2	0.16

Notes: Includes only job titles C, D, G, K, and Q, and the promotions to Level 2 of workers who entered the firm at Level 1. Z-statistics in parentheses below each estimate. Statistical significance at the 10%, 5%, and 1% levels denoted by *, **, and ***. All effects are evaluated at the means for all covariates. Marginal effects displayed for continuous covariates. For dummy covariates, cell entries are the differences in predicted probabilities when the dummy equals 1 and when it equals 0.