Unobserved Worker Quality and Inter-Industry Wage Differentials[†]

Suqin Ge
Virginia Tech

João Macieira
U.S. Department of Transportation

May 20, 2023

Abstract

This study quantitatively assesses two alternative explanations for inter-industry wage differentials: worker heterogeneity in the form of unobserved quality and firm heterogeneity in the form of a firm's willingness to pay (WTP) for workers' productive attributes. Building on hedonic models of differentiated product demand, we develop an empirical hedonic model of labor demand and apply a two-stage nonparametric procedure to recover worker and firm heterogeneities. In the first stage we recover unmeasured worker quality by estimating market-specific hedonic wage functions nonparametrically. In the second stage we infer each firm's WTP parameters for worker attributes by using first-order conditions from the demand model. We apply our approach to quantify inter-industry wage differentials on the basis of individual data from the NLSY79 and find that worker quality accounts for approximately two thirds of the inter-industry wage differentials.

Keywords: hedonic models, inter-industry wage differentials, labor quality, wage determination.

JEL Codes: J31, J24, C51, M51

[†]We would like to thank Carlos Madeira, Ken Wolpin, and the seminar and conference participants from various institutions for their valuable comments and suggestions. Contact information: Ge, Department of Economics, Virginia Tech, email: ges@vt.edu; Macieira, U.S. Department of Transportation, Office of the Inspector General, email: Joao.Macieira@oig.dot.gov. The views expressed herein are those of the authors and do not necessarily reflect the positions of the U.S. Department of Transportation Office of the Inspector General or any agency of the U.S. Government.

1 Introduction

Substantial evidence exists on large and persistent wage differentials among industries for workers with the same observed productivity characteristics, such as education and experience (Dickens and Katz, 1987). The (unexplained) inter-industry wage differentials have attracted the attention of economists for decades because these differentials are used to examine the alternative theories of wage determination and the underlying forces of wage structural change.[1] Explanations for inter-industry wage differentials largely fall into two categories. The first one emphasizes the role of worker-specific productive abilities not measured in data (Murphy and Topel, 1987a, 1987b). The second one emphasizes the importance of firm-specific heterogeneity in the form of compensating wage differences (Rosen, 1986), efficiency wage (Katz, 1986; Krueger and Summers, 1988), and rent sharing (Katz and Summers, 1989; Nickell and Wadhwani, 1990; Van Reenen, 1996). Gibbons and Katz (1992) empirically assess both explanations by following a sample of (approximately) exogenously displaced workers but remain agnostic that either explanation alone can fit the empirical evidence on inter-industry wage differentials.

Debate persists over how much observed inter-industry wage differentials can be explained by unobserved worker or firm characteristics. To disentangle simultaneous worker- and firm-level heterogeneity in wage determination, microdata that match the characteristics of firms to those of their workers are preferred (Abowd et al., 1999). Several recent studies (Abowd et al., 2012; Card et al., 2013; Song et al., 2019; Lachowska et al., 2022) have decomposed inter-industry (or between-firm) wage differences into a worker fixed effect and a firm fixed effect by using extensive matched employer–employee panel data.[2] However, such matched employer–employee panels are usually difficult for researchers to access. Moreover, the decomposition of inter-industry wage differences by using a worker fixed effect assumes unobserved worker characteristics to be time-invariant and equally valued by all industries, but this assumption may not hold in practice. For example, if labor quality evolves over time as a result of learning-by-doing, a worker fixed effect cannot fully capture the effects of unmeasured quality on wages.

In this study, we develop an empirical hedonic model of labor demand and apply a two-stage nonparametric procedure to recover unobserved worker and firm heterogeneity. We

[1] Thaler (1989) reviews the debate on whether residual inter-industry wage differentials can emerge from a competitive equilibrium or simply reflect non-competitive forces, such as efficiency wage. Katz and Autor (1999) provide a comprehensive survey on changes in wage structure.

[2] In a related paper, Fox and Smeets (2011) use matched employer–employee panel data to explain productivity dispersion across firms. Recent advances in the estimation of matching games also highlight the importance of unobserved heterogeneities, such as those of firms and workers, on quantifying equilibrium matching outcomes (Fox, 2018; Fox et al., 2018).

model labor demand as an optimal choice of worker attributes. Worker quality is modeled as a worker attribute unobserved by econometricians but valued by employers. First, we nonparametrically recover unobserved worker quality by using an estimator based on Bajari and Benkard (2005), Imbens and Newey (2009), and Norets (2010) without explicit assumptions on supply-side behavior.[3] This estimator exploits the uniqueness of the equilibrium wage function in each labor market and its monotonicity in unobserved attributes to identify worker quality while allowing for quality to be correlated with other observed worker characteristics, such as education and experience. We separate unobserved worker quality from other unobserved factors by exploiting the fact that worker quality is specific to the worker but not to the industry the individual works in. We control for possible selection bias due to worker self-selection into industries using semi-parametric methods (Dahl 2002). We build on recent identification results for related models (Torgovitsky 2015, D'Haultfoeuille and Fevrier 2015) to estimate our model using instrumental variables. Second, we nonparametrically infer firm-specific willingness to pay (WTP) with respect to both observed and unobserved worker attributes by using model results relating WTP and first-order conditions for profit maximization. The WTP of worker characteristics are estimated as random coefficients in a hedonic wage function. Once unobserved worker and firm effects are identified, we can quantitatively assess their importance in accounting for inter-industry wage differentials on the basis of widely available individual data.

We estimate the labor demand model using individual data from the National Longitudinal Survey of Youth 1979 (NLSY79) to explore the importance of worker and firm effects in wage determination. We estimate the model for two different years and seven different industries, and we identify unobserved worker quality in each year and firm WTP for productive characteristics in each market. Our estimates show that the worker effect captured by unobserved worker quality is statistically more important in explaining wages than the firm effect measured by firm WTP. Unmeasured worker quality accounts for approximately two thirds of the inter-industry wage differentials. Although worker quality is persistent, it evolves over time and cannot be captured by a worker fixed effect alone. Observed worker characteristics that are supposed to account for productivity differences typically explain no more than 30 to 40 percent of wage variations across workers. Considerable residual variance suggests differences in unmeasured worker ability: highly skilled workers earn high wages. Our empirical analysis reveals that the percentage of explained wage differentials across workers nearly doubles when log wage regressions on observed worker attributes are

[3] A minimum set of assumptions about the supply side must be in place so that an hedonic equilibrium exists. We illustrate this type of assumptions below. Nonetheless, we observe that such assumptions are quite weak. For example, the researcher does not need to specify if the supply-side behavior is static or dynamic.

augmented by estimated unobserved worker quality.

Using matched employer–employee panel data from France, Abowd, et al. (1999) also find that worker effects are more important than firm effects in explaining inter-industry wage differentials. However, these authors assume both worker and firm effects to be fixed over time. The more recent literature has begun to explore extensions that allow worker and/or firm effects to drift over time. For example, using administrative wage records from Washington State, Lachowska et al. (2022) find firm effects to be highly persistent with an intertemporal correlation of 0.74. Our study complements this literature by estimating time variant worker and firm effects in a hedonic labor demand model using individual data.

This paper builds on the classic hedonic model (Rosen, 1974) and borrows insights from recent work on estimating demand for differentiated products.[4] The literature on demand for differentiated products has been able to estimate heterogeneities in taste for product attributes and in unobserved product quality since the seminal work by Berry et al. (1995).[5] These models usually assume a finite set of discrete choices as the computation becomes intractable when the set of possible choices is too large (McFadden et al., 1987). Recent developments in using hedonic approach to estimate differentiated product demand models have alleviated this concern (e.g., Bajari and Benkard, 2005; Bayer et al., 2007), but they typically focus on one single market. In this paper, we extend the existing approach to allow for data to be drawn from multiple labor markets varying by industry and by time period.

Our paper is related to many applications involving demand for differentiated products and the identification of unobserved product quality, for example, in the markets for computers (Bajari and Benkard, 2005), housing (Bayer, Ferreira and McMillan, 2007; Bajari and Khan 2005), ready-to-eat cereal (Nevo, 2001), imports and exports (Khandelwal, 2010; Amiti and Khandelwal, 2013), and internet service (Krasnokutskaya et al., 2020). Several recent papers also apply models of differentiated products to the labor market. Card et al. (2018) develop a model of wage setting in which workers with idiosyncratic tastes view workplaces as differentiated products to understand firms and labor market inequality. Azar

[4]The classic hedonic model considers a market with a continuum of products and perfect competition and assumes all product characteristics to be perfectly observed. Rosen's estimation strategy is criticized by Brown and Rosen (1982), Epple (1987), and Bartik (1987), who argue that preference estimates are biased because consumers who strongly prefer a product characteristic purchase more of that characteristic. Bajari and Benkard (2005) relax some of these assumptions and propose a hedonic model of demand for differentiated products; this model accounts for unobserved product characteristics and heterogeneous consumers. Ekeland (2010) and Chiappori et al. (2010) show the existence and uniqueness of equilibrium in hedonic markets under weak assumptions. Ekeland et al. (2004) and Heckman et al. (2010) thoroughly discuss identification issues in estimating hedonic models.

[5]Fox et al. (2012) formally establish the identification of differentiated product demand models without supply-side assumptions. Berry and Haile (2014) present identification results for nonparametric models of differentiated product markets using market level data. They show that nonparametric identification primarily relies on the existence of valid instruments.

et al. (2022) estimate a demand model for differentiated job vacancies to measure firm- and market-level labor supply elasticities. They identify a variable representing job attributes observed by workers but not by researchers, such as working conditions or employment-related reputation of the firm. We consider a similar variable in the labor demand model to identify worker quality observed by employers but not by researchers.

This paper is organized as follows. In section 2 we present a hedonic model of labor demand and discuss its properties. In section 3 we outline the estimation methods used to recover unobserved worker quality and employer preferences for worker attributes. In section 4 we describe the data used in our empirical analysis. Section 5 presents and discusses the estimation results. Section 6 concludes and outlines possible extensions for future research. All derivations and auxiliary results can be found in the appendices.

2 A Model of Labor Demand

This section describes a labor demand model for heterogeneous workers where firms (buyers) view workers (sellers) as differentiated products. Consider an economy in which labor markets are indexed by $a = 1, ..., A$. Each market $a = (l, t)$ is located in industry $l = 1, ..., L$ at time $t = 1, ..., T$, and the total number of labor markets $A = L \times T$. Each market has a continuum of job vacancies, denoted by V_a, with positive measure υ_a. Each job vacancy $i \in V_a$ is a single-worker firm, which decides whether to hire a worker to fill the vacancy.[6]

There is a continuum of workers in each labor market a, denoted by Ξ_a, with positive measure μ_a. Each worker $j \in \Xi_a$ is represented by a set of characteristics that potential employers value differently. M characteristics can be observed by both the employer and the researcher. Let $X_{j,a}$ denote a $1 \times M$ vector of worker j's observed characteristics in market a.[7] Examples of observed worker characteristics include education, work experience, and gender. We use a scalar $\xi_{j,t}$ to represent unobserved worker characteristics valued by all employers (regardless of industry) but unobserved by the researcher, such as productive abilities, communication skills, and career ambition. For simplicity, we interpret the variable

[6] The restriction that each firm hires one worker is just for convenience. The hedonic model can be extended for a firm (the buyer) to hire multiple workers with the same characteristics, as discussed in Rosen (1974). In addition, one can also view a "firm" as a collection of vacancies where hiring decisions are made by independent decision makers within the firm. Assuming that the firm splits profit-maximizing decisions and delegates each decision to a "local manager" is also present in other work (e.g. Aguirregabiria and Ho, 2012).

[7] It would also be accurate to index X using time t instead of using market a, as in typical panel data applications. However, this distinction between a and t will be important when discussing identification of unobserved worker quality, as we assume that quality may vary by time but not by industry. For this reason, we index our panel using j for workers and a for observed time-industry combinations unless indexing by time t is pertinent.

$\xi_{j,t}$ as representing worker j's unmeasured quality at time t, which varies over time but does not vary across industries.

Employers are profit maximizers that choose labor input $E_{i,j,a}$ to fill job vacancy i. $E_{i,j,a}$ represents labor efficiency units of worker j, and it measures the different skill levels of labor in terms of different quantities of efficiency unit.[8] The employer's problem is

$$\max_{E_{i,j,a}\in\mathbb{R}_0^+} \pi_{i,a}\left(E_{i,j,a}\right) = R_{i,a}(E_{i,j,a}) - w_{j,a}, \tag{1}$$

where $R_{i,a}(E_{i,j,a})$ is the employer-specific revenue per worker net of non-labor cost and $w_{j,a}$ is the wage rate.[9]

We model a worker's labor efficiency units as a function of his or her characteristics such that $E_{i,j,a} = E_{i,a}(X_{j,a}, \xi_{j,t})$. The employer's decision then becomes a problem of choosing worker attributes to maximize profit on the job vacancy:

$$\max_{X_{j,a},\xi_{j,t}} \pi_{i,a}(X_{j,a}, \xi_{j,t}) = R_{i,a}(X_{j,a}, \xi_{j,t}) - w_{j,a}. \tag{2}$$

Note that the vacancy profit function in (2) is quasi-linear in wage, a key property that facilitates our results as in the related literature on hedonic equilibrium (e.g. Ekeland 2010, Chiappori et al. 2010). In addition, we assume that the employer will leave the vacancy unfilled if no worker generates profits higher than the value of not hiring. The option of not hiring is denoted by $j = 0$.

In the proposed heterogeneous labor demand model, a wage function in each market $a = (l, t)$ maps the set of worker characteristics onto the set of wages. If an equilibrium wage exists for each market a, the structure of our labor demand model yields the following wage function properties under weak conditions: (1) there is one wage for each set of worker characteristics in each market a, and (2) for each market a, the equilibrium wage function increases in unobserved worker quality. The following proposition establishes these results.

Proposition 1 *Suppose that for each market $a = 1, ..., A$, $R_{i,a}(X_{j,a}, \xi_{j,t})$ is (i) Lipschitz continuous in $(X_{j,a}, \xi_{j,t})$ and (ii) strictly increasing in $\xi_{j,t}$ for all employers $i \in V_a$ in market a, then there exists a unique Lipschitz-continuous equilibrium wage function $w_a(X_{j,a}, \xi_{j,t})$ that is strictly increasing in $\xi_{j,t}$ for each market $a = 1, ..., A$.*

The proof is provided in Appendix B. We follow a similar strategy taken by Bajari and Benkard (2005) in their demand model for differentiated products. The wage function is not

[8] Sattinger (1980, pp. 15–20) provides a review and discussion on the efficiency unit assumption.

[9] The employer's problem (1) can be derived from standard firm primitives, such as production function, input and output prices. We illustrate this point in Appendix A.

additively separable a priori because we have limited information about its form. Similar to the results of Ekeland (2010), the uniqueness result of Proposition 1 applies to employed workers and therefore nothing can be said about equilibrium wages for workers not matched to a vacancy.

Given the wage function $w_a(X_{j,a}, \xi_{j,t})$, the firm problem in (2) becomes

$$\max_{X_{j,a}, \xi_{j,t}} \pi_{i,a}(X_{j,a}, \xi_{j,t}) = R_{i,a}(X_{j,a}, \xi_{j,t}) - w_a(X_{j,a}, \xi_{j,t}). \tag{3}$$

Suppose that worker characteristic m, denoted by $x^c_{j,m,a}$, is a continuous variable and that worker j^* maximizes profit for employer i. The following first-order conditions hold:

$$\frac{\partial R_{i,a}(X_{j^*,a}, \xi_{j^*,t})}{\partial x^c_{j,m,a}} = \frac{\partial w_a(X_{j^*,a}, \xi_{j^*,t})}{\partial x^c_{j,m,a}}, \tag{4}$$

$$\frac{\partial R_{i,a}(X_{j^*,a}, \xi_{j^*,t})}{\partial \xi_{j,t}} = \frac{\partial w_a(X_{j^*,a}, \xi_{j^*,t})}{\partial \xi_{j,t}}. \tag{5}$$

Thus, with a firm's optimal labor demand, the value the firm derives from the last unit of each worker characteristic is equal to the implicit price it has to pay for that unit. Otherwise, the firm can increase its profits by employing an alternative worker with a different set of worker attributes.

Some restrictions on the revenue-per-worker function $R_{i,a}(X_{j,a}, \xi_{j,t})$ are required for model identification. We allow each firm to have a unique set of preference parameters in market a, denoted by $\boldsymbol{\beta}_{i,a}$, for its revenue-per-worker function and use the following log-linear specification for the revenue function:

$$R_a(X_{j,a}, \xi_{j,t}; \boldsymbol{\beta}_{i,a}) \equiv \beta_{i,a,0} + \ln(X_{j,a}) \cdot \boldsymbol{\beta}_{i,a,X} + \beta_{i,a,\xi} \ln(\xi_{j,t}). \tag{6}$$

In this specification, each firm i's revenue is linear in the logarithms of worker attributes $(X_{j,a}, \xi_{j,t})$.[10] Coefficients $\boldsymbol{\beta}_{i,a,X}$ and $\beta_{i,a,\xi}$ represent employer i's preference for characteristic vector $X_{j,a}$ and $\xi_{j,t}$, respectively. When the optimal choice is not hiring, all coefficients in the revenue function are equal to zero. Similar specifications are used to estimate preference parameters in hedonic models of demand for differentiated products (Bajari and Benkard, 2005; Bajari and Kahn, 2005). These random coefficient models are considerably more flexible than standard logit or probit models, where preference parameters are assumed to

[10] Without loss of generality and for ease of exposition, we assume that all observed characteristics are strictly positive. The log-linear form in (6) can accommodate binary variables by adding linear functions on the levels of these variables (e.g. as in Bajari and Benkard 2005 and Bajari and Khan 2005). We include binary variables such as gender, race and marital status in our empirical application.

be identical across individuals. Although seemingly arbitrary, the log-linearity assumption can be derived under mild conditions on model primitives.[11] Appendix A shows how the log-linear revenue function can be derived from common specifications of labor efficiency and the production function.

Given the parametric form in (6), the employer's problem in Equation (3) becomes

$$\max_{X_{j,a},\xi_{j,t}} \beta_{i,a,0} + \ln(X_{j,a}) \cdot \boldsymbol{\beta}_{i,a,X} + \beta_{i,a,\xi} \ln(\xi_{j,t}) - w_a(X_{j,a}, \xi_{j,t}). \tag{7}$$

The firm's first-order conditions in Equations (4) and (5) on any continuous characteristic $x^c_{j,m,a}$ and $\xi_{j,t}$ evaluated at the observed optimal choice j^* become

$$\beta_{i,a,x^c_{j,m,a}} = \frac{\partial w_a(X_{j^*,a}, \xi_{j^*,t})}{\partial x^c_{j,m,a}} x^c_{j^*,m,a} = \frac{\partial w_a(X_{j^*,a}, \xi_{j^*,t})}{\partial x^c_{j,m,a}/x^c_{j^*,m,a}}, \tag{8}$$

$$\beta_{i,a,\xi} = \frac{\partial w_a(X_{j^*,a}, \xi_{j^*,t})}{\partial \xi_{j,t}} \xi_{j^*,t} = \frac{\partial w_a(X_{j^*,a}, \xi_{j^*,t})}{\partial \xi_{j,t}/\xi_{j^*,t}}. \tag{9}$$

Therefore, we can interpret parameter vector $\boldsymbol{\beta}_{i,a}$ as firm i's (approximate) marginal WTP for a percentage increase in worker characteristics in market a.

For worker characteristics that take on discrete values we cannot point-identify the coefficients of these characteristics using first-order conditions similar to those in Equation (8).[12] Instead, we can establish bounds for these coefficients by using the condition that firm i's choice of the discrete characteristic observed in the data maximizes profit in Equation (3). For example, suppose that firm i hires worker j^*. Let $\hat{X}_{j^*a}$ and $\bar{X}_{j^*a}$ denote the vectors of observed characteristics with $female = 1$ and $female = 0$, respectively, and all other elements equal the corresponding observed attributes in vector X_{j^*a}. The implicit price faced by employer i for a female worker is then $w_a(\hat{X}_{j^*a}, \xi_{j^*t}) - w_a(\bar{X}_{j^*a}, \xi_{j^*t})$. $\beta_{i,a,,f}$ is denoted as the coefficient for the female dummy in the revenue function. Profit maximization implies that $\beta_{i,a,f} > w_a(\hat{X}_{j^*a}, \xi_{j^*t}) - w_a(\bar{X}_{j^*a}, \xi_{j^*t})$ if worker j^* is female and $\beta_{i,a,f} \leq w_a(\hat{X}_{j^*a}, \xi_{j^*t}) - w_a(\bar{X}_{j^*a}, \xi_{j^*t})$ otherwise. That is, if employer i hires a female worker, then i's WTP for this characteristic exceeds the implicit price for the characteristic.[13]

[11] The proposed functional form is not required for identification, and other parametric specifications may be considered. When we use an alternative linear–in-levels specification, its performance in explaining inter-industry wage differentials is similar to the linear-in-logs specification used in the present study. The linear-in-logs case allows for a clear interpretation of $\boldsymbol{\beta}_i$ as discussed below.

[12] Note, however, that having discrete worker attributes does not undermine the existence of hedonic equilibrium. Ekeland (2010) explicitly allows for this possibility when demonstrating equilibrium existence in hedonic markets where both sellers and buyers have quasi-linear payoffs. Our assumptions for demand and supply meet these conditions.

[13] Bajari and Khan (2005) provide a similar example in the context of their hedonic housing demand model, where similar identification concerns arise. Thus, the lack of point identification of WTP for discrete

As in Bajari and Benkard (2005), Proposition 1 is based on demand-side arguments and implicitly assumes the existence of an equilibrium price function. While further structure on the supply side is not necessary to identify and estimate worker quality and firm heterogeneity, it is important to discuss supply-side assumptions that can guarantee the existence of the hedonic equilibrium. In our model, firms are represented by a measurable continuum of vacancies and that they take the hedonic wage function as given. On the supply side we have a measurable continuum of workers that also take the hedonic wage function as given.[14] Following Ekeland (2010) and Chiappori et al (2010), let each worker $j \in \Xi_a$ in market a maximizes an utility function that is quasi-linear in wage by choosing among job vacancies in the market. The choice of a job vacancy is equivalent to supplying the attributes required by the vacancy. Without loss of generality and consistent with (6), we let firm heterogeneity be summarized by $\boldsymbol{\beta}_{i,a}$. Each worker j solves the following problem

$$\max_{X_{j,a},\xi_{j,t}} U_{j,a}(X_{j,a}, \xi_{j,t}; \boldsymbol{\beta}_{i,a}) = w_a(X_{j,a}, \xi_{j,t}) - C_a(X_{j,a}, \xi_{j,t}; \boldsymbol{\chi}_{j,a}, \boldsymbol{\beta}_{i,a}), \tag{10}$$

where $C(.)$ is a market-specific cost function, and $\chi_{j,a}$ represents worker heterogeneity not valued in job vacancies, such as preferences for job amenities.

For a given wage function $w_a(X, \xi)$, we define the labor demand for productive attributes (X, ξ) in market a by a firm of type $\boldsymbol{\beta}_{i,a}$ as the solution to the firm's problem in (3), denoted by the vector $\Lambda^d(\boldsymbol{\beta}_{i,a})$. We define the labor supply in market a by a worker of type $\boldsymbol{\chi}_{j,a}$ analogously as the solution to the worker's problem in (10), denoted by $\Lambda^s(\boldsymbol{\chi}_{j,a})$. An hedonic equilibrium in market a consists of a wage function $w_a^*(X, \xi)$ such that, for each profile of productive worker attributes (X, ξ), the density of attributes' demand is equal to the density of attributes' supply. Intuitively, given the equilibrium wage function $w_a^*(X, \xi)$, both firms and workers choose optimal (X, ξ), yielding $\Lambda^d(\boldsymbol{\beta}_{i,a})$ and $\Lambda^s(\boldsymbol{\chi}_{j,a})$ for each firm $i \in V_a$ and worker $j \in \Xi_a$, respectively. For each given $(\widetilde{X}, \widetilde{\xi})$, integrating over the measure υ of firms (or, equivalently, the probability density function of $\boldsymbol{\beta}_{i,a}$) on the set $\{\boldsymbol{\beta}_{i,a} : \Lambda^d(\boldsymbol{\beta}_{i,a}) = (\widetilde{X}, \widetilde{\xi})\}$ gives the density of firms demanding workers with attributes $(\widetilde{X}, \widetilde{\xi})$. Similar considerations hold for workers supplying $(\widetilde{X}, \widetilde{\xi})$. The wage equilibrium function must result in market clearing, that is, for each $(\widetilde{X}, \widetilde{\xi})$ that has positive supply and demand densities, the aggregate mass of workers must equal the aggregate mass of vacancies. Ekeland (2010) and Chiappori et al. (2010) provide conditions for the existence and uniqueness of an hedonic equilibrium

attributes is an issue that our framework has in common with other applications of hedonic models.

[14] As discussed in Rosen (1974), the price function in an hedonic equilibrium is defined by the supply of a product with given attributes being equal to the demand of that product. In turn, both supply and demand depend on the entire price function.

under very general conditions.[15]

In addition to allowing for existence and uniqueness of hedonic equilibrium in each market, the quasi-linearity of (3) and (10) in wages also allows for additional assumptions about equilibrium wages by market that facilitate our empirical implementation. If the probability density functions of both firm types $\beta_{i,a}$ and worker types $\chi_{j,a}$ for each given market a can be represented by probability density functions (PDF) conditional on variables representing that market, such as time and industry dummies, then equilibrium wage functions can be parametrized by those variables instead of being indexed by a. This follows from the definition of equilibrium wage function, which requires that, for each given $(\widetilde{X}, \widetilde{\xi})$, the integral of the PDF of firm types in market a choosing $(\widetilde{X}, \widetilde{\xi})$ must equal the integral of the PDF of worker types in market a supplying $(\widetilde{X}, \widetilde{\xi})$.[16] However, we do not pose any structure on how workers sort themselves into each market a prior to match with an employer in that market. In the next section we discuss how model estimation takes into consideration this self-selection of workers into industries in each period t.

3 Estimation of Labor Demand Model

The market-specific wage functions implied by our hedonic model is of the nonseparable form $Y = g(X, \varepsilon)$, where Y is the product price, X is a vector of observed characteristics, and ε is a variable representing unobserved attributes. A large body of literature examines the estimation and identification of nonseparable functions (e.g., Matzkin, 2003; Chesher, 2003; Chernozhukov et al., 2007). Although most estimators proposed in this literature allow for at most one variable in X to be correlated with ε (e.g., Bajari and Benkard, 2005; Imbens and Newey, 2009), our application considers multiple variables in X to be correlated with unobserved attributes in ε. In addition, we face two additional challenges: (i) our unobserved worker attribute of main interest (worker quality) is time- and worker-specific, and (ii) selection bias may be present due to self-selection into industries. We separate worker quality from market-specific regression residuals by integrating industry effects out after applying the selection bias correction of Dahl (2002). We exploit the insight that quality

[15] For an in-depth discussion on the conditions for identification and estimation of hedonic models, see Ekeland et al. (2004), Heckman et al. (2010).

[16] In other words, this corresponds to replacing the PDF of worker types $f(\beta_{i,a})$ with $f(\beta_i | D_I = l, D_P = t)$, where D_I is a variable representing industry and D_P represents time period, so that market $a \equiv (l, t)$ is represented by D_I and D_P. After doing an analogous replacement for the PDF of worker types $\chi_{j,a}$, the equilibrium wage function in market a becomes a function not only of $(\widetilde{X}, \widetilde{\xi})$ but also of D_I and D_P. The quasi-linearity of firm and worker payoffs in each market ensures uniqueness of this function (see Ekeland (2010) for details).

is worker-specific and not a function of the industry an individual works in.[17]

Our estimation strategy proceeds in two stages. In the first stage, we recover unobserved worker quality up to a normalization after estimating a triangular system of simultaneous equations for each market using nonparametric methods. To consider the potential correlation between worker quality and other observed worker characteristics, we use the identification results of Torgovitsky (2015) combined with both an extended version of the estimator of Imbens and Newey (2009) and the selection correction of Dahl (2002).[18] In the second stage, we use the first-order conditions in Equations (8) and (9) to infer firm-specific parameters on their WTP for continuous worker characteristics.

Although we make some parametric assumptions about the model, we use nonparametric methods to estimate model parameters. The first-stage estimation involves estimating functions parametrized by industry and time dummies, but the Normal mixture estimator that we apply is nonparametric as the number of Normal mixtures is determined by a data-based criterion (this is analogous to choosing the bandwidth parameter for nonparametric kernel estimators using a data-based formula). Moreover, our second-stage estimation quantifies the firm-specifc parameters $\beta_{i,a}$ using Local Linear Regression, a nonparametric estimator. This estimator quantifies the wage function derivatives in (8) and (9) using kernel-based regressions to estimate $\beta_{i,a}$ for each firm i in each given market a.

3.1 Estimation of Unobserved Worker Quality

Because unobserved worker quality has no inherent units, we normalize $\xi_{j,t}$ to lie in the interval $[0,1]$ by using a monotonic transformation $F_{\xi,t}(\xi_{j,t})$, where $F_{\xi,t}(.)$ is the cumulative distribution function (CDF) of $\xi_{j,t}$ at period t. If the observed characteristics $X_{j,t}$ are uncorrelated with $\xi_{j,t}$, and data come from a single market, then we can recover the unobserved quality by using estimates of wage CDF conditional on worker characteristics (e.g. as in Bajari and Benkard, 2005). In the context of our labor demand model, however, observable worker characteristics, such as education and experience, are likely correlated with unobserved worker quality. To confront the endogeneity problem, we develop an estimator that allows for multiple endogenous variables, following Imbens and Newey (2009).

Our estimator for unobserved worker quality involves estimation of a triangular system

[17] That is, we apply two different corrections related to industry affiliation. First, we account for possible selection bias, as workers may self-select into industries. This is accomplished by applying the approach of Dahl (2002). Second, we observe that, unlike wage $w_{j,a}$, worker quality $\xi_{j,t}$ depends only on time t, not on market $a=(l,t)$. This is accomplished by integrating out industry affiliation using the Law of Total Probability.

[18] We follow the empirical guidelines for the implementation of Dahl's correction using the results of Bourguignon et. al (2007).

of equations in each market. Let X_0 and X_1 be the sub-vectors of the vector of the observed characteristics such that $X = (X_0, X_1)$.[19] In addition, let $X_0 = (x_{0,1}, .., x_{0,M_0})$ represent the variables in X that may be correlated with unobserved quality ξ, where M_0 denotes the number of endogenous variables in X_0. We assume that the researcher also observes a vector Z of instruments correlated with X_0 but uncorrelated with ξ.[20] Sub-vector X_1 represents the vector of exogenous variables.

In each market a, the observed wage for a worker is determined by

$$w = \tilde{w}_a(X_0, X_1, \delta_a), \tag{11}$$

where $\tilde{w}_a(.)$ is an unknown, market-specific wage function that is strictly increasing in a scalar δ_a for each X.[21] As in Torgovitsky (2015), we also assume that there exist market-specific, reduced-form functions $h_{a,m}$ such that

$$x_{0,m} = h_{a,m}(X_1, Z, \eta_{a,m}),\ m = 1, ..., M_0, \tag{12}$$

for each endogenous regressor $x_{0,m} \in X_0$. $\eta_{a,m}$ is an error term such that $(\delta_a, \eta_{a,1}, ..., \eta_{a,M_0})$ are jointly independent of (X_1, Z), and each $h_{a,m}(.)$ is an unknown function that is strictly increasing in $\eta_{a,m}$. The reduced-form, nonparametric functions $h_{a,m}(.)$ are analog to first-stage linear regressions in two-stage least squares (2SLS) estimation, with the difference that they allow for nonlinearity in the exogenous characteristics X_1, the instruments Z, and the first-stage residual $\eta_{a,m}$. This flexible form for $h_{a,m}(.)$ avoids modelling the endogenous regressors in X_0 as a linear function of X_1 and Z. Avoiding this strong assumption is desirable in applications similar to ours, where agent decisions such as schooling are highly nonlinear in unobservables and neglecting this nonlinearity has empirical consequences (Card 2001, Torgovitsky 2017). Nonetheless, a direct 2SLS estimation of the nonlinear equations in (11) becomes intractable. The control function method outlined below deals with this issue by using the first-stage residual $\eta_{a,m}$ instead of the first-stage regression fit to control for the

[19] To simplify notation, we suppress the individual subscript j and the time subscript t whenever it is possible.

[20] In our empirical application the vector Z has a dimension of $G \geq M_0$, which satisfies the traditional requirement of using at least as many instruments as endogenous variables. For a comprehensive discussion of conditions on Z to identify nonseparable triangular systems, see Torgovitsky (2015) and D'Haultfoeuille and Fevrier (2015).

[21] We use the notation $\tilde{w}_a(.)$ instead of $w_a(.)$ because this is a regression equation where the residual δ_a includes ξ along with other error components. Below we pose an additional monotonicity assumption to separate ξ from this residual. Moreover, an implicit assumption here is that the expected value of w conditional on observing $(X_{j,a}, \xi_{j,t})$ must be equal to the hedonic equilibrium wage function $w_a(X_{j,a}, \xi_{j,t})$. We make use of this assumption when proposing an estimator for firm heterogeneity parameters in the next section.

endogeneity in X_0.[22]

Under the additional assumption that the random variables $w|(X_0 = x_0, Z = z)$ and $X_0|(Z = z)$ are continuously distributed for all x_0 and z, the market-specific equations (11) and (12) form a triangular system that is point identified (Torgovitsky 2015). Nonetheless, we need additional assumptions to deal with two challenges. First, unobserved worker quality needs to be comparable across markets, so that industry-specific effects can be averaged out when computing a worker quality measure. Second, we only observe worker wage and attributes for her chosen job, raising industry selection bias concerns (Heckman 1974, 1979).

We address the first concern by assuming that, for each market $a = (l, t)$, the market wage equations in (11) can be written as a single function that also depends on industry l and time t:

$$\tilde{w}_a(X_0, X_1, \delta_a) = \tilde{w}(X_0, X_1, D_I = l, D_P = t, \delta) \tag{13}$$

where D_I and D_P are dummy variables for industry and time period, respectively. This assumption implies that the shapes of the wage functions $\tilde{w}_a(.)$ can be controlled with industry and time dummies.[23] In addition, it allows the comparison of the wage rates of any two workers with similar attributes that are employed in different industries or time periods.[24]

We also simplify (12) by assuming, for each market $a = (l, t)$, that they can be written as a single function that also depends on time t:

$$h_{a,m}(X_1, Z, \eta_{a,m}) = h_m(X_1, Z, D_P = t, \eta_m),\ m = 1, ..., M_0.^{25} \tag{14}$$

[22] We note that 2SLS estimation and control function estimation are equivalent when both the main equation and the first-stage regression are linear (Imbens and Wooldridge, 2007). The control function approach is convenient when the equations involved are not linear in the endogenous regressors.

[23] We thank a referee for pointing out that this assumption is comparable to the homogeneization approach often adopted in the empirical auction literature (e.g. Haile, Hong and Shum, 2003). The homogeneization approach solves the problem of observing comparable but not identical goods sold at different auctions by assuming that each bidder's expected valuation of the good is separable in the observed attributes of both the good and the auction. This assumption allows bids to become comparable across auctions. In our aplication, this assumption makes worker wages conditional on observed attributes comparable across markets by also conditioning on market attributes (i.e. industry and time dummies). We observe that the employer's valuation of a worker is controlled by the revenue function in (2). This function is typically assumed separable, as in (6). Appendix A illustrates how the revenue function can be derived from firm primitives.

[24] Another implication of this property is that the quantiles of the wage distribution conditional on worker attributes become comparable across industries and time periods. For example, we can compare the median wage of two industries in a given year for a worker with certain attributes by modifying the values in the industry dummies in D_I. This property will play an important role below when computing worker quality, as it involves integrating industry effects across conditional wage distributions.

[25] Unlike the wage equation in (13), we do not simplify the reduced-form equations in (12) to depend on industry dummies. Industry dummies may be endogenous regressors due to potential worker self-selection into industries and therefore should not be used for reduced-form IV regressions. Nonetheless, the dependence of endogenous variables on exogenous worker attributes X_1, instruments Z, time dummies D_P and unobservables η_m is consistent with dynamic models for worker decisions on market attributes such as schooling

Substituting (13) and (14) for the market-specific equations (11) and (12) results in a single triangular system where the unobservable scalar δ and the error term vector η are not market-specific:

$$
\begin{aligned}
w &= \tilde{w}(X_0, X_1, D_I, D_P, \delta) && (15)\\
x_{0,m} &= h_m(X_1, Z, D_P, \eta_m),\ m = 1, ..., M_0. && (16)
\end{aligned}
$$

We address the second concern by applying the semi-parametric method of Dahl (2002) to control for industry self-selection. This method extends the selection bias correction of Heckman (1974) to the case of multinomial self-selection. In the original approach of Heckman (1974), the researcher observes wages only for workers who chose to participate in the labor market. Heckman (1974) proposes a two-step correction for self-selection into the labor market by first estimating a probit for labor market participation and then augment the original wage regression with an Inverse Mills Ratio term derived from the participation probit. Dahl (2002) extends this approach for cases where the self-selection is not binary by replacing the Inverse Mills Ratio with a function of the selection probabilities derived from a multinomial logit. Applying Dahl's approach to our model requires the additional assumption that industry choice probabilities can be modelled using a multinomial logit model and a vector of additional instruments Z_S impacting industry affiliation that does not overlap with X_1 and Z. Let $\tilde{\eta}_l$ denote the probability that a worker's job is affiliated to industry l.[26] In addition, let X_1 denote a vector of exogenous variables. Then the industry choice probability is

$$
\tilde{\eta}_l = \frac{\exp\left(\theta^S_{0,l} + X_1 \cdot \boldsymbol{\theta}^S_{X_1,l} + Z_S \cdot \boldsymbol{\theta}^S_{Z_S,l} + D_P \cdot \boldsymbol{\theta}^S_{D_P,l}\right)}{1 + \sum_{q=1}^{L} \exp\left(\theta^S_{0,q} + X_1 \cdot \boldsymbol{\theta}^S_{X_1,q} + Z_S \cdot \boldsymbol{\theta}^S_{Z_S,q} + D_P \cdot \boldsymbol{\theta}^S_{D_P,q}\right)},\ l = 1, ..., L. \qquad (17)
$$

We assume that $(\delta, \eta_1, ..., \eta_{M_0})$ are jointly independent of (X_1, Z, D_P), each $h_m(.)$ is an unknown function that is strictly increasing in η_m, and the controls for the possible endogeneity of D_I are identified by the multinomial logit system in (17). Furthermore, the unobserved scalar δ may mix worker quality ξ with other unmeasured factors net of worker-specific effects, denoted by ϵ. We assume that $\delta = s(\xi, \epsilon)$, where $s(., .)$ is a strictly increasing function in the first argument.[27]

(Card 1995, 2001).

[26] We use the notation $\tilde{\eta}_l$ because these probabilities are sufficient statistics to control for industry self-selection probability in a way similar to the reduced-form residuals $\eta_{a,m}$ in (12) serving as control functions for endogenous regressors. While both serve the purpose of controlling for endogeneity, they are derived under a different set of assumptions. For more details, see Dahl (2002) and Bourguignon et al. (2007).

[27] This assumption is intuitive by Proposition 1, where we establish that the equilibrium wage in each market is strictly increasing in unobserved worker quality. For example, $\delta = \xi + \epsilon$ corresponds to the

A control variable is a variable conditional on which X_0 and δ are independent. The first step of our estimation builds on estimators conditional on control variables as an alternative to traditional IV estimators to deal with endogenous regressors (e.g., Blundell and Powell, 2003, 2004; Imbens and Newey, 2009; Bajari and Benkard, 2005; Petrin and Train, 2010; Farre, Klein and Vella, 2013). Theorem 1 of Imbens and Newey (2009) shows that when $M_0 = 1$, the researcher can form a control variable using the CDF of the single endogenous regressor x_{01} conditional on X_1 and Z. We consider an extended setup for an arbitrary number of endogenous regressors. In what follows, we denote the vector of errors defined in (16) by $\boldsymbol{\eta} = (\eta_1, ..., \eta_{M_0})$. The following proposition shows that $\boldsymbol{\eta}$ is a vector of control variables that can be used to estimate unobserved worker quality ξ.

Proposition 2 *Let $F_{x_{0m}|X_1,Z,D_P}(.|.)$ denote the CDF of the endogenous characteristic $x_{0,m}$ conditional on the vector of exogenous characteristics X_1, an instrument set Z, and dummy variables D_P controlling for time period t. If each η_m is normalized to lie in the interval $[0,1]$ such that for each $m = 1, ..., M_0$, $\eta_m = F_{x_{0m}|X_1,Z,D_P}(x_{0m}|X_1, Z, D_P)$, then X and ξ are independent conditional on both $\boldsymbol{\eta} = (\eta_1, ..., \eta_{M_0})$ and Dahl's (2002) controls $\tilde{\eta} = (\tilde{\eta}_1, ..., \tilde{\eta}_L)$. Moreover, unobserved worker quality at period t is given by*

$$\xi = \int_{\boldsymbol{\eta}\in[\mathbf{0},\mathbf{1}]^{M_0}} \int_{\tilde{\boldsymbol{\eta}}\in[\mathbf{0},\mathbf{1}]^{L}} \left\{ \sum_{l=1}^{L} F_{w|X,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I,D_P}(w|X, \eta, \tilde{\eta}, D_I = l, D_P = t) \Pr\left(D_I = l|X, \eta, \tilde{\eta}, D_P = t\right) \right\} dG(\eta, \tilde{\eta}), \tag{18}$$

where $\mathbf{G}(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$ is the joint CDF of all controls, and $\Pr(D_I = l|X, \boldsymbol{\eta}, \tilde{\boldsymbol{\eta}}, D_P = t)$ is the probability for a worker with characteristics X and control variables $(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$ to work in industry l at time t.

Our proof (Appendix C) extends Theorem 1 of Imbens and Newey (2009) and Theorem 4 of Bajari and Benkard (2005) by allowing for multiple endogenous characteristics.[28] Note that Equation (18) involves taking expectations over all industries $l = 1, ..., L$ given individuals'

additive separable case where market-specific error adds to worker quality. The multiplicatively separable case where $f(\xi, \epsilon) = f_1(\xi) \times f_2(\epsilon)$ and $f_2(\epsilon) \neq 0$ for every ϵ would also work if $f_1(\xi)$ is strictly monotone. Berry and Haile (2018) extend the identification results of Matzkin (2008) for general additive separable structures and document their applicability in many settings, including differentiated product markets. We discuss identification of unobserved worker quality below.

[28] We note that allowing for multiple endogenous characteristics in X_0 is not central to the significance of Proposition 2. To the best of our knowledge, allowing for multiple endogenous characteristics was not considered in the literature but it is conceptually straightforward. The significance of Proposition 2 lies on identifying unobserved worker quality out of estimable functions derived from data representing multiple markets. Intuitively, this is achieved by exploiting the worker-industry-time structure of our panel data by conditioning on a specific time period t and then integrating industry effects out, resulting in a worker- and time-specific quantity net of other effects.

observed attributes X and control values $(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$ at time period t.[29] Thus, ξ is a highly nonlinear function that is unconditional on industry affiliation, reflecting unobserved worker attributes valued in all industries.

Unobserved worker quality can be recovered in four steps empirically. First, for each endogenous variable indexed by $m = 1, ..., M_0$, we estimate the values of η_m by using an empirical analog of $F_{x_{0m}|X_1,Z,D_P}(.|.)$. We also estimate Dahl's (2002) controls $\tilde{\eta}$ in this step by estimating the multinomial logit model defined in (17). Second, we use the recovered series $(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$ to nonparametrically estimate $F_{w|X,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I,D_P}(.|.)$, the integrand function in Equation (18). Third, we estimate $\Pr(D_I|X, \boldsymbol{\eta}, \tilde{\boldsymbol{\eta}}, D_P)$ by using proportions of workers across industries conditional on worker characteristics X, the recovered controls, and the time dummies D_P. Fourth, the integrand is estimated by integrating $(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$ out. This integration can be done numerically by averaging across all the observations of estimated series $(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$.

Several nonparametric methods, such as the kernel method and series estimators, have been proposed to estimate conditional CDFs. Imbens and Newey (2009) find that series estimators are preferable in empirical frameworks similar to ours. Among series estimators, mixtures of normal distributions are frequently used nonparametric estimators (e.g., Bajari, et al., 2007; Bajari et al., 2011) because of their desirable approximation and consistency properties (e.g., Norets, 2010). We use this type of estimator because it fits the data well and is computationally more tractable for the numeric integration in Equation (18) than other methods. We discuss these estimators and their merits for numeric integration in Appendix D.

We can use the empirical analogs of the conditional CDFs to estimate the unobserved quality of each worker j at period t by using Equation (18):

$$\hat{\xi}_{jt} = \int\limits_{\boldsymbol{\eta}\in[\mathbf{0},\mathbf{1}]^{M_0}} \int\limits_{\tilde{\boldsymbol{\eta}}\in[\mathbf{0},\mathbf{1}]^{L}} \left\{ \sum_{l=1}^{L} \hat{F}_{w|X,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I,D_P}(w_{j,a}|X_{j,a}, \eta, \tilde{\eta}, D_I = l, D_P = t; \hat{\theta}_w) \lambda_l(X_{j,a}, \eta, \tilde{\eta}, D_P = t; \hat{\theta}_{\lambda,t}) \right\} d\hat{G}(\eta, \tilde{\eta}) \tag{19}$$

where $\hat{\mathbf{G}}(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$ represents the empirical analog of $\mathbf{G}(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$.[30]

[29] Observe that this probability of industry affiliation, $\Pr(D_I = l|X, \eta, \tilde{\eta}, D_P = t)$ is different from $\tilde{\eta}_l$ defined in (17). This is because $\tilde{\eta}_l$ is a function of (X_1, Z_S, D_P) whereas this extended probability of industry affiliation depends on $(X, \eta, \tilde{\eta}, D_P)$. We revisit this distinction in Appendix D when discussing estimation details.

[30] We use the notation $\hat{G}(\eta, \tilde{\eta})$ to point that an empirical analog of $G(\eta, \tilde{\eta})$ is needed to estimate quality. However, averaging over the estimated series for $(\eta, \tilde{\eta})$ or a random sample of this series will do this integration without the need for estimating $G(\eta, \tilde{\eta})$ directly.

3.2 Estimation of Firm WTP Parameters

The labor demand problem described in Equation (3) is characterized by the revenue-per-worker function $R_{i,a}(X_{j,a}, \xi_{j,t})$. As discussed in the previous section, we consider a log-linear function for $R_a(X_{j,a}, \xi_{j,t}; \boldsymbol{\beta}_{i,a})$ (Equation 6). Under this model specification, Equation (8) suggests that if we recover an estimate of $\partial w_a(X_{j^*a}, \xi_{j^*t})/\partial x^c_{j,m,a}$, then we can learn a firm's WTP for worker characteristic m. As we observe each worker's characteristics in our data, we can flexibly estimate $\partial w_a(X_{j^*a}, \xi_{j^*t})/\partial x^c_{j,m,a}$ by using nonparametric methods. After we recover unobserved worker quality, we can also estimate a firm's WTP for unobserved quality based on $\partial w_a(X_{j^*a}, \xi_{j^*t})/\partial \xi_{jt}$, following Equation (9).

A practical, flexible way to quantify wage function derivatives at each point in data is to apply local linear regression methods to data on wages, observed worker attributes, and unobserved quality estimates. Bajari and Khan (2005) use this approach to estimate a hedonic price function in the housing market and quantify derivatives of the pricing function. However, two important differences are observed. First, Bajari and Khan (2005) assume that ξ is independent of all observed characteristics X. Although this assumption is acceptable in their housing demand model, it is unreasonable for our application because of endogeneity concerns about schooling and experience. Second, their direct application of local linear regression to housing data does not separate the derivative $\partial w_a(X_{j^*a}, \xi_{j^*t})/\partial \xi_{jt}$ from ξ_{jt}. We separate the two values by first quantifying unobserved worker quality through the methods described above and then treating the estimated $\xi_{j,t}$ as an extra regressor for local linear regression.

In what follows, we follow the exposition of Fan and Gijbels (1996) on the multivariate local linear regression estimator adapted to our notation.[31] We are interested in estimating the expected wage in market a conditional on observing (X, ξ) and its derivatives with respect to (X, ξ). By taking expectations on equation (11), this conditional expectation equals $w_a(X, \xi)$. The conventional Nadaraya-Watson kernel estimator applied to a sample of observations on (w, X, ξ) from market a would estimate $w_a(X, \xi)$ for each observation, but it would not quantify the derivatives of this function needed to compute empirical analogs of equations (8) and (9). Consider a first-order Taylor expansion of $w_a(X, \xi)$ in a neighborhood of a data point $(X_{j^*,a}, \xi_{j^*,a})$ for observation j^*

$$w_a(X, \xi) \approx b_{j^*,a,0} + b_{j^*,a,1}(x_1 - x_{j^*,1,a}) + ... + b_{j^*,a,M}(x_M - x_{j^*,M,a}) + b_{j^*,a,\xi}(\xi - \xi_{j^*,a}) \quad (20)$$

[31] For detailed discussions, see Fan and Gijbels (1996). The multivariate case followed here is presented in section 7.8.1., pp. 297-301. Nonetheless, Chapter 3 in this reference discusses the advantages of local linear regression as a generalization of the Nadaraya-Watson kernel that also quantifies derivatives at specific data points.

where

$$
\begin{aligned}
b_{j^*,a,0} &= w(X_{j^*,a}, \xi_{j^*,a}), & (21)\\
b_{j^*,a,m} &= \frac{\partial w_a}{\partial x_m}(X_{j^*,a}, \xi_{j^*,a}), & (22)\\
b_{j^*,a,\xi} &= \frac{\partial w_a}{\partial \xi}(X_{j^*,a}, \xi_{j^*,a}). & (23)
\end{aligned}
$$

Denoting the number of observations for market a by J_a, and given the values for observation j^* in that market, the local linear regression estimator minimizes

$$\sum_{i=1}^{J_a} \left(w_{i,a} - b_{j^*,a,0} - \sum_{m=1}^{M} b_{j^*,a,m}(x_{i,m,a} - x_{j^*,m,a}) - b_{j^*,a,\xi}(\xi_{i,a} - \xi_{j^*,a}) \right)^2 K_{\mathbf{H}_a}(\psi_{i,j^*,a}) \quad (24)$$

where $K_{\mathbf{H}_a}(.)$ is a multivariate kernel function with smoothing parameter matrix H_a, and $\psi_{i,j^*,a}$ is a vector stacking differences between data point i and data point j, that is,

$$\psi_{i,j^*,a} \equiv [(x_{i,1,a} - x_{j^*,1,a}), ..., (x_{i,M,a} - x_{j^*,M,a}), (\xi_{i,a} - \xi_{j^*,a})]. \quad (25)$$

Fan and Gijbels (1996) provide a formula for the coefficients in (24) for each observation j^*. The $J_a \times 1$ vector of all observed wages in market a is denoted by $\mathbf{w}_a$, and the vector that stacks all coefficients is denoted by $\mathbf{b}_{j^*,a}$, which is solved according to

$$\mathbf{b}_{j^*,a} = \left(\Psi_{j^*,a}^T \Omega_{j^*,a} \Psi_{j^*,a}\right)^{-1} \Psi_{j^*,a}^T \Omega_{j^*,a} \mathbf{w}_a, \quad (26)$$

where $\Psi_{j^*,a}$ and $\Omega_{j^*,a}$ are matrices defined as

$$
\begin{aligned}
\Psi_{j^*,a} &= \begin{bmatrix}\mathbf{1} & \boldsymbol{\psi}_{j^*,a}\end{bmatrix} = \begin{bmatrix} 1 & (x_{1,1,a} - x_{j^*,1,a}) & ... & (x_{1,M,a} - x_{j^*,M,a}) & (\xi_{1,a} - \xi_{j^*,a}) \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & (x_{J_a,1,a} - x_{j^*,1,a}) & ... & (x_{J_a,M,a} - x_{j^*,M,a}) & (\xi_{J_a,a} - \xi_{j^*,a}) \end{bmatrix} & (27)\\
\Omega_{j^*,a} &= \mathbf{diag}\left(K_{\mathbf{H}_a}(\boldsymbol{\psi}_{j^*,a})\right), & (28)
\end{aligned}
$$

where $K_{\mathbf{H}_a}(\psi_{j^*,a})$ is a vector stacking the values of $K_{\mathbf{H}_a}(\psi_{i,j^*,a})$ for all observations $i = 1, ..., J_a$. We choose $K_{\mathbf{H}_a}$ to be the multivariate standard normal density of dimension $M+1$.

Fan and Gijbels (1996) provide asymptotically optimal methods for bandwidth matrix choice. However, these approaches may be unreliable for applications that use several covariates, such as ours and Bajari and Khan (2005). In addition, the number of observations in our data for some markets is not large, raising precision concerns. We deal with these concerns by first computing an optimal bandwidth matrix and then imposing intuitive shape restrictions to our nonparametric estimator in (26).[32] Namely, we impose that expected wages are non-negative and that the derivatives of wages with respect to quality, schooling and experience are also non-negative. We adjust (26) using the framework proposed by Du, Parmeter and Racine (2013). For nonparametric regression estimators that can be written as a matrix product of the form $A(z) \times w$ (e.g., the Nadaraya-Watson kernel estimator, local linear regression estimators such as (26)), the extended estimator is $A(z) \times (w. * p)$, where $p = (p_1, ..., p_{J_a})$ is a $J_a \times 1$ vector of parameters and the operator ".*" represents the Hadamard matrix element-by-element product. For each market $a = 1, ..., A$, the parameter vector $\hat{\mathbf{p}}_a$ solves the following quadratic problem:

$$\begin{aligned} & \hat{\mathbf{p}}_a = \underset{p_1,...,p_{J_a}}{\arg\min} \textstyle\sum_{j=1}^{J_a} \left(1/J_a - p_j\right)^2 \\ s.t. \quad & \textstyle\sum_{j=1}^{J_a} p_j = 1 \\ & B_s(\boldsymbol{\psi}_{j^*,a}, \mathbf{w}_a) \times \mathbf{p} \geq 0, \end{aligned} \tag{29}$$

where $B(\boldsymbol{\psi}_{j^*,a}, \mathbf{w}_a) = ([\left(\Psi_{j^*,a}^T \Omega_{j^*,a} \Psi_{j^*,a}\right)^{-1} \Psi_{j^*,a}^T \Omega_{j^*,a}]^T). * \mathbf{w}_a$, and $B_s(.)$ denotes the specific rows of $B(\boldsymbol{\psi}_{j^*,a}, \mathbf{w}_a)$ that we want to restrict. Our adjusted estimator for the coefficients in (24) is $\hat{b}_{j^*,a} = B(\boldsymbol{\psi}_{j^*,a}, \mathbf{w}_a) \times \hat{\mathbf{p}}_a$. In addition to proving consistency, Du, Parmeter and Racine (2013) show that $B(\boldsymbol{\psi}_{j^*,a}, \mathbf{w}_a) \times \mathbf{p}$ is equivalent to (26) when $p_j = 1/J_a$, $j = 1, ..., J_a$. The quadratic program in is a restricted least squares problem that can be solved using the cyclic projection algorithm of Dykstra (1983) using the R package Dykstra.

According to the first-order conditions in (8) and (9), each firm's preference parameter for a continuous attribute must equal to the product of that attribute's value and the derivative of the wage function for that attribute. Therefore, for each observation j^* in market a, our estimate for firm quality preference parameter $\beta_{i,a,\xi}$ is the product of the estimated quality value for that observation, $\xi_{j^*,t}$, and the corresponding wage function derivative in $\hat{b}_{j^*,a,\xi}$. We obtain firm preference parameters for education and experience analogously. Bajari and Khan (2005) provide an exact formula for the marginal WTP from increasing an attribute x from an initial value x_0 to x_1 using the estimated preference parameters (while keeping all other attributes constant). The formula for marginal WTP under log-linear specification is $\beta_{i,a,x} \cdot (\ln x_1 - \ln x_0)$. As in that reference, we can only identify variations in WTP, which

[32] We compute the optimal bandwidth matrix that minimizes the asymptotic mean integrated squared error (AMISE) for the joint density of $\psi_{j^*,a}$ using the R package *ks* available from the CRAN project website (Duong 2007). As in other practical applications of local linear regression with several covariates, the bandwidth matrix H is selected by inspection of the estimates. Consistent with the recommendations in Duong (2007), the bandwidth matrices obtained via the Smoothed Cross Validation (SCV) of Hall, et al. (1992) was our final choice. We tried alternative criteria for bandwidth matrices, such as Normal Scale (HNS option in ks package based), with no significant changes in our results.

depend on the firm's initial choice x_0 by construction.[33] Identifying the levels of firms' WTP at each combination of worker attributes is beyond the scope of this paper and it would require additional structural assumptions on the firm's problem.

A firm's WTP for a discrete worker characteristic is not point-identified even if the researcher assumes a parametric distribution. This lack of point identification precludes the usage of firm WTP for discrete attributes in our statistical analysis of inter-industry wage differentials. Thus, we focus on firm WTP on continuous attributes, including education, work experience, and unobserved worker quality.

3.3 Identification

In this subsection we discuss the identification of the unobserved worker quality and firm WTP parameters. Our estimator for unobserved worker quality in (19) is a weighted sum of estimates obtained from each industry at a given period t. Thus, we start by discussing the identification of the unobserved worker quality conditional on a given market defined by industry and time period. For ease of exposition, we first consider the case where all observed worker attributes X are exogenous.

When X is independent of ξ, unobserved worker quality is identified by variations in wages while keeping observed worker attributes constant. Intuitively, a professional painter is paid a higher hourly wage in the construction industry than a handyman with the same vector X because the former is more skilled than the latter. Proposition 1 formalizes this intuition by establishing a market-specific unique wage function that is strictly increasing in ξ. If the data represents a single market, ξ is identified up to a monotone transform by inverting the wage function.[34]

However, our empirical application uses data from different labor markets. In this situation, it is convenient to index data points using worker-industry-time indexes instead of using worker-market notation. Without loss of generality and to establish a link with the regression defining inter-industry wage differentials presented below, consider the simple wage

[33] Two job vacancies where the employer-specific parameter $\beta_{i,a,x}$ is the same for attribute x may have different marginal WTP values due to differences in x_0. For example, if vacancy A is filled by a worker with 2 years of experience, the marginal WTP for one extra year of experience (keeping everything else constant) is $\beta_{i,a,x} \cdot (\ln 3 - \ln 2)$. If a worker with 6 years of experience is hired for vacancy B, the marginal WTP is $\beta_{i,a,x} \cdot (\ln 7 - \ln 6)$. For this reason, we use the formula $\beta_{i,a,x} \cdot (\ln x_1 - \ln x_0)$ only to fix an interpretation for our results on firm WTP parameters $\beta_{i,a}$. We use the estimates of $\beta_{i,a}$ at each observation for our empirical and graphical analysis.

[34] Theorem 2 of Bajari and Benkard (2005) formally establishes this identification result. In their empirical application, the CDF of price conditional on X suffices to point-identify ξ as quantiles are invariant to monotone transforms.

regression problem

$$w_{j,l,t} = X_{j,l,t}\beta + \delta_{j,l,t}, \tag{30}$$

Our identifying assumption is that unobserved worker quality varies by worker j and time t, but not by industry l. Therefore, unobserved worker quality could be identified using a worker-time fixed effect $\xi_{j,t}$.[35] Replacing $\delta_{j,l,t} = \xi_{j,t} + \epsilon_{j,l,t}$ yields

$$w_{j,l,t} = X_{j,l,t}\beta + \xi_{j,l,t} + \epsilon_{j,l,t}. \tag{31}$$

When the researcher does not have enough data variation to identify fixed effects, one solution is to replace the fixed effect with a linear function of sample averages of $X_{j,l,t}$ averaged across dimension l, denoted $\bar{X}_{j,t}$, plus a random effect (Mundlak, 1978). This solution is also valid when the fixed effects are present in nonlinear estimators (e.g. Papke and Wooldridge, 2008). While we cannot implement this solution as we only have observations for the industry in which the worker is employed, these results indicate that we can identify $\xi_{j,t}$ if we replace simple averaging across industries with suitable expectations taken using estimated probabilities of industry affiliation. Our identification approach outlined below exploits this point by posing monotonicity assumptions analog to $\delta_{j,l,t} = \xi_{j,t} + \epsilon_{j,l,t}$.[36] In addition, we require structural model assumptions to identify outcomes outside the industry where the worker is employed.[37]

We separate ξ from unmeasured factors net of worker-specific effects, ϵ, by assuming that the residuals in the wage regression (15) are strictly increasing in unobserved worker quality.[38] In addition to keeping consistency with the results in Proposition 1, this restric-

[35] Decomposing a residual into fixed effects when the data is a panel of 3 or more dimensions is commonly found in the Industrial Organization literature. Examples include airline industry studies where the researcher uses data indexed by origin-destination-quarter (e.g. Aguirregabiria and Ho 2012, Li et al. 2022). In these studies the researcher is interested in controlling for example, for origin-destination fixed-effects. Other examples include differentiated product sales indexed by product-city-week where the researcher estimates product-week fixed effects (e.g. Nevo, 2001).

[36] In an ideal setting, the decomposition $\delta_{j,l,t} = \xi_{j,t} + \epsilon_{j,l,t}$ would be applied to the wage differentials equation below by considering worker-time fixed effects as unobserved worker quality controls. While this is not doable due to observing worker data only at specific $a = (l, t)$ combinations, we note that the hedonic model used to recover $\xi_{j,t}$ is not a mere nonlinear version of a wage differentials regression. In equilibrium, hedonic models set prices given attributes valued by consumers (in our case, firms value X and ξ), whereas the inter-industry wage differentials regression considers additional observables as extra regressions. We turn to this point when discussing inter-industry wage differentials estimation.

[37] We are not the first to resort to structural modeling to identify time-varying unobservables when endogeneity and selection bias concerns are present. For example, Olley and Pakes (1996) estimate time-varying productivity shocks in the telecommunications industry by posing assumptions on firm investment, market exit and productivity dynamics.

[38] An additive shock assumption such as $\delta_{j,l,t} = \xi_{j,t} + \epsilon_{j,l,t}$ is intuitive but not necessary to separate unobserved worker quality from other factors. For example, suppose that the wage residual δ mixing unobserved worker quality ξ and unmeasured factors ϵ is given by $s(\xi, \epsilon) = -b + c\xi^d$ where $\epsilon = (b, c, d)$ is a vector of

tion guarantees that the rank order of the residual series is the same as the rank order of unobserved quality series in each market.[39] Thus, the CDF of ξ conditional on market a equals the CDF of wages conditional on X and market a. Replacing conditioning on market $a = (l, t)$ with conditioning on time t would result in $\xi_{j,t}$ normalized in the interval $[0, 1]$ as the CDF of ξ conditional on time t. This can be done if the CDF of wages conditional on X and market a, denoted by $F_a(w|X)$, can be written as $F(w|X, D_I = l, D_P = t)$. Our assumption in (13) that market wage functions can be parametrized by industry and time dummies yields this equivalence. Intuitively, it also identifies probabilities and expectations of wages for markets other than the one where the worker is employed. $F(w|X, D_I = l, D_P = t)$ can be estimated nonparametrically from a sample where the researcher observes data on worker wages, observed attributes X, industry affiliation D_I and time period D_P. Identification of the CDF of ξ conditional only on period t follows from applying the Law of Total Probability to make that CDF unconditional on worker industry affiliation. We characterize the probability required to integrate D_I out, denoted $\Pr(D_I = l|X, D_P = t)$, by a reduced-form multinomial logit model. The parameters of this model are identified by the joint variation in data for the variables in (D_I, X, D_P).[40]

The same identification approach to isolate $\xi_{j,t}$ from other factors applies when at least one variable in X is correlated with ξ, but with some differences. First, the estimation of each CDF of wages conditional on X and on market a must also condition on enough controls in η to address endogeneity concerns. In addition, this CDF must also condition on controls $\tilde{\eta}$ to address industry self-selection. Second, we need to integrate $(\eta, \tilde{\eta})$ out in our final step of quality estimation. Therefore we focus attention on the identification of the vector of control variables $(\eta, \tilde{\eta})$ from data variations.[41] We assume that the control functions η

strictly positive factors. Then $\xi = \exp((\log(\delta + b) - \log(c))/d)$. Thus, keeping ϵ constant, for any two workers k and j in market a, if $\delta_{k,a} < \delta_{j,a}$ then $\xi_{k,t} < \xi_{j,t}$. The function $s(\xi, \epsilon)$ does not need to have a closed form as long as it is invertible in the first argument. Intuitively, we assume that employers observe factors in ϵ and can infer the quality of any two workers k and j given their observed attributes. Thus, employers from all markets can summarize valued worker characteristics unobserved by the researcher up a common scalar metric ξ. Torgovitzky (2015) discusses the identification power of this type of rank invariance assumption and provides additional examples.

[39] In other words, keeping other unobservables $\epsilon_{j,l,t}$ constant, $\Pr(\delta \leq \delta_{j,l,t})$ and $\Pr(\xi \leq \xi_{j,t})$ are equivalent because $\delta = s(\xi, \epsilon)$ is monotone in the first argument and therefore applying the inverse function to $\delta_{j,l,t}$ while keeping $\epsilon_{j,l,t}$ results in $s^{-1}(s(\xi_{j,t}, \epsilon_{j,a}), \epsilon_{j,a}) = \xi_{j,t}$. A more detailed demonstration is presented in Appendix C. Krasnokutskaya et al. (2020) also exploit rank order properties to identify unobserved worker quality in their labor procurement model.

[40] Technically, the parameters of this model are identified by the score conditions of the MLE problem of fitting a multinomial logit model to all observations dated from period t on industry affiliation and observed worker attributes. Using the notation of Appendix D, U is replaced with X and D_P, and it is augmented with control functions $(\boldsymbol{\eta}, \tilde{\boldsymbol{\eta}})$ for the cases where some of the regressors in X and industry affiliation are endogenous.

[41] Theorems 2 and S2 in Torgovitsky (2015) establish the technical identification of triangular systems similar to the market-specific equations (15) and (16). For this reason, we focus on data variations identifying

address endogeneity concerns for X_0 in both the conditional wage CDF and the multinomial logit model for the probabilities $\Pr(D_I = l|X, \eta, \tilde{\eta}, D_P = t)$ necessary to integrate D_I out in the final step of worker quality computation. The estimator of each control variable in η is the empirical analog of $F_{x_{0m}|X_1,Z,D_P}(.|.)$, which is set to be a Normal mixture for each endogenous variable in X_0. The number of Normal mixtures used and the corresponding MLE problem are discussed in Appendix D. Joint data variations in X_1, Z_{jt}, D_P and X_0 identify the Normal mixture parameters. After applying the estimated conditional CDFs to this data, we also observe the control variables η for each worker in our sample. The self-selection controls proposed by Dahl (2002), $\tilde{\eta}$, are identified by the multinomial logit structure in (17) and the corresponding parameters are estimable by MLE using sample variation in (D_I, X_1, Z_S, D_P).[42] Moreover, the joint variation in observed wages, worker attributes X, industry and time dummies (D_I, D_P) and control variables $(\eta, \tilde{\eta})$ identifies the parameters of the PDF of wages conditional on those observables. Our estimator for the CDF of wages conditional on these variables, $\hat{F}_{w|X,\boldsymbol{\eta},D_l,D_P}(w_{jt}|X_{jt}, \eta, D_l = l, D_P = t; \hat{\theta}_w)$, is a Normal mixture model that is estimated by the same MLE procedure as the controls η. After normalizing ξ so that its marginal distribution at period t is uniform, unobserved worker quality in that period is identified as all the components of the right-hand side of (19) are identified.

The series of firm preference parameters for schooling, experience and unobserved worker quality are identified using the parametric form for vacancy revenue in (6) and the employer's problem in Equation (3). These assumptions result in the formulas for firm preference parameters in (8) and (9). The computation of these formulas at each data point involve both the worker's attribute values as well as the wage function derivatives at that data point. After computing unobserved worker quality for each observation using (19), we observe all the needed worker attributes. The nonparametric estimator in (26) provides a matrix formula for wage function derivatives at each data point involving only wage and worker attributes data. This establishes the technical identification of firm preference parameters at each data point.[43] The data sources of identification of wage function derivatives at a

model parameters.

[42] We observe that this multinomial logit estimate is different from the industry affiliation probability used to integrate out industry effects in the final step of quality computation. By definition, $\tilde{\eta}_l = \Pr(D_I = l|X_1, \mathbf{Z}_S, D_P = t)$, whereas the probability used when applying the Law of Total Probability is defined as $\Pr(D_I = l|X, \boldsymbol{\eta}, \tilde{\boldsymbol{\eta}}, D_P = t)$. The latter is more general as includes all $\tilde{\boldsymbol{\eta}}$,$X_0$ and all endogeneity controls $\boldsymbol{\eta}$ as extra regressors.

[43] While we implement a version of this estimator proposed by Du, Parmeter and Racine (2013), the data requirements are the same. Moreover, we impose that expected wages are non-negative and that the derivatives of wages with respect to quality, schooling and experience are also non-negative. These intuitive restrictions are additional sources of identification and efficiency, but are not necessary for identification of derivatives. See Fan and Gijbels (1996) for an extensive discussion of the properties of local linear regression.

specific data point j^* are the data observations on (w, X, ξ) that are in a neighborhood of $(w_{j^*,a}, X_{j^*,t}, \xi_{j^*,t})$. Intuitively, the distances from $w_{j^*,a}$ to neighboring wage observations in market a inform about a derivative's numerator. An analogous remark applies to the derivative's denominator for each worker attribute.

4 Data

The micro data used in our empirical analysis come from the 1990 and 1993 waves of the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14–22 years old when they were first surveyed in 1979. The NLSY79 data contain rich information on employment and demographic characteristics. For each individual, the NLSY79 reports age, gender, race, education, marital status, region of residence, employment status, occupation, and earnings. In addition, the NLSY79 asks questions on individual background and employer characteristics. We obtain information on parental education, Armed Force Qualification Test (AFQT) score, and each worker's industrial affiliation.

Data on individuals' usual earnings (inclusive of tips, overtime, and bonuses but before deductions) have been collected during every survey year on the first five jobs since the last interview date in the NLSY79. Combining the amount of earnings with information on the applicable unit of time (e.g., per hour, per day, or per week) yields the hourly wage rate. The earnings variable used in this study is the hourly wage for the CPS job, that is, the current or most recent job. We consider hourly wage less than \$1.00 and greater than \$250.00 to be outliers and eliminate them from the sample.

We construct the work experience variable from the week-by-week NLSY79 Work History Data. The usual hours worked per week at all jobs are available from January 1, 1978. Annual hours are computed by aggregating weekly hours in each calendar year. An individual accumulates one year of experience if she works for at least 1,000 hours a year. We restrict our sample to those with complete history of work experience. The sample we analyze contains 4,266 observations from the 1990 survey and 3,522 observations from the 1993 survey.

We use our NLSY data to estimate a standard cross-section Mincer wage equation to examine industrial wage premiums. Columns (1) and (5) of Table 1 report the raw differences in log hourly wages by industry for both the 1990 and 1993 observations. These differences are computed from cross-section regressions of log wage on a set of industry dummy variables by using one digit Census Industry Classification (CIC) Codes.[44] We use two cross-section

[44] The service industry is used as the reference industry. Because the wage regressions include a constant, we treat the service industry as having zero effect on wages.

wage observations so that we can check the consistency of our results over time and across different points in the career path. A simple summary measure of the importance of industry coefficients is their standard deviation. We report both weighted and unweighted standard deviations of estimates of the industry coefficient. Unweighted standard deviation measures the difference in wages between a randomly chosen industry and the average industry, whereas weighted standard deviation (by employment) measures the difference in wages between a worker in a given industry and the average worker. Both statistics demonstrate substantial variation in wages across industries.

In Columns (2) and (6) we examine the extent to which the raw inter-industry wage differentials persist once the usual human capital controls are added. Our strategy is to control for worker characteristics as well as possible, and then analyze the effects of industry dummy variables. We estimate industry wage differentials from the cross-section wage function

$$w = X\zeta + D\tau + \varepsilon, \tag{32}$$

where w is the logarithm of the hourly wage, X is a vector of individual attributes, D is a vector of industry dummy variables, and ε is a random error term. The controls are education, work experience, gender, race, marital status, occupation, location dummies, union status, veteran status, and several interaction terms.

The industry dummy variables are statistically significant in both years, substantial in magnitude, and similar to those estimated with data from the 1970s and 1980s (e.g., Blackburn and Neumark, 1992; Krueger and Summers, 1988). For example, earnings in construction, transportation, communication, and public utilities, are substantially higher than those in the wholesale and retail trade and service industries, even with controls for years of schooling, experience, gender, and race. Adding human capital controls reduces inter-industry wage differentials, as measured by their standard deviations, by 8%–10% in 1990 and 15%–20% in 1993.

Even after various human capital controls are included, the coefficient estimates on industry dummies in Equation (32) may pick up the differences in unobserved worker quality across industries. Previous research has attempted to correct unobserved quality bias in estimated industry effects by including proxies for worker quality, such as test scores in wage regressions (Blackburn and Neumark, 1992). In Columns (3) and (7), we include AFQT scores as additional independent variables in the wage equations. Compared with the estimates from Columns (2) and (6), the standard deviations of the industry effects decline slightly for both the 1990 (from 0.136 to 0.133, unweighted) and 1993 regressions (from 0.115 to 0.114, unweighted). Furthermore, including parental education in the wage regressions only slightly affects the standard deviations of the industry effects, as shown in Columns (4) and (8) of Table 1. These results fail to support the unobserved quality explanation for industry wage differentials, consistent with the conclusion reached by Blackburn and Neumark (1992).

Another approach to solving the problem of unobserved labor quality is to analyze longitudinal data and estimate the first-difference specification of wage equations (e.g., Gibbons and Katz, 1992; Krueger and Summers, 1988; Murphy and Topel, 1987a, 1987b). When we pool the 1990 and 1993 samples, 877 of the workers report changes in their one digit industry from 1990 to 1993. Column (9) of Table 1 reports the first-difference estimates of the wage regression. The industry variables are jointly significant. For example, the first-difference results show that workers who join the construction sector gain a 23.1% pay increase. These results are consistent with the findings by Krueger and Summers (1988), who interpret their findings as evidence that differences in labor quality cannot explain inter-industry wage differentials.[45]

One potential problem with using test scores and family background as proxies to remove omitted-quality bias is that test scores and family background are only partly correlated with the types of ability rewarded in labor markets. The ability to do well in standard tests may differ from the motivation and perseverance necessary to succeed in the workplace. On the other hand, first-difference estimates rely on the assumption that unobserved quality is time invariant and equally rewarded in all industries and can therefore be differenced out as an individual fixed effect. If labor quality evolves over time, perhaps through learning-by-doing, or if it is valued differently across firms, then an individual fixed effect can no longer capture its effect on wages. Therefore, we cannot conclude from Table 1 that inter-industry wage differentials are not attributable to variations in unobserved labor quality.

5 Empirical Results

This section presents estimates of our hedonic labor demand model. We first discuss the validity of our instruments and outline estimation results and robustness for the unobserved worker quality recovered in our first stage of estimation. We then present firm WTP parameter estimates based on our model specification. Finally, we assess how much unobserved worker quality and firm WTP for education, work experience, and quality account for inter-

[45] One notable difference between our first-difference results and those of previous studies (e.g., Gibbons and Katz, 1992; Krueger and Summers, 1988) is that they attempt to correct for selection bias from industry changes by using samples of displaced workers. Such a sample of displaced workers is not available from the NLSY79. However, our estimates yield similar results to those of analyzing non-displaced longitudinal data in Krueger and Summers (1988).

industry wage differentials.

5.1 Unobserved Worker Quality

We use the NLSY79 data on wages and observed worker characteristics to estimate unobserved worker quality based on Equation (19). We estimate the labor demand model separately for two years (1990 and 1993) and each one of the seven one-digit industries. Our approach is flexible enough to allow unobserved worker quality to evolve over time and allow firms to reward both observed worker attributes and unobserved labor quality differently. The variables of observed worker characteristics, represented by the vector X, include years of schooling, years of work experience, and dummy variables on gender, race, and marital status. Out of these variables, years of schooling and experience are potentially correlated with unobserved worker quality and constitute the sub-vector X_0.[46] All other observed characteristics are included in sub-vector X_1.

Our approach to estimating the unobserved worker quality requires instrumental variables for two purposes: (i) instruments Z to control for the endogenous variables in X_0, and (ii) instruments Z_S to address self-selection into industries. To estimate the control variables for education and experience, we use an instrument vector Z that includes three variables: a dummy for the existence of a local college, a dummy that equals one if the local college information is missing and is zero otherwise, and worker's age.[47] As shown by Card (1993), Kling (2001), and others, the existence of a local college would reduce the cost of college and affect schooling outcomes and therefore, it has been widely used as an instrument for years of schooling. The local college information is missing for about a quarter of our sample. Thus, we create a variable flagging missing local college information and use it as an additional instrument. Finally, age is obviously exogenous as it cannot be influenced by the worker, and it is also related to experience (Mincer, 1974). Therefore, we use age as an instrument for work experience. The local college instrument is binary and therefore does not satisfy the "large support" condition in Imbens and Newey (2009). However, as

[46] We experiment with alternative specifications of vector X containing other worker characteristics observed in the NLSY79. The results on quality estimates and subsequent wage differential analysis do not change significantly. In addition, some variables other than schooling and experience, such as marital status, are also potentially endogenous. However, marital status is less likely to be correlated with worker quality that is valued in the labor market compared to schooling and experience. Given that more variables in X or X_0 would increase computational cost drastically, we focus our analysis on the current parsimonious specification of X and X_0 without loss of generality.

[47] NLSY geocode is used to identify each individual's county and state of residence, and we match them with local school information. Annual data on location, type of institution, and other characteristics associated with all colleges in the U.S. are available from the Department of Education's annual IPEDS "Institutional Characteristics" surveys. We construct a dummy variable for the presence of any 2-year or 4-year college in the county of residence at age 18, following Ge (2011).

shown in Torgovitsky (2015), this condition is unnecessary to identify the control function variables when the unobservable in the estimating equation (i.e. unobserved worker quality) is a scalar. Torgovitsky (2017) also uses the same instrument and identification argument to estimate returns to schooling. For a rigorous yet intuitive discussion on the identification proof, see Torgovitsky (2015, pp.1188-1192).

To control for self-selection into industries, we need a vector of additional instruments Z_S that affect industry affiliation but do not overlap X_1 or Z. We include two variables on mothers' and fathers' years of schooling in Z_S. Intuitively, we expect parental education to be associated with individuals' career choices. For example, if a worker's parents have low levels of schooling, the worker may have had limited information on career prospects and therefore feel attracted only for professions related to parental occupations. At the other end of the spectrum, for example, if parents have graduated from law school, the worker may feel motivated to become a legal professional. We estimate a multinomial logit model of industry selection where the regressors are the variables in X_1 and a dummy for year 1993 along with Z_S. The predicted probabilities are used as our control functions for industry self-selection, as in Dahl (2002).

Although the nonparametric control function approach used in this paper does not have a direct analog to first-stage regressions in traditional 2SLS estimation because our model is nonlinear, we show the identifying power of our instruments by presenting the correlations between the endogenous variables and our instruments. Table 2 presents the linear regression results of schooling and experience on the instruments Z, the exogenous variables in X_1 and a time dummy for year 1993. We interpret these regressions as a linear approximation to our reduced-form equations in (16). All instruments in Z are statistically significant and the F-test reject the null hypothesis of no significance. The coefficient for local college is positive for education, and age has positive coefficients for both experience and schooling. These results provide some confidence for our instruments. Table 3 presents the results from estimating the multinomial logit model for industry selection. The parental education variables are significant at 5% and 10% levels for most of the one-digit industries considered and the Chi-square statistic is significant at 1% level. Therefore, the parental education variables help explaining industry selection, validating their use as instruments.

We use Equation (19) to estimate the unobserved worker quality. Table 4 shows the joint distribution between some of the observed worker characteristics and the unobserved worker quality. As for worker attributes on human capital, average worker quality increases with educational attainment, work experience, and AFQT scores. Across industries, we also observe substantial differences in average worker quality; transportation and public utilities, finance, and construction have higher average worker quality than wholesale and retail trade and service.

The top two panels of Table 5 report correlations between the estimated quality and human capital variables in each year. The correlations of these variables are positive but relatively low; all six correlations are less than 0.40. The estimated quality is less significantly correlated with experience than AFQT score and education in both years. The correlations between the estimated quality and AFQT score are 0.387 and 0.337 in 1990 and 1993, respectively. The relatively low correlations imply that worker quality rewarded in labor markets may not reflect completely in the AFQT score. Therefore, explicitly incorporating AFQT scores into wage regressions cannot fully account for variations in unobserved worker quality across industries. The bottom panel of Table 5 reports the correlation between the quality estimates in 1990 and 1993 to be fairly high at 0.710. Worker quality is by no means fixed over time according to our estimates, even though it is highly persistent. The evolution of labor quality over the career path may be related to post-school human capital investment, such as learning-by-doing on the job and specific human capital. According to our estimates, the correlation between the 1990 quality and the 1993 quality is 0.753 for those who stay in the same industry for both years, whereas the correlation is lower at 0.560 for those who switch industry between 1990 and 1993. The difference may be explained by the loss of industry-specific human capital for switchers. In conclusion, standard first-difference estimators cannot difference out the effects of unobserved quality on wages.

5.2 Distributions of WTP Parameters

We estimate the structural model of labor demand presented in Section 2 for both 1990 and 1993. This estimation yields for each firm a WTP parameter for schooling, experience, and unobserved worker quality, respectively. We present histograms of WTP parameters for these attributes for the 1990 and 1993 firms with the estimated kernel densities. In each figure, we plot the distribution of WTP parameters for firms across all industries, followed by the distribution of the same parameters in each one-digit industry. WTP considerably varies for both observed education and work experience, and unobserved worker quality. All the distributions are right-skewed and are not normally distributed.

Panel A of Figure 1 presents the histogram of firm-specific preference/WTP parameter ($\beta_{i,a,x}$ in equation (8)) for education in all industries in 1990. The distribution has a long right tail, with a mean of 490 and a standard deviation of 1182. Each firm's marginal WTP for a worker's education can be computed based on the estimated preference parameter. For example, if a firm's preference parameter is equal to the mean value of 490 and it currently hires a worker with 6 years of education, then the firm is willing to pay an additional \$0.76

(= \$4.90 × (ln 7 − ln 6)) per hour on top of the worker's current hourly wage to hire a worker with 7 years of education while keeping all other worker attributes constant. Similarly, an increase from 7 to 8 years in education would result in an additional \$0.65 per hour.[48]

Panels B to H of Figure 1 present histograms of firm WTP parameter for education in each one-digit industry. Mining, finance, insurance, and real estate industries have the highest mean WTP parameters for education, whereas wholesale and retail trade and service industries have the lowest mean WTP parameters for education. All industry-specific distributions are right-skewed. Specifically, the distribution in the finance, insurance and real estate industry has the longest tail with a standard deviation of 1931, and the distribution in the wholesale and retail trade industry is the least dispersed with a standard deviation of 719.

Figure 2 presents the histograms of 1990 firm-specific WTP parameters for work experience in all industries in Panel A and in each one-digit industry in Panels B to H. The average WTP parameter for work experience is lower than the average WTP parameter for education (204 vs. 490), and the WTP parameter for experience is less dispersed with a standard deviation of 492. Firms in the construction industry value work experience the most, with a mean WTP parameter equal to 356, whereas experience is the least valued in the wholesale and retail trade industry with a mean WTP parameter of 121. In terms of dispersion, the construction industry has the longest right tail, and the distribution of WTP parameter for experience is most concentrated in the wholesale and retail trade industry.

Firm-specific WTP parameters for worker quality in all industries and in each one-digit industry in 1990 are presented in Figure 3. Because worker quality has no intrinsic units and is normalized between 0 and 1, the values of WTP parameters for quality are unimportant; thus, we focus on their relative levels across industries. Based on Panels B to H, (unobserved) worker quality is less valuable to firms in the wholesale and retail trade and mining industries than to firms in the manufacturing industry. The distribution of WTP parameter for quality is most dispersed in the manufacturing industry and least dispersed in the mining industry.

Similarly, we present the distributions of WTP parameters for education, work experience, and worker quality from 1993 in Figures 4 to 6. Firms in most industries (except for mining) value education more in 1993 than in 1990. The 1993 distributions of WTP parameters for education in Figure 4 are also more dispersed than the 1990 distributions in Figure

[48] These patterns are consistent with recent evidence on returns to schooling using NSLY data and non-separable models (e.g. Torgovitsky 2017). Our examples here focus on marginal WTP for one additional year of schooling by a firm while keeping all other worker attributes constant. In practice, the strong positive correlation between education and worker quality shown in Table 4 implies that one additional year of schooling is associated with an increase in worker quality. As firms also value worker quality, the wage increase associated with one additional year of schooling would be much larger than these numbers.

1. Likewise, Figure 5 and Figure 6 show that firms in all industries value worker experience and quality more highly in 1993 than in 1990, and the distributions of WTP parameters for experience and quality are also more dispersed in 1993, as indicated by the higher means and variances of WTP parameters in Figure 5 and Figure 6 than those in Figure 2 and Figure 3. These results are consistent with the increasing returns to both observed human capital and unobserved ability documented in literature.

5.3 Inter-industry Wage Differentials

Columns (2) and (6) of Table 6 present estimates of coefficient τ in Equation (32) by adding recovered worker quality as an extra control variable in the 1990 and 1993 cross-section wage regressions. For comparison, Columns (1) and (5) report the same estimates with all controls, including AFQT scores and family background, but without estimated quality. The coefficient on worker quality is high and statistically significant. The magnitude of the coefficients on industry dummies declines, and many of them become statistically insignificant after worker quality is included. The standard deviation of the unweighted inter-industry wage differentials decreases by 87% from 0.133 to 0.017 in 1990 and from 0.114 to 0.008 in 1993. The weighted standard deviation of wage differentials declines by a similar magnitude. These results suggest that unmeasured worker quality is an important driving force of inter-industry wage differentials. Worker quality also accounts for a large portion of the overall wage variation as the adjusted R^2 of the log wage regression increases from 0.356 to 0.889 in 1990 and from 0.376 to 0.873 in 1993 once worker quality is included in the regressions.[49]

Columns (3) and (7) of Table 6 present estimates of τ coefficients in Equation (32) by adding recovered firm-specific WTP to education, experience, and quality as additional control variables. The industry wage premiums in both years decrease but remain significant. The standard deviation of the unweighted inter-industry wage differentials decreases from 0.133 to 0.109 in 1990 and from 0.114 to 0.084 in 1993. The adjusted R^2 of the log wage regression increases from 0.356 to 0.540 in 1990. Compared with worker quality (columns 2 and 6), firm WTP can account for a smaller portion of the inter-industry wage differentials and overall wage dispersion. When both worker quality and firm WTP are included in the OLS wage regressions in columns (4) and (8), the standard deviations of industry wage differentials almost stay the same as in the regressions that control only for worker quality. In all the specifications including worker quality or firm WTP in Table 6, we bootstrap the standard errors of parameters.

[49] Using a different dataset and different methodology, Abowd, Kramarz and Margolis (1999) also find that wage regressions that include person effects can explain between 77% to 83% of wage variance, whereas regressions that exclude person effects can explain only between 30% to 55% of the variance.

We further decompose the contribution of worker heterogeneity (in terms of unobserved labor quality) and firm heterogeneity (measured by WTP for human capital attributes) to inter-industry wage differentials. We estimate inter-industry wage differentials by regressing (32) with two-digit industry dummies while controlling for education, years of experience and its square, gender, race, marital status, union and veteran status, region dummies, occupation, parental education, AFQT test score, and several interaction terms.[50] Table 7 uses the industry-level averages of worker quality and firm-specific WTP parameters to account for the industry wage differentials, and we present bootstrapped standard errors in parentheses. The first column of Table 7 shows the separate influence of worker heterogeneity on explaining industry effects by regressing the estimated industry wage premiums on industry-average worker quality alone. Similarly, column (2) of Table 7 presents industry-level regressions using industry-average firm WTP parameters alone. Industry-average worker quality alone accounts for approximately two thirds of observed inter-industry wage variation, whereas the explanatory power of industry-average firm WTP parameters is relatively low. Therefore, individual effects, as measured by average worker quality, are more important than firm effects, as measured by WTP parameters, for explaining inter-industry wage differentials.[51] The combination of worker quality and firm WTP can explain close to 70% of the overall variations in inter-industry wage differentials in both years.

5.4 Robustness Checks

We use our unobserved worker quality estimates to assess their power on explaining inter-industry wage differentials. One concern is that the unobserved worker quality is recovered from the wage residuals in (32) and therefore, this term could capture the firm-worker match value or part of the willingness to pay by firms because of different monopsony power. In this section, we conduct robustness checks to provide evidence that our measure of unobserved worker quality is separated from other candidate factors.

Without longitudinal matched employer-employee data, we are not able to estimate firm-worker match effects directly. Instead, we use job satisfaction measures from the NLSY79 to infer match quality following the literature (Gielen, 2008; Ferreira and Taylor, 2011). We also use measures on firm size to examine the potential confounding effect of monopsony (Green,

[50] These results are available from the authors upon request.

[51] Using matched employer–employee data from France, Abowd, Kramarz, and Margolis (1999) find that individual heterogeneity alone explains 84%–92% of the inter-industry wage variation, whereas firm heterogeneity alone explains only 7%–25%. Thus, they reach the same conclusion as ours that individual effects are more important than firm effects for explaining inter-industry wage differentials. However, our approach does not require the use of matched employer–employee data and does not impose the assumption that unobserved labor quality is fixed over time.

Machin and Manning, 1996; Burdett and Mortensen, 1998). In our data, job satisfaction is measured by a scalar from 1 ("very satisfied") to 4 ("very dissatisfied") and firm size is measured by dummies on small, medium, and large firms.[52] We regress the estimated unobserved worker quality on dummies of job satisfaction and firm size in each year and present the regression results in Table 8. Columns (1) and (3) report results when industry dummies are not included for 1990 and 1993, respectively. The coefficients on job satisfaction and firm size dummies are mostly statistically significant and have the expected signs. Both job satisfaction and firm size are positively associated with worker quality. However, these two factors jointly explain 4.1% of worker quality variations in 1990 and 5.7% of worker quality variations in 1993. Adding the industry dummies in columns (2) and (4) increases the adjusted-R^2 for both years, yet the portion of explained worker quality variations is less than 10%. Although we cannot complete rule out the effects of match quality and monopsony power, we provide evidence that these factors only explain a small portion of the variations in unobserved worker quality.

While our exercise provides suggestive evidence that the variations in unobserved worker quality is not driven by other factors unrelated to worker skill, it is important to discuss the differences between the wage residuals in (32) and the unobserved worker quality recovered by our model. For all specifications in this paper, the wage regression in (32) controls not only for the productive attributes vector X in our model, but also for cross products and other controls commonly used when estimating inter-industry wage differentials (e.g. industry dummies, AFQT, parental education, veteran and union status, location and occupation dummies). In addition, the estimation of Equation (32) using OLS does not address the endogeneity of some regressors or self-selection of industry. In contrast, our unobserved worker quality is derived using our structural model of demand where we condition on X while controlling for selection bias, instrumenting for endogeneity, and integrating out industry effects. Heuristically, our unobserved worker quality approximates worker- and time-effects when the researcher has observations at the worker-industry-time level (because each market is an industry-time combination and both industry and time period are observable). Therefore, by replacing worker-time fixed effects with a structural estimate, our measure for unobserved worker quality approximates the portion of the residuals in (32) that could be explained by worker-time fixed effects.

[52] A small firm is defined as a firm with less than 100 workers, a medium-sized firm employs between 100 and 499 workers, whereas a large firm has 500 or more workers.

6 Concluding Remarks

In this paper we propose an alternative approach to explain inter-industry wage differentials by using a hedonic model of labor demand. The model allows the nonparametric identification of unobserved worker quality as well as employer-specific WTP for worker attributes. Our approach complements a growing literature that adapts models of demand for differentiated products to identify firm quality observed by workers but not by the researcher (e.g., Azar, et al. 2022, Card et al. 2018). Our framework does not require the use of matched employer–employee panels to separate the worker effect and the firm effect in inter-industry wage differentials. Instead, we can rely on widely available household or individual micro data sets. Using data from the NLSY79, we find that unmeasured worker quality accounts for most of inter-industry wage differentials and that unmeasured worker quality varies over one's career despite its high degree of persistence.

The assumption that unobserved worker quality can be summarized by a composite scalar is central to our empirical strategy by allowing us to use the identification results of Torgovitsky (2015). Modeling skills as multidimensional is pioneered by Willis and Rosen (1979) and Heckman and Sedlacek (1985). Recent studies, such as Lise and Postel-Vinay (2020), suggest that different types of observed skills are very different productive attributes. Relaxing the assumption of single-dimensional unobserved skill is possible (Matzkin 2003), but requires either a "large support" condition (Imbens and Newey, 2009) or "measurable separability" with continuous instruments (Florens et al., 2008).

An important caveat to the effects of firm WTP on industry wage premiums is that the hedonic labor demand model does not point-identify employer-specific WTP for discrete worker characteristics, such as gender, race, and marital status, even if the researcher makes strong assumptions about the distribution of WTP parameters. Our framework shares this feature with other related models (e.g., Bajari and Benkard, 2005; Bajari and Khan, 2005). Therefore, we cannot identify which portion of inter-industry wage differentials can be explained by WTP for discrete attributes. Finding a set of mild assumptions that can point-identify employer WTP for discrete attributes is beyond the scope of this study and is thus left for future work.

As in the hedonic model of differentiated products proposed by Bajari and Benkard (2005), supply-side assumptions on worker behavior are not required to estimate our labor demand model. An interesting extension of our framework is to explicitly model labor supply behavior and allow workers to choose which firm to work for. For example, Lamadon, Mogstad and Setzler (2022) identify and estimate an equilibrium model of the labor market where workers with heterogeneous preferences and productivities choose over firms with different productivities and non-wage job characteristics to quantify the importance of imperfect competition. An important topic for future research is to introduce supply-side behavior into our labor demand model and separately identify compensating wage differentials from WTP parameters in an equilibrium model.

References

[1] Abowd, J., F. Kramarz, P. Lengermann, K.L. McKinney, and S. Roux (2012), "Persistent Inter-industry Wage Differences: Rent Sharing and Opportunity Costs," *IZA Journal of Labor Economics*, 1:7, doi:10.1186/2193-8997-1-7.

[2] Abowd, J., F. Kramarz, and D. Margolis (1999), "High Wage Workers and High Wage Firms," *Econometrica* 67: 251–334.

[3] Aguirregabiria, V. and C. Y. Ho, (2012), "A dynamic oligopoly game of the US airline industry: Estimation and policy experiments," *Journal of Econometrics*, 168(1), 156-173.

[4] Amiti, M., and A. Khandelwal (2013), "Import Competition and Quality Upgrading," *Review of Economics and Statistics* 95(2): 476–490.

[5] Azar, J., Berry, S. T. and Marinescu, I. E. (2022), "Estimating Labor Market Power," No. w30365. National Bureau of Economic Research.

[6] Bajari, P., and C.L. Benkard (2005), "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy* 113: 1239–1276.

[7] Bajari, P., J. Fox, K. Kim and S. Ryan (2011), "A Simple Nonparametric Estimator for the Distribution of Random Coefficients," *Quantitative Economics* 2: 381–418.

[8] Bajari, P., J. Fox and S. Ryan (2007), "Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients," *American Economic Review Papers and Proceedings* 97: 459–463.

[9] Bajari, P., and M.E. Kahn (2005), "Estimating Housing Demand With an Application to Explaining Racial Segregation in Cities," *Journal of Business and Economic Statistics* 23: 20–33.

[10] Bartik, T.J. (1987), "The Estimation of Demand Parameters in Hedonic Price Models," *Journal of Political Economy* 95: 81–88.

[11] Bayer, P., F. Ferreira and R. McMillan (2007), "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy* 115(4): 588–638.

[12] Berry, S. and P. Haile (2014), "Identification in Differentiated Product Markets using Market Level Data," *Econometrica* 82(5): 1749–1797.

[13] Berry, S. and P. Haile (2018), "Identification of Nonparametric Simultaneous Equation Models with a Residual Index Structure," *Econometrica* 86(1): 289-315.

[14] Berry, S., J. Levinsohn and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica* 63(4): 841–890.

[15] Bhat, C. (2001), "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research Part B: Methodological* 35: 677–693.

[16] Blackburn, M. and D. Neumark (1992), "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics* 107: 1421–1436.

[17] Blundell, R. and J. Powell (2003), "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics*, Vol. II, ed. by Dewatripont, L. Hansen and S. Turnovsky. Cambridge: Cambridge University Press, pp. 312-357.

[18] Blundell, R. and J. Powell (2004), "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies* 71: 581–913.

[19] Bourguignon, F., Fournier, M., & Gurgand, M. (2007), "Selection Bias Corrections Based on the Multinomial Logit Model: Monte Carlo Comparisons," *Journal of Economic Surveys,* 21(1), 174-205.

[20] Brown, J.N. and H.S. Rosen (1982), "On the Estimation of Structural Hedonic Price Models," *Econometrica* 50: 765–768.

[21] Burdett, K., & Mortensen, D. T. (1998), "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 257-273.

[22] Card, D. (1993), "Using Georgraphic Variation in College Proximity to Estimate the Return to Schooling," *NBER working paper 4483.*

[23] Card, D. (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127-1160.

[24] Card, D., A. R. Cardoso, J. Heining and P. Kline (2018), "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics* 36: S13-S70.

[25] Card, D., J. Heining and P. Kline (2013), "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics* 128: 967–1015.

[26] Chernozhukov, V., G. Imbens and W. Newey (2007), "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics* 139: 4–14.

[27] Chesher, A. (2003), "Identification in Nonseparable Models," *Econometrica* 71: 1405–1441.

[28] Chiappori, P.-A., R. McCann and L. Nesheim (2010), "Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology and Uniqueness" *Economic Theory* 42: 317–354.

[29] Dahl, G. B. (2002), "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367-2420.

[30] D'Haultfoeuille, X. and P. Fevrier. (2015), "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica* 83: 1199–1210.

[31] Dickens, William T. and Lawrence F. Katz (1987), "Inter-Industry Wage Differencesand Industry Charachteristics," *Unemployment and the Structure of Labor Markets*, edited by K. Lang and J. Leonard, pp. 48-89. New York: Basil Blackwell.

[32] Du, P., C. Parmeter and J. Racine (2013), "Nonparametric Kernel Regression with Multiple Shape Constraints," *Statistica Sinica* 23: 1347–1371.

[33] Duong, T. (2007), "ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R," *Journal of Statistical Software* Vol. 21 Issue 7.

[34] Dykstra, R. L. (1983), "An Algorithm for Restricted Least Squares Regression," *Journal of the American Statistical Association*, 78(384), 837-842.

[35] Ekeland, I. (2010), "Existence, Uniqueness and Efficiency of Equilibrium in Hedonic Markets with Multidimensional Types," *Economic Theory* 42: 275–315.

[36] Ekeland, I., J. Heckman and L. Nesheim (2004), "Identification and Estimation of Hedonic Models," *Journal of Political Economy* 112: S60–S109.

[37] Epple, D. (1987), "Hedonic Prices and Implicit Markets: Estimating Supply and Demand Functions for Differentiated Products," *Journal of Political Economy* 95: 59–80.

[38] Fan, J. and I. Gijbels (1996), *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability*. London, Chapman & Hall.

[39] Farre, L. R. Klein and F. Vella (2013), "A Parametric Control Function Approach to Estimating the Returns to Schooling in the Absence of Exclusion Restrictions: an Application to the NSLY," *Empirical Economics*, 44: 111–133.

[40] Ferreira, P., & Taylor, M. (2011), "Measuring Match Quality using Subjective Data," *Economics Letters*, 113(3): 304-306.

[41] Florens, J., J. Heckman, C. Meghir and E. Vytlacil (2008), "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica* 76: 1191–1206.

[42] Fox, J. (2018), "Estimating Matching Games With Transfers," *Quantitative Economics* 9: 1–38.

[43] Fox, J. and V. Smeets (2011), "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review* 52: 961–989.

[44] Fox, J., K. Kim, S. Ryan and P. Bajari (2012), "The Random Coefficients Logit Model is Identified," *Journal of Econometrics* 166(2): 204–212.

[45] Fox, J., C. Yang and D. Hsu (2018), "Unobserved Heterogeneity in Matching Games," *Journal of Political Economy* 126(4): 1339–1373.

[46] Ge, Suqin (2011), "Women's College Decision: How Much Does Marriage Matter?" *Journal of Labor Economics* 29: 773–818.

[47] Gielen, A. C. (2013), "Repeated Job Quits: Stepping Stones or Learning about Quality?" *IZA Journal of European Labor Studies* 2(1): 1-22.

[48] Gibbons, R. and L. Katz (1992), "Does Unmeasured Ability Explain Inter-Industry Wage Differentials," *Review of Economic Studies* 59: 515–535.

[49] Green, F., Machin, S., & Manning, A. (1996), "The Employer Size-Wage Effect: Can Dynamic Monopsony Provide an Explanation?" *Oxford Economic Papers*, 48(3): 433-455.

[50] Haile, P., Hong, H., and Shum, M. (2003), "Nonparametric Tests for Common Values at First-Price Sealed-Bid Auctions." *NBER working paper* 10105.

[51] Hall, P., J. Marron and B. Park (1992), "Smoothed Cross-Validation," *Probability Theory and Related Fields*, 92, 1–20.

[52] Heckman, J. (1974), "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679–694.

[53] Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

[54] Heckman, J., R. and B. Honore (1990), "The Empirical Content of the Roy Model," *Econometrica* 58, No. 5, 1121-1149.

[55] Heckman, J., R. Matzkin and L. Nesheim (2010), "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica* 78: 1569–1591.

[56] Heckman, J. and G. Sedlacek (1985), "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy* 93: 1077–1125.

[57] Imbens, G. and W. Newey (2009), "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica* 77: 1481–1512.

[58] Imbens, G. and J. Wooldridge (2007), "Control Function and Related Methods." *What's New in Econometrics.*

[59] Katz, L. (1986), "Efficiency Wage Theories: A Partial Evaluation," *NBER Macroeconomics Annual* 1: 235–276.

[60] Katz, L. and D. Autor (1999), "Changes in the Wage Structure and Earnings Inequality," in *Handbook of Labor Economics*, O. Ashenfelter and D. Card, edt. vol. 3, Part A, 1463–1555. Elsevier.

[61] Katz, L. and L. Summers (1989), "Industry Rents: Evidence and Implications," *Brookings Papers on Economic Activity: Microeconomics*, 209–275.

[62] Khandelwal, A. (2010), "The Long and Short (of) Quality Ladders," *Review of Economic Studies* 77: 1450–1476.

[63] Kling, J. R. (2001), "Interpreting instrumental variables estimates of the returns to schooling," *Journal of Business & Economic Statistics* 19, no. 3: 358-364.

[64] Krasnokutskaya, E., and K. Song and X. Tang (2020), "The Role of Quality in Internet Service Markets," *Journal of Political Economy* 128(1): 75-117.

[65] Krueger, A.B., and L.H. Summers (1988), "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica* 56: 259–294.

[66] Lachowska, M., A. Mas, R.D. Saggio, and S.A. Woodbury (2022), "Do Firm Effects Drift? Evidence from Washington Administrative Data," *Journal of Econometrics.*

[67] Lamadon, T., Mogstad, M., and Setzler, B. (2022), "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112(1), 169-212.

[68] Li, S., Mazur, J., Park, Y., Roberts, J., Sweeting, A. and J. Zhang (2022), "Repositioning and Market Power after Airline Mergers," *The RAND Journal of Economics*, 53(1), 166-199.

[69] Lise, J. and F. Postel-Vinay (2020), "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 110(8), 2328-2376.

[70] Matzkin, R. L. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica* 71: 1332–1375.

[71] Matzkin, R. L. (2008), "Identification in Nonparametric Simultaneous Equations," *Econometrica*, 76: 945-978.

[72] McFadden, D., K. Train and M. Ben-Akiva (1987), "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices," *RAND Journal of Economics* 18: 109–123.

[73] Mincer, J. (1974), Schooling, Experience and Earnings. Columbia University Press, New York.

[74] Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 69-85.

[75] Murphy, K.M., and R. Topel (1987a), "Efficiency Wages Reconsidered: Theory and Evidence," in *Advances in the Theory and Measurement of Unemployment,* Y. Weiss and G. Fishelson (eds.), New York: St. Martin's Press.

[76] Murphy, K.M. and R. Topel (1987b), "Unemployment, Risk and Earnings," in *Unemployment and the Structure of Labor Markets*, K. Lang and J. Leonard (eds.), Oxford: Basil Blackwell.

[77] Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica* 69(2): 307–342.

[78] Nickell, S. and S. Wadhwani (1990), "Insider Forces and Wage Determination," *Economic Journal* 100: 496–509.

[79] Norets, A. (2010), "Approximation of Conditional Densities by Smooth Mixtures of Regressions," *Annals of Statistics* 38: 1733–1766.

[80] Papke, L. E., and J. Wooldridge(1996), "Econometric Methods for Fractional Response Variables with an Application to 401 (k) Plan Participation Rates," *Journal of Applied Econometrics*, 11(6), 619-632.

[81] Papke, L. E., and J. Wooldridge (2008), "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of econometrics*, 145, 121-133.

[82] Olley, G. S., and A. Pakes. (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica* 64: 1263–1297.

[83] Petrin, A. and K. Train (2010), "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research* 47: 3–13.

[84] Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 82: 34–55.

[85] Rosen, S. (1986), "The Theory of Equalizing Differences," in *Handbook of Labor Economics*, vol. 1, edited by Orley Ashenfelter and Richard Layard. Amsterdam: North-Holland.

[86] Roy, A. D. (1951), "Some Thoughts on the Distribution of Earnings," *Oxord Economic Papers*, 3, 135-146.

[87] Sattinger, M. (1980), *Capital and the Distribution of Labor Earnings*. North-Holland.

[88] Solon, G., Haider, S. J., and Wooldridge, J. M. (2015), "What are We Weighting for?," *Journal of Human Resources*, 50(2), 301-316.

[89] Song, J., D.J. Price, F. Guvenen, N. Bloom, and T. Wachter (2019), "Firming Up Inequality," *Quarterly Journal of Economics* 134: 1–50.

[90] Thaler, R.H. (1989), "Anomalies: Interindustry Wage Differentials," *Journal of Economic Perspectives* 3: 181–193.

[91] Torgovitsky, A. (2015), "Identification of Nonseparable Models Using Instruments with Small Support," *Econometrica* 83: 1185–1197.

[92] Torgovitsky, A. (2017), "Minimum Distance from Independence Estimation of Nonseparable Instrumental Variables Models," *Journal of Econometrics* 199 (3): 35–48.

[93] Van Reenen, John (1996), "The Creation and Capture of Rents: Wages and Innovation in a Panel of U.K. Companies," *Quarterly Journal of Economics* 111: 195–226.

[94] Willis, R.J. and S. Rosen (1979), "Education and Self-Selection," *Journal of Political Economy* 87: S7–S36.

[95] Wooldridge, J. M. (1999), "Asymptotic Properties of Weighted M-estimators for Variable Probability Samples," *Econometrica*, 67(6), 1385-1406.

[96] Wooldridge, J. M. (2001), "Asymptotic Properties of Weighted M-estimators for Standard Stratified Samples," *Econometric Theory*, 17(2), 451-470.

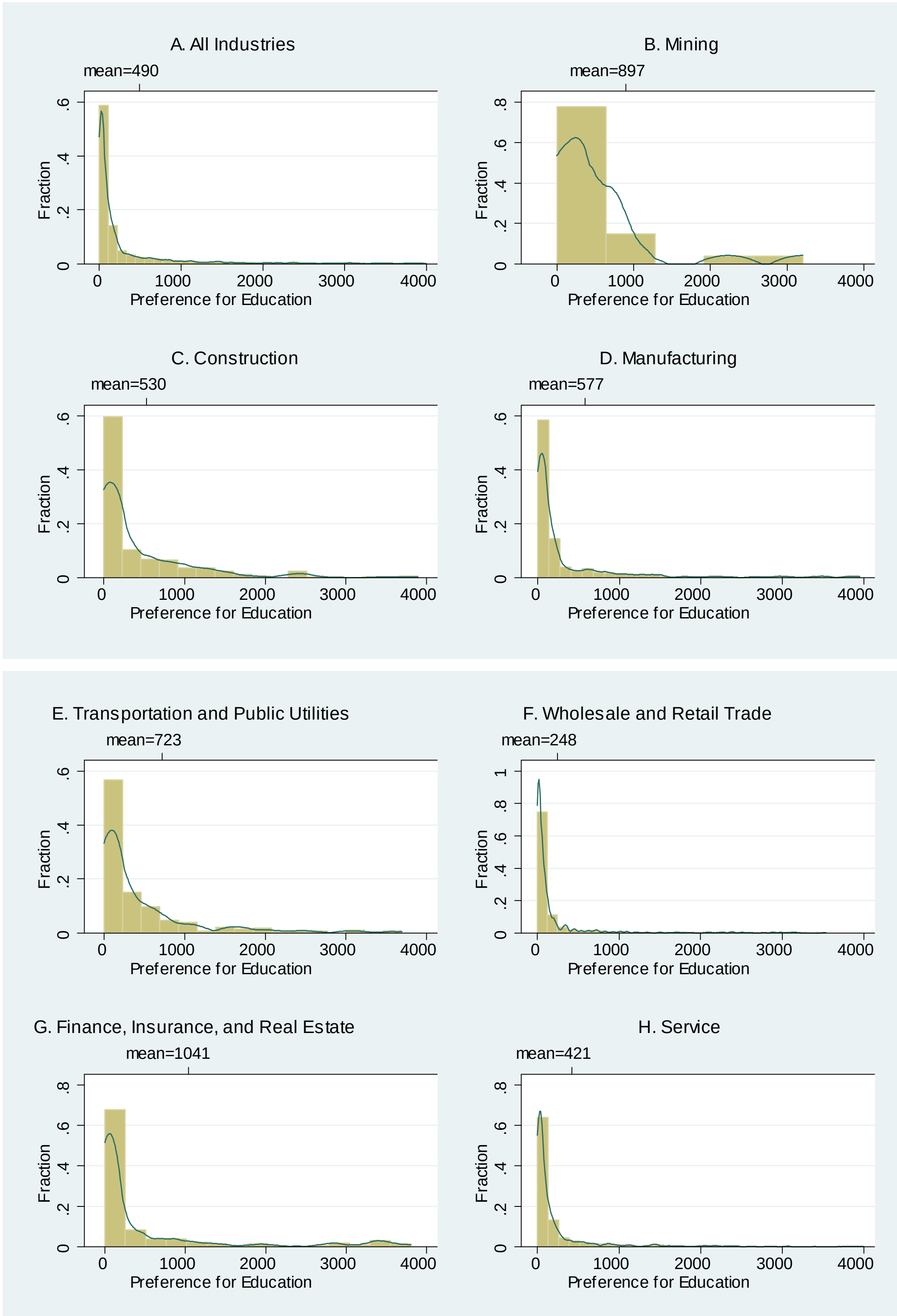

Figure 1: Firm Preference for Education Across Industries, 1990

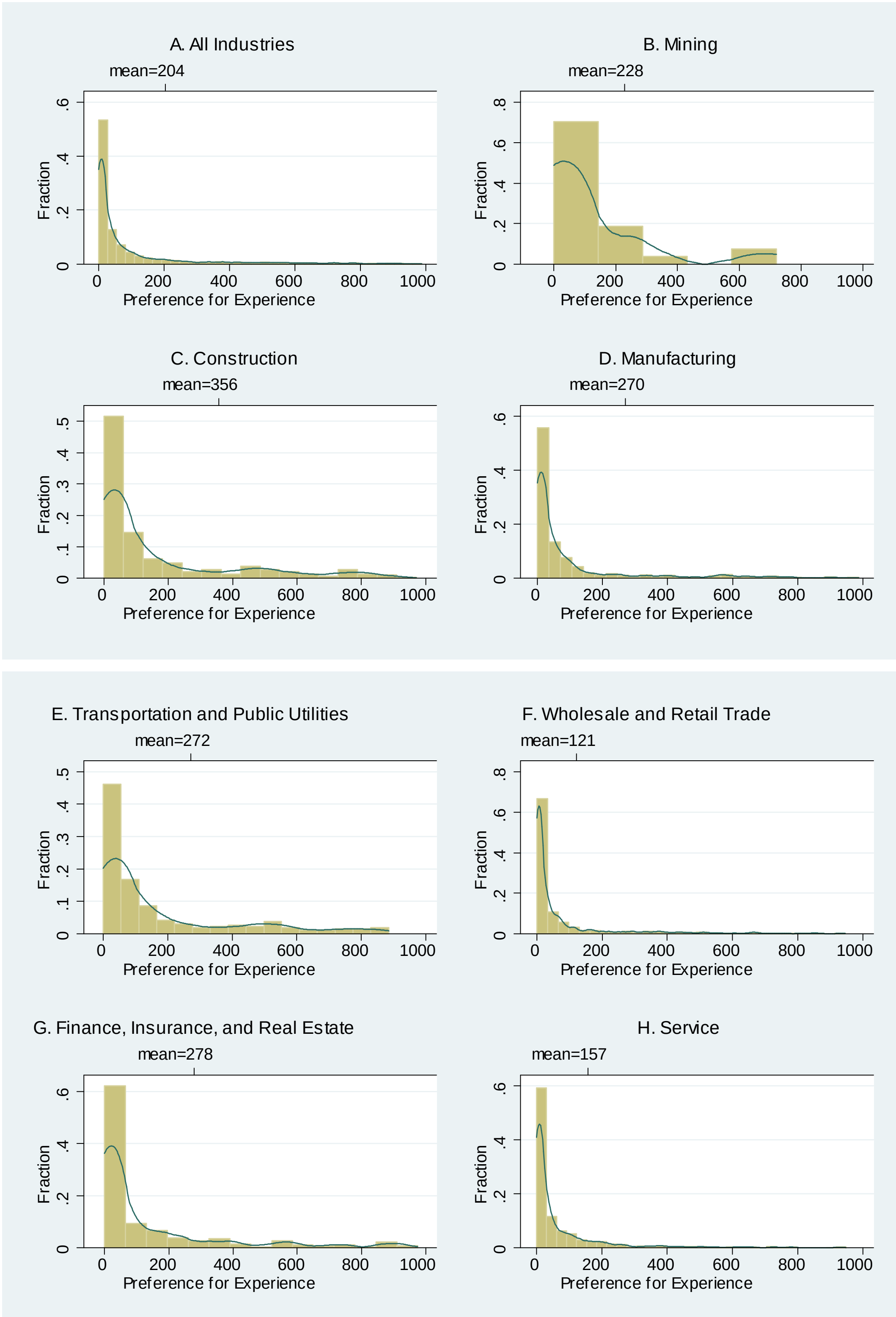

Figure 2: Firm Preference for Work Experience Across Industries, 1990

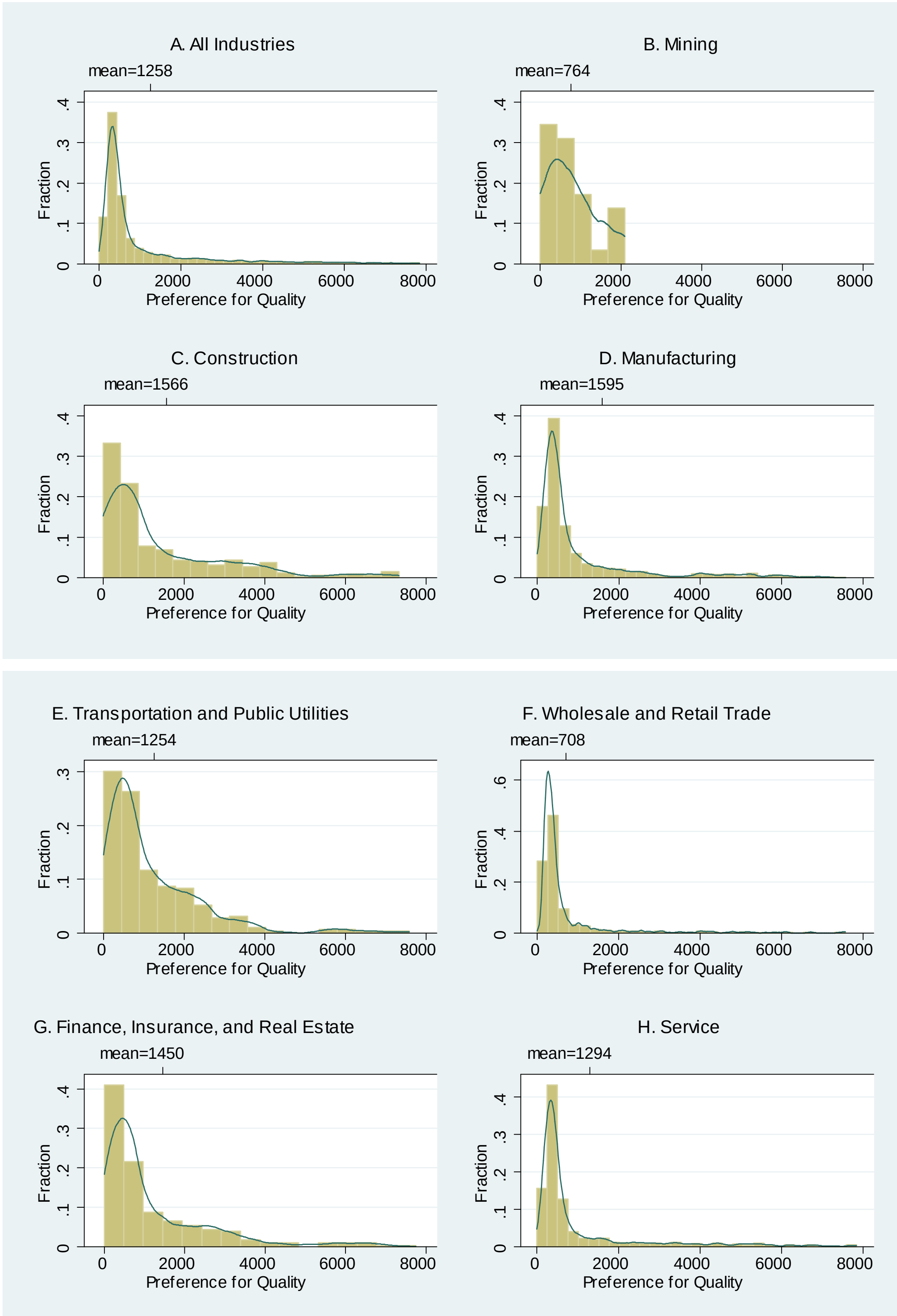

Figure 3: Firm Preference for Worker Quality Across Industries, 1990

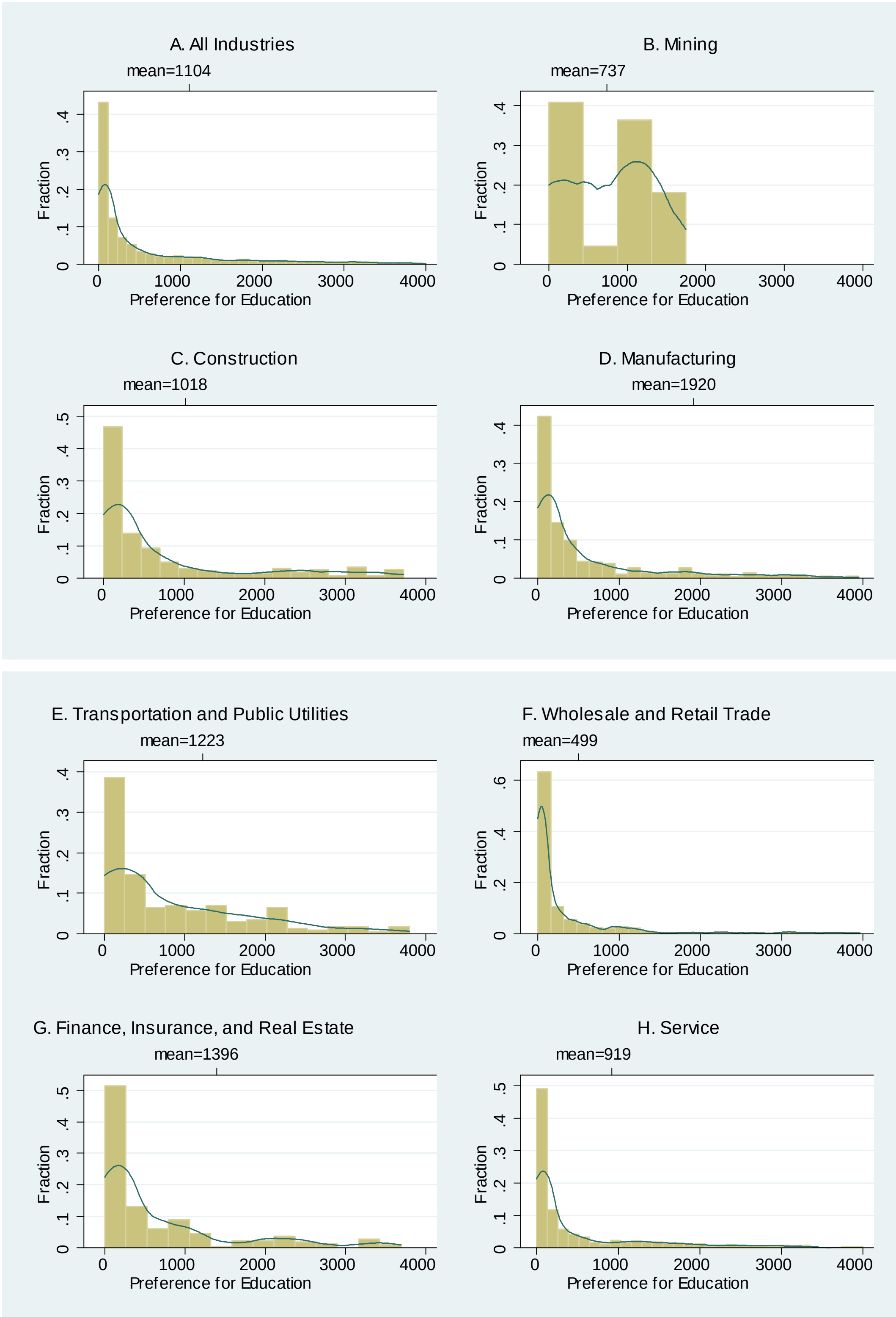

Figure 4: Firm Preference for Education Across Industries, 1993

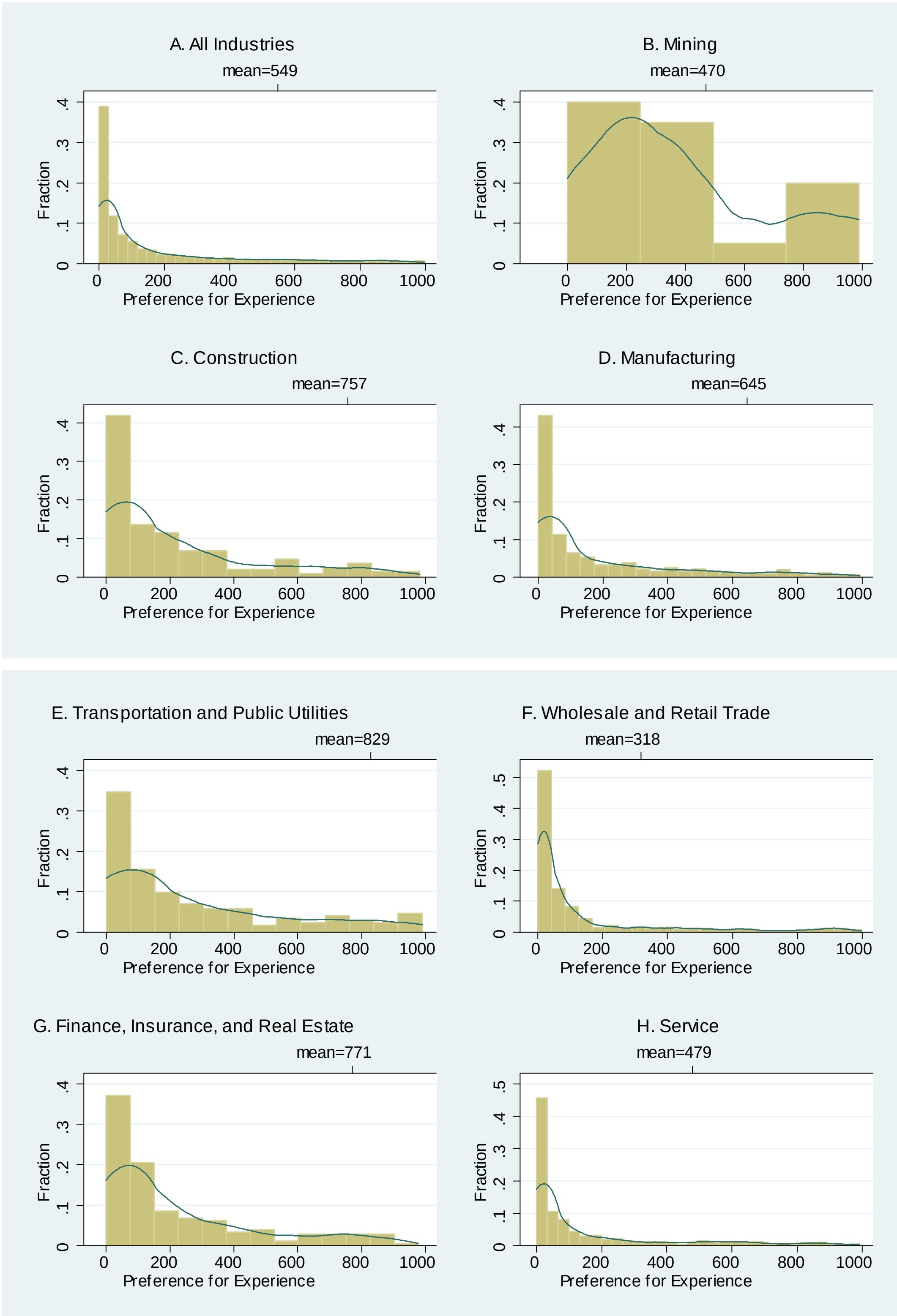

Figure 5: Firm Preference for Work Experience Across Industries, 1993

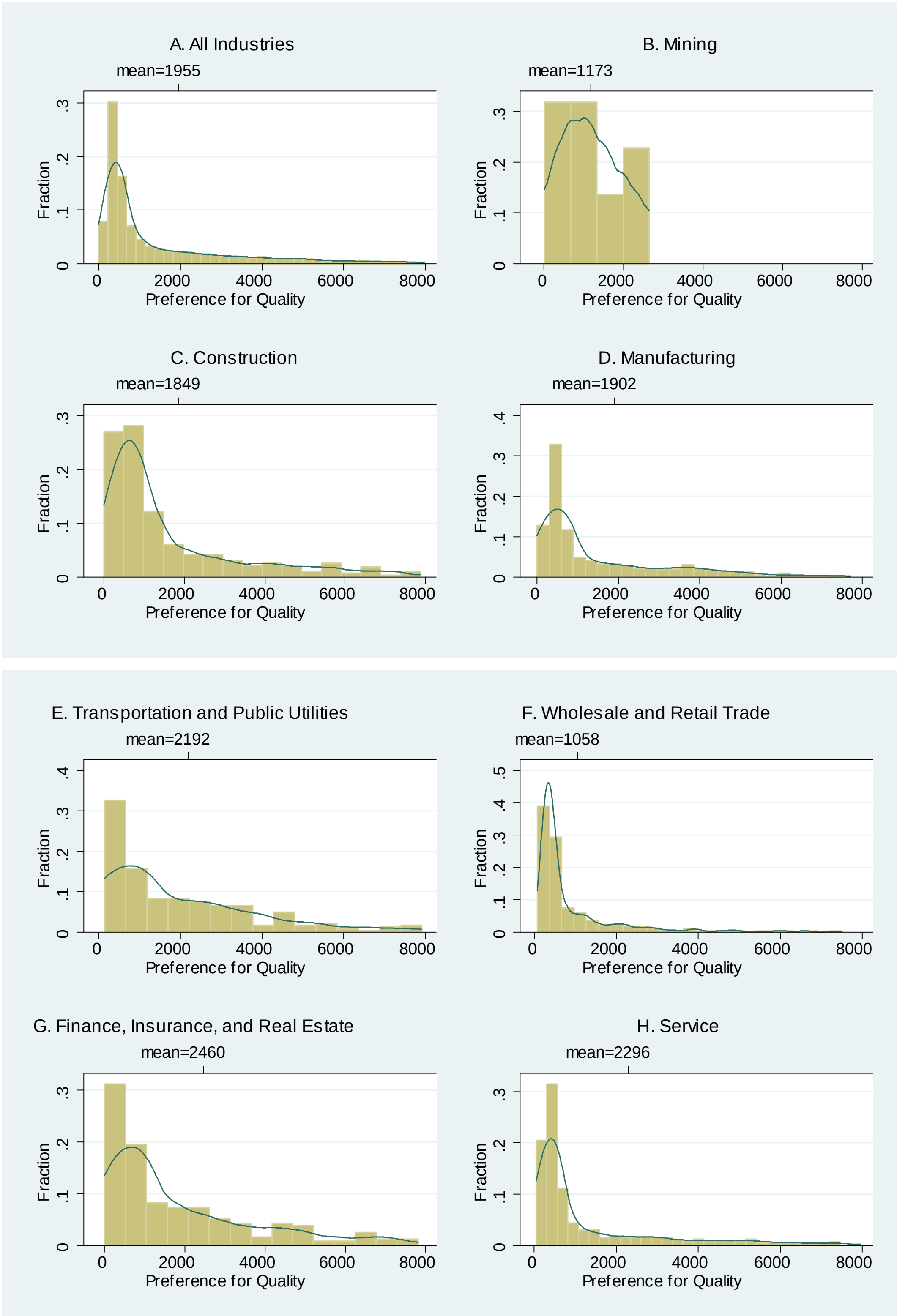

Figure 6: Firm Preference for Worker Quality Across Industries, 1993

Table 1. Estimated Wage Differentials for One-Digit Industries, NLSY79
(Standard Errors in Parentheses)

	1990 Cross Section				1993 Cross Section				Fixed Effects
Industry	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mining	0.211	0.287	0.275	0.276	0.112	0.126	0.118	0.120	0.159
	(0.097)	(0.082)	(0.081)	(0.081)	(0.119)	(0.098)	(0.097)	(0.097)	(0.211)
Construction	0.215	0.273	0.266	0.265	0.153	0.216	0.214	0.212	0.231
	(0.032)	(0.028)	(0.028)	(0.028)	(0.037)	(0.033)	(0.033)	(0.033)	(0.064)
Manufacturing	0.101	0.160	0.158	0.159	0.103	0.139	0.138	0.139	0.161
	(0.022)	(0.019)	(0.019)	(0.019)	(0.026)	(0.022)	(0.022)	(0.022)	(0.051)
Transportation, Communication,	0.208	0.178	0.174	0.172	0.224	0.168	0.164	0.163	0.065
Public Utilities	(0.033)	(0.028)	(0.028)	(0.028)	(0.038)	(0.032)	(0.032)	(0.032)	(0.065)
Wholesale and Retail Trade	-0.159	-0.083	-0.084	-0.086	-0.178	-0.118	-0.118	-0.120	-0.036
	(0.022)	(0.019)	(0.019)	(0.019)	(0.026)	(0.022)	(0.022)	(0.022)	(0.048)
Finance, Insurance, and Real Estate	0.228	0.173	0.166	0.166	0.233	0.142	0.143	0.142	0.133
	(0.034)	(0.029)	(0.029)	(0.029)	(0.038)	(0.033)	(0.033)	(0.033)	(0.071)
Other Control Variables	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
AFQT	No	No	Yes	Yes	No	No	Yes	Yes	No
Parental Education	No	No	No	Yes	No	No	No	Yes	No
Unweighted St.d. of Differentials	0.147	0.136	0.133	0.133	0.143	0.115	0.114	0.114	0.096
Weighted St.d. of Differentials	0.052	0.047	0.046	0.046	0.052	0.044	0.043	0.043	0.038
No. of Observations	4,266	4,266	4,266	4,266	3,522	3,522	3,522	3,522	877
Adjusted R Squared	0.056	0.345	0.355	0.356	0.048	0.370	0.376	0.376	0.044

Notes. The dependent variable is log (hourly wage). The reported estimates are the coefficient values for the industry dummy variables. The reference industry is service. Other control variables are education, years of experience and its square, gender dummy, race dummy, ever married dummy, union and veteran status, four region dummies, three occupation dummies, marriage and gender interaction, education and gender interaction, education squared and gender interaction, age and gender interaction, and a constant.

Table 2. First-Stage Regressions: Schooling and Experience

	Schooling		Experience	
	Coef.	(St.d.)	Coef.	(St.d.)
Local college	0.522	(0.102)	0.428	(0.123)
Missing local college	0.539	(0.137)	-0.499	(0.164)
Age	0.080	(0.021)	0.532	(0.026)
Female	0.292	(0.058)	-1.417	(0 .070)
Black	-0.419	(0.067)	-0.850	(0.081)
Ever married	-0.570	(0.064)	0.597	(0.077)
Year 1993	0.058	(0.084)	0.873	(0.100)
Constant	10.548	(0.588)	-8.065	(0.706)
F-Statistics for IVs	28.47		348.82	
Prob > F	0.000		0.000	
Observations	7,788		7,788	
Adjusted R Squared	0.024		0.238	

Table 3. Multinomial Logit Regressions for Industry Selection
(Standard Errors in Parentheses)

	Industries					
	(1)	(2)	(3)	(4)	(5)	(6)
Mother's schooling	0.046	-0.041	-0.015	0.030	-0.016	0.001
	(0.058)	(0.020)	(0.013)	(0.021)	(0.013)	(0.021)
Father's schooling	-0.120	-0.047	-0.061	-0.023	-0.019	0.038
	(0.047)	(0.016)	(0.011)	(0.016)	(0.011)	(0.017)
Female	-1.982	-3.140	-1.154	-1.319	-0.635	-0.066
	(0.335)	(0.155)	(0.066)	(0.099)	(0.065)	(0.100)
Black	-1.265	-0.380	-0.150	0.067	-0.301	-0.224
	(0.479)	(0.127)	(0.075)	(0.108)	(0.076)	(0.116)
Ever married	0.712	0.127	0.216	0.177	-0.093	0.002
	(0.361)	(0.102)	(0.072)	(0.105)	(0.070)	(0.106)
Year 1993	-0.172	-0.061	-0.095	-0.050	-0.187	0.010
	(0.287)	(0.095)	(0.065)	(0.095)	(0.064)	(0.095)
Constant	-2.647	0.403	0.727	-1.229	0.314	-2.090
	(0.587)	(0.189)	(0.137)	(0.217)	(0.139)	(0.227)
Chi-square statistic			1186.50			
Prob > Chi-square			0.000			
No. of Observations			7,788			
Pseudo R Squared			0.047			

Notes. Column (1) corresponds to mining industry; column (2) corresponds to construction; column (3) corresponds to manufacturing; column (4) corresponds to transportation, communication and public utilities; column (5) corresponds to wholesale and retail trade; column (6) corresponds to finance, insurance and real estate. The reference industry is service.

Table 4. Conditional Worker Quality Distribution

	1990		1993	
Normalized Worker Quality	Mean	(St.d.)	Mean	(St.d.)
All workers	0.461	(0.277)	0.507	(0.261)
By education				
High school incompletes	0.350	(0.250)	0.399	(0.241)
High school graduates	0.404	(0.257)	0.448	(0.251)
Some college	0.485	(0.266)	0.524	(0.253)
College graduates	0.636	(0.255)	0.658	(0.224)
By work experience				
0-4 years	0.355	(0.277)	0.328	(0.264)
5-9 years	0.485	(0.276)	0.500	(0.276)
10+ years	0.526	(0.242)	0.554	(0.227)
By AFQT percentile scores				
$AFQT \leq 25$	0.351	(0.241)	0.411	(0.248)
$25 < AFQT \leq 50$	0.458	(0.266)	0.518	(0.252)
$50 < AFQT \leq 75$	0.539	(0.271)	0.567	(0.248)
$AFQT > 75$	0.630	(0.266)	0.639	(0.237)
By industry				
Mining	0.549	(0.258)	0.542	(0.212)
Construction	0.552	(0.274)	0.555	(0.229)
Manufacturing	0.483	(0.263)	0.532	(0.242)
Transportation, Communication, Public Utilities	0.565	(0.270)	0.597	(0.236)
Wholesale and Retail Trade	0.352	(0.245)	0.398	(0.241)
Finance, Insurance, and Real Estate	0.561	(0.262)	0.609	(0.237)
Service	0.454	(0.284)	0.500	(0.276)

Table 5. Correlations of Estimated Quality and Observed Human Capital Variables

	1990 cross-section		
	Education	Experience	AFQT
Estimated quality	0.354	0.249	0.387
	1993 cross-section		
	Education	Experience	AFQT
Estimated quality	0.349	0.249	0.337
	1990 and 1993 pooled		
	Estimated quality in 1993		
Estimated quality in 1990	0.710		

Table 6. Estimated Wage Differentials for One-Digit Industries with Quality and WTP Estimates, NLSY79
(Standard Errors in Parentheses)

	1990 Cross Section				1993 Cross Section			
Industry	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mining	0.276	0.034	0.260	0.044	0.120	-0.013	0.174	0.027
	(0.081)	(0.028)	(0.068)	(0.024)	(0.097)	(0.018)	(0.064)	(0.019)
Construction	0.265	0.027	0.186	0.022	0.212	0.009	0.158	0.018
	(0.028)	(0.012)	(0.024)	(0.012)	(0.033)	(0.015)	(0.026)	(0.013)
Manufacturing	0.159	0.038	0.093	0.031	0.139	0.012	0.105	0.016
	(0.019)	(0.008)	(0.015)	(0.008)	(0.022)	(0.010)	(0.017)	(0.009)
Transportation, Communication,	0.172	-0.006	0.153	-0.000	0.163	-0.002	0.115	0.004
Public Utilities	(0.028)	(0.010)	(0.023)	(0.010)	(0.032)	(0.015)	(0.026)	(0.013)
Wholesale and Retail Trade	-0.086	0.012	-0.056	0.011	-0.120	0.004	-0.057	0.011
	(0.019)	(0.007)	(0.016)	(0.008)	(0.022)	(0.009)	(0.019)	(0.009)
Finance, Insurance, and Real Estate	0.166	0.022	0.134	0.024	0.142	0.005	0.102	0.010
	(0.029)	(0.010)	(0.021)	(0.009)	(0.033)	(0.018)	(0.025)	(0.015)
Worker Quality	No	Yes	No	Yes	No	Yes	No	Yes
Firm's Willingness to Pay (WTP)	No	No	Yes	Yes	No	No	Yes	Yes
Unweighted St.d. of Differentials	0.133	0.017	0.109	0.016	0.114	0.008	0.084	0.009
Weighted St.d. of Differentials	0.046	0.006	0.033	0.005	0.043	0.002	0.029	0.003
No. of Observations	4,266	4,266	4,266	4,266	3,522	3,522	3,522	3,522
Adjusted R Squared	0.356	0.889	0.540	0.895	0.376	0.873	0.633	0.901

Notes. The dependent variable is log (hourly wage). The reported estimates are the coefficient values for the industry dummy variables. The reference industry is service. Other control variables are education, years of experience and its square, gender dummy, race dummy, ever married dummy, union and veteran status, four region dummies, three occupation dummies, marriage and gender interaction, education and gender interaction, education squared and gender interaction, age and gender interaction, mother's schooling, father's schooling, AFQT test score, and a constant. The standard errors in the specifications including worker quality or firm's willingness to pay are bootstrapped.

Table 7. Decomposition of Inter-Industry Wage Differentials

	(1)	(2)	(3)
	1990 two-digit industry premiums		
Quality	1.034		1.177
	(0.101)		(0.151)
Firm preferences	No	Yes	Yes
R squared	0.686	0.393	0.726
Adjusted R squared	0.678	0.345	0.697
	1993 two-digit industry premiums		
Quality	1.068		0.848
	(0.112)		(0.262)
Firm preferences	No	Yes	Yes
R squared	0.700	0.654	0.745
Adjusted R squared	0.692	0.627	0.717

Table 8. Robustness Checks on Unobserved Worker Quality

	(1)	(2)	(3)	(4)
	1990 worker quality		1993 worker quality	
Job satisfaction dummies				
Very satisfied	0.123	0.117	0.107	0.101
	(0.025)	(0.024)	(0.027)	(0.027)
Satisfied	0.086	0.080	0.065	0.059
	(0.025)	(0.024)	(0.027)	(0.026)
Dissatisfied	0.032	0.029	0.054	0.053
	(0.029)	(0.028)	(0.031)	(0.030)
Firm size dummies				
Medium	0.068	0.063	0.076	0.073
	(0.011)	(0.011)	(0.011)	(0.011)
Large	0.123	0.112	0.159	0.142
	(0.011)	(0.011)	(0.012)	(0.012)
Constant	0.330	0.426	0.387	0.487
	(0.024)	(0.055)	(0.027)	(0.030)
One-digit industry dummies	No	Yes	No	Yes
R squared	0.042	0.095	0.058	0.100
Adjusted R squared	0.041	0.092	0.057	0.097

Notes. The reference job satisfaction dummy is "very dissatisfied," and the reference firm size dummy is small firm.

Appendix A: An Example of Deriving Log Linear Revenue Function

In what follows, we illustrate how a linear revenue function can be derived from common specifications of labor efficiency and production function. The output of worker j at employer i in market a is given by the production function $F_{i,a}(E_{i,j,a}, K_{i,a})$, where $E_{i,j,a}$ is the labor efficiency units of worker j when working for employer i in market a, and $K_{i,a}$ is the composite non-labor input, including all intermediate inputs and capital.

Employers are profit maximizers that choose labor input $E_{i,j,a}$ and non-labor input $K_{i,a}$ given market wage rate $w_{j,a}$, rental price $r_{i,a}$ of non-labor input $K_{i,a}$, and output price $p_{i,a}$. Formally, employer i's problem is

$$\max_{(E_{i,j,a},K_{i,a})\in\mathbb{R}_0^{2,+}} \pi_{i,a} = p_{i,a}F_{i,a}(E_{i,j,a}, K_{i,a}) - w_{j,a} - r_{i,a}K_{i,a}, \tag{33}$$

where the production function $F_{i,a}(E_{i,j,a}, K_{i,a})$ is assumed to be continuously differentiable and strictly increasing in $K_{i,a}$. The first-order condition on $K_{i,a}$ implicitly defines a unique employer-specific optimal choice of the composite non-labor input given its rental price, a labor efficiency level, and the output price.

$$\frac{\partial \pi_{i,a}}{\partial K_{i,a}} = p_{i,a}\frac{\partial F_{i,a}}{\partial K_{i,a}} - r_{i,a} = 0 \Longrightarrow K^*_{i,a} = K^*_{i,a}(E_{i,j,a}, p_{i,a}, r_{i,a}). \tag{34}$$

Replacing the optimal choice of non-labor input in (33) simplifies the employer's problem to

$$\max_{E_{i,j,a}\in\mathbb{R}_0^{+}} \pi_{i,a}\left(E_{i,j,a}\right) = R_{i,a}(E_{i,j,a}) - w_{j,a}, \tag{35}$$

where $R_{i,a}(E_{i,j,a})$ is the employer-specific revenue per worker net of non-labor cost; that is

$$R_{i,a}(E_{i,j,a}) = p_{i,a}F_{i,a}(E_{i,j,a}, K^*_{i,a}(E_{i,j,a}, p_{i,a}, r_{i,a})) - r_{i,a}K^*_{i,a}(E_{i,j,a}, p_{i,a}, r_{i,a}). \tag{36}$$

Without loss of generality, we focus on continuous, strictly positive worker attributes, and we suppress the market subindex $a \equiv (l, t)$ for ease of exposition. Consider the following specification for the labor efficiency units at employer i of worker j with characteristic vector $\left(X_j, \xi_j\right) = \left(x_{j,1}, x_{j,2}, \cdots, x_{j,M}, \xi_j\right)$:

$$E_{i,j} = \rho_{i,0} + \ln(X_j)\cdot \boldsymbol{\rho}_{i,X} + \rho_{i,\xi}\ln(\xi_j), \quad \forall j. \tag{37}$$

In addition, consider a CES production function:

$$F_i(E_{i,j}, K_i) = [\lambda_i E_{i,j}^{\sigma_i} + (1-\lambda_i)K_i^{\sigma_i}]^{1/\sigma_i},$$

where λ_i governs the income shares between labor and non-labor inputs and σ_i determines the elasticity of substitution between inputs.

The first-order condition of the employer's problem with respect to K_i implies that its optimal demand takes the form of $K_i^* = \delta_i E_{i,j}$, where

$$\delta_i = \left[\frac{\lambda_i}{\left(\frac{r_i}{p_i(1-\lambda_i)}\right)^{\sigma_i/(1-\sigma_i)} - (1-\lambda_i)} \right]^{1/\sigma_i}.$$

The profit from hiring worker j, given the optimal choice of non-labor input, becomes

$$\pi_{ij} = p_i F_i(E_{i,j}, \delta_i E_{i,j}) - w_j - r_i \delta_i E_{i,j}.$$

Therefore, the revenue function (net of capital costs) assumes the form $R_i(E_{i,j}) = \gamma_i E_{i,j}$, where is given by γ_i is given by

$$\gamma_i = p_i \left[\lambda_i + (1-\lambda_i)\delta_i^{\sigma_i}\right]^{1/\sigma_i} - r_i \delta_i.$$

Intuitively, γ_i represents the dollar value of the marginal productivity of labor efficiency units for employer i. Given the specification for labor efficiency (37), the revenue per worker function has the following parametric form

$$R_i(X_j, \xi_j; \boldsymbol{\beta}_i) = \gamma_i E_{i,j} = \beta_{i,0} + \ln(X_j) \cdot \boldsymbol{\beta}_{i,X} + \beta_{i,\xi} \ln(\xi_j),$$

where the coefficient vector β_i is the product of the vector of efficiency unit coefficients $\rho_i = (\rho_{i,0}, \rho_{i,X}, \rho_{i,\xi})$ and γ_i.

Appendix B: Proof of Proposition 1

Proposition 1 is illustrated as follows. For any two workers j and j' employed in market a, three conditions hold:

(1) If $X_{j,a} = X_{j',a}$ and $\xi_{j,t} = \xi_{j',t}$, then $w_{j,a} = w_{j',a}$.

(2) If $X_{j,a} = X_{j',a}$ and $\xi_{j,t} > \xi_{j',t}$, then $w_{j,a} > w_{j',a}$.

(3) $|w_{j,a} - w_{j',a}| \leq E \times (|X_{j,a} - X_{j',a}| + |\xi_{j,t} - \xi_{j',t}|)$ for some $E < \infty$.

Suppose that $w_{j,a} > w_{j',a}$ for some market a in which both workers j and j' are employed and $X_{j,a} = X_{j',a}$ and $\xi_{j,t} = \xi_{j',t}$. Then $R_{i,a}(X_{j,a}, \xi_{j,t}) - w_{j,a} < R_{i,a}(X_{j',a}, \xi_{j',t}) - w_{j',a}$ for all employers $i \in V_a$. This observation implies that no one would hire worker j in market a and is thus a contradiction.

Suppose that $w_{j,a} \leq w_{j',a}$ for some market a in which both workers j and j' are employed and $X_{j,a} = X_{j',a}$ and $\xi_{j,t} > \xi_{j',t}$. Given that $R_{i,a}(X_{j,a}, \xi_{j,t})$ strictly increases in $\xi_{j,t}$, $R_{i,a}(X_{j,a}, \xi_{j,t}) - w_{j,a} > R_{i,a}(X_{j',a}, \xi_{j',t}) - w_{j',a}$ for all employers $i \in V_a$. This observation implies that no one would hire worker j' in market a and is thus a contradiction.

The assumption that $R_{i,a}(X_{j,a}, \xi_{j,t})$ is Lipschitz-continuous in $(X_{j,a}, \xi_{j,t})$ implies that for any two workers j and j' differing in at least one characteristic,

$$|R_{i,a}(X_{j,a}, \xi_{j,t}) - R_{i,a}(X_{j',a}, \xi_{j',t})| \leq E \times (|X_{j,a} - X_{j',a}| + |\xi_{j,t} - \xi_{j',t}|),$$

for some $E < \infty$. Given that $|R_{i,a}(X_{j,a}, \xi_{j,t}) - R_{i,a}(X_{j'a}, \xi_{j't})| = |[(R_{i,a}(X_{j,a}, \xi_{j,t}) - w_{j,a}) - (R_{i,a}(X_{j',a}, \xi_{j',t}) - w_{j',a})] + (w_{j,a} - w_{j',a})|$,

$$\begin{aligned} & |[(R_{i,a}(X_{j,a}, \xi_{j,t}) - w_{j,a}) - (R_{i,a}(X_{j',a}, \xi_{j',t}) - w_{j',a})] + (w_{j,a} - w_{j',a})| \\ \leq \ & E \times (|X_{j,a} - X_{j',a}| + |\xi_{j,t} - \xi_{j',t}|). \end{aligned}$$

Assuming that without loss of generality $w_{j,a} > w_{j',a}$, then the second term on the left-hand side, $w_{j,a} - w_{j',a}$, is positive. Because the demand for worker j is positive, the first term on the left-hand side must be positive for some employer i. For these employers, we can ignore the absolute sign.

$$\begin{aligned} & |[(R_{i,a}(X_{j,a}, \xi_{j,t}) - w_{j,a}) - (R_{i,a}(X_{j',a}, \xi_{j',t}) - w_{j',a})] + (w_{j,a} - w_{j',a})| \\ = \ & [(R_{i,a}(X_{j,a}, \xi_{j,t}) - w_{j,a}) - (R_{i,a}(X_{j',a}, \xi_{j',t}) - w_{j',a})] + (w_{j,a} - w_{j',a}) > w_{j,a} - w_{j',a}. \end{aligned}$$

Therefore,

$$w_{j,a} - w_{j',a} \leq E \times (|X_{j,a} - X_{j',a}| + |\xi_{j,t} - \xi_{j',t}|) \text{ for employer } i \text{ that prefers } j \text{ over } j'.$$

In this instance, we use the fact that both workers have positive demand to limit how much their wages can vary.

Appendix C: Proof of Proposition 2

We use the assumption that each function $h_m(., \eta_m)$ is strictly monotonic in η_m to define $h_m^{-1}(x_{0,m}, X_1, Z, D_P)$ as the inverse of $h_m(X_1, Z, D_P, \eta_m)$. According to the proof of Lemma

1 of Matzkin (2003), for each $m = 1, ..., M_0$,

$$\begin{aligned} F_{X_{0,m}|X_1,Z,D_P}(x_{0,m}|x_1,z,t) &= \Pr(X_{0,m} \leq x_{0,m}|X_1 = x_1, Z = z, D_P = t) \\ &= \Pr(h_m(x_1,z,t,\eta_m) \leq x_{0,m}|X_1 = x_1, Z = z, D_P = t) \\ &= \Pr(\eta_m \leq h_m^{-1}(x_{0,m},x_1,z,t)|X_1 = x_1, Z = z, D_P = t) \\ &= \Pr(\eta_m \leq h_m^{-1}(x_{0,m},x_1,z,t)) \\ &= F_{\eta_m}(h_m^{-1}(x_{0,m},x_1,z,t)) = h_m^{-1}(x_{0,m},x_1,z,t) = \eta_m, \end{aligned} \tag{38}$$

where the second equality follows from the definition of the function $h_m(.)$, the third equality follows from the monotonicity assumption, the fourth equality follows from the independence between (X_1, Z, D_P) and η_m, and the last equality is the result of normalizing η_m so that it lies in $U[0,1]$.

Next, we show that the vector $\boldsymbol{\eta} \equiv (\eta_1, ..., \eta_{M_0})$ consists of control variables conditional on which (and on Dahl's controls $\tilde{\boldsymbol{\eta}}$) X and δ are independent by adapting the proof of Theorem 1 of Imbens and Newey (2009) for multiple endogenous variables. For any bounded function $p(x_0, x_1, l, t)$, it follows from the independence of (X_1, Z, D_P) and $(\delta, \boldsymbol{\eta})$ that

$$\begin{aligned} E[p_a(x_0,x_1,l,t)|\delta,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}] &= E[p(h_1(x_1,z,t,\eta_1),...,h_{M_0}(x_1,z,t,\eta_{M_0}),x_1,l,t)|\delta,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}] \\ &= \int \sum_{l=1}^{L} \{p(h_1(x_1,z,t,\eta_1),...,h_{M_0}(x_1,z,t,\eta_{M_0}),x_1,l,t) \times \tilde{\eta}_l\} \\ &\quad dF_{X_1,Z,,D_P}(x_1,z,t) \\ &= E[p(x_0,x_1,l,t)|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}], \end{aligned} \tag{39}$$

where the second equality follows from the definition of the expectation operator and that $\tilde{\eta}_l = \Pr(l|x_1, z_S, t)$ also by definition, for all industries $l = 1, ..., L$. This equality highlights that δ does not affect any of the arguments in the function $p(x_0, x_1, l, t)$, hence the independence established in the third equality. Thus, for any bounded function $b(\delta)$, it follows from the Law of Iterated Expectations that

$$\begin{aligned} E[p(x_0,x_1,l,t)b(\delta)|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}] &= E[b(\delta)E[p(x_0,x_1,l,t)|\delta,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}]|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}] \\ &= E[b(\delta)E[p(x_0,x_1,l,t)|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}]|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}] \\ &= E[b(\delta)|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}]E[p(x_0,x_1,l,t)|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}], \end{aligned}$$

which indicates the independence between (X, D_I, D_P) and δ conditional on $\boldsymbol{\eta}$ *and* $\tilde{\boldsymbol{\eta}}$.

In addition, the conditional CDF of wages in (18) evaluated at worker j's observables $(w_{j,a}, X_{j,a}, \boldsymbol{\eta}_{j,a}, \tilde{\boldsymbol{\eta}}_{j,a}, D_{I,j,t}, D_{P,j,t})$ at market $a = (l, t)$ simplifies to

$$
\begin{aligned}
&F_{w|X,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I,D_P}(w_{j,a}|X_{j,a},\boldsymbol{\eta}_{j,a},\tilde{\boldsymbol{\eta}}_{j,a},l,t) \\
&= \Pr(\widetilde{w}_a(X,\delta_{j,a}) \leq w_{j,a}|X = X_{j,a},\boldsymbol{\eta} = \boldsymbol{\eta}_{j,a},\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_{j,a},D_I = l, D_P = t) \\
&= \Pr(\delta \leq \widetilde{w}_a^{-1}(X,w_{j,a})|X = X_{j,a},\boldsymbol{\eta} = \boldsymbol{\eta}_{j,a},\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_{j,a},D_I = l, D_P = t) \\
&= \Pr(s(\xi,\epsilon) \leq \widetilde{w}_a^{-1}(X_{j,a},w_{j,a})|X = X_{j,a},\boldsymbol{\eta} = \boldsymbol{\eta}_{j,a},\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_{j,a},D_I = l, D_P = t) \\
&= \Pr(s(\xi,\epsilon) \leq \delta_{j,a}|X = X_{j,a},\boldsymbol{\eta} = \boldsymbol{\eta}_{j,a},\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_{j,a},D_I = l, D_P = t) \\
&= \Pr(\xi \leq s^{-1}(s(\xi_{j,t},\epsilon_{j,a}),\epsilon_{j,a})|X = X_{j,a},\boldsymbol{\eta} = \boldsymbol{\eta}_{j,a},\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_{j,a},D_I = l, D_P = t) \\
&= \Pr(\xi \leq \xi_{jt}|X = X_{j,a},\boldsymbol{\eta} = \boldsymbol{\eta}_{j,a},\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_{j,a},D_I = l, D_P = t),
\end{aligned}
\tag{40}
$$

where the first equality follows from the definition of our conditional CDF, the second one follows from the monotonicity of the wage function on δ, the third one results from both $X = X_{jt}$ and the assumption that δ is a function of both unobserved quality ξ and ϵ, the fourth one follows from the equivalence in (13) and the invertibility of the wage function at data point $(X_{j,a}, w_{j,a})$, denoted $\delta_{j,a}$, the fifth one follows from the monotonicity of $s(\xi,\epsilon)$ in the first argument, and the last equality results from applying the inverse function $s^{-1}(.,\epsilon)$ to the function $s(.,\epsilon)$ when the first argument equals $\xi_{j,t}$.

Our final step involves taking expectations with respect to industries $l = 1, ..., L$ and then integrating with respect to $\boldsymbol{\eta}$ and $\tilde{\boldsymbol{\eta}}$.. We replace the conditional wage CDF in (18) with its equivalent conditional CDF on ξ just derived. Given the independence between X and δ conditional on $(\boldsymbol{\eta},\tilde{\boldsymbol{\eta}})$ and the normalizations of ξ and $\boldsymbol{\eta}$, it follows from the Law of Total Probability that

$$
\begin{aligned}
&\int\limits_{\boldsymbol{\eta}\in[\mathbf{0},\mathbf{1}]^{M_0}}\int\limits_{\tilde{\boldsymbol{\eta}}\in[\mathbf{0},\mathbf{1}]^{L}} \left\{\sum_{l=1}^{L} F_{w|X,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I,D_P}(w_{j,a}|X_{j,a},\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I = l, D_P = t)\Pr(D_I = l|X_{j,a},\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_P = t)\right\}\mathbf{dG}(\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}) \\
&= \int\limits_{\boldsymbol{\eta}\in[\mathbf{0},\mathbf{1}]^{M_0}}\int\limits_{\tilde{\boldsymbol{\eta}}\in[\mathbf{0},\mathbf{1}]^{L}} \left\{\sum_{l=1}^{L}\Pr(\xi \leq \xi_{jt}|X_{j,a},\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I = l, D_P = t)\Pr(D_I = l|X_{j,a},\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_P = t)\right\}\mathbf{dG}(\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}) \\
&= \int\limits_{\boldsymbol{\eta}\in[\mathbf{0},\mathbf{1}]^{M_0}}\int\limits_{\tilde{\boldsymbol{\eta}}\in[\mathbf{0},\mathbf{1}]^{L}} \left\{\sum_{l=1}^{L}\Pr(\xi \leq \xi_{jt}, D_I = l|X_{j,a},\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_P = t)\right\}\mathbf{dG}(\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}) \\
&= \int\limits_{\boldsymbol{\eta}\in[\mathbf{0},\mathbf{1}]^{M_0}}\int\limits_{\tilde{\boldsymbol{\eta}}\in[\mathbf{0},\mathbf{1}]^{L}} \Pr(\xi \leq \xi_{jt}|X_{j,a},\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_P = t)\mathbf{dG}(\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}) \\
&= \int\limits_{\boldsymbol{\eta}\in[\mathbf{0},\mathbf{1}]^{M_0}}\int\limits_{\tilde{\boldsymbol{\eta}}\in[\mathbf{0},\mathbf{1}]^{L}} \Pr(\xi \leq \xi_{jt}|\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_P = t)\mathbf{dG}(\boldsymbol{\eta},\tilde{\boldsymbol{\eta}}) \\
&= F_{\xi,t}(\xi_{jt}) \\
&= \xi_{jt}.
\end{aligned}
\tag{41}
$$

Appendix D: Details of Estimation Steps

Our nonparametric estimators of conditional CDFs and PDFs are used in multiple instances. In what follows, we denote a random variable by Y and conditioning variables by U for the sake of generality. Specifically, our estimator for the conditional PDF $\hat{f}$ of a variable Y, given a $1 \times H$ vector of covariates U, is a weighted mixture of normal densities:

$$\hat{f}(Y|U;\boldsymbol{\theta}) \equiv \sum_{r=1}^{R(N)} \alpha_r(U, \boldsymbol{\theta}^\alpha)\phi(Y|\mu_r, \sigma_r), \tag{42}$$

where $R(N)$ represents the (integer) number of normal densities as an (increasing) function of sample size N, $\boldsymbol{\theta}$ is a vector of the parameters of the density function, and $\phi(.|\mu_r, \sigma_r)$ is a normal density with mean μ_r and standard deviation σ_r. The corresponding conditional CDF of Y is

$$\hat{F}(Y|U;\boldsymbol{\theta}) \equiv \sum_{r=1}^{R(N)} \alpha_r(U, \boldsymbol{\theta}^\alpha)\Phi(Y|\mu_r, \sigma_r), \tag{43}$$

where $\Phi(.|\mu_r, \sigma_r)$ denotes the CDF of the same normal distribution. Each normal density in Equation (42) is weighted by a multinomial logit function $\alpha_r(U, \boldsymbol{\theta}^\alpha)$ with an $(H+1) \times 1$ parameter vector $\boldsymbol{\theta}^\alpha$ defined as

$$\alpha_r(U;\boldsymbol{\theta}^\alpha) = \begin{cases} \dfrac{1}{1+\sum_{q=2}^{R(N)} \exp\left(\theta^\alpha_{0,q} + U \cdot \boldsymbol{\theta}^\alpha_{U,q}\right)} & if\ r=1, \\ \dfrac{\exp\left(\theta^\alpha_{0,r} + U \cdot \boldsymbol{\theta}^\alpha_{U,r}\right)}{1+\sum_{q=2}^{R(N)} \exp\left(\theta^\alpha_{0,q} + U \cdot \boldsymbol{\theta}^\alpha_{U,q}\right)} & if\ r=2,...,R(N). \end{cases} \tag{44}$$

Norets (2010) demonstrates that this specification approximates well the true conditional PDF of Y given U. We also use a multinomial logit function to model the fraction of workers in each industry $l = 1, ..., L$ given worker attributes U as

$$\lambda_l(U;\boldsymbol{\theta}^\lambda) = \begin{cases} \dfrac{1}{1+\sum_{s=2}^{L} \exp\left(\theta^\lambda_{0,s} + U \cdot \boldsymbol{\theta}^\lambda_{U,s}\right)} & if\ l=1, \\ \dfrac{\exp\left(\theta^\lambda_{0,l} + U \cdot \boldsymbol{\theta}^\lambda_{U,l}\right)}{1+\sum_{s=2}^{L} \exp\left(\theta^\lambda_{0,s} + U \cdot \boldsymbol{\theta}^\lambda_{U,s}\right)} & if\ l=2,...,L. \end{cases} \tag{45}$$

Our maximum likelihood estimator for the PDF of an endogenous attribute $x_{0,m}$, conditional on exogenous worker characteristics X_1 and an instrument set Z, is defined as

$$\hat{\boldsymbol{\theta}}_{x_{0,m}} \equiv \arg\max_{\boldsymbol{\theta}} \sum_{a=1}^{A}\sum_{j=1}^{J_a} \log\{\hat{f}(x_{0,m,j,a}|X_{1,j,a}, Z_{j,a}, D_{P,j,a};\boldsymbol{\theta})\}, \tag{46}$$

where J_a is the number of workers sampled in market a.[53]

We implement the estimation of (46) in R using the package *flexmix*. This package estimates mixtures of distributions and facilitates the selection of the number of mixtures by computing information criteria, such as Akaike's AIC and Swartz's BIC. This selection is analogous to the choice of smoothing parameters of other nonparametric estimators such as kernels or local linear regressions. We follow the standard practice of choosing the number of mixtures that achieves the lowest BIC.[54]

After $\hat{\boldsymbol{\theta}}_{x_{0,m}}$ is estimated for each $m = 1, ..., M_0$, the corresponding estimate for the control variable for each worker j in market a is

$$\eta_{j,a,m} = \hat{F}(x_{0,m,j,a}|X_{1,j,a}, Z_{j,a}, D_{P,j,a}; \hat{\boldsymbol{\theta}}_{x_{0,m}}). \tag{47}$$

We estimate Dahl's (2002) controls $\tilde{\eta}$ defined by the multinomial logit in (17) by maximum likelihood using the probability formulas in (45) with $U = (X_1, Z_S, D_P)$.[55]

Our maximum likelihood estimator for the PDF of wages conditional on observed worker attributes X, control variables $(\eta, \tilde{\eta})$ and industry and time dummies is

$$\hat{\boldsymbol{\theta}}_w \equiv \arg\max_{\boldsymbol{\theta}} \sum_{a=1}^{A} \sum_{j=1}^{J_a} \log\{\hat{f}(w_{j,a}|X_{j,a}, \boldsymbol{\eta}_{j,a}, D_{I,j,a}, \tilde{\boldsymbol{\eta}}_{j,a}, D_{P,j,a}; \boldsymbol{\theta})\}. \tag{48}$$

With control variable estimates of $\eta_{j,a,m}$ for all m and Dahl's (2002) controls $\tilde{\eta}_{j,a,m}$ for all industries $l = 1, ..., L$, $\hat{\theta}_w$ is obtained by solving Equation (48) using the same approach as for (46). Finally, we estimate the multinomial model for the probability of industry affiliation $\Pr(D_I = l|X, \eta, \tilde{\eta}, D_P)$ by maximum likelihood using the probability formulas in (45) with $U = (X, \eta, \tilde{\eta}, D_P)$. We observe that these estimated probabilities of industry affiliation is different from Dahl's controls $\tilde{\eta}$ discussed above because the estimation of $\tilde{\eta}$ does include (X_0, η) as regressors. The estimation of $\tilde{\eta}$ addresses endogeneity from industry self-selection, whereas $\Pr(D_I = l|X, \eta, \tilde{\eta}, D_P)$ is used to integrate industry effects out using the Law of Total Probability in the second line of (41).

Aside from the approximation and consistency properties (Norets, 2010), Normal density mixtures have an advantage over other nonparametric estimators in regards to numerical

[53]The data used in our application includes weights corresponding to the individual inverse-probability weights in each cross-section of the NSLY from year t. However, consistent with the results of Wooldridge (1999, 2001), we did not find meaningful differences in model estimates with individual weights.

[54]This is done by starting estimation with two mixtures and increasing the number of mixtures until the BIC seems convergent to a minimum. In addition, the package *flexmix* will estimate a lower number of mixtures in case the requested number of mixtures proves excessive for the data at hand. Intuitively, this corresponds to eliminating those normal distributions in the mixture with weights very close to zero.

[55]We implement this estimation in R using the function *multinom* of the package *nnet*.

integrations needed to calculate unobserved worker quality. Replacing integration with averaging (19) over sample values of controls $(\eta, \tilde{\eta})$ yields the empirical formula

$$\hat{\xi}_{jt} = \frac{1}{S}\sum_{s=1}^{S}\left\{\sum_{l=1}^{L}\hat{F}_{w|X,\boldsymbol{\eta},\tilde{\boldsymbol{\eta}},D_I,D_P}(w_{j,a}|X_{j,a},\boldsymbol{\eta}_s,\tilde{\boldsymbol{\eta}}_s,D_I=l,D_P=t;\hat{\boldsymbol{\theta}}_w)\lambda_l(X_{j,a},\boldsymbol{\eta}_s,\tilde{\boldsymbol{\eta}}_s,D_P=t;\hat{\boldsymbol{\theta}}_{\lambda,t})\right\} \tag{49}$$

where S represents the size of all market samples combined for which estimated values of $(\eta, \tilde{\eta})$ are available. Each value of this sum can be easily evaluated using the formulas in (43) and (45) at the estimated parameters $(\hat{\theta}_w, \hat{\theta}_{\lambda,t})$ and at each given data point $(w_{j,a}, X_{j,a}, D_{P,j,a})$ using all S draws for $(\eta, \tilde{\eta})$. Conventional nonparametric alternatives such as kernel estimators would involve a first round of averaging kernel functions over all observations in the data for each given observation of interest (e.g., as in local linear estimation in (26), for each sample value of controls $(\eta_s, \tilde{\eta}_s)$ $s = 1, ..., S$. And then the researcher would need another round of averaging across S draws to perform integration. In contrast, (43) only requires averaging across S draws after evaluating simple functions, yielding lower computational burden and simpler programming.