Comparative Advantage, Learning, and Sectoral Wage Determination

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation	Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent. 2005. Comparative advantage, learning, and sectoral wage determination. 2005. Journal of Labor Economics 23(4): 681-724.
Published Version	http://dx.doi.org/10.1086/491606
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:2766651
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Comparative Advantage, Learning, and Sectoral Wage Determination

Robert Gibbons, *Massachusetts Institute of Technology and National Bureau of Economic Research*

Lawrence F. Katz, *Harvard University and National Bureau of Economic Research*

Thomas Lemieux, *University of British Columbia and National Bureau of Economic Research*

Daniel Parent, *McGill University*

We develop a model in which a worker's skills determine the worker's current wage and sector. The market and the worker are initially uncertain about some of the worker's skills. Endogenous wage changes and sector mobility occur as labor market participants learn about these unobserved skills. We show how the model can be estimated using nonlinear instrumental variables techniques. We apply our methodology to study wages and allocation of workers across occupations and industries using individual-level panel data from the National Longitudinal Survey of Youth. We find that high-wage sectors employ high-skill workers and offer high returns to workers' skills.

I. Introduction

We analyze the theoretical and econometric implications of comparative advantage and learning for the wages and sector affiliations of individuals

and for changes in these variables over workers' careers. After developing the theory and econometrics, we turn to two empirical applications of our methodology concerning the wages and allocations of workers across occupations and across industries.

Our focus on comparative advantage is motivated by a large and established literature. Many have found that the average characteristics of individuals vary by sector.[1] Furthermore, several have found that the measured returns to individuals' observable characteristics vary by sector.[2] Finally, Heckman and Scheinkman (1987) rejected the hypothesis that the returns to individuals' time-invariant unmeasured characteristics are constant across sectors.

Our focus on learning is motivated by a smaller and more recent literature. While Jovanovic (1979), Harris and Holmstrom (1982), and others showed long ago that learning models could provide new interpretations for important features of the data (such as the return to seniority and the increase in the variance of wages with experience), recent work has built on these foundations to derive and test novel implications, many of which have survived confrontations with data.[3]

Our theoretical model emphasizes the role of worker skills that cannot be measured by an econometrician. To clarify the exposition of the econometrics, we develop the theory in stages. We begin with two models in which workers' skills are equally valued in all sectors. In the first of these models, all labor market participants have perfect information about workers' skills; in the second, information is initially imperfect, but output observations convey additional information over time and so endogenize

[1] For example, Dickens and Katz (1987) found differences in average education levels by industry, and Blackburn and Neumark (1992) found sorting by test scores across industries.

[2] For example, Mincer and Higuchi (1988) found differences in returns to tenure and experience across industries in Japan and the United States, and Freeman and Medoff (1984) found differences in returns to education and experience for union and nonunion workers.

[3] For example, Farber and Gibbons (1996) derive and test the prediction that the residual from a regression of an individual's score on an ability test (the Armed Forces Qualifying Test [AFQT]) on observable characteristics and the first wage should have increasing explanatory power for subsequent wages as experience increases. Continuing in this vein, Altonji and Pierret (2001) derive and test the prediction that the effect of observable characteristics (like education) should decrease with experience, while the effect of initially unobservable characteristics like AFQT (not only the AFQT residual) should increase with experience. Chiappori, Salanié, and Valentin (1999) derive and test a new prediction from the Harris-Holmstrom model, that comparing two individuals in the same job in period 1 and the same (higher) job in period 3, future wage and promotion prospects are brighter for the individual who was promoted later (i.e., in period 3 rather than period 2).

wage changes. We then develop two other models in which different sectors place different values on workers' skills and workers sort themselves into different sectors on the basis of comparative advantage. In the first of these latter two models, labor market participants again have perfect information about workers' skills; in the second, information is again imperfect, so learning endogenizes not only wage changes but now also sector mobility.

Our richest model, with comparative advantage and learning, resembles the learning and matching models of wages and turnover pioneered by Jovanovic (1979), Ross, Taubman, and Wachter (1981), and MacDonald (1982). In Jovanovic's model, a worker's performance is independent across jobs, whereas in our model (like those of Ross et al. and MacDonald), a worker's performance in one job determines not only the expected value of staying in that job but also the expected value of moving to a given new job. We differ from Ross et al. and MacDonald by introducing a one-dimensional notion of ability that determines a worker's productivity in every sector, much as in Murphy (1986). The resulting model of learning and sorting is a natural generalization of the two-period, two-sector, two-type model in Gibbons and Katz (1992).

As is well known, in our simplest theoretical model (in which worker skills are equally valued in all sectors and there is no learning by labor-market participants), the returns to time-varying worker characteristics can be estimated using first differences to eliminate individual fixed effects that are unmeasured by the econometrician. Similarly, in this simplest model, first differences can be used to estimate sectoral wage differentials without bias from unmeasured fixed effects. Unfortunately, first-difference estimation is not appropriate for any of the three other theoretical models we develop. Simply put, in these models, a worker's fixed ability does not translate into a fixed effect in a wage equation, so first-differencing the wage equation does not correct ability bias. The contribution of this article is to move beyond merely warning of this problem (e.g., as in Gibbons and Katz 1992) to proposing a solution.

Our theoretical models rely heavily on the assumption of normality. Many models that rely on normality can be estimated by maximum likelihood or two-step methods, but estimating our dynamic model of wage determination and sector affiliation would be computationally difficult because it entails more than two sectors and more than two periods. In addition, it is not necessary to estimate the full model when the parameters of interest are those that determine the returns to skills and wage differences across sectors. We therefore undertake the more modest task of estimating the wage equations in each sector.[4]

---

[4] Other results exist on the identifiability of related models in the absence of normality. For example, Heckman and Honoré (1990) show that the Roy model is

We show that our richest theoretical model produces a random coef-
ficients econometric model in a panel data setting, which can be estimated
using a nonlinear instrumental variables technique. Even in this richest
model, consistent estimates of the effects of both measured and unmea-
sured skills on wages require neither distributional assumptions nor stan-
dard exclusion restrictions. (That is, we use normal distributions to de-
velop the full theory, but we do not need these assumptions to estimate
the parameters of interest related to sectoral wage differentials and sector-
specific returns to skills.) Instead, the estimation strategy utilizes natural
restrictions available in panel data with three or more observations per
person. Our econometric approach is similar to other panel data models
in which first-differenced estimates are inconsistent, such as Holtz-Eakin,
Newey, and Rosen (1988) and Lemieux (1998).

After developing the theory and econometrics, we undertake two em-
pirical investigations concerning sorting and wage differentials across oc-
cupations and industries using individual-level panel data from the Na-
tional Longitudinal Survey of Youth. Our richest theoretical model is
consistent with several familiar facts about wage determination: a typical
individual's wage increases with experience, the variance of the wage dis-
tribution across individuals increases with experience, and the skewness
of the wage distribution increases with experience.[5] But variants of
Mincer's (1974) theory can also explain these basic facts, so we focus on
our model's further predictions concerning the returns to skills and the
resulting allocation of workers across sectors. For both occupations and
industries, we find important variation in sector-specific returns to ob-
served and unobserved skills. Furthermore, in both cases, high-wage sec-
tors employ high-skill workers and offer high returns to workers' skills,
so estimates of sectoral wage differences that do not account for sector-
specific returns to skill and the sorting of workers across sectors on the
basis of unmeasured skills are misleading and difficult to interpret.

Although our empirical work explores two standard definitions of sec-
tors (i.e., occupations and industries), other definitions are possible. For
example, sectors could be jobs inside a firm (Lluis, in this issue; Gibbons
and Waldman 1999), states or regions within a country (Borjas, Bronars,
and Trejo 1992), or entire countries (Borjas 1987). In fact, the individuals
in our model need not be workers. Instead, with some changes to the
model, the workers could be reinterpreted as firms, where what we call
worker ability is reinterpreted as firm productivity, much as Jovanovic
(1982) reinterpreted Jovanovic (1979).

---

identified with panel data and exogenous shifts in the price of skills over time. But
Heckman and Honoré focus on the estimation of a sequence of static models; they
do not address learning and job mobility.

[5] An illustrative discussion of these implications of the model and a comparison
with alternative labor market models is presented in Neal and Rosen (2000).

II. Theory and Econometrics

The four theoretical models analyzed below are special cases of the following model. If worker i is employed in sector j at time t, the worker's output is

$$y_{ijt} = \exp(X_{it}\beta_j + \psi_{ijt}), \tag{1}$$

where X_{it} is a vector of human capital and demographic variables measured by the econometrician and ψ_{ijt} represents determinants of productivity that are not measured by the econometrician. The worker characteristics X_{it} and the slope vector β_j are known by all labor market participants at the beginning of period t; the realized output y_{ijt} is observed by all labor market participants at the end of period t. The error term ψ_{ijt} has the components

$$\psi_{ijt} = Z_i + b_j(\eta_i + \varepsilon_{ijt}) + c_j, \tag{2}$$

where Z_i denotes the portion of worker i's productive ability that is equally valued in all sectors, η_i denotes the portion that is differentially valued across sectors, and ε_{ijt} is a random error. The coefficients $\{b_j, c_j; j = 1, \dots, J\}$ are fixed and known to all labor market participants. The noise terms ε_{ijt} are normal with zero mean and precision h_ε (i.e., variance $\sigma_\varepsilon^2 = 1/h_\varepsilon$) and are independent of each other and of all the other random variables in the model.

In developing the theory and econometrics, we treat Z_i and η_i differently. We assume throughout that Z_i is observed by all labor market participants; this is the standard case of a fixed effect that the econometrician cannot observe but market participants can. For η_i, however, we consider two cases: perfect information (no learning by market participants, as with Z_i) and imperfect information (learning). One could also imagine investigating learning about Z_i. Farber and Gibbons (1996) study this problem in the absence of sector-specific returns to ability (i.e., $b_j = b$ and $\beta_j = \beta$ for every j, so that a worker's unmeasured ability is $Z_i + b\eta_i$ and is equally valued in every sector), but the combined problem of learning about Z_i and about η_i (with b_j varying across sectors) awaits future research.

In the imperfect information case, all labor market participants share symmetric but imperfect information about η_i. In particular, given their initial information (Z_i and X_{i1}), all participants in the labor market share the prior belief that η_i is normal with mean m and precision h. Subsequent productivity observations, y_{ijt}, refine this belief. Information in the labor market therefore remains symmetric and improves over time. For simplicity, we assume that subsequent realizations of measured skills, X_{it}, are conditionally independent of η_i given Z_i and X_{i1}. (This assumption is not

only convenient but realistic because the major time-varying element of X_{it} is experience.) Thus, market participants can compute

$$s_{it} = \frac{\ln y_{ijt} - X_{it}\beta_j - Z_i - c_j}{b_j}, \tag{3}$$

which yields $s_{it} = \eta_i + \varepsilon_{ijt}$, a noisy signal about the worker's ability that is independent of the worker's sector during period t. We call s_{it} the worker's normalized productivity observation for period t. Let $s_i^t = (s_{i1}, \dots, s_{it})$ denote the history of the worker's normalized productivity observations through period t. Then (from chap. 9 of DeGroot 1970) the posterior distribution of η_i given history s_i^t is normal with mean

$$m_t(s_i^t) = \frac{hm + h_\varepsilon(s_{i1} + \dots + s_{it})}{h + th_\varepsilon} \tag{4}$$

and precision $h_t = h + th_\varepsilon$.

To close the model, we assume that workers are risk neutral and that there is no cost to firms or workers at the beginning or end of a job (i.e., no hiring, firing, or mobility costs), so we can restrict attention to single-period compensation contracts. For simplicity, we further restrict attention to contracts that specify the period's wage before the period's production occurs (as opposed to piece-rate contracts). Competition among firms causes each firm in a given sector to offer a given worker a wage equal to the expected value of the worker's output in that sector, given the worker's observed characteristics and history of previous output realizations.

It is not controversial that workers' productive abilities are imprecisely measured in standard micro data sets. But if unmeasured skills are to explain estimated sector wage differentials, then these skills must be non-randomly allocated across sectors. This, too, is plausible, for example because different sectors use different technologies that require workers' skills in different proportions. But if this unmeasured skill explanation for measured sectoral wage differentials is correct, it suggests that the few skills that are measured in standard micro data sets (hereafter "measured skills") should be systematically related to the sector in which the worker is employed. We investigate this prediction about measured skills in our empirical work on occupations and industries in Sections III and IV. In this section's discussion of econometric issues, however, we confine our attention to estimating the role of unmeasured skills.

A. Sorting without Comparative Advantage

In this subsection, we ignore the possibility of comparative advantage by assuming that $b_j = b$ for every j, so that a worker's unmeasured ability is $Z_i + b\eta_i$ and is equally valued in every sector. Continuing in this vein,

we also assume in this section that $\beta_j = \beta$ for every j. But we allow the intercepts c_j to vary by sector, in keeping with the possibility that measured sector premia may reflect true sector effects. Of course, all else constant, jobs in sectors with high values of c_j may be more attractive (depending on the source of c_j, such as rent sharing vs. compensating differentials). If some sectors are more attractive, issues such as queuing and rationing arise. Because our main interest is in the richer model with comparative advantage in Section II.C, we do not formally address queuing or rationing here.

In the perfect information case without comparative advantage, all firms know that the worker's ability is $Z_i + b\eta_i$. As always, the wage offered to worker i by firms in sector j in period t is the worker's expected output in that sector, but the only uncertainty in this case is the error term $b\varepsilon_{ijt}$ in (2). Recall that if $\log \theta$ is normally distributed with mean μ and variance σ^2, then $E(\theta) = \exp\{\mu + (1/2)\sigma^2\}$. Therefore, the log wage offered to worker i in sector j in period t is

$$\ln w_{ijt} = X_{it}\beta + Z_i + b\eta_i + c_j + (1/2)b^2\sigma_\varepsilon^2. \tag{5}$$

Turning to the imperfect information case without comparative advantage, in each period, firms in sector j bid worker i's wage up to the worker's expected output in that sector (conditional on the publicly observable information available at that date), so the log wage is

$$\ln w_{ijt} = X_{it}\beta + Z_i + bm_{i,t-1} + c_j + (1/2)b^2\sigma_t^2, \tag{6}$$

where $m_{i,t-1}$ is shorthand for $m_{t-1}(s_i^{t-1})$ and $\sigma_t^2 = (h + th_\varepsilon)/\{h_\varepsilon[h + (t-1)h_\varepsilon]\}$. Note that σ_t^2 converges to $\sigma_\varepsilon^2 (= 1/h_\varepsilon)$, the corresponding variance term in equation (5), as the number of periods t goes to infinity. Also, in both the perfect and the imperfect information cases, the worker's ability η_i is unmeasured by the econometrician (as is Z_i); in the latter case, η_i is also unobserved by labor market participants (unlike Z_i). Note that, since t represents the number of years of experience in the model, the error component $(1/2)b^2\sigma_t^2$ will be captured by a function in labor market experience that we include in all estimated models.

B. Estimation without Comparative Advantage

In the absence of both learning and comparative advantage, the source of possible bias in conventional cross-section estimates of sectoral log wage differentials is the potential partial correlation between sector affiliation and unmeasured skills (Z_i and η_i) conditional on measured skills (X_{it}). In this simplest case, the worker's fixed ability ($Z_i + b\eta_i$) creates a worker fixed effect in the wage regression, which can be eliminated in standard fashion. For example, a first-differenced regression eliminates the fixed effect $Z_i + b\eta_i$ in (5).

Even without comparative advantage, however, learning implies that fixed ability is not a fixed effect in the earnings equation. The key property of our learning model is that Bayesian beliefs are a martingale. That is, the conditional expectation $m_t(s_i^t)$ in (4) obeys the law of motion

$$m_{it} = m_{i,t-1} + \xi_{it}, \tag{7}$$

where ξ_{it} is a noise term orthogonal to $m_{i,t-1}$. In somewhat more intuitive terms, the market begins period t with the information contained in s_i^{t-1} and then extracts new information about η_i from the output observation y_{ijt} (or, equivalently, s_{it}). But the new information that can be extracted from y_{ijt} is precisely the part that could not be forecasted from s_i^{t-1}. Hence, the innovation ξ_{it} is orthogonal to the prior belief $m_{i,t-1}$.

Farber and Gibbons (1996) explored some of the implications of this martingale property. But they focused on several specific predictions that can be derived regarding regressions in which the dependent variable is the level of earnings, not the log of earnings. In this article, in contrast, we use the log of earnings as the dependent variable, so the specific Farber-Gibbons predictions do not hold, but the martingale property of the market's beliefs continues to create endogeneity problems, as follows.[6]

Formally, a first-differenced regression eliminates Z_i but not $m_{i,t-1}$ from (6). Instead, first-differencing (6) for a worker who switches from sector j to sector j' yields

$$\ln w_{it} - \ln w_{i,t-1} = (X_{it} - X_{i,t-1})\beta + b(m_{i,t-1} - m_{i,t-2})$$
$$+ (c_{j't} - c_j) + (1/2)b^2(\sigma_t^2 - \sigma_{t-1}^2), \tag{8}$$

where $m_{i,t-1} - m_{i,t-2} = \xi_{i,t-1}$. But $\xi_{i,t-1}$ may be correlated with the change in sector affiliation through whatever (unmodeled) process led unmeasured ability to be correlated with sector affiliation in the first place.[7] Thus, with learning, first-differenced estimates of sectoral wage differ-

---

[6] Relative to Farber and Gibbons, we also use the more specific production function (1), the more specific error structure (2), and the more specific distributional assumptions given in the text below (2). We impose these more specific assumptions in order to explore several issues related to the returns to skills across sectors that Farber and Gibbons could not address with their more general model.

[7] For example, suppose that there are only two levels of ability, high and low, but that sectors differ in the proportion of high-ability workers they employ. Consider a high-ability worker whose employment exogenously ends in sector j. Suppose that such a worker is equally likely to be reemployed in any of the economy's jobs for high-ability workers. Then there is some probability that the worker's new job is again in sector j, but if the worker changes sectors, then it is likely that the new job is in a sector with a large number of high-ability jobs. In this case, positive information about a worker's ability will tend to be associated with shifts to high-wage sectors (where high-skill jobs are more plentiful), and the reverse for negative information.

entials are biased if the change in the residual is correlated with the change in sector affiliation. Fortunately, this endogeneity problem is simple to correct because the new information summarized in $\xi_{i,t-1}$ is not related to wage, skill, or sector information in period $t - 1$ or earlier. (See Sec. II.D for more discussion of this issue.) For example, equation (8) can be estimated by two-stage least squares using the interaction of the worker's (publicly observable) score on an ability test (taken before the worker entered the labor market) and the worker's sector affiliation at $t - 1$ as a valid instrumental variable for changes (between $t - 1$ and t) in sector affiliation.

C. Sorting with Comparative Advantage

In this section, we relax the assumption that a worker's ability is equally valued in every sector. By introducing comparative advantage, we endogenize sector affiliation. By subsequently introducing learning, we endogenize not only wage changes but also sector mobility.

To analyze comparative advantage, we now return to the production function specified in (1) and (2), where the slope coefficients β_j in (1) and b_j in (2) vary by sector. We index the J sectors so that b_j strictly increases in j: sector $j + 1$ values the worker's ability η_i more than does sector j. In keeping with the notion that ability is productive, we assume that $b_1 > 0$. Given a fixed X_{it}, there exist critical values of η_i that determine the efficient assignment of workers to sectors. Denoting these critical values by $\{v_j(X_{it}); j = 0, 1, \ldots, J\}$, the efficient assignment rule assigns worker η_i to sector j if and only if $v_{j-1}(X_{it}) < \eta_i < v_j(X_{it})$, where $v_0(X_{it}) = -\infty$, $v_n(X_{it}) = \infty$, and $v_j(X_{it})$ strictly increases in j. See figure 1 for a graphical representation of this efficient assignment rule.

We again analyze first perfect and then imperfect information. In the perfect information case, firms in sector j bid worker i's wage up to the expected output in that sector:

$$\ln w_{ijt} = X_{it}\beta_j + Z_i + b_j\eta_i + c_j + (1/2)b_j^2\sigma_\varepsilon^2, \tag{9}$$

analogous to (5) but with the sector-specific returns β_j and b_j.[8] If the worker faces no mobility constraints, worker i will choose to work in sector j if $v_{j-1}(X_{it}) < \eta_i < v_j(X_{it})$. Thus, taking the model literally, sector mobility in the perfect information case is driven entirely by changes in $X_{it}\beta_j$. One could envision exogenous shocks to sector demand that pro-

---

[8] In this model, the sector-specific intercepts are now c_j plus the term $(1/2)b_j^2\sigma_\varepsilon^2$. This additional term accounts for the fact that when wages are paid in advance, differences in the variance of productivity across sectors (due to differences in the returns to skill b_j) lead to systematic differences in mean log wages because of the log normal transformation mentioned earlier. We view this as just one among several other possible sources of systematic wages differences across sectors. Other possible factors include compensating differentials, efficiency wages, and rents.

$E(\ln y_{ijt})$

$E(\ln y_{i,j+1,t}) =$
$X_{it}\beta_{j+1} + Z_i + b_{j+1}\eta_i + c_{j+1}$

$E(\ln y_{ijt}) =$
$X_{it}\beta_j + Z_i + b_j\eta_i + c_j$

$E(\ln y_{i,j-1,t}) =$
$X_{it}\beta_{j-1} + Z_i + b_{j-1}\eta_i + c_{j-1}$

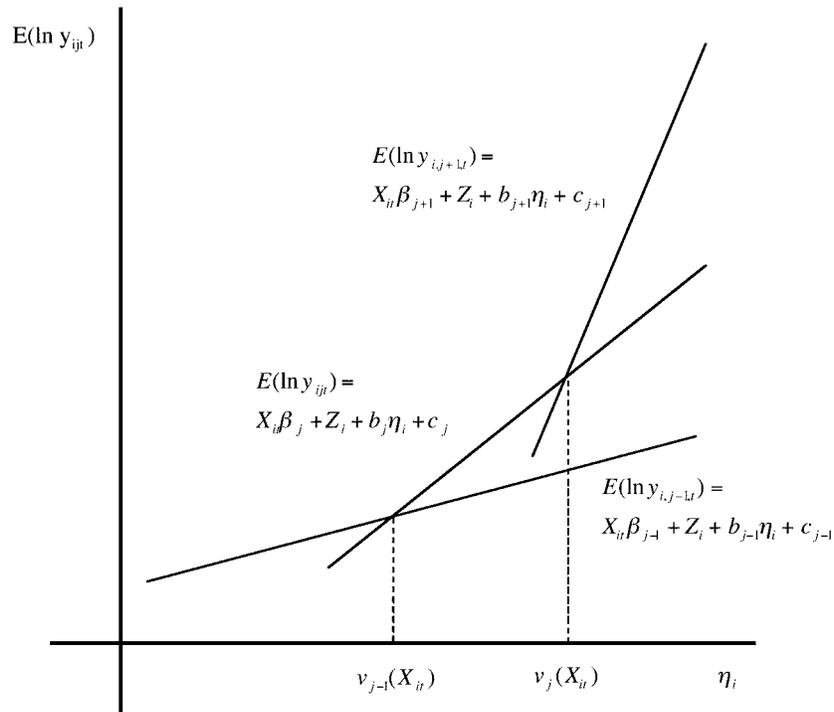$v_{j-1}(X_{it})$ $v_j(X_{it})$ η_i

Fig. 1.—Efficient assignment with comparative advantage

duce additional sector mobility in this model, but we will not formally model such shocks for the same reason that we did not model queues or rationing in Section II.A: our ultimate interest is in the model with comparative advantage and learning, which gives a coherent account of sectoral mobility without reference to queues, rationing, or sectoral shocks. Whatever the reason that worker i is employed in sector j in period t, in the perfect information case with comparative advantage, we assume that the worker's wage is given by (9).

In the imperfect information case, we again assume that information in the labor market is symmetric but imperfect, as described above. As in the model of learning without comparative advantage, all participants in the labor market share the prior belief that η_i is normal with mean m and precision h, conditional on their initial information Z_i and X_{i1}. Inferences from the productivity observations, y_{ijt}, are greatly simplified because the information content of an output observation is constant across sectors; that is, (2) involves $b_j(\eta_i + \varepsilon_{ijt})$ rather than $b_j\eta_i + \varepsilon_{ijt}$. This functional form is what allows us to define the normalized productivity observation for worker i in period t, s_{it} from (3), to be a noisy signal

about the worker's ability that is independent of the worker's sector during period t. Relaxing this assumption about the functional form of (2) would complicate the analysis because workers' sector choices would then depend on the benefit from faster learning, as well as on the benefit from increased expected output given current beliefs. Relaxing the assumption that all labor market participants observe Z_i (so that there could be learning about both Z_i and η_i) would cause similar complications. Under our assumptions, the posterior distribution of η_i given the history s_i^t is normal with mean m_{it} given by (4) and precision $h_t = h + th_\varepsilon$, regardless of the worker's history of sector affiliations.

In this fourth model, with learning and comparative advantage, we finally have an internally consistent account for sector affiliation, wages, sector mobility, and wage changes, as follows. In each period, firms in sector j bid a worker's wage up to the worker's expected output in that sector, conditional on the publicly observable information about the worker available at that date:

$$\ln w_{ijt} = X_{it}\beta_j + Z_i + b_j m_{i,t-1} + c_j + (1/2)b_j^2\sigma_t^2, \tag{10}$$

which is analogous to (6) but with sector-specific returns β_j and b_j. The model also includes sector-specific (experience) effects since the posterior variance σ_t^2, which declines with time (labor market experience), is interacted with b_j^2. The worker chooses to work in the sector offering the highest current wage. Thus, worker i chooses sector j in period $t + 1$ if $v_{j-1}(X_{i,t+1}) < m_{it} < v_j(X_{i,t+1})$. In all of our models, including this richest one, if the parameters $\{\beta_j, b_j, c_j; j = 1, \dots, J\}$ and the measured characteristics X_{it} take on certain values, then one or more sectors may lie below the upper envelope in figure 1 for all values of the unmeasured characteristics, in which case no workers with such measured characteristics should be employed in such sectors.

D. Estimation with Comparative Advantage

We now develop a nonlinear instrumental variables procedure to estimate the parameters $\{\beta_j, b_j, c_j; j = 1, \dots, J\}$ in (9) and (10). This procedure does not rely on normality and can be implemented using standard computer packages. To discuss the estimation of the model, define the sector indicators D_{ijt}, where D_{ijt} equals one if person i is employed in sector j at time t and zero otherwise.

The wage equation (10) for each sector j can then be written as a single equation where measurement error μ_{it} is assumed to be independent of sector affiliation:

$$\ln w_{it} = \sum_j D_{ijt} c_j + \sum_j D_{ijt} X_{it}\beta_j + Z_i$$
$$+ \sum_j D_{ijt} b_j m_{i,t-1} + (1/2) \sum_j D_{ijt} b_j^2 \sigma_t^2 + \mu_{it}. \tag{11}$$

Estimates of the sector slopes and intercepts $\{\beta_j, c_j; j = 1, \dots, J\}$ obtained by estimating equation (11) with ordinary least squares (OLS) are inconsistent. The problem is that expected ability influences sector affiliation, so $m_{i,t-1}$ is correlated with the set of sector dummies $\{D_{ijt}, j = 1, \dots, J\}$. If the worker's ability were fixed, known, and equally valued in all sectors (as was the case with $Z_i + b\eta_i$ in Sec. II.B), then a first-differenced regression would eliminate this ability bias. But the endogeneity problem in equation (11) is different from the usual fixed-effect case for two reasons. First, as noted in Section II.B, $m_{i,t-1}$ is a martingale rather than a fixed effect. This martingale property does not depend on the normality assumptions in our theoretical model; all Bayesian beliefs are martingales. In the absence of comparative advantage, we could handle this martingale problem as described in Section II.B. But, second, comparative advantage causes $m_{i,t-1}$ to be interacted with the set of sector dummies $\{D_{ijt}, j = 1, \dots, J\}$ in (11).

Other panel data models in which first-differenced estimates are inconsistent have been considered in the literature. For example, Holtz-Eakin et al. (1988) discuss the estimation of models in which the fixed effect is interacted with year dummies. They show that consistent estimates can be obtained by quasi-differencing the equation of interest and then using appropriate instrumental variables techniques. Similarly, Lemieux (1998) estimates a model in which the return to a time-invariant unobserved characteristic is different in the union and the nonunion sectors.

The estimation strategy we follow also relies on quasi-differencing combined with instrumental variables (IV) techniques. But the general case of our estimation strategy—equation (13) below—is hard to interpret, so we first provide intuition via the following two-sector example. To simplify the exposition, we suppress the independent variables X_{it}, the fixed effect Z_i, and the measurement error μ_{it}. We also assume that all workers are in sector 1 in period $t - 1$. Finally, we set $c_1 = 0$ and $b_1 = 1$, so the wage equation for period $t - 1$ is extremely simple:

$$\ln w_{i,t-1} = m_{i,t-2} + (1/2)\sigma_{t-1}^2.$$

The wage paid to worker i employed in sector k in period t is then

$$\ln w_{it} = c_k + b_k m_{i,t-1} + (1/2)b_k^2 \sigma_t^2,$$

where $k = 1$ or 2, depending on whether the worker changes sectors. Multiplying the former equation by b_k, subtracting it from the latter equation, and recalling from (7) that $m_{i,t-1} = m_{i,t-2} + \xi_{i,t-1}$ then yields

$$\ln w_{it} - b_k \ln w_{i,t-1} = c_k + b_k [\xi_{i,t-1} - (1/2)\sigma_{t-1}^2] + (1/2)b_k^2 \sigma_t^2.$$

Thus, for workers who remain in sector 1, we have a first-differenced

wage equation ($b_1 = 1$), but for workers who move to sector 2, we have a quasi-differenced wage equation ($b_2 \neq 1$).

To estimate this equation, let D_{it} represent employment in sector 2 in period t. We can then write

$$
\ln w_{it} - \ln w_{i,t-1} = c_2 D_{it} + (b_2 - 1) D_{it} \ln w_{i,t-1}
$$

$$
+ \; \xi_{i,t-1}[1 + (b_2 - 1)D_{it}]
$$

$$
+ \; (1/2)|(b_2^2 - 1)D_{it} + 1|\sigma_t^2
$$

$$
- \; (1/2)|(b_2 - 1)D_{it} + 1|\sigma_{t-1}^2.
$$

This now looks more like a standard equation in first differences, but running a simple regression of the first difference in wages on D_{it} and $D_{it} \ln w_{i,t-1}$ will not yield consistent estimates of the coefficients c_2 and b_2. Although the error component $\xi_{i,t-1}$ is not correlated with $\ln w_{i,t-1}$, $\xi_{i,t-1}$ is still positively correlated with the sector affiliation dummy D_{it} because workers who get a positive innovation are more likely to switch to sector 2 (in the case where $b_2 > 1$, or negatively correlated if $b_2 < 1$). Furthermore, the composite error term $\xi_{i,t-1}[1 + (b_2 - 1)D_{it}]$ is mechanically correlated with $D_{it} \ln w_{i,t-1}$. These problems can be resolved by rearranging terms and estimating the quasi-differenced equation

$$
\frac{\ln w_{it}}{1 + (b_2 - 1)D_{it}} - \ln w_{i,t-1} = (c_2/b_2)D_{it} + \xi_{i,t-1}
$$

$$
+ \; (1/2)\big[\sigma_t^2 + (b_2 - 1)D_{it}\sigma_t^2 - \sigma_{t-1}^2\big]
$$

by nonlinear instrumental variables using appropriate instruments for D_{it}. This equation represents a simplified special case of our quasi-differenced estimating equation (13) below. As discussed earlier, the innovation term $\xi_{i,t-1}$ is uncorrelated with variables from period $t - 1$ and earlier. Any variables from period $t - 1$ and earlier (including sector affiliation histories) can thus be used as instruments. We also address the issue of the sector-varying term $D_{it}\sigma_t^2$ in this estimating equation by including interactions of sector dummies and experience; these interaction terms are treated as endogenous and instrumented for using interactions of lagged sector affiliation dummies and experience.

In our general model, we proceed analogously. First, solving (11) for $m_{i,t-1}$ yields

$$
m_{i,t-1} = \frac{\ln w_{it} - \sum_j D_{ijt}c_j - \sum_j D_{ijt}X_{it}\beta_j - Z_i - (1/2)\sum_j D_{ijt}b_j^2\sigma_t^2 - \mu_{it}}{\sum_j D_{ijt}b_j}. \tag{12}
$$

The lagged version of equation (12) yields a similar expression for

$m_{i,t-2}$. Substituting the expressions for both $m_{i,t-1}$ and $m_{i,t-2}$ into the law of motion $m_{i,t-1} = m_{i,t-2} + \xi_{i,t-1}$ yields:[9]

$$
\frac{\ln w_{it}}{\sum_j D_{ijt} b_j} = \frac{\ln w_{i,t-1}}{\sum_j D_{ij,t-1} b_j}
$$

$$
+ \frac{\sum_j D_{ijt} c_j + \sum_j D_{ijt} X_{it} \beta_j + (1/2) \sum_j D_{ijt} b_j^2 \sigma_t^2}{\sum_j D_{ijt} b_j}
$$

$$
- \frac{\sum_j D_{ij,t-1} c_j - \sum_j D_{ij,t-1} X_{i,t-1} \beta_j + (1/2) \sum_j D_{ij,t-1} b_j^2 \sigma_{t-1}^2}{\sum_j D_{ij,t-1} b_j} + e_{it},
$$

$$(13)$$

where

$$
e_{it} = \xi_{i,t-1} + \frac{Z_i + \mu_{it}}{\sum_j D_{ijt} b_j} - \frac{Z_i + \mu_{i,t-1}}{\sum_j D_{ij,t-1} b_j}.
$$

The estimating equations include interactions of sector dummies with experience and of lagged sector dummies with lagged experience to capture the sector-varying, experience-varying terms involving $D_{ijt}\sigma_t^2$ and $D_{ij,t-1}\sigma_{t-1}^2$, respectively. The sector dummies D_{ijt} are correlated with $\xi_{i,t-1}$ in equation (13) because expected ability influences sector affiliation. To handle this problem, we need instrumental variables for the set of sector dummies $\{D_{ijt}, j = 1, \dots, J\}$ (and for $\ln w_{i,t-1}$ and for the interactions of sector dummies and experience capturing $D_{ijt}\sigma_t^2$). Such instrumental variables must, of course, be orthogonal to the error term e_{it} in equation (13). In particular, they must be orthogonal to the innovation term $\xi_{i,t-1}$.

The most obvious candidate instrumental variables are skill or sector information from period $t - 1$ or earlier. The second lag of the wage ($\ln w_{i,t-2}$) is also a potential instrumental variable, as are interactions among these various variables. Since the evolution of wages and sector affiliation over time is driven by the evolution of m_{it}, these wage, skill, and sector histories should help predict $m_{i,t-1}$ and thus $\ln w_{i,t-1}$ and $\{D_{ijt}, j = 1, \dots, J\}$. We chose the interaction between sector affiliation at time $t - 1$ and $t - 2$, $\{D_{ij,t-1}, j = 1, \dots, J\}$ and $\{D_{ij,t-2}, j = 1, \dots, J\}$, as our main instrumental variables. These interactions between sector affiliation at

[9] In an earlier version of this article (Gibbons et al. 2002), we used a slightly different quasi-differences procedure that amounts to multiplying both sides of eq. (13) by $\sum D_{ijt} b_j$. Unfortunately, this alternative procedure yields inconsistent estimates in the presence of learning (but consistent estimates when there is comparative advantage but no learning). We are very grateful to Robert Topel and Derek Neal for pointing out this problem.

$t - 1$ and $t - 2$ are uncorrelated with the error term e_{it} in equation (13) given the model's assumption that sector affiliation is determined only by perceptions about the sector-sensitive components of ability (X_{it} and η_i) and is independent of any part of ability that is not differentially valued across sectors (Z_i). In appendix B we discuss in more detail why the model suggests using these variables as instruments. We also show evidence of their predictive power. For efficiency reasons discussed below, we also include a set of interactions of sector affiliation at time $t - 2$ with the explanatory variables X_{it} (as summarized by a skill index and years of experience) in the instrument set.[10]

Equation (13) is not a standard wage equation since some of the parameters are on both sides of the equation. We therefore estimate the parameters in equation (13) using nonlinear instrumental variables (NLIV) techniques. Consider e, a vector in which all the individual error terms e_{it} are stacked, and V, a matrix in which the individual instrument vector v_{it} (e.g., sector histories) are stacked. Since the error terms e should be uncorrelated with the instruments V, the orthogonality condition $(1/N)e'V = 0$ should hold. The NLIV method consists of setting the sample analogs of $(1/N)e'V$ as close as possible to zero by finding the values of the parameters c_j, b_j, and β_j (for $j = 1, \dots, J$) that minimize the quadratic form

$$S = (1/N)(e'V)M(e'V)', \qquad (14)$$

where M is a weighting matrix. Note that the parameters $\{\beta_j, b_j, c_j; j = 1, \dots, J\}$ are implicitly included in S because the elements e_{it} of e are computed as the difference between $\ln w_{it}$ and the explanatory factors on the right-hand side of equation (13). Under the assumption that e_{it} is homoskedastic and uncorrelated, it is well known that the most efficient estimate is obtained by using the inverse of the variance of $e'V$ as a weighting matrix, $M = (V'V)^{-1}$.[11] Furthermore, it is easily shown (Hansen 1982) that

[10] Since both the terms $X_{it}\beta_j$ and $(1/2)b_j^2\sigma_t^2$ require interactions with sector affiliation, which is endogenous, it is natural to include some instruments for sector affiliation interacted with those terms in the instrument set. In the estimation we replace $X_{it}\beta_j$ by a skill index discussed below and use experience to proxy for σ_t^2. This leads to adding interactions between the second lag of sector affiliation (the instruments for sector affiliation) and the skill index and experience to the main set of instruments (interactions between sector affiliation at $t - 1$ and $t - 2$).

[11] Equation (13) shows that e_{it} is a relatively complex function of the sector dummies. So e_{it} may be heteroskedastic even if the "structural" error terms ξ_{it}, Z_i, and μ_{it} are homoskedastic. We could allow for heteroskedasticity by using an efficient generalized method of moments (GMM) two-step procedure in which the parameters are first estimated using $M = (V'V)^{-1}$. These consistent but inefficient parameters can then be used to compute a heteroskedasticity-robust variance matrix Σ of $e'V$. Efficient GMM estimates are then obtained using $M = \Sigma^{-1}$. Empirically, however, we had more convergence problems with the efficient GMM than with the

N times the minimized value of S follows a χ^2 distribution with q degrees of freedom, where q is the number of overidentifiying restrictions (the difference between the number of instruments and the number of parameters).[12]

In the linear case where e_{it} is homoskedastic and uncorrelated, NLIV is just the well-known two-stage least squares estimator. One difficulty with NLIV is that since we project a nonlinear function of the model variables and of the parameters (e) into a linear set of instruments (V), the instruments must be chosen in a way that predicts sufficiently well the explanatory right-hand side of equation (13).[13] In addition to the sector histories discussed above, we thus include as instruments a set of interactions between the explanatory variables X_{it} (as summarized by a skill index and years of experience) and the period $t - 2$ dummies for sector affiliation $\{D_{ij,t-2}, j = 1, \dots, J\}$.

In the perfect information case (but still with sorting), where unmeasured ability η_i is observed by labor market participants, the quasi-differenced equation (13) remains the same except that the innovation term $\xi_{i,t-1}$ drops from the error term e_{it}. The remaining endogeneity problem is due to the correlation between $\ln w_{i,t-1}$ and the error component $\mu_{i,t-1}$. In this case, we simply use the full set of interactions between the sector dummies at time t and $t - 1$ as instruments for $\ln w_{i,t-1}$.[14]

III. Data

The data set used in this article is the National Longitudinal Survey of Youth (NLSY). Individuals in the NLSY were between the ages of 14

---

inefficient, but consistent, estimates based on the weighting matrix $M = (V'V)^{-1}$. This mirrors the findings of Altonji and Segal (1996), who found serious small-sample problems with the optimally weighted GMM (or optimal minimum distance) estimator.

[12] In complicated nonlinear models like ours, however, this overidentification test is better thought of as an omnibus specification test than as a standard test of the validity of instruments commonly performed for linear models. These tests should be interpreted with care. We discuss this issue in detail in an earlier version of the article (Gibbons et al. 2002).

[13] See Newey (1990) for more discussion and proposed (nonparametric) solutions to this problem. Note that choosing the functional form or the number of instruments can also be problematic in the linear model (Donald and Newey 2001).

[14] In the absence of learning, either the interactions between sector affiliation at time t and $t - 1$ or at time $t - 1$ and $t - 2$ can be used as instruments. In practice, this choice has little impact on the results since both sets of instruments predict very well the wage (see app. B). Since sector affiliation is exogenous in this model, we do not need to include the additional interaction terms between the skill index, experience, and sector affiliation at time $t - 2$ discussed above. Note that Lemieux (1998) uses an identical strategy to estimate union wage differentials when unmeasured ability is known to all labor market participants but is differently rewarded in the union and nonunion sectors: the interaction between the union status at time t and $t - 1$ is used as an instrument for the lagged wage.

and 21 on January 1, 1979. We use up to 17 yearly observations per worker (from 1979 to 1996).[15] One advantage of the NLSY is that it allows us to follow workers from the time they make their first long-term transition to the labor force.

We use the same sample selection criteria as those used by Farber and Gibbons (1996). We classify individuals as having made a long-term transition to the labor force when they spend at least 3 consecutive years primarily working, following a year spent primarily not working. Someone is classified as primarily working if she or he has worked at least half the weeks since the last interview and averaged at least 30 hours per week during the working weeks. Note that the "last interview" does not necessarily refer to the previous calendar year if an individual had not been interviewed the year before. Self-employed workers are deleted, as are members of the NLSY military subsample. Readers are referred to appendix 1 in Farber and Gibbons (1996) for more details on the criteria used to construct our NLSY sample.

Farber and Gibbons used NLSY data from 1979 to 1991 interview years, whereas our data are through 1996. Except for the longer sampling frame, the only noteworthy difference between our sample and Farber and Gibbons's has to do with union coverage in 1994. For some reason, the question on union coverage in the current or most recent job at interview time (job number 1 in the work history file) was not asked in that year. Although the error was caught and fixed during the field period, many respondents were simply not asked this question even though they should have been. Consequently, the raw data shows a large number of "valid skips."[16] We provide a correction of our own to partially fix this problem and recover quite a few of those missing observations. More precisely, if an individual in 1994 is working for the same employer as the one he worked for in the previous interview, we assign the value of the union coverage dummy for the previous interview year to the current one. If the individual interviewed in 1994 has started working for a new employer since the last interview, we check to see whether she or he is still working for that employer in 1996. If so, we assign the value of the union coverage dummy for that year to the 1994 interview.

From this NLSY sample, we focus on the subsample of observations at which the individual was working at the interview date for at least the previous 3 years. This sample restriction enables us to use the first and second lags of various variables in the estimation, as explained in Section II.D. We exclude workers in agricultural jobs. Since we (later) divide manufacturing into durable and nondurable goods manufacturing, we also

---

[15] There was no interview in 1995.

[16] Personal communication from Steve McClaskie of the Center for Human Resource Research.

exclude a few workers who hold jobs in manufacturing industries that are hard to classify as producing durable as opposed to nondurable goods.[17] We are left with a sample of 35,438 observations on 5,904 workers that satisfy these sample selection criteria.

To summarize the relationship between the wage premia and observable skills, we construct a "skill index" for each worker. We first estimate a flexible log (hourly) wage equation using our sample.[18] The base explanatory variables used in the log wage equation are years of education; education category dummies (dropout, high school graduates, some college, and college degree); (actual) experience; experience squared; dummy variables for race, gender, marital status, and union status; and a set of dummies for year, industry, and occupation. We include sets of pairwise interactions between the education category dummies, gender, and race, as well as interactions between gender and experience, gender and marital status, and race and experience. We then use the estimated coefficients from that equation to predict the wage of each worker. The skill index is the predicted wage based solely on the education and experience of the worker. That is, although characteristics such as occupation, industry, union status, and demographic characteristics are included in the initial wage equation, they are not used to construct the skill index for the worker. We normalize the skill index to have zero mean.

We have also run an expanded skill-index model including the Armed Forces Qualifying Test (AFQT) score as a broad measure of premarket skills. In this case, the predicted wage for the skill index also uses the AFQT score. The detailed skill index regression results are reported in table A1 (in app. A) with and without the AFQT variable included. We present estimates for our models of occupation and industry wage differentials using the skill index without AFQT since there exists substantial evidence that the skills measured by AFQT are not fully observed by firms at labor market entry and represent some of the ability component learned about by firms as workers gain labor market experience (Altonji and Pierret 2001; Lange 2005). The findings are quite similar for models using the skill index with and without AFQT.[19]

IV. Wages and Returns to Skills across Occupations

We believe that the concepts of sorting and comparative advantage are likely to play a more important role for occupations than for industries,

---

[17] These industries are stone, clay, and glass; tobacco manufacturing; leather and leather products; and not specified manufacturing. Workers in these industries represent less than 1% of the full sample.

[18] The wage variable in all estimated models is the hourly wage on the current job at the time of the survey.

[19] Gibbons et al. (2004) present a full set of estimates using the skill index with AFQT.

so we first estimate our models for occupations. As we mention in Section V, other factors such as compensating wage differences and rent-sharing may mask the importance of comparative advantage in the case of industries. Furthermore, our one-factor model is well suited to cases where there is a natural ordering of sectors from least skill sensitive to most skill sensitive. We believe that this ordering is more likely to apply to occupations (e.g., going from operatives to craft workers to managers) than to industries.

A. Occupational Wage Premia without Comparative Advantage

Throughout this article, we divide workers into seven conventional occupation aggregates.[20] In table 1 we report the raw occupation log wage differentials (relative to the service occupation) and the average values of measured skills (education and experience) and other measured characteristics (race, sex, and marital status) by occupation. As is well known, there are large differences across occupations in mean wages and in mean values for education and other characteristics. There is also a strong link between these two variables: the correlation between the raw wage premium and the mean level of education is 0.81 (bottom row of table 1).

The mean skill index for each occupation is reported in column 7 of table 1. In keeping with the positive correlation between the wage premium and mean education, we find that the correlation between the raw wage premium and the mean skill index is 0.96. But the cross-occupation variation in mean log wages in column 1 (SD = 0.181) is almost twice as large as the cross-occupation variation in the skill index in column 7 (SD = 0.099), suggesting that there may be more to the story than just observable skills. In this spirit, column 1 of table 2 reports an OLS regression of the log wage on the skill index and six occupation dummies (operatives and laborers are the base occupation). All the models reported in table 2 also include controls for industry affiliation (nine dummy variables), gender, race, marital status, union status, and a full set of year dummies.

The skill index is highly significant and has a coefficient of one (by construction), but the occupation coefficients remain highly significant, although smaller than the raw wage differentials reported in table 1. Of course, such a regression merely replicates the common finding that the occupation coefficients are significant even after controlling for measured characteristics. We report it as our point of departure.

In this OLS model, the standard deviation of the estimated occupational wage premia is .092. Column 2 of table 2 reports first-differenced estimates

---

[20] Using a more detailed classification does not alter our basic findings and comes at the cost of less precise estimates of the occupation effects. Precision is an issue for the some of the nonlinear instrumental variables models presented below.

Table 1
Average Characteristics by Occupation

	Raw Wage Differences[a] (1)	Years of Education (2)	Years of Experience (3)	Proportion Female (4)	Proportion Nonwhite (5)	Proportion Married (6)	Skill Index (7)	Sample Proportion (8)
Occupation:								
Professionals	.522	15.66	7.38	.534	.136	.583	.1479	.212
Managers	.425	14.38	8.31	.432	.137	.562	.1049	.138
Sales occupations	.388	14.38	7.49	.400	.118	.550	.0828	.052
Clerical occupations	.116	13.17	7.27	.786	.233	.503	−.0907	.203
Craft workers	.290	12.17	7.96	.077	.157	.499	−.0358	.118
Operative and laborers	.090	11.92	7.44	.244	.265	.489	−.0979	.171
Service occupations	.000	12.73	7.10	.532	.310	.424	−.0989	.108
Standard deviation (across occupations)	.181	1.26	.39	.210	.069	.050	.099	
Correlation with raw wage difference	1.00	.81	.53	−.12	−.94	.93	.96	

Note.—This table is based on a sample of 35,438 observations for 5,904 workers. See text for further details.
[a] Mean log wage in the industry relative to service occupations (mean log wage in service occupations is 1.431).

Table 2
Estimates of Occupation Wage Differentials without Comparative Advantage

	No Learning				Learning	
	OLS/All (1)	FD/All (2)	FE/All (3)	FD/New Jobs (4)	FDIV/All (5)	FDIV/New Jobs (6)
Occupation effects:						
Professionals	.242*	.042*	.042*	.095*	.017	.001
	(.007)	(.007)	(.008)	(.016)	(.014)	(.030)
Managers	.238*	.022*	.049*	.052*	−.002	.047
	(.008)	(.007)	(.007)	(.015)	(.013)	(.025)
Sales occupations	.211*	.006	.018	.044*	−.014	.053
	(.010)	(.009)	(.010)	(.018)	(.016)	(.031)
Clerical occupations	.062*	−.003	−.019*	.010	.006	.049*
	(.007)	(.006)	(.007)	(.014)	(.013)	(.025)
Craft workers	.154*	.038*	.038*	.085*	.010	.041*
	(.008)	(.006)	(.007)	(.013)	(.010)	(.021)
Service occupations	.028*	−.020*	−.046*	.000	.009	.022
	(.008)	(.007)	(.008)	(.014)	(.014)	(.023)
Operatives and laborers	.000	.000	.000	.000	.000	.000
Skill index	1.000*	.676*	.866*	.630*	.690*	.663*
	(.011)	(.050)	(.034)	(.092)	(.043)	(.093)
R-squared	.381	.043	.740	.069	.031	.061
Observations	35,438	35,438	35,438	9,198	35,438	9,198
Test of equality of occupation effects (p-value)	.0000	.0001	.0001	.0001	.4695	.1156
Standard deviation of occupation effects	.092	.024	.038	.038	.011	.020
Adjusted standard deviation	.092	.023	.037	.035	. . .	. . .

NOTE.—All specifications also include controls for gender, race, marital status, year effects, industry (nine dummies), and a dummy for collective bargaining coverage. In cols. 5 and 6, the instrumental variables (for changes in occupation) are the full set of interactions between the occupation dummies at time $t-1$ and $t-2$. The adjusted standard deviation is the actual standard deviation of the estimated occupation effects corrected for the sampling variation in these estimates. OLS = ordinary least squares estimation. FD = first-differences estimation. FE = fixed effects estimation. Standard errors are in parentheses.

* Significantly different from zero at the 95% confidence level.

of these premia; their standard deviation falls to .024. Of course, these first-differenced estimates might be attenuated by false transitions. One approach to the false transitions problem is to estimate a fixed effect regression rather than a first-differenced regression.[21] We present fixed-effect estimates in column 3 of table 2; the occupational wage premia have a standard deviation of .038. Another approach to the false transitions problem is to recompute the first-differenced estimates on the subsample of observations in which the worker reports taking a new job (with a new employer). The resulting wage premia (in col. 4) have a standard deviation of .038. In sum, the estimates in columns 3 and 4 are consistent with the view that more than half of the variation in occupational wage premia (after controlling for measurable skills) may be due to unmeasured ability bias, even in our simplest model without comparative advantage or learning.

In Columns 5 and 6 of table 2 we explore the possibility of further bias associated with learning (but not comparative advantage). As described at the end of Section II.B, the problem is that learning about ability may be correlated with the change in sector affiliation (such as where job loss is exogenous but reemployment is not). As suggested in Section II.B, we can use wage, skill, and sector information from period $t - 1$ or earlier to instrument for the change in sector affiliation between periods $t - 1$ and t. In columns 5 and 6 of table 2 we use as instruments the full set of interactions between occupation dummies at times $t - 1$ and $t - 2$. For the full sample (col. 5), none of the individual occupation effects is significant. Furthermore, the standard deviation of the estimated occupation effects is quite small (.011), and we cannot reject that these premia are all zero (p-value $= .47$). The results for the subsample of new jobs (col. 6) are relatively similar. Now some of the estimated occupation effects are individually significant, but we still cannot reject the null hypothesis that all premia are jointly equal to zero (p-value $= .12$).

The estimates of the model for the sample of new jobs are much less precise than when all observations are being used. As we will see in the next section, limiting the analysis to new jobs appears to be a much more efficient way of eliminating false transitions in the case of industries than occupations. The problem is that people can clearly change occupation by being promoted or reassigned to a different task while staying with the same employer. We lose these legitimate changes when we focus on new jobs only. By contrast, it is much more difficult for an employee to change industry while staying with the same employer. This means we

---

[21] Fixed-effect estimates use information from both first differences and longer differences and so are less affected by measurement error than first-difference estimates are (Griliches and Hausman 1986).

should lose little legitimate information by focusing on new jobs in the case of industries.[22]

In sum, our results suggest that accounting for both unmeasured ability and learning eliminates most of the occupational wage premia. The results in columns 2–4 of table 2 indicate that controlling for measured and unmeasured skills explains 80% of the raw standard deviation of wages across occupations (0.181). The remaining premia are no longer significant when learning is accounted for in columns 5 and 6 of table 2, though these results are less precise than in the more standard models of columns 1–4.

B. Occupational Wage Premia and Occupational Skill Premia

Our exploration of occupation wage premia without comparative advantage strongly suggests that learning combined with the sorting of both measured and unmeasured skills accounts for the bulk of occupational wage premia. In this section, we explore the sources of this sorting by adding comparative advantage to the analysis. We indeed find important differences in the returns to measured and unmeasured skills across occupations. This finding suggests caution in interpreting the standard occupational wage premia reported in table 2 (and elsewhere in the literature). In addition, as we describe below, such occupation-specific returns to skill make estimated occupational wage premia difficult to interpret, even after controlling for differences in returns to skills across sectors. As a result, we now shift our focus to these differential returns to skill. In particular, we investigate whether high-skill workers are concentrated in high-return occupations, as our theory suggests.

Table 3 extends our analysis of occupational wage premia in table 2 by reporting not only these premia but also occupation-specific returns to skills. All models reported in table 3 also include the same set of additional controls (gender, race, year dummies, etc.) as in table 2. Column 1 of table 3 reports OLS estimates of the wage premia, while column 2 reports the occupation-specific returns to observable skill. Most of the estimated

---

[22] An alternative way of reducing false transitions is to replace the actual industry at a given point in time with the modal industry over the whole duration of the job (i.e., over all observations of a given employer-employee match). When this imputation procedure is used, industry can only change when workers change jobs. The results obtained using this alternative procedure are very similar to those obtained by simply limiting the sample to new jobs only. Note also that the problem of false transitions has been reduced since 1994 in the NLSY with the introduction of a dependent interviewing procedure for industry and occupation (respondents are first asked whether they have changed industry and occupation since the last interview, while this check was not performed prior to 1994). Given our focus on learning, however, we would lose most of the period during which learning presumably matters by limiting our analysis to the post-1993 period.

Table 3
Estimates of Occupation Wage Differentials with Comparative Advantage

	No Learning				Learning	
	OLS/All		NLIV/All		NLIV/All	
	Main (1)	Interacted (2)	Main (3)	Interacted (4)	Main (5)	Interacted (6)
Occupation effects:						
Professionals	.256*	1.280†	.090*	1.231†	−.029	.924
	(.009)	(.040)	(.023)	(.061)	(.041)	(.082)
Managers	.236*	1.534†	.073*	1.217†	−.008	.973
	(.009)	(.042)	(.024)	(.057)	(.045)	(.085)
Sales occupations	.208*	1.754†	.096*	1.250†	.054	.998
	(.012)	(.055)	(.029)	(.064)	(.055)	(.239)
Clerical occupations	.028*	1.145†	.039	1.127†	−.071	.836†
	(.008)	(.041)	(.022)	(.053)	(.039)	(.073)
Craft workers	.214*	1.284†	.032	1.013	−.098*	.817†
	(.009)	(.050)	(.019)	(.047)	(.033)	(.065)
Service occupations	.028*	1.259†	.053*	1.158†	−.037	.919
	(.010)	(.048)	(.025)	(.054)	(.048)	(.084)
Operatives and laborers	.000	1.000	.000	1.000	.000	1.000
Endogenous variables	. . .		$w1$		$w1, D0, SK \times D0, E \times D0$	
Instrumental variables			$D0 \times D1$		$D1 \times D2, SK \times D2, E \times D2$	
Number of observations	35,438		35,438		35,438	
Test of equality of occupation effects (p-value)	.000		.0016		.0009	
Test of equality of interaction slopes (p-value)	.000		.0001		.0002	
Standard deviation of occupation effects	.105		.027		.053	
Adjusted standard deviation	.105		.012		.029	

NOTE.—All specifications also include controls for gender, race, marital status, year effects, industry (nine dummies), collective bargaining coverage, and (except in cols. 1 and 2) occupation-specific experience effects. The specification in cols. 3–6 also includes the first lag of the occupation dummies interacted with the first lag of experience. In cols. 1 and 2, the interaction terms indicate the effect of measured skills in the occupation relative to operatives and laborers (effect normalized to one for operatives and laborers). In cols. 3–6, the interaction terms indicate the effect of both measured and unmeasured skills in the occupation, again relative to operatives and laborers, but these occupation-specific slopes for measured and unmeasured skills are constrained to be proportional across all occupations. In the endogenous variables, $w1$ stands for the lagged wage and $D0$ stands for contemporaneous values of the occupation dummies $\{D_{ijt}; j = 1, \dots ,6\}$. In the instrumental variables, $D1$ and $D2$ stand for the first and second lag of occupation dummies, SK stands for the skill index, and E stands for experience. The adjusted standard deviation is the actual standard deviation of the estimated occupation effects corrected for the sampling variation in these estimates. Standard errors are in parentheses.
* Estimated coefficient for this main effect is significantly different from zero at the 95% level.
† Estimated coefficient for this interaction is significantly different from one at the 95% level.

returns to skill are quite plausible. For example, all occupations have a significantly larger return to skill than operatives and laborers. Managers and sales occupations have the largest returns to skill, although the returns for professionals may be a bit smaller than expected.

In spite of these significant differences in occupation-specific returns to observable skills (p-value of .00 on the joint test of equality of returns), the associated occupational wage premia are quite similar to those from column 1 of table 2 (which did not allow for occupation-specific returns to skill). For example, the standard deviation of the estimated occupational wage premia is .105—just slightly larger than the .092 in column 1 of table 2. But our analysis in table 2 suggested an important role for un-

measured skills, so we next investigate occupation-specific returns to unobservable skills.

The remaining models reported in columns 3–6 of table 3 allow for occupation-specific returns to both measured and unmeasured skill. In all models, we include (but do not show in the table) a set of interactions between occupation and experience to capture the term $(1/2)b_j^2\sigma_t^2$ in equation (13).[23] We allow returns to measured and unmeasured skill to be different but proportional. In terms of the parameters of the model, this means that $\beta_j = kb_j$ for all occupations j, where k is a proportionality parameter.

In the first model, reported in columns 3 and 4 of table 3, we analyze the model with comparative advantage but without learning, so the only endogenous variable is the lagged wage. In these models we use the full set of interactions between occupational affiliation at time t and $t-1$ as instrumental variables. Relative to the OLS model of columns 1 and 2, two features of the results in columns 3 and 4 are striking. First, the occupational wage premia become much smaller once occupation-specific returns to unmeasured skills are accounted for in the estimation. Second, most of the occupation-specific returns to skill remain significantly different from one (the normalized return to skill for operatives and laborers). Furthermore, the pattern of returns to skill across occupations now shows professionals, managers, and sales occupations with the highest returns.

The joint tests at the bottom of table 3 confirm this pattern of results. The null hypothesis that the occupational returns to skill are all the same can be strongly rejected (p-value $=$ .0001). The null hypothesis that the wage premia are all zero cannot be rejected either, though the p-value is a little higher than in the case of the returns to skill (p-value $=$ .0016).

As a final step, columns 5 and 6 of table 3 report estimates of our richest theoretical model—equation (13), which allows for both comparative advantage and learning, so that both the lagged wage and the current occupation are endogenous. As discussed earlier, we use the full set of interactions between occupational affiliation at time $t-1$ and $t-2$ as instruments (plus interactions between the skill index and occupational affiliation at time $t-1$ and $t-2$). The results reported in columns 5 and 6 are relatively similar to the corresponding model without learning (cols. 3 and 4) for comparisons of the six occupations, excluding the base occupation of operatives and laborers. The joint test that all occupational wage premia are the same can still be strongly rejected (p-value $=$ .0009)

---

[23] Strictly speaking, this term should appear in only the learning model. We include it in all models with unmeasured skills for the sake of comparability across specifications.

as can the null hypothesis that returns to skill are all the same (p-value = .0002).

The individual estimates by occupation are a little harder to interpret because the base group (operative and laborers) has relatively higher returns to skills and a higher occupational wage premium than in the other models (cols. 2 and 4). As a result, all the estimated returns to skill and occupation main effects look puzzlingly small relative to the base occupation. Among occupations other than operative and laborers, however, there is still some heterogeneity in returns to skills. For example, higher wage occupations such as managers or sales exhibit higher returns to skill than a lower wage occupation like clerical occupations.

C. Interpretation

The evidence reported in tables 1–3 strongly suggests that comparative advantage and sorting based on observable and unobservable skills play important roles in explaining raw occupational wage premia. Table 1 shows strong and systematic sorting of the highly skilled into highly paid occupations (correlation coefficient = .96). Perhaps not surprisingly, table 2 shows that controlling for measured and unmeasured skills in conventional ways (OLS and first differences) successively reduces the standard deviation of occupational wage premia from 0.181 to 0.092 and to between 0.024 and 0.038 (depending on the estimator used to control for time-invariant unmeasured skills). The standard deviation of the remaining occupational wage premia remains at .053 once comparative advantage and learning are explicitly accounted for by introducing occupation-specific returns to measured and unmeasured skills (cols. 5 and 6 of table 3). The pattern of occupation-specific returns to skill is strongly consistent with measured skill sorting across occupations with the exception of the surprisingly high relative returns to skills for operatives and laborers in the full model.

The correlation between average measured skills and returns to skill across our seven occupational categories is 0.74 and 0.34 in the models without and with learning, respectively.[24] In terms of the main effects, introducing learning does not change the results substantially (with the exception of operatives and laborers) once comparative advantage is properly accounted for in table 3. One possible explanation for this finding is that though learning about ability may be quite important in the first few years in the labor market, it may not be as important further into workers' careers (Neal 1999). This may explain why learning plays a limited role in our NLSY sample where we have up to 15 years of labor market observations per worker.

---

[24] The correlation rises to 0.65 excluding operatives and laborers in the model with learning.

V. Wages and Returns to Skills across Industries

A substantial literature has established that there are large and persistent wage differentials among industries, even after controlling for a wide variety of worker and job characteristics (Katz 1986; Dickens and Katz 1987; Krueger and Summers 1987, 1988). One possibility is that these interindustry wage differentials largely reflect differences in workers' productive abilities that are not captured by the variables available in standard individual-level data sets. An alternative explanation is that measured interindustry wage differences are "true wage differentials" reflecting compensating differentials, noncompetitive rent-sharing, or efficiency-wage considerations. Vigorous debate has centered on the extent to which industry wage differences reflect competitive factors such as unmeasured ability and compensating differentials (Murphy and Topel 1987, 1990) as opposed to labor market rents and on whether such measured wage differentials potentially may justify certain types of industrial or trade policies (Katz and Summers 1989; Topel 1989).

In this context, our model with comparative advantage and learning can be viewed as a renewed attempt at "explaining" interindustry wage differentials by the systematic allocation of unmeasured skills across industries. For reasons mentioned earlier, we nonetheless expect comparative advantage to play less of a role in explaining sectoral wage differences across industries than across occupations.

A. Industry Wage Premia without Comparative Advantage

We divide workers into 10 conventional industry aggregates.[25] In table 4, we report the raw industry log wage differentials (relative to the retail trade industry) and the average values of measured skills (education and experience) and other measured characteristics (race, sex, and marital status) by industry. Like others, we find large differences across industries in mean wages and in mean values for education and other characteristics. Like Dickens and Katz (1987), we find substantial correlation between these raw wage premia and these mean characteristics. For example, the correlation between the wage premium and the mean level of education is .49. To move beyond individual skill measures such as education, we use the same skill index as in the previous section. The mean skill index for each industry is reported in the final column of table 4. The correlation between the wage premium and the mean skill index is .85. So, at first pass, sorting on observable skill appears to play a slightly smaller role in explaining interindustry wage differences than it did for occupations (when the correlation coefficient was .96). A related point is that that the cross-industry variation

---

[25] Using a more detailed classification does not substantially alter our basic findings and comes at the cost of less precise estimates of the industry effects.

Table 4
Average Characteristics by Industry

	Raw Wage Differences[a] (1)	Years of Education (2)	Years of Experience (3)	Proportion Female (4)	Proportion Nonwhite (5)	Proportion Married (6)	Skill Index (7)	Sample Proportion (8)
Industry:								
Mining and durable manufacturing	.359	13.15	7.76	.303	.153	.577	.0106	.142
Nondurable manufacturing	.203	13.02	7.40	.405	.170	.529	−.0304	.094
Construction	.283	11.99	8.06	.085	.166	.486	−.0457	.062
Transportation, communication, and utilities	.387	13.44	8.09	.333	.253	.529	.0261	.066
Finance, insurance, and real estate	.348	14.10	7.43	.659	.172	.505	.0140	.089
Professional and business services	.397	14.44	7.52	.471	.173	.525	.0628	.084
Personal services	−.032	12.63	7.26	.400	.236	.436	−.0685	.041
Wholesale trade	.222	14.39	7.34	.659	.219	.545	.0239	.216
Retail trade	.000	12.92	7.16	.455	.190	.426	−.0654	.145
Public administration	.392	14.11	8.06	.484	.307	.593	.0494	.062
Standard deviation (across industries)	.151	.78	.34	.160	.047	.051	.045	
Correlation with raw wage difference	1.00	.49	.70	−.02	.04	.80	.85	

NOTE.—This table is based on a sample of 35,438 observations on 5,904 workers. See the text for further details.
[a] Mean log wage in the industry relative to retail trade (mean log wage in the retail trade industry is 1.445).

Table 5
Estimates of Industry Wage Differentials Without Comparative Advantage

	No Learning				Learning	
	OLS/All (1)	FD/All (2)	FE/All (3)	FD/ New Jobs (4)	FDIV/ All (5)	FDIV/ New Jobs (6)
Industry effects:						
Mining and durable						
manufacturing	.285*	.111*	.149*	.165*	.065*	.146*
	(.008)	(.007)	(.008)	(.014)	(.015)	(.023)
Nondurable manufacturing	.184*	.075*	.105*	.108*	.036*	.071*
	(.008)	(.008)	(.009)	(.015)	(.016)	(.024)
Construction	.270*	.135*	.151*	.173*	.103*	.145*
	(.010)	(.009)	(.010)	(.017)	(.020)	(.030)
Transportation, communi-						
cations, and utilities	.282*	.092*	.130*	.144*	.055*	.112*
	(.009)	(.009)	(.010)	(.017)	(.021)	(.029)
Finance, insurance, and						
real estate	.257*	.061*	.111*	.089*	−.017	.061*
	(.008)	(.010)	(.010)	(.016)	(.024)	(.027)
Professional and business						
services	.255*	.064*	.089*	.084*	−.004	.049*
	(.009)	(.008)	(.009)	(.014)	(.015)	(.022)
Personal services	.006	−.012	−.001	−.001*	−.016	.014
	(.011)	(.009)	(.010)	(.017)	(.017)	(.026)
Wholesale trade	.090*	.054*	.067*	.078*	.035*	.066*
	(.007)	(.007)	(.007)	(.012)	(.014)	(.021)
Retail trade	.000	.000	.000	.000	.000	.000
Public administration	.252*	.100*	.149*	.150*	.028	.081*
	(.011)	(.010)	(.011)	(.022)	(.024)	(.039)
Skill index	1.000*	.811*	1.185*	.617*	.686*	.636*
	(.013)	(.050)	(.046)	(.092)	(.043)	(.092)
R-squared	.405	.035	. . .	.072	. . .	. . .
Number of observations	35,438	35,438	35,438	9,198	35,438	9,198
Test of equality of industry						
effects (p-value)	.0000	.0000	.0000	.0000	.0001	.0001
Standard deviation of indus-						
try effects	.108	.040	.054	.059	.037	.047
Adjusted standard deviation	.108	.039	.054	.057	.032	.039

NOTE.—All specifications also include controls for gender, race, marital status, year effects, occupation (six dummies), and collective bargaining coverage. Instrumental variables (for changes in industry affiliation) are the full set of interactions between the industry affiliation dummies at time $t-1$ and $t-2$. The adjusted standard deviation is the actual standard deviation of the estimated industry effects corrected for the sampling variation in these estimates. Standard errors are in parentheses.
[a] Estimated coefficient is significantly different from zero at the 95% level.

in mean log wages (.151) is much larger than the cross-industry variation in the skill index (.045), whereas the cross-occupation variation in the skill index (0.099) represented more than half the cross-occupation variation in wages (0.181).

   Table 5 reports the estimates of the models without comparative advantage. All the models reported in table 5 (and table 6) also include controls for occupational affiliation (six dummy variables), gender, race, marital status, union status, and a full set of year dummies. The results

without learning reported in columns 1–4 are relatively similar to those obtained by Krueger and Summers (1988) and others. For instance, OLS estimates of the interindustry wage differentials in column 1 are large and significant, with a standard deviation of .108. Furthermore, more than half of the standard deviation of the OLS wage premia across industries remains when unmeasured skills are controlled for using fixed effects (.054 in col. 3) or first differences for new jobs (.059 in col. 4). As discussed for occupations, the smaller standard deviation of industry wage premia obtained from first differences for all workers (.040) is likely due to false transitions among workers staying with the same employer.[26]

Columns 5 and 6 of table 5 report first-differenced IV models, to allow for the possibility of learning (but not comparative advantage). The instrumental variables used in this model are the full set of interactions between industry affiliation at time $t-1$ and $t-2$. The standard deviation of the estimated interindustry wage differentials falls somewhat (from about .056 in cols. 3 and 4 to about .042 in cols. 5 and 6). Unlike the case of occupations, however, the null hypothesis of no industry wage premia is still strongly rejected (p-value $=$ .0001), even in these IV estimates.

Other interesting patterns emerge from the comparison of results for industry and occupations. For example, the standard deviation of raw wages differences is smaller for industries than for occupations (in col. 1 of tables 1 and 4), but the standard deviation of the wage premia across sectors is at least as large for industries as for occupation in the standard models without comparative advantage or learning (OLS, first-difference, and fixed-effect estimates in cols. 1–4 of tables 2 and 5). This comparison suggests that the sorting of measured and unmeasured skills across sectors plays more of a role in explaining the raw wage differentials across occupations than across industries.

A more subtle point is that controlling for skills has much more impact for some industries than others. Take the case of two relatively "high-wage" industries, construction and professional and business services (PBS). Despite high wages, construction has relatively low measured skills, while PBS has the highest measured skills of all industries (table 4). The raw log wage differences indicate that PBS pays 0.114 more than construction. Just controlling for measured skills reverses this pattern. The

[26] Our results are also consistent with Krueger and Summers's (1988) finding that first-differenced estimates of the industry wage effects can be significantly biased downward because of misclassification errors in industry affiliation. Standard first-differenced estimates are misspecified when the whole sample is used but well specified for job changers. Since misclassification errors in industry changes are much less likely to occur when a job change is observed than otherwise, we believe misclassification errors are the primary source of misspecification in the first-differenced estimates for the whole sample.

OLS estimates indicate that construction now pays .015 more than PBS (col. 1 of table 5). Controlling for unmeasured skills increases the gap in favor of construction to between 0.062 and 0.089, depending on the estimator being used (cols. 3 and 4).

This differential effect of controlling for skills (even without learning or comparative advantage) suggests that no single theory can likely explain the wage premia for all industries. In sectors like PBS, the systematic sorting of skills that follows from our model of comparative advantage likely accounts for a large share of the premium; in sectors like construction, compensating wage differences and unionization (rent-sharing) are more plausible explanations. We next explore this hypothesis formally by introducing comparative advantage in the estimated models.

B. Industry Wage Premia and Industry Skill Premia

Table 6 extends our analysis by reporting not only the industry wage premia but also industry-specific returns to skills. Columns 1 and 2 report OLS estimates of the wage premia and returns to measurable skills, respectively. As in the case of occupations, there is substantial heterogeneity in the returns to skill across industries. In spite of this heterogeneity in industry-specific returns to skills, controlling for this heterogeneity slightly increases the standard deviation of the estimated industry wage premia from .108 (in col. 1 of table 5) to .109. Roughly speaking, industries with high wage premia tend to exhibit high return to skill, though construction is an important exception.

The standard deviation of industry effects is reduced by about 40% to .061 when industry-specific returns to both measured and unmeasured skills are introduced in columns 3 and 4 of table 6.[27] (The reduction in the adjusted standard deviation of industry effects—to adjust for imprecision in the estimates—is similar when restricting the sample to new jobs in cols. 5 and 6.) Most of the industry wage premia remain quite large and significant in the models when allowing for industry-specific returns to unmeasured skills. This contrasts with our earlier finding (in table 3) of a three-quarters reduction in the magnitude of the standard deviation of occupation effects when allowing occupation-specific returns to measured and unmeasured skills. Such results are consistent with our interpretation that comparative advantage plays a more important role in ex-

---

[27] As in the case of occupations, we constrain returns to measured and unmeasured skills to be proportional in all the models with industry-specific returns to skill. The instrumental variables are also selected in the same fashion as in the models for occupation. In the models without learning, we use interactions between industry affiliation at time t and $t-1$ as instruments. In the models with learning, the instruments used are the interactions between industry affiliation at time $t-1$ and $t-2$ and the interactions of industry affiliation at time $t-2$ with the skill index and experience.

Table 6
Estimates of Interindustry Wage Differentials with Comparative Advantage

	No Learning						Learning			
	OLS/All		NLIV/All		NLIV/New Jobs		NLIV/All		NLIV/New Jobs	
	Main (1)	Interacted (2)	Main (3)	Interacted (4)	Main (5)	Interacted (6)	Main (7)	Interacted (8)	Main (9)	Interacted (10)
Industry effects:										
Mining and durable manufacturing	.285*	1.145†	.182*	1.124†	.260*	1.109	.176*	1.173†	.188*	.979
	(.008)	(.039)	(.021)	(.042)	(.043)	(.082)	(.039)	(.083)	(.060)	(.121)
Nondurable manufacturing	.189*	1.212†	.111*	1.098†	.163*	1.223†	.158*	1.275†	.265*	1.625*
	(.009)	(.044)	(.022)	(.042)	(.048)	(.100)	(.048)	(.100)	(.135)	(.339)
Construction	.254*	.861†	.180*	1.103	.322*	1.368†	.375*	1.804†	.327*	1.329
	(.011)	(.058)	(.030)	(.066)	(.082)	(.183)	(.114)	(.288)	(.135)	(.302)
Transportation, communications, and utilities	.281*	1.017	.156*	1.208†	.219*	1.368†	.182*	1.725†	.212	1.647
	(.009)	(.048)	(.029)	(.062)	(.066)	(.149)	(.080)	(.232)	(.130)	(.339)
Finance, insurance, real estate	.246*	1.320†	.084*	1.217†	.111	1.353†	.053	1.263	.007	1.231
	(.009)	(.043)	(.028)	(.060)	(.058)	(.130)	(.055)	(.136)	(.078)	(.199)
Professional and business services	.241*	1.270†	.086*	1.186†	.109*	1.168†	.093	1.148	.087	1.010
	(.009)	(.043)	(.024)	(.050)	(.044)	(.092)	(.091)	(.093)	(.058)	(.102)
Personal services	−.020	.626†	.010	1.010	.093	1.297	.191*	1.398	.258	1.422
	(.012)	(.061)	(.039)	(.092)	(.100)	(.242)	(.098)	(.223)	(.186)	(.401)

Wholesale trade	.101*	.829†	.091*	1.135†	.079*	1.012	.072*	1.105	.069	.836
	(.007)	(.036)	(.021)	(.045)	(.035)	(.072)	(.036)	(.084)	(.047)	(.089)
Retail trade	.000	1.000	.000	1.000	.000	1.000	.000	1.000	.000	1.000
	...	...	...	...	...	...	...	...	...	...
Public administration	.258*	.922	.147*	1.222†	.181*	1.395	.136	1.468†	−.082	1.034
	(.010)	(.049)	(.030)	(.068)	(.079)	(.204)	(.071)	(.211)	(.085)	(.237)
Endogenous variables	...		$w1$		$w1$		$w1$, D0, $SK \times D0$, $E \times D0$		$w1$, D0, $SK \times D0$, $E \times D0$	
Instrumental variables	...		$D0 \times D1$		$D0 \times D1$		$D1 \times D2$, $SK \times D2$, $E \times D2$		$D1 \times D2$, $SK \times D2$, $E \times D2$	
Number of observations	35,438		35,438		9,198		35,438		9,198	
Test of equality of industry effects (p-value)	.0000		.0000		.0000		.0004		.0048	
Test of equality of industry slopes (p-value)	.0000		.0154		.0234		.0140		.0078	
Standard deviation of industry effects	.109		.061		.090		.098		.129	
Adjusted standard deviation	.109		.055		.066		.067		.075	

Note.—All specifications also include controls for gender, race, marital status, year effects, occupation (six dummies), collective bargaining coverage, and industry-specific experience effects (except in the model of cols. 1–2). The specifications in cols. 3–10 include interactions of the first lag of industry dummies with the first lag of experience. In cols. 1–2, the interaction terms indicate the effect of measured skills in the industry relative to retail trade (normalized to one in retail trade). In cols. 3–10, the interaction terms indicate the effect of both measured and unmeasured skills in the industry, again relative to retail trade, but these industry-specific slopes for measured and unmeasured skills are constrained to be proportional across all industries. In the endogenous variables, $w1$ stands for the lagged wage and $D0$ stands for contemporaneous values of the industry dummies $\{D_{ijt}, j = 1, \dots, 10\}$. In the instrumental variables, $D1$ and $D2$ stand for the first and second lag of industry dummies, SK stands for the skill index, and E stands for experience. The adjusted standard deviation is the actual standard deviation of the estimated industry effects corrected for the sampling variation in these estimates. Standard errors are in parentheses.

* Estimated coefficient for this main effect is significantly different from zero at the 95% level.

† Estimated coefficient for this interaction is significantly different from one at the 95% level.

plaining sectoral wage premia for occupations than for industries. But this is not to say that comparative advantage plays no role in wage and affiliation decisions across industries. For example, comparing column 3 to column 1 of table 6, the three "high-wage" industries that experience the largest decrease in estimated wage premia are finance, insurance, and real estate (FIRE); professional and business services (PBS); and transportation, communications, and utilities. These three industries also happen to have the largest estimated returns to skill in column 4 and relatively high skill levels (table 4).

Broadly speaking, the models reported in the remaining columns of table 6 are qualitatively similar to the model for all workers without learning of columns 3–6. In all cases, the joint test of equality of industry wage premia is strongly rejected. The results for returns to skill are more mixed. Even though the estimates with learning (cols. 7–10) are less precisely estimated, there is still significant heterogeneity in returns to skill.

C. Interpretation

The existing literature on interindustry wage differentials suggests that neither a simple unmeasured ability explanation (in which ability is equally valued in all industries and market perceptions of worker quality are time invariant) nor a pure rent-based explanation appears fully consistent with evidence from longitudinal analyses of the wage changes of industry switchers (Krueger and Summers 1988) or the pre- and postdisplacement wages of workers displaced by plant closings (Gibbons and Katz 1992). These findings have motivated recent work that has focused on econometric approaches for estimating industry wage differentials while accounting for heterogeneous matches between workers (Neal 1995; Kim 1998; Bils and McLaughlin 2001). Our approach is also in this vein.[28]

Our results reinforce the view that a single explanation does not fit all industries. For instance, the industry wage premia in mining, manufacturing, and construction remain large and statistically significant even in our richest model with comparative advantage and learning. By contrast, introducing these two factors essentially eliminates the wage premia in industries such as FIRE and PBS.

VI. Conclusion

We develop a model of wages and sector choices that generalizes the static model of sorting with perfect information to the case in which some skills are unobserved by both the market and the worker. Wage changes and sector mobility arise endogenously as the market and the incumbent

---

[28] A complementary approach focuses on correlations between ability and investments in sector-specific skills (Neal 1998).

firm learn about a worker's skills. We show how this model can be estimated using nonlinear instrumental variables techniques.

We illustrate our theoretical and econometric approach by studying both occupations and industries. Broadly speaking, the results suggest that the measured occupational wage differentials in a cross-section regression are largely due to unmeasured and unobserved worker skills. We find evidence that the sorting of skills into "high-wage" occupations is explained by high returns to skills in these occupations. Although comparative advantage appears to play a fundamental role in occupational wage differences, the role of learning is more limited. One possible explanation for this finding is that though learning may be quite important in the first few years in the labor market, it may not be as important later on.

The results for industries are mixed, which is consistent with the existing literature. Our richest model with comparative advantage and learning explains relatively well the cross-sectional premia in industries like finance, insurance, and real estate and professional and business services. More traditional explanations like compensating differences and rent-sharing seem to be better suited for industries such as mining, manufacturing, and construction.

Appendix A

Table A1
Log Wage Equations Estimated to Construct the Skill Index

Variables Used to Construct the Skill Index	(1)	(2)
AFQT		.272
		(.010)
Years of education	.038	.028
	(.003)	(.003)
Education category:		
High school dropout	−.126	−.046
	(.028)	(.028)
High school graduate	−.140	−.095
	(.018)	(.018)
Some college	−.138	−.124
	(.017)	(.017)
Experience	.050	.047
	(.004)	(.003)
Experience squared (÷100)	−.054	−.058
	(.018)	(.018)
Other control variables:		
Female	−.100	−.090
	(.012)	(.012)
Married	.083	.079
	(.006)	(.005)
Nonwhite	−.071	−.008
	(.017)	(.017)
Union	.168	.176
	(.005)	(.005)

Table A1 (Continued)

Variables Used to Construct the Skill Index	(1)	(2)
Year dummies:	Yes	Yes
Industry dummies	Yes	Yes
Occupation dummies	Yes	Yes
Interaction terms:		
Female × High school dropout	−.001	−.012
	(.018)	(.018)
Female × High school graduate	.002	.005
	(.012)	(.012)
Female × Some college	.021	.037
	(.013)	(.013)
Female × Experience	−.000	−.001
	(.001)	(.001)
Female × Married	−.081	−.076
	(.008)	(.008)
Female × Nonwhite	.021	.021
	(.010)	(.010)
Experience × High school dropout	−.015	−.014
	(.002)	(.002)
Experience × High school graduate	−.007	(−.007)
	(.002)	(.001)
Experience × Some college	−.001	−.001
	(.002)	(.002)
Nonwhite × High school dropout	.087	.050
	(.019)	(.019)
Nonwhite × High school graduate	.001	−.006
	(.014)	(.014)
Nonwhite × Some college	−.004	−.004
	(.015)	(.015)
Nonwhite × Experience	−.002	−.003
	(.002)	(.002)
Constant	1.046	1.060
	(.049)	(.049)
R-squared	.405	.418

Note.—All models are estimated on a sample of 35,438 observations for 5,904 workers. The skill index used in tables 1–6 is constructed by computing a predicted wage from col. 1 using only education (both years of education and education categories) and experience and holding the other variables (female, married, nonwhite, union, year, industry, and occupation) at their average sample values.

Appendix B

Choice of Instruments

This appendix explains the choice of the interaction between the sector affiliation at time $t - 1$ and $t - 2$ as the main set of instruments in the model with comparative advantage and learning. The intuition for this choice is most easily understood using figure 1. First, note that in the most general model with comparative advantage and learning, figure 1 remains as is except that the unmeasured ability η_i is replaced by its expected value conditional on the available information, $m_{i,t-1}$. Consider the case of sectors j and $j - 1$ in figure 1. There are four possible "sector histories" in these two sectors. Workers can either "stay" in sector $j - 1$ or j in both time periods or switch (from j to $j - 1$ or from $j - 1$ to j). Workers with expected ability close to the critical value $v_{j-1}(X)$ are more likely to switch sector than either workers with expected ability clearly lower than $v_{j-1}(X)$ (sector $j - 1$ stayers) or clearly higher than $v_{j-1}(X)$ (sector $j - 1$ stayers). In terms of expected ability and wages, "switchers" currently in sector $j - 1$ should earn more than other workers in sector $j - 1$, while "switchers" currently in sector j should earn less than other workers in sector j.

Tables B1 (occupations) and B2 (industries) show that the data are broadly consistent with this prediction. These tables show the average log wages as a function of current and previous sector affiliations. For example, consider operatives and laborers as sector $j - 1$ and craft workers as sector j. As expected, "stayers" in the craft occupation earn more (0.08) than craft workers who switched from operatives and laborers to craft (−0.09). By contrast, workers who "switch down" from craft to operatives and laborers earn more (−0.11) than "stayers" in the operative and laborers occupation (−0.17).

This example shows how the interaction between sector affiliation at time t and $t - 1$ helps predict wages even after controlling for the current sector affiliation. Similar reasoning can be used to show how the interaction between sector affiliation at time $t - 1$ and $t - 2$ can be used to predict sector affiliation at time t. Recall that like the lagged wage, the current sector affiliation is endogenous in the model with comparative advantage and learning.

Consider, for example, the choice of sector at time t of individuals observed to be in sector j at time $t - 1$. Remember from figure 1 that the expected ability of sector j workers (at time $t - 1$) who were in sector $j - 1$ at time $t - 2$ should be lower (close to $v_{j-1}(X)$) than the expected ability of sector j workers who were also in sector j at time $t - 2$. Consider a positive productivity signal that increases expected ability m_i. Since sector j stayers (at time $t - 1$ and $t - 2$) are closer to the upper critical value $v_j(X)$ than the switchers who just came from sector $j - 1$, the stayers are more likely to move to sector $j + 1$ at time t than the switchers. This shows how the interaction between sector affiliation at time $t - 1$ and $t - 2$ can be used to predict sector affiliation at time t.

Table B1
Average Value of Log Wage in Year t (Relative to the Mean) as a Function of Occupations in Years t and $t-1$

Occupation Categories	First Lag of Occupation Categories							
	Professional	Managers	Sales Occupations	Clerical Occupations	Craft Workers	Operatives and Laborers	Service Occupations	All
Professionals	.30	.36	.26	.07	.18	.02	−.05	.26
Managers	.36	.21	.26	.01	.12	−.16	−.22	.16
Sales occupations	.23	.22	.24	−.09	−.20	−.35	−.45	.13
Clerical occupations	.05	−.05	−.14	−.15	−.10	−.27	−.41	−.15
Craft workers	.15	.13	−.12	−.12	.08	−.09	−.21	.03
Operatives and laborers	.01	−.22	−.25	−.25	−.11	−.17	−.39	−.17
Service occupations	−.06	−.26	−.51	−.45	−.27	−.39	−.24	−.26
All	.27	.16	.14	−.13	.03	−.17	−.25	.00

Table B2
Average Value of Log Wage (Relative to the Mean) in Year t as a Function of Industry in Year t and $t - 1$

Industry Categories	Mining and Durables	Nondurables	Construction	Transportation and Utilities	Finance, Insurance, and Real Estate	Business and Professional Services	Personal Services	Wholesale Trade	Retail Trade	Public Administration	All
						First Lag of Industry Categories					
Mining and durables	.15	.03	−.05	.15	.09	.24	−.10	.05	−.21	.24	.11
Nondurables	.01	−.03	−.22	−.08	.03	.02	−.22	−.08	−.18	−.05	−.05
Construction	−.05	−.12	.09	.06	−.01	−.04	−.15	−.11	−.18	−.13	.03
Transportation and utilities	.17	.08	−.04	.20	.17	.10	−.18	−.09	−.19	.10	.14
Finance, insurance, and real estate	.02	−.12	.00	−.08	.13	.16	−.33	−.08	−.22	.05	.10
Business and professional services	.21	.02	−.11	.03	.16	.25	−.18	.01	−.22	.02	.15
Personal services	−.18	−.41	−.16	−.17	−.40	−.26	−.26	−.34	−.38	−.01	−.28
Wholesale trade	.05	−.12	−.19	−.04	−.11	−.04	−.30	−.00	−.22	.01	−.03
Retail trade	−.29	−.31	−.39	−.23	−.33	−.24	−.31	−.29	−.23	−.25	−.25
Public administration	−.01	.05	−.07	.07	.00	.16	−.05	−.01	−.08	.19	.14
All	.10	−.05	.02	.14	.10	.16	−.25	−.02	−.23	.15	−.00

Table B3
***F*-Test of the Predictive Power of Instruments**

	Lagged Wage	Current Sector (Range of Values)	Number of Excluded Instruments
Occupations	16.27	16.27–38.69	60
Industries	11.46	17.45–26.68	117

NOTE.—As mentioned in the text and the tables, in the case with comparative advantage but no learning we use the sector affiliations at time t and $t-1$ (instead of $t-1$ and $t-2$) as instruments. In this case, the F-statistics are 10.70 and 8.27 for occupations and industries, respectively, which largely exceed the critical values.

In summary, sector histories (interaction between sector dummies at time $t-1$ and $t-2$) should help predict both the wage at $t-1$ and the sector affiliation at time t in the model with comparative advantage and learning. Tables B1 and B2 suggest that they do so in the case of the wage. A more formal test consists of testing the predictive power of the sector histories on the lagged wage and current affiliation after controlling for all other exogenous variables of the model. In a linear model, this test is just the standard F-test of the predictive power of the excluded instruments in the first-stage equation. Testing for the predictive power of instruments is more complicated in a nonlinear context. In the linear model $y = x\beta + e$, the "first stage" consists of projecting all the right-hand-side variables (x) on the set of instruments, where the x variables also happen to be (minus) the derivative (or gradient) of e with respect to the parameters β (since $e = y - x\beta$). In the nonlinear model, the equivalent of the first stage is thus a regression of the gradient of e with respect to the structural parameters. Those derivatives will typically be nonlinear combinations of various x variables. One way to measure the predictive power of the instruments is thus to compute F-tests on the excluded instruments in these "gradient" regressions.

The more intuitive approach we follow is to linearize the gradients as a function of the various explanatory variables. Once this linearization is performed, we can simply compute the usual F-test on the "first-stage" equations for the lagged wage and the current sector affiliation. To be consistent with the estimated models of table 3 and table 6, our excluded instruments are the interactions between sector affiliation at time $t-1$ and $t-2$ as well as interactions of sector affiliation at time $t-2$ with the skill index and experience. The F-statistics are reported in table B3. In all cases, they largely exceed the critical values (at the 95% confident level) of 1.26 for industries and 1.39 for occupations.

References

Altonji, Joseph G., and Charles R. Pierret. 2001. Employer learning and statistical discrimination. *Quarterly Journal of Economics* 116 (February): 313–50.

Altonji, Joseph G., and Lewis Segal. 1996. Small sample bias in GMM estimation of covariance structures. *Journal of Business and Economics Statistics* 14 (July): 353–66.

Bils, Mark, and Kenneth J. McLaughlin. 2001. Inter-industry mobility and the cyclical upgrading of labor. *Journal of Labor Economics* 19 (January): 94–135.

Blackburn, McKinley, and David Neumark. 1992. Unobserved ability, efficiency wages, and interindustry wage differentials. *Quarterly Journal of Economics* 107 (November): 1421–36.

Borjas, George J. 1987. Self-selection and the earnings of immigrants. *American Economic Review* 77 (September): 531–53.

Borjas, George J., Stephen G. Bronars, and Stephen J. Trejo. 1992. Self-selection and internal migration in the United States. *Journal of Urban Economics* 32 (September): 159–85.

Chiappori, Pierre-André, Bernard Salanié, and Julie Valentin. 1999. Early starters versus late beginners. *Journal of Political Economy* 107 (August): 731–60.

DeGroot, Morris. 1970. *Optimal statistical decisions*. New York: McGraw Hill.

Dickens, William T., and Lawrence F. Katz. 1987. Inter-industry wage differences and industry characteristics. In *Unemployment and the structure of labor markets*, ed. Kevin Lang and Jonathan Leonard, 48–89. London: Basil Blackwell.

Donald, Stephen G., and Whitney Newey. 2001. Choosing the number of instruments. *Econometrica* 69 (September): 1161–91.

Farber, Henry S., and Robert Gibbons. 1996. Learning and wage dynamics. *Quarterly Journal of Economics* 111 (November): 1007–47.

Freeman, Richard B., and James Medoff. 1984. *What do unions do?* New York: Basic Books.

Gibbons, Robert, and Lawrence F. Katz. 1992. Does unmeasured ability explain inter-industry wage differences? *Review of Economic Studies* 59 (July): 515–35.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux, and Daniel Parent. 2002. Comparative advantage, learning, and sectoral wage determination. Working Paper no. 8889, National Bureau of Economic Research, Cambridge, MA.

———. 2004. Comparative advantage, learning, and sectoral wage determination. Unpublished manuscript, MIT, Cambridge, MA.

Gibbons, Robert, and Michael Waldman. 1999. A theory of wage and promotion dynamics in internal labor markets. *Quarterly Journal of Economics* 114 (November): 1321–58.

Griliches, Zvi, and Jerry A. Hausman. 1986. Errors in variables in panel data. *Journal of Econometrics* 31, no. 1:93–118.

Hansen, Lars. 1982. Large sample properties of generalized methods of moments estimators. *Econometrica* 50 (July): 1029–54.

Harris, Milton, and Bengt Holmstrom. 1982. A theory of wage dynamics. *Review of Economic Studies* 49 (July): 315–33.

Heckman, James J., and Bo Honoré. 1990. The empirical content of the Roy model. *Econometrica* 58 (September): 1121–49.

Heckman, James J., and Jose Scheinkman. 1987. The importance of bundling in a Gorman-Lancaster model of earnings. *Review of Economic Studies* 54 (April): 243–55.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey Rosen. 1988. Estimating vector autoregressions in panel data. *Econometrica* 56 (November): 1371–96.

Jovanovic, Boyan. 1979. Job matching and the theory of turnover. *Journal of Political Economy* 87 (October): 972–90.

———. 1982. Selection and the evolution of industry. *Econometrica* 50 (May): 649–70.

Katz, Lawrence F. 1986. Efficiency wage theories: A partial evaluation. In *NBER macroeconomics annual*, no. 1, ed. Stanley Fischer, 235–75. Cambridge, MA: MIT Press.

Katz, Lawrence F., and Lawrence H. Summers. 1989. Industry rents and industrial policy. *Brookings papers on economic activity: Microeconomics*, 209–75.

Kim, Dae Il. 1998. Reinterpreting industry premiums: Match specific productivity. *Journal of Labor Economics* 16 (July): 479–504.

Krueger, Alan B., and Lawrence H. Summers. 1987. Reflections on the inter-industry wage structure. In *Unemployment and the structure of labor markets*, ed. Kevin Lang and Jonathan Leonard, 17–47. London: Basil Blackwell.

———. 1988. Efficiency wages and the inter-industry wage structure. *Econometrica* 56 (March): 259–94.

Lange, Fabian. 2005. The returns to schooling and ability during the early career: Evidence on employer learning and post-school investment. Unpublished manuscript, Yale University, New Haven, CT.

Lemieux, Thomas. 1998. Estimating the effects of unions on wage inequality in a panel data model with comparative advantage and nonrandom selection. *Journal of Labor Economics* 16 (April): 261–91.

Lluis, Stéphanie. In this issue. The role of comparative advantage and learning in wage dynamics and intrafirm mobility: Evidence from Germany.

MacDonald, Glenn. 1982. A market equilibrium theory of job assignment and sequential accumulation of information. *American Economic Review* 72 (December): 1038–55.

Mincer, Jacob. 1974. *Schooling, experience, and earnings*. New York: Columbia University Press.

Mincer, Jacob, and Y. Higuchi. 1988. Wage structures and labor turnover in the United States and Japan. *Journal of the Japanese and International Economy* 2 (June): 97–133.

Murphy, Kevin J. 1986. Incentives, learning, and compensation: A theoretical and empirical investigation of managerial labor contracts. *Rand Journal of Economics* 17 (Spring): 59–76.

Murphy, Kevin M., and Robert Topel. 1987. Unemployment, risk, and earnings. In *Unemployment and the structure of labor markets*, ed. Kevin Lang and Jonathan Leonard, 103–40. London: Basil Blackwell.

———. 1990. Efficiency wages reconsidered: Theory and evidence. In *Advances in the theory and measurement of unemployment*, ed. Yoram Weiss and Gideon Fishelson, 204–40. London: Macmillan.

Neal, Derek. 1995. Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics* 13 (October): 653–77.

———. 1998. The link between ability and specialization. *Journal of Human Resources* 33 (Winter): 174–200.

———. 1999. The complexity of job mobility among young males. *Journal of Labor Economics* 17 (April): 237–61.

Neal, Derek, and Sherwin Rosen. 2000. Theories of the distribution of earnings. In *Handbook of income distribution*, ed. Anthony B. Atkinson and François Bourguignon, 1:379–428. Amsterdam: North-Holland.

Newey, Whitney. 1990. Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58 (July): 809–37.

Ross, Stephen A., Paul Taubman, and Michael L. Wachter. 1981. Learning by observing and the distribution of wages. In *Studies in labor markets*, ed. Sherwin Rosen, 359–71. Chicago: University of Chicago Press.

Topel, Robert. 1989. Discussion of Katz and Summers. *Brookings papers on economic activity: Microeconomics*, 283–88.