Unobserved Worker Quality and Inter-Industry Wage Differentials†

Suqin Ge
Virginia Tech

João Macieira
U.S. Department of Transportation

March 11, 2020

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. 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). 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. Recent advances in the estimation of matching games further highlight the importance of unobserved heterogeneities, such as those of firms and workers, on quantifying equilibrium matching outcomes (Fox, 2018; Fox, Yang and Hsu, 2018).

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. Several studies (Abowd, Kramarz and Margolis, 1999; Abowd et al., 2005) have decomposed inter-industry 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 model of labor demand and apply a two-stage nonparametric procedure to recover unobserved worker and firm heterogeneity. First, 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 from Denmark to explain productivity dispersion across firms.

nonparametrically recover unobserved worker quality by using an estimator based on Bajari and Benkard (2005), Imbens and Newey (2009), and Norets (2010). 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 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. 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.

Since the pioneer work of Rosen (1974), hedonic models have been widely used in empirical literature. Our approach builds on the classic hedonic model and borrows insights from recent work on estimating demand for differentiated products in industrial organization.[3] 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, Levinsohn and Pakes (1995). Fox et al. (2012) formally establish the identification of differentiated product demand models without supply-side assumptions. These models have been applied widely to quantify the role of unobserved product quality on market outcomes.[4] However, 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, Train and Ben-Akiva, 1987). Recent developments in using hedonic approach to estimate differentiated product demand models have alleviated this concern (e.g., Bajari and Benkard, 2005; Bayer, Ferreira and McMillan, 2007), but they typically focus on one single

[3]Most of hedonic literature 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, Heckman, and Nesheim (2004) and Heckman, Matzkin, and Nesheim (2010) thoroughly discuss identification issues in estimating hedonic models.

[4]For example, Khandelwal (2010) uses a model of differentiated product demand to estimate heterogeneity in product markets' scope for quality differentiation. Amiti and Khandelwal (2013) use a similar approach to measure product quality in US imports and estimate the relationship between import tariff and quality upgrading.

market.

In this paper, we 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. Literature on industrial organization has proposed nonparametric methods to identify product characteristics observed by consumers but not by researchers without explicit assumptions on supply-side behavior (e.g., Bajari and Benkard, 2005).[5] We use these methods to recover unobserved worker quality and extend the existing approach to allow for data to be drawn from multiple labor markets varying by industry and by time period. As in hedonic literature, the marginal prices of worker characteristics are estimated as random coefficients in a hedonic wage function.

Our labor demand model is estimated on the basis of individual data from the National Longitudinal Survey of Youth 1979 (NLSY79) to explore the importance of worker and firm effects in wage determination.[6] We estimate the model separately 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.[7]

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

[5] 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.

[6] Our application is related to many applications involving demand for differentiated products in the industrial organization literature. For example, Nevo (2001) estimates the demand for ready-to-eat cereal to study the importance of product quality differentiation, multi-product pricing and potential firm collusion in price-cost margins. While there are clear differences between differentiated product markets and labor markets, there are similarities in the demand side behavior that we exploit. It is reasonable to assume that a firm chooses among candidates to fill a vacancy agreeing with the firm's preferences. Similar to product label inspection and in-store sample tasting, the firm can learn about job candidates' attributes and unobserved quality from inspecting CVs and interviews. Supply-side assumptions are necessary for policy counterfactuals, but unnecessary to identify unobserved quality and preferences for attributes. Recent applications to markets with substantially different supply sides, such as computers (Bajari and Benkard 2005) and housing (Bayer, Ferreira and McMillan, 2007; Bajari and Khan 2005), confirm the validity of this approach.

[7] Using matched employer–employee panel data from France, Abowd, Kramarz, and Margolis (1999) also find that worker effects are more important than firm effects in explaining inter-industry wage differentials. However, those authors assume both work and firm effects to be fixed over time, whereas we allow them to vary.

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 augmented by estimated unobserved worker quality.

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. 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.

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 $X_{j,t}$ denote a $1 \times M$ vector of worker j's observed characteristics at time t. 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 $\xi_{j,t}$ as representing worker j's unmeasured quality at time t, which is rewarded in all labor markets. The observed worker characteristics $X_{j,t}$ and the unobserved worker quality $\xi_{j,t}$ are both worker- and time-specific, but they do not vary across industries.

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. The variable $E_{i,j,a}$ measures the different skill levels of labor in terms of different quantities of efficiency unit.[8].

Employers are profit maximizers that choose labor input $E_{i,j,a}$ and non-labor input $K_{i,a}$

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

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

$$\max_{(E_{i,j,a},K_{i,a})\in\mathbb{R}_0^{2,+}} \pi_{i,a} = p_{i,a}F_{i,a}(E_{i,j,a}, K_{i,a}) - w_{j,a} - r_{i,a}K_{i,a}, \tag{1}$$

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.

$$\frac{\partial \pi_{i,a}}{\partial K_{i,a}} = p_{i,a}\frac{\partial F_{i,a}}{\partial K_{i,a}} - r_{i,a} = 0 \Longrightarrow K^*_{i,a} = K^*_{i,a}(E_{i,j,a}, p_{i,a}, r_{i,a}). \tag{2}$$

Replacing the optimal choice of non-labor input in (1) simplifies the employer's problem to

$$\max_{E_{i,j,a}\in\mathbb{R}_0^{+}} \pi_{i,a}\left(E_{i,j,a}\right) = R_{i,a}(E_{i,j,a}) - w_{j,a}, \tag{3}$$

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

$$R_{i,a}(E_{i,j,a}) = p_{i,a}F_{i,a}(E_{i,j,a}, K^*_{i,a}(E_{i,j,a}, p_{i,a}, r_{i,a})) - r_{i,a}K^*_{i,a}(E_{i,j,a}, p_{i,a}, r_{i,a}). \tag{4}$$

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

$$\max_{X_{j,t},\xi_{j,t}} \pi_{i,a}(X_{j,t}, \xi_{j,t}) = R_{i,a}(X_{j,t}, \xi_{j,t}) - w_{j,a}. \tag{5}$$

Note that the vacancy profit function in (5) 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 $a = 1, ..., A$, $R_{i,a}(X_{j,t}, \xi_{j,t})$ is (i) Lipschitz continuous in $(X_{j,t}, \xi_{j,t})$ 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 $w_a(X_{j,t}, \xi_{j,t})$ that is strictly increasing in $\xi_{j,t}$ for each market $a = 1, ..., A$.*

The proof is provided in Appendix A. We follow a similar strategy taken by Bajari and Benkard (2005) in their demand model for differentiated products. The wage function is not 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 $w_a(X_{j,t}, \xi_{j,t})$, the firm problem in (5) becomes

$$\max_{X_{j,t}, \xi_{j,t}} \pi_{i,a}(X_{j,t}, \xi_{j,t}) = R_{i,a}(X_{j,t}, \xi_{j,t}) - w_a(X_{j,t}, \xi_{j,t}). \tag{6}$$

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

$$\frac{\partial R_{i,a}(X_{j^*,t}, \xi_{j^*,t})}{\partial x^c_{j,m,t}} = \frac{\partial w_a(X_{j^*,t}, \xi_{j^*,t})}{\partial x^c_{j,m,t}}, \tag{7}$$

$$\frac{\partial R_{i,a}(X_{j^*,t}, \xi_{j^*,t})}{\partial \xi_{j,t}} = \frac{\partial w_a(X_{j^*,t}, \xi_{j^*,t})}{\partial \xi_{j,t}}. \tag{8}$$

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 $R_{i,a}(X_{j,t}, \xi_{j,t})$ 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:

$$R_a(X_{j,t}, \xi_{j,t}; \boldsymbol{\beta}_{i,a}) \equiv \beta_{i,a,0} + \ln(X_{j,t}) \cdot \boldsymbol{\beta}_{i,a,X} + \beta_{i,a,\xi} \ln(\xi_{j,t}). \tag{9}$$

In this specification, each firm i's revenue is linear in the logarithms of worker attributes $(X_{j,t}, \xi_{j,t})$.[9] Coefficients $\boldsymbol{\beta}_{i,a,X}$ and $\beta_{i,a,\xi}$ represent employer i's preference for characteristic

[9] Without loss of generality and for ease of exposition, we assume that all observed characteristics are strictly positive. The log-linear form in (9) can accommodate binary variables by adding linear functions on

vector $X_{j,t}$ 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 be identical across individuals. Although seemingly arbitrary, the log-linearity assumption can be derived under mild conditions on model primitives.[10] Appendix B 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 (9), the employer's problem in Equation (6) becomes

$$\max_{X_{j,t},\xi_{j,t}} \beta_{i,a,0} + \ln(X_{j,t}) \cdot \boldsymbol{\beta}_{i,a,X} + \beta_{i,a,\xi} \ln(\xi_{j,t}) - w_a(X_{j,t}, \xi_{j,t}). \tag{10}$$

The firm's first-order conditions in Equations (7) and (8) on any continuous characteristic $x^c_{j,m,t}$ and $\xi_{j,t}$ evaluated at the observed optimal choice j^* become

$$\beta_{i,a,x^c_{j,m,t}} = \frac{\partial w_a(X_{j^*,t}, \xi_{j^*,t})}{\partial x^c_{j,m,t}} x^c_{j^*,m,t} = \frac{\partial w_a(X_{j^*,t}, \xi_{j^*,t})}{\partial x^c_{j,m,t}/x^c_{j^*,m,t}}, \tag{11}$$

$$\beta_{i,a,\xi} = \frac{\partial w_a(X_{j^*,t}, \xi_{j^*,t})}{\partial \xi_{j,t}} \xi_{j^*,t} = \frac{\partial w_a(X_{j^*,t}, \xi_{j^*,t})}{\partial \xi_{j,t}/\xi_{j^*,t}}. \tag{12}$$

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.[11]

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 (11).[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 (6). For example, suppose that firm i hires worker j^*. Let $\hat{X}_{j^*t}$ and $\bar{X}_{j^*t}$ denote the

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.

[10] 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.

[11] For a discussion on how to interpret similar coefficients in the context of housing demand, see Bajari and Khan (2005).

[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.

vectors of observed characteristics with $female = 1$ and $female = 0$, respectively, and all other elements equal the corresponding observed attributes in vector X_{j^*t}. The implicit price faced by employer i for a female worker is then $w_a(\hat{X}_{j^*t}, \xi_{j^*t}) - w_a(\bar{X}_{j^*t}, \xi_{j^*t})$. $\beta_{i,a,,f}$ is denoted as the coefficient for the female dummy in the revenue function. Profit maximization implies that $\beta_{i,a,f} > w_a(\hat{X}_{j^*t}, \xi_{j^*t}) - w_a(\bar{X}_{j^*t}, \xi_{j^*t})$ if worker j^* is female and $\beta_{i,a,f} \leq w_a(\hat{X}_{j^*t}, \xi_{j^*t}) - w_a(\bar{X}_{j^*t}, \xi_{j^*t})$ 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]

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 (9), we let firm heterogeneity be summarized by $\boldsymbol{\beta}_{i,a}$. Each worker j solves the following problem

$$\max_{X_{j,t}, \xi_{j,t}} U_{j,a}(X_{j,t}, \xi_{j,t}; \boldsymbol{\beta}_{i,a}) = w_a(X_{j,t}, \xi_{j,t}) - C_a(X_{j,t}, \xi_{j,t}; \boldsymbol{\chi}_{j,a}, \boldsymbol{\beta}_{i,a}), \tag{13}$$

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 a 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 (6), 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 (13), 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

[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 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.

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 under very general conditions.[15]

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, Imbens and Newey, 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 the additional challenge that our unobserved worker attribute of main interest (worker quality) is time- and worker-specific. We separate worker quality from market-specific regression residuals by integrating industry effects out, exploiting the insight that quality is worker-specific and not a function of the industry an individual works in.

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 a 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 an extended version of the estimators proposed by Imbens and Newey (2009).[16] In the second stage, we use the first-order conditions in Equations (11) and (12) to infer firm-specific parameters on their WTP for continuous worker characteristics.

[15] For an in-depth discussions on hedonic market equilibrium and the conditions for identification and estimation, see Heckman et al. (2010).

[16] This extension builds on Matzkin (2003), which demonstrates that the unobserved component in a nonseparable function is identified only up to a normalization. See Appendix C for details.

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 of equations in each market. Let X_0 and X_1 be the sub-vectors of the vector of the observed characteristics such that $X = (X_0, X_1)$.[17] In what follows, we use the notation $F_{Y|X,a}(Y|X)$ to denote the CDF of Y conditional on X in a market $a \equiv (l,t)$. 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.[18] Sub-vector X_1 represents the vector of exogenous variables.

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

$$w = w_a(X_0, X_1, \delta_a), \tag{14}$$

where $w_a(.)$ is an unknown, market-specific wage function that is strictly increasing in a scalar δ_a for each X. Let $\delta_a = \xi + \epsilon_a$, where ξ denotes unobserved worker quality, and ϵ_a denotes other unmeasured factors net of worker-specific effects.[19] As in Torgovitsky (2015), we also assume that there exist market-specific functions $h_{a,m}$ such that

$$x_{0,m} = h_{a,m}(X_1, Z, \eta_{a,m}),\ m = 1, ..., M_0, \tag{15}$$

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

[18] 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).

[19] As wage functions are market-specific, the unobserved scalar δ_a may mix worker quality with other unobservables. Additive separability is not necessary to establish our results. More generally, any function $f : \mathbb{R}^2 \to \mathbb{R}$ for which $\delta_a = f(\xi, \epsilon_a)$ is invertible in the first argument would work. This also nests the multiplicatively separable case, where $f(\xi, \epsilon_a) = f_1(\xi) \times f_2(\epsilon_a)$ and $f_2(\epsilon_a) \neq 0$ for every ϵ_a.

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 $\eta_{a,m}$. 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 (14) and (15) form a triangular system that is point identified (Torgovitsky 2015).

A control variable is a variable conditional on which X and δ_a 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 (15) by $\boldsymbol{\eta}_a = (\eta_{a,1}, ..., \eta_{a,M_0})$. The function $h_{a,m}(.)$ is defined for each market a, and therefore estimations involving this function only use data from market a. The following proposition shows that $\boldsymbol{\eta}_a$ is a vector of control variables that can be used to estimate unobserved worker quality ξ.

Proposition 2 *Let $F_{x_{0m}|X_1,Z,a}(.|.)$ denote the CDF of the endogenous characteristic $x_{0,m}$ conditional on the vector of exogenous characteristics X_1 and an instrument set Z in a given market $a \equiv (l, t)$. If each $\eta_{a,m}$ is normalized to lie in the interval $[0, 1]$ such that for each $m = 1, ..., M_0$, $\eta_{a,m} = F_{x_{0m}|X_1,Z,a}(x_{0m}|X_1, Z)$, then in each market $a = 1, ..., A$, X and ξ are independent conditional on $\boldsymbol{\eta}_a = (\eta_{a,1}, ..., \eta_{a,M_0})$. Moreover, unobserved worker quality is given by*

$$\xi = \sum_{l=1}^{L} \left\{ \int_{\boldsymbol{\eta}_a \in [\mathbf{0},\mathbf{1}]^{M_0}} F_{w|X,\boldsymbol{\eta},a}(w|X, \boldsymbol{\eta}_a) \mathbf{dG}_a(\boldsymbol{\eta}_a) \right\} \Pr(l|X, t), \tag{16}$$

where $\mathbf{G}_a(\boldsymbol{\eta}_a)$ is the joint CDF of the control variables in market a, and $\Pr(l|X, t)$ is the probability for a worker with characteristics X 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.[20] Note that Equation (16) involves taking expectations over all industries $l = 1, ..., L$ given individuals' observed attributes X at time period t. Thus, ξ is a highly nonlinear function that is

[20] Torgovitsky (2015, 2017) also defines a vector of control variables equivalent to $\boldsymbol{\eta}_a$. Here we focus on identifying and estimating unobserved worker quality out of estimable functions. Intuitively, this is achieved 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 observables.

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 $m = 1, ..., M_0$ and in each market $a = 1, ..., A$, we estimate the values of $\eta_{a,m}$ by using an empirical analog of $F_{x_{0m}|X_1,Z,a}(.|.)$. Second, we use the recovered series of $\eta_{a,m}$ to nonparametrically estimate $F_{w|X,\boldsymbol{\eta},a}(.|.)$, the integrand function in Equation (16). Third, the integrand is estimated by integrating $\boldsymbol{\eta}_a$ out by using Halton draws of an M_0-dimensional unit cube.[21] Fourth, we estimate $\Pr(l|X, t)$ by using proportions of workers across industries conditional on worker characteristics X at time t.

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, Fox and Ryan, 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 (16) than other methods.

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:

$$\hat{f}(Y|U;\boldsymbol{\theta}) \equiv \sum_{r=1}^{R(N)} \alpha_r(U, \boldsymbol{\theta}^{\alpha})\phi(Y|\mu_r, \sigma_r), \tag{17}$$

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

$$\hat{F}(Y|U;\boldsymbol{\theta}) \equiv \sum_{r=1}^{R(N)} \alpha_r(U, \boldsymbol{\theta}^{\alpha})\Phi(Y|\mu_r, \sigma_r), \tag{18}$$

[21] Halton draws consist of a sequence of numbers within the unit interval that uses a prime number as its base (Halton, 1960). For example, the first eight numbers in the sequence corresponding to base 3 are 1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, and 8/9. To span the domain of the M_0-dimensional unit cube, Halton draws can be formed by using different prime numbers for each dimension. Halton draws exhibit advantages over random draws from $U[0, 1]$ in terms of low variance and few draws (Bhat, 2001; Petrin and Train, 2010). To integrate with respect to $\mathbf{G}_a(\boldsymbol{\eta}_a)$, the Halton draws can be used directly after replacing $\mathbf{dG}_a(\boldsymbol{\eta}_a)$ with $\mathbf{g}(\boldsymbol{\eta}_a)\mathbf{d}\boldsymbol{\eta}_a$ if $\mathbf{G}_a(\boldsymbol{\eta}_a)$ is differentiable, where $\mathbf{g}(\boldsymbol{\eta}_a)$ is the joint PDF of $\boldsymbol{\eta}_a$.

where $\Phi(.|\mu_r, \sigma_r)$ denotes the CDF of the same normal distribution. Each normal density in Equation (17) 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

$$\alpha_r(U; \boldsymbol{\theta}^\alpha) = \begin{cases} \dfrac{1}{1 + \sum_{q=2}^{R(N)} \exp\left(\theta_{0,q}^\alpha + U \cdot \boldsymbol{\theta}_{U,q}^\alpha\right)} & if\ r = 1, \\ \dfrac{\exp\left(\theta_{0,r}^\alpha + U \cdot \boldsymbol{\theta}_{U,r}^\alpha\right)}{1 + \sum_{q=2}^{R(N)} \exp\left(\theta_{0,q}^\alpha + U \cdot \boldsymbol{\theta}_{U,q}^\alpha\right)} & if\ r = 2, ..., R(N). \end{cases} \quad (19)$$

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

$$\lambda_l(U; \boldsymbol{\theta}^\lambda) = \begin{cases} \dfrac{1}{1 + \sum_{s=2}^{L} \exp\left(\theta_{0,s}^\lambda + U \cdot \boldsymbol{\theta}_{U,s}^\lambda\right)} & if\ l = 1, \\ \dfrac{\exp\left(\theta_{0,l}^\lambda + U \cdot \boldsymbol{\theta}_{U,l}^\lambda\right)}{1 + \sum_{s=2}^{L} \exp\left(\theta_{0,s}^\lambda + U \cdot \boldsymbol{\theta}_{U,s}^\lambda\right)} & if\ l = 2, ..., L. \end{cases} \quad (20)$$

In each market $a \equiv (l, t)$, 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

$$\hat{\boldsymbol{\theta}}_{x_{0,m}} \equiv \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^{J_a} \omega_{jt} \log\{\hat{f}(x_{0,m,j,t}|X_{1,j,t}, Z_{jt}; \boldsymbol{\theta})\}, \quad (21)$$

where J_a is the number of workers sampled from market a in the data, and ω_{jt} is a sampling weight reflecting how many workers worker j represents in the population at time t.[22] 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

$$\eta_{j,a,m} = \hat{F}_a(x_{0,m,j,t}|X_{1,j,t}, Z_{jt}; \hat{\boldsymbol{\theta}}_{x_{0,m}}). \quad (22)$$

[22] We need to select $R(N)$ in order to obtain estimates of distribution parameters in Equation (21). This is analogous to the selection of smoothing parameters of other nonparametric estimators such as kernels or local linear regressions. Following Bajari and Benkard (2005) and Bajari and Khan (2005), among others, we guide our choice by visual inspection of the estimates. Our starting point for choosing the number of normal distribution in the mixture is $R(N) = int(\sqrt{N/2})$, a rule of thumb proposed by Mardia, Kent and Bibby (1979). We then decrease the value for $R(N)$ to obtain a model as parsimonious as possible provided that it does not change the estimated distribution significantly. Intuitively, this corresponds to eliminating those normal distributions in the mixture with weights close to zero. In our application, the weights ω_{jt} correspond to the individual weights in each cross-section of the NSLY from year t.

Our maximum likelihood estimator for the PDF of wages conditional on observed worker attributes X and control variables $\boldsymbol{\eta}_{j,a}$ in each market a is

$$\hat{\boldsymbol{\theta}}_w \equiv \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^{J_a} \omega_{jt} \log\{\hat{f}(w_{jt}|X_{jt}, \boldsymbol{\eta}_{j,a}; \boldsymbol{\theta})\}. \tag{23}$$

With control variable estimates of $\eta_{j,a,m}$ for all m, $\hat{\boldsymbol{\theta}}_w$ is obtained by solving Equation (23). Finally, we estimate the multinomial model for $\Pr(l|X,t)$ by maximum likelihood in each period t :

$$\hat{\boldsymbol{\theta}}_{\lambda,t} \equiv \arg\max_{\boldsymbol{\theta}^{\lambda}} \sum_{a=1}^{A} \mathbf{1}(a \in B(t)) \left[\sum_{j=1}^{J_a} \omega_{jt} \left\{ \sum_{l=1}^{L} \mathbf{1}(D_{jt} = l) \log(\lambda_l(X_{jt}, \boldsymbol{\theta}^{\lambda})) \right\} \right]. \tag{24}$$

where $B(t)$ represents the set of markets where the time period is equal to t, and D_{jt} is an indicator of the industrial affiliation of individual j at time t.[23]

We can then estimate the unobserved quality of each worker j in market a by using Equation (16):

$$\hat{\xi}_{jt} = \sum_{l=1}^{L} \left\{ \int_{\boldsymbol{\eta}_a \in [\mathbf{0},\mathbf{1}]^{M_0}} \hat{F}_{w|X,\boldsymbol{\eta},a}(w_{jt}|X_{jt}, \boldsymbol{\eta}_a; \hat{\boldsymbol{\theta}}_w) \mathbf{d}\hat{\mathbf{G}}_a(\boldsymbol{\eta}_a) \right\} \lambda_l(X_{jt}; \hat{\boldsymbol{\theta}}_{\lambda,t}), \tag{25}$$

where $\hat{\mathbf{G}}_a(\boldsymbol{\eta}_a)$ represents the empirical analog of $\mathbf{G}_a(\boldsymbol{\eta}_a)$.[24]

3.2 Estimation of Firm WTP Parameters

The labor demand problem described in Equation (6) is characterized by the revenue-per-worker function $R_{i,a}(X_{j,t}, \xi_{j,t})$. As discussed in the previous section, we consider a log-linear function for $R_a(X_{j,t}, \xi_{j,t}; \boldsymbol{\beta}_{i,a})$ (Equation 9), where we consider a function linear in logarithms of continuous variables and in levels of discrete variables. Under this model specification, Equation (11) suggests that if we recover an estimate of $\partial w_a(X_{j^*t}, \xi_{j^*t})/\partial x^c_{j,m,t}$, then we can learn a firm's WTP for worker characteristic m. As we observe each worker's characteristics

[23] Although there could be efficiency gains from pooling observations from different markets in the estimation, we choose to estimate the model for each market a separately for several reasons. First, if wages differ across markets due to differences in market equilibrium, then pooling observations from different markets is invalid. Second, if the market index t represents years, estimating the model by year allows one to identify unmeasured worker quality without imposing structure on its evolution over time.

[24] We use Halton draws after replacing $\mathbf{dG}_a(\boldsymbol{\eta}_a)$ with $\mathbf{g}(\boldsymbol{\eta}_a)\mathbf{d}\boldsymbol{\eta}_a$. We have considered two alternative estimators for $\mathbf{g}(\boldsymbol{\eta}_a)$: a multivariate uniform density, and a multivariate kernel density estimator, as in Duong (2007). In our application, both methods generate similar worker quality estimates.

in our data, we can flexibly estimate $\partial w_a(X_{j^*t}, \xi_{j^*t})/\partial x^c_{j,m,t}$ by using nonparametric methods. After we recover unobserved worker quality, we can also estimate a firm's WTP for unobserved quality based on $\partial w_a(X_{j^*t}, \xi_{j^*t})/\partial \xi_{jt}$, following Equation (12).

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 $\partial w_a(X_{j^*t}, \xi_{j^*t})/\partial \xi_{jt}$ 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.

Specifically, for a given market a, the wage function at each data observation $j^* \in \Xi_a$ (locally) satisfies the equation

$$w_{j^*,a} = b_{j^*,a,0} + b_{j^*,a,1} x_{j^*,1,t} + ... + b_{j^*,a,M} x_{j^*,M,t} + b_{j^*,a,\xi} \xi_{j^*,t}, \tag{26}$$

where each coefficient $b_{j^*,a,m}$, $m = 1, ..., M$, represents the derivative of $w_a(.)$ with respect to characteristic m at point j^*. Intuitively, this corresponds to the fact that by a first-order Taylor expansion argument, a function $w_a(.)$ at point (X_{j^*t}, ξ_{j^*t}) is well approximated by a tangent hyperplane in a neighborhood of the function value at that point, $w_{j^*,a}$.[25]

In the context of nonparametric regression, Fan and Gijbels (1996) provide a formula for the coefficients in Equation (26) 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

$$\mathbf{b}_{j^*,a} = \left(\Psi^T_{j^*,a} \Omega_{j^*,a} \Psi_{j^*,a}\right)^{-1} \Psi^T_{j^*,a} \Omega_{j^*,a} \mathbf{w}_a, \tag{27}$$

[25] See Judd (1998) and Fan and Gijbels (1996) for a discussion.

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

$$\Psi_{j^*,a} = \begin{bmatrix}\mathbf{1} & \boldsymbol{\psi}_{j^*,a}\end{bmatrix} = \begin{bmatrix} 1 & (x_{1,1,a} - x_{j^*,1,a}) & \dots & (x_{1,M,a} - x_{j^*,M,a}) & (\xi_{1,a} - \xi_{j^*,a}) \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & (x_{J_a,1,a} - x_{j^*,1,a}) & \dots & (x_{J_a,M,a} - x_{j^*,M,a}) & (\xi_{J_a,a} - \xi_{j^*,a}) \end{bmatrix} \quad (28)$$

$$\Omega_{j^*,a} = \mathbf{diag}\left(K_{\mathbf{H}_a}(\boldsymbol{\psi}_{j^*,a})\right), \quad (29)$$

where $K_{\mathbf{H}_a}(\boldsymbol{\psi}_{j^*,a})$ is a multivariate kernel function with smoothing parameter matrix $\mathbf{H}_a$, and $K_{\mathbf{H}_a}$ is a 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 (27).[26] 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 (27) 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 (27)), 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:

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

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 (26) 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

[26] 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, Marron and Park (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.

Racine (2013) show that $B(\boldsymbol{\psi}_{j^*,a}, \mathbf{w}_a) \times \mathbf{p}$ is equivalent to (27) when $p_j = 1/J_a,\ j = 1, ..., J_a$.

According to the first-order conditions in (11) and (12), 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 depend on the firm's initial choice x_0 by construction. 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.

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.[27] We use two cross-section 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

$$w = X\zeta + D\tau + \varepsilon, \tag{30}$$

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.

[27] 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.

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.

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 (30) 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.[28]

One potential problem with using test scores and family background as proxies to remove

[28] 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).

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 outline estimation results 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-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 (25). 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.[29] To estimate the control variables for education and experience, we use an instrument vector Z that includes worker age and the existence of a local college. All other observed characteristics are included in

[29] 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.

sub-vector X_1.

Table 2. Conditional Worker Quality Distribution

	1990		1993	
Normalized Worker Quality	Mean	(St.d.)	Mean	(St.d.)
All workers	0.449	(0.262)	0.462	(0.254)
By education				
High school incompletes	0.336	(0.217)	0.339	(0.202)
High school graduates	0.389	(0.230)	0.391	(0.230)
Some college	0.467	(0.249)	0.477	(0.249)
College graduates	0.634	(0.263)	0.647	(0.220)
By work experience				
0-4 years	0.356	(0.244)	0.296	(0.214)
5-9 years	0.466	(0.263)	0.441	(0.249)
10+ years	0.516	(0.253)	0.518	(0.246)
By AFQT percentile scores				
$AFQT \leq 25$	0.340	(0.210)	0.354	(0.214)
$25 < AFQT \leq 50$	0.437	(0.241)	0.460	(0.240)
$50 < AFQT \leq 75$	0.522	(0.259)	0.529	(0.247)
$AFQT > 75$	0.636	(0.279)	0.640	(0.243)
By industry				
Mining	0.518	(0.254)	0.476	(0.214)
Construction	0.528	(0.270)	0.504	(0.251)
Manufacturing	0.470	(0.257)	0.490	(0.251)
Transportation, Communication, Public Utilities	0.540	(0.262)	0.556	(0.256)
Wholesale and Retail Trade	0.348	(0.224)	0.362	(0.224)
Finance, Insurance, and Real Estate	0.544	(0.259)	0.550	(0.242)
Service	0.443	(0.264)	0.454	(0.255)

As shown by Card (1993) and others, the existence of a local college would reduce the cost of college and affect schooling outcomes.[30] The local college instrument is binary and therefore does not satisfy the "large support" condition in Imbens and Newey (2009). However, as 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).

Table 2 shows the joint distribution between some of the observed worker characteristics and 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.

Table 3. Correlations of Estimated Quality and Observed Human Capital Variables

	1990 cross-section		
	Education	Experience	AFQT
Estimated quality	0.395	0.238	0.424
	1993 cross-section		
	Education	Experience	AFQT
Estimated quality	0.427	0.285	0.421
	1990 and 1993 pooled		
	Estimated quality in 1993		
Estimated quality in 1990	0.741		

The top two panels of Table 3 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 0.45. The estimated quality is less significantly correlated with experience than AFQT score and education in both years, but worker quality

[30] 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).

becomes more correlated with experience over time. Learning-by-doing may explain the increasing correlation between worker quality and experience. The correlations between the estimated quality and AFQT score are 0.424 and 0.421 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 3 reports the correlation between the quality estimates in 1990 and 1993 to be fairly high at 0.741. 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. Thus, 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 (11)) for education in all industries in 1990. The distribution has a long right tail, with a mean of 142 and a standard deviation of 518. 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 142 and it currently hires a worker with 6 years of education, then the firm is willing to pay an additional \$0.22 ($= \$1.42 \times (\ln 7 - \ln 6)$) 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 \$0.19 per hour.[31]

[31] 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 3 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.

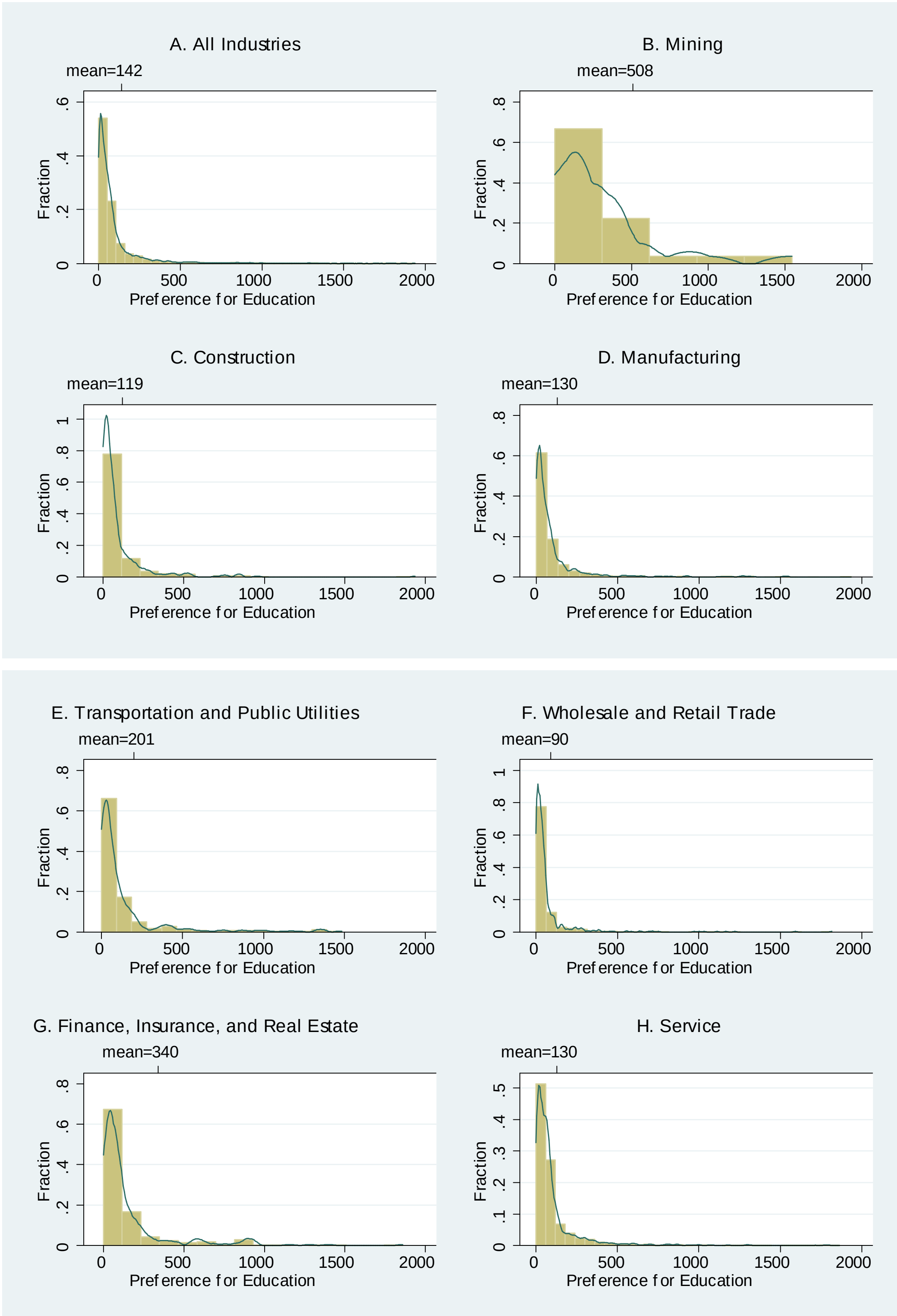

Figure 1: Firm Preference for Education Across Industries, 1990

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 1523, and the distribution in the service industry is the least dispersed with a standard deviation of 276.

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 (60 vs. 142), and the WTP parameter for experience is less dispersed with a standard deviation of 194. Firms in the finance, insurance, and real estate industry value work experience the most, with a mean WTP parameter equal to 137, whereas experience is the least valued in the wholesale and retail trade and service industries with a mean WTP parameter of 45. In terms of dispersion, the finance, insurance and real estate industry has the longest right tail, and the distribution of WTP parameter for experience is most concentrated in the service 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 finance, insurance and real estate industry. The distribution of WTP parameter for quality is most dispersed in the service 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 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.

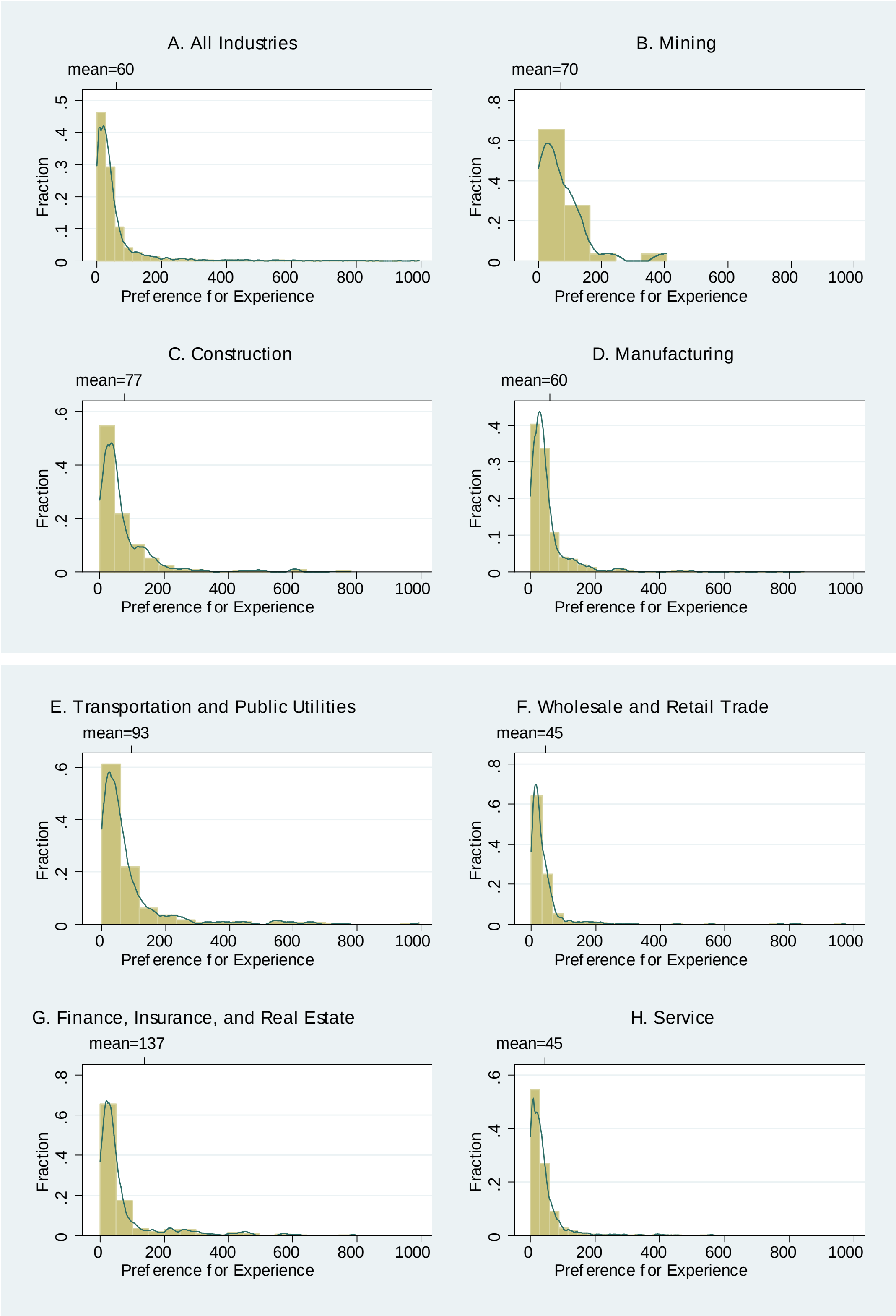

Figure 2: Firm Preference for Work Experience Across Industries, 1990

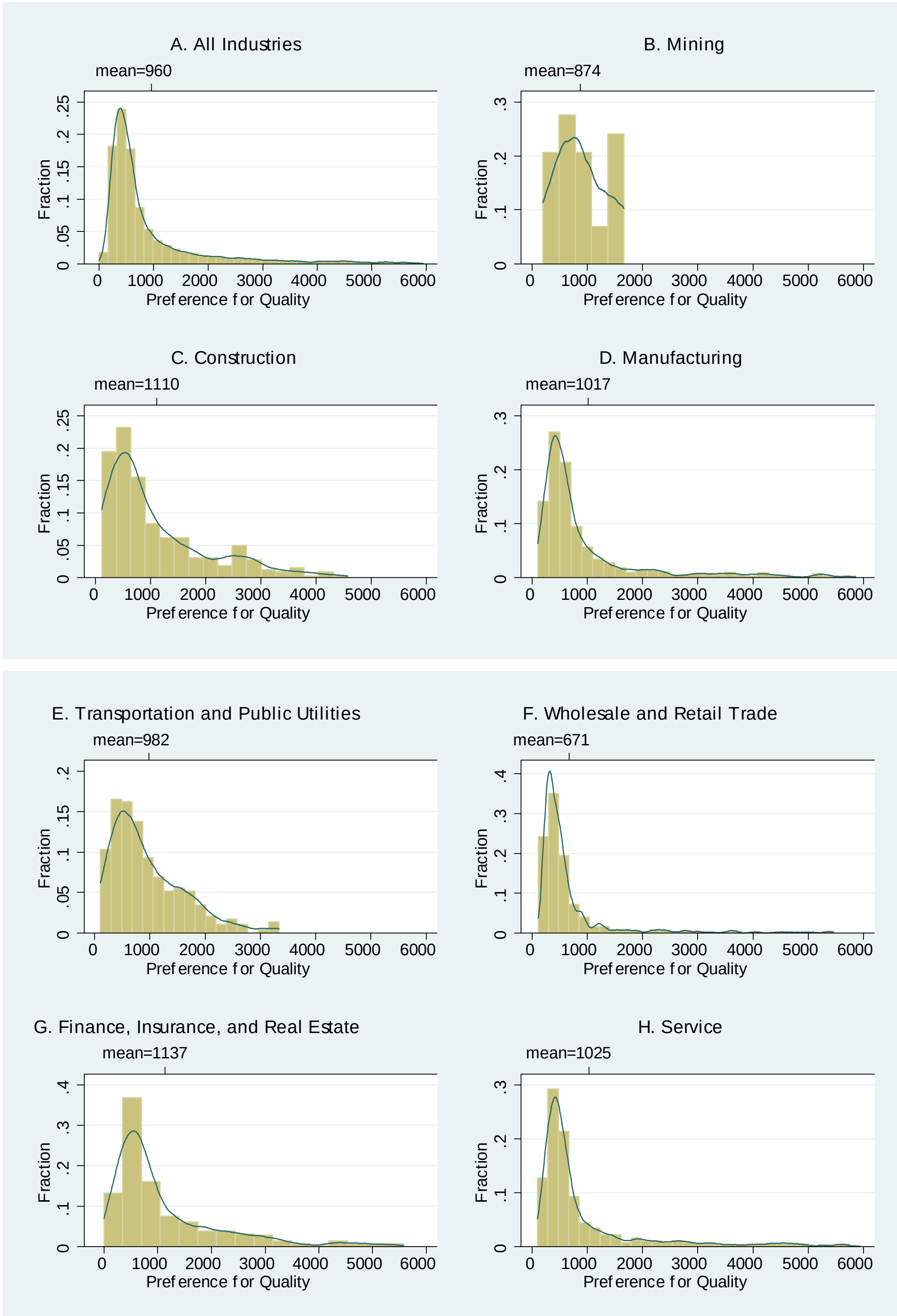

Figure 3: Firm Preference for Worker Quality Across Industries, 1990

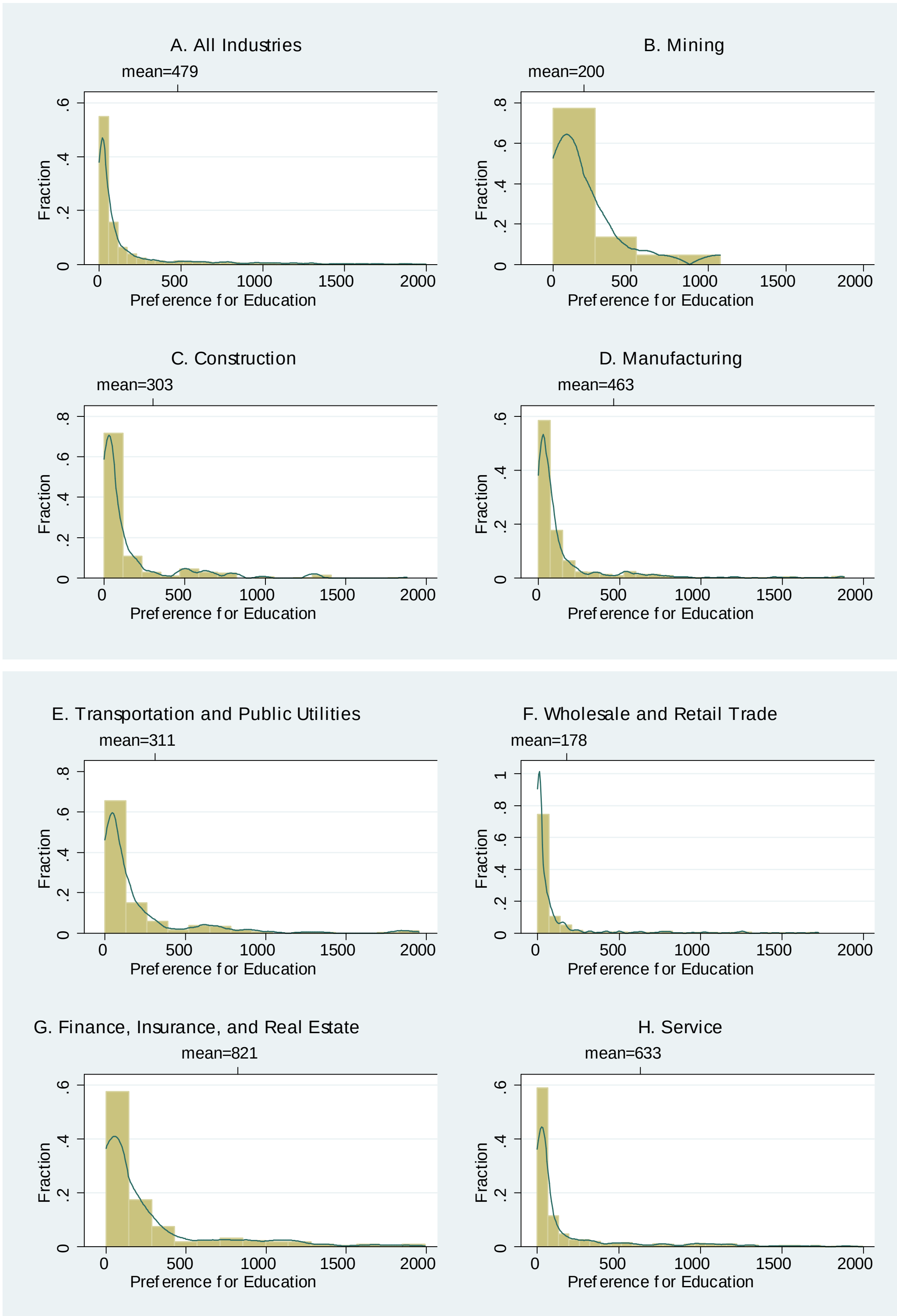

Figure 4: Firm Preference for Education Across Industries, 1993

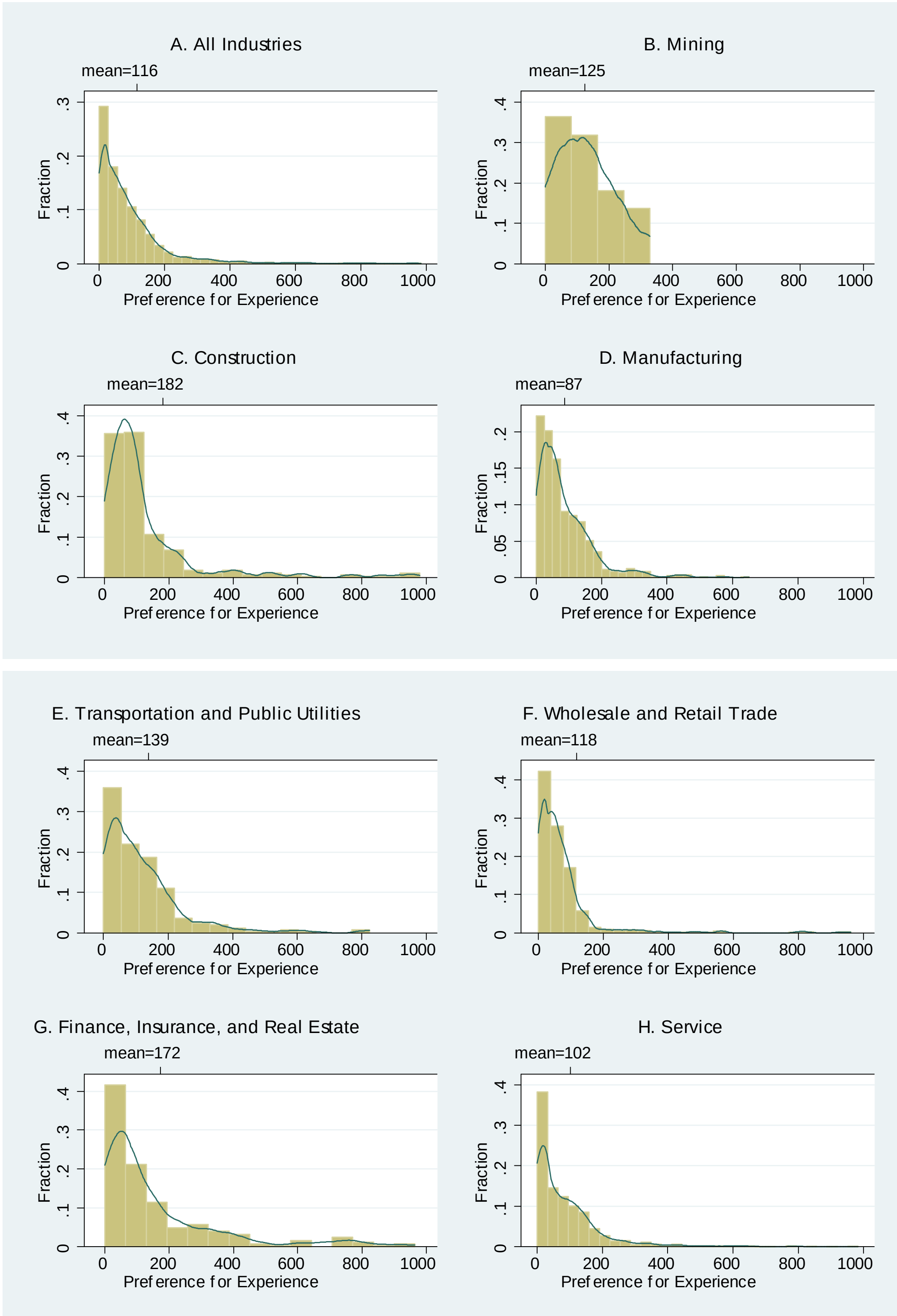

Figure 5: Firm Preference for Work Experience Across Industries, 1993

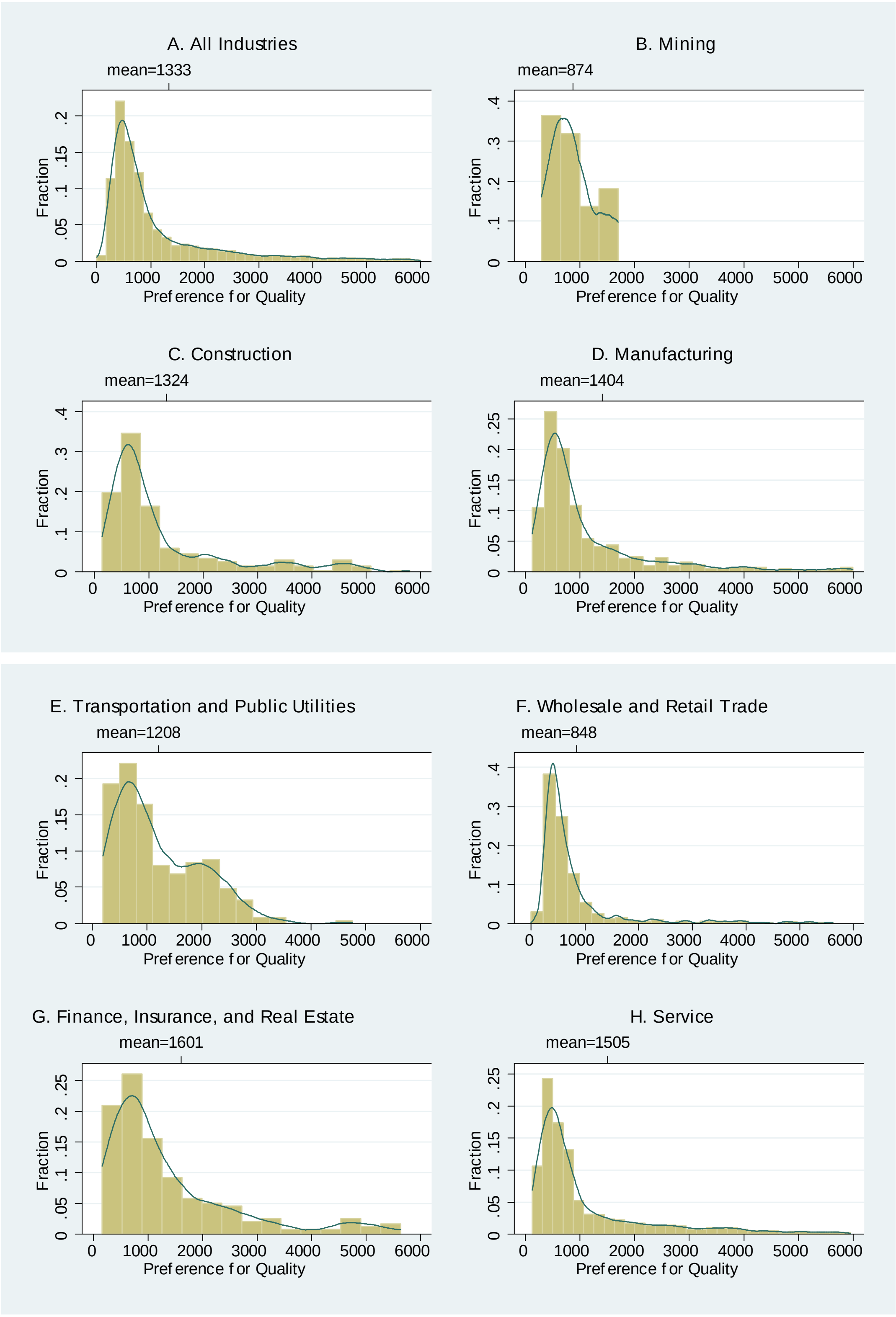

Figure 6: Firm Preference for Worker Quality Across Industries, 1993

Firm WTP parameters across workers' human capital attributes are not independently distributed. Table 4 reports the correlation matrix of WTP parameters across worker attributes on education, experience, and quality in each year. In both years, firm WTP parameters for all human capital attributes have strong positive correlations with each other.

Table 4. Correlation Matrix of WTP Parameters by Year

	Education	Experience	Quality
1990			
Education	1		
Experience	0.777	1	
Quality	0.379	0.308	1
1993			
Education	1		
Experience	0.427	1	
Quality	0.809	0.460	1

5.3 Inter-industry Wage Differentials

Columns (2) and (6) of Table 5 present estimates of coefficient τ in Equation (30) 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 88% from 0.133 to 0.016 in 1990 and from 0.114 to 0.014 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 0.901 in 1990 and from 0.376 to 0.882 in 1993 once worker quality is included in the regressions.[32]

[32] 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.

Table 5. 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	0.038	0.260	0.032	0.120	0.024	0.191	0.044
	(0.081)	(0.021)	(0.061)	(0.020)	(0.097)	(0.020)	(0.056)	(0.018)
Construction	0.265	0.017	0.176	0.017	0.212	-0.001	0.162	0.008
	(0.028)	(0.011)	(0.019)	(0.011)	(0.033)	(0.014)	(0.026)	(0.013)
Manufacturing	0.159	0.034	0.118	0.034	0.139	0.006	0.126	0.014
	(0.019)	(0.008)	(0.014)	(0.008)	(0.022)	(0.010)	(0.018)	(0.009)
Transportation, Communication,	0.172	-0.002	0.176	-0.002	0.163	-0.020	0.198	0.004
Public Utilities	(0.028)	(0.009)	(0.019)	(0.009)	(0.032)	(0.012)	(0.025)	(0.012)
Wholesale and Retail Trade	-0.086	0.006	-0.025	0.006	-0.120	-0.005	-0.056	0.001
	(0.019)	(0.007)	(0.014)	(0.007)	(0.022)	(0.009)	(0.017)	(0.009)
Finance, Insurance, and Real Estate	0.166	0.018	0.151	0.017	0.142	0.015	0.155	0.019
	(0.029)	(0.008)	(0.017)	(0.008)	(0.033)	(0.016)	(0.027)	(0.016)
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	0.016	0.102	0.014	0.114	0.014	0.099	0.015
Weighted St.d. of Differentials	0.046	0.005	0.032	0.005	0.043	0.003	0.036	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	0.901	0.670	0.901	0.376	0.882	0.625	0.894

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.

Columns (3) and (7) of Table 5 present estimates of τ coefficients in Equation (30) 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 0.102 in 1990 and from 0.114 to 0.099 in 1993. The adjusted R^2 of the log wage regression increases from 0.356 to 0.670 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 5, we bootstrap the standard errors of parameters.

Table 6. Decomposition of Inter-Industry Wage Differentials

	(1)	(2)	(3)
	1990 two-digit industry premiums		
Quality	1.059		1.545
	(0.115)		(0.231)
Firm preferences	No	Yes	Yes
R squared	0.641	0.358	0.699
Adjusted R squared	0.632	0.307	0.666
	1993 two-digit industry premiums		
Quality	1.032		1.122
	(0.105)		(0.268)
Firm preferences	No	Yes	Yes
R squared	0.672	0.513	0.682
Adjusted R squared	0.664	0.475	0.648

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 (30) 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.[33] Table 6 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 6 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 6 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.[34] 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.

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 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 (2019), suggest that different types of observed skills are very different productive attributes. Relaxing the assumption of single-dimensional unobserved skill is possible (Matzkin 2003),

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

[34] 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.

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. In such an equilibrium model, compensating differences may be separately identified from WTP parameters, but such exercise involves various challenges in identification and remains an important topic for future research.

Appendix A: 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 $X_{j,t} = X_{j',t}$ and $\xi_{j,t} = \xi_{j',t}$, then $w_{j,a} = w_{j',a}$.

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

(3) $|w_{j,a} - w_{j',a}| \leq E \times (|X_{j,t} - X_{j',t}| + |\xi_{j,t} - \xi_{j',t}|)$ 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 $X_{j,t} = X_{j',t}$ and $\xi_{j,t} = \xi_{j',t}$. Then $R_{i,a}(X_{j,t}, \xi_{j,t}) - w_{j,a} < R_{i,a}(X_{j',t}, \xi_{j',t}) - w_{j',a}$ 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 $X_{j,t} = X_{j',t}$ and $\xi_{j,t} > \xi_{j',t}$. Given that $R_{i,a}(X_{j,t}, \xi_{j,t})$ strictly increases in $\xi_{j,t}$, $R_{i,a}(X_{j,t}, \xi_{j,t}) - w_{j,a} > R_{i,a}(X_{j',t}, \xi_{j',t}) - w_{j',a}$ 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 $R_{i,a}(X_{j,t}, \xi_{j,t})$ is Lipschitz-continuous in $(X_{j,t}, \xi_{j,t})$ implies that for any two workers j and j' differing in at least one characteristic,

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

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

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

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.

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

Therefore,

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

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

Appendix B: 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. We suppress the market subindex $a \equiv (l, t)$ in our notation for ease of exposition. Without loss of generality, we focus on continuous, strictly positive worker attributes.

Consider the following specification for the labor efficiency units at employer i of worker j with characteristic vector $\left(X_j, \xi_j\right) = \left(x_{j,1}, x_{j,2}, \cdots, x_{j,M}, \xi_j\right)$:

$$E_{i,j} = \rho_{i,0} + \ln(X_j) \cdot \boldsymbol{\rho}_{i,X} + \rho_{i,\xi} \ln(\xi_j), \quad \forall j. \tag{31}$$

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 (31), 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 $\boldsymbol{\beta}_i$ is the product of the vector of efficiency unit coefficients $\boldsymbol{\rho}_i = (\rho_{i,0}, \boldsymbol{\rho}_{i,X}, \rho_{i,\xi})$ and γ_i.

Appendix C: Proof of Proposition 2

We use the assumption that, in each market $a = 1, ..., A$, each function $h_{a,m}(., \eta_{a,m})$ is strictly monotonic in $\eta_{a,m}$ to define $h_{a,m}^{-1}(x_{0,m}, X_1, Z)$ as the inverse of $h_{a,m}(X_1, Z, \eta_{a,m})$. According to the proof of Lemma 1 of Matzkin (2003), for each $m = 1, ..., M_0$ and $a = 1, ..., A$,

$$\begin{aligned} F_{X_{0,m}|X_1,Z,a}(x_{0,m}|x_1, z) &= \Pr(X_{0,m} \leq x_{0,m}|X_1 = x_1, Z = z, a) \\ &= \Pr(h_{a,m}(x_1, z, \eta_{a,m}) \leq x_{0,m}|X_1 = x_1, Z = z, a) \\ &= \Pr(\eta_{a,m} \leq h_{a,m}^{-1}(x_{0,m}, x_1, z)|X_1 = x_1, Z = z, a) \\ &= \Pr(\eta_{a,m} \leq h_{a,m}^{-1}(x_{0,m}, x_1, z)) \\ &= F_{\eta_{a,m}}(h_{a,m}^{-1}(x_{0,m}, x_1, z) = h_{a,m}^{-1}(x_{0,m}, x_1, z) = \eta_{a,m}, \end{aligned}$$

where the second equality follows from the definition of the function $h_{a,m}(.)$ in a given market a, the third equality follows from the monotonicity assumption, the fourth equality follows from the independence between (X_1, Z) and $\eta_{a,m}$, and the last equality is the result of normalizing $\eta_{a,m}$ so that it lies in $U[0, 1]$.

Next, we show that, for each market a, the vector $\boldsymbol{\eta}_a \equiv (\eta_{a,1}, ..., \eta_{a,M_0})$ consists of control variables conditional on which X and δ_a are independent by adapting the proof of Theorem 1 of Imbens and Newey (2009) for multiple endogenous variables. For any bounded function $p_a(x_0, x_1)$, it follows from the independence of (X_1, Z) and $(\delta_a, \boldsymbol{\eta}_a)$ that

$$\begin{aligned} E[p_a(x_0, x_1)|\delta_a, \boldsymbol{\eta}_a] &= E[p_a(h_{a,1}(x_1, z, \eta_{a,1}), ..., h_{a,M_0}(x_1, z, \eta_{a,M_0}), x_1)|\delta_a, \boldsymbol{\eta}_a] \\ &= \int p_a(h_{a,1}(x_1, z, \eta_{a,1}), ..., h_{a,M_0}(x_1, z, \eta_{a,M_0}), x_1) dF_{X_1,Z,}(x_1, z) \\ &= E[p_a(x_0, x_1)|\boldsymbol{\eta}_a]. \end{aligned}$$

Thus, for any bounded function $b_a(\delta_a)$, it follows from the Law of Iterated Expectations that

$$\begin{aligned} E[p_a(x_0,x_1)b_a(\delta_a)|\boldsymbol{\eta}_a] &= E[b_a(\delta_a)E[p_a(x_0,x_1)|\delta_a,\boldsymbol{\eta}_a]|\boldsymbol{\eta}_a] \\ &= E[b_a(\delta_a)E[p_a(x_0,x_1)|\boldsymbol{\eta}_a]|\boldsymbol{\eta}_a] \\ &= E[b_a(\delta_a)|\boldsymbol{\eta}_a]E[p_a(x_0,x_1)|\boldsymbol{\eta}_a], \end{aligned}$$

which indicates the independence between X and δ_a conditional on $\boldsymbol{\eta}_a$, for each $a = 1, ..., A$.

Finally, the integral in (16) simplifies to

$$\begin{aligned} &\int_{\boldsymbol{\eta}_a\in[\mathbf{0},\mathbf{1}]^{M_0}} F_{w|X,\boldsymbol{\eta},a}(w_{j,a}|X_{jt},\boldsymbol{\eta}_a)\mathbf{dG}_a(\boldsymbol{\eta}_a) \\ &= \int_{\boldsymbol{\eta}_a\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(w_a(X,\delta_a)\le w_{j,a}|X=X_{jt},\boldsymbol{\eta}_a,a)\mathbf{dG}_a(\boldsymbol{\eta}_a) \\ &= \int_{\boldsymbol{\eta}_a\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\delta_a\le w_a^{-1}(X,w_{j,a})|X=X_{jt},\boldsymbol{\eta}_a,a)\mathbf{dG}_a(\boldsymbol{\eta}_a) \\ &= \int_{\boldsymbol{\eta}_a\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\xi+\epsilon_a\le w_a^{-1}(X_{jt},w_{j,l,t})|X=X_{jt},\boldsymbol{\eta}_a,a)\mathbf{dG}_a(\boldsymbol{\eta}_a) \\ &= \int_{\boldsymbol{\eta}_a\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\xi+\epsilon_a\le \xi_{jt}+\epsilon_a|X=X_{jt},\boldsymbol{\eta}_a,a)\mathbf{dG}_a(\boldsymbol{\eta}_a) \\ &= \int_{\boldsymbol{\eta}_a\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\xi\le \xi_{jt}|X=X_{jt},\boldsymbol{\eta}_a,a)\mathbf{dG}_a(\boldsymbol{\eta}_a), \end{aligned}$$

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_a, the third one results from both $X = X_{jt}$ and the separability of δ_a between unobserved quality ξ and ϵ_a in market a, and the fourth one follows unobserved quality of worker j being separable from ϵ_a for any worker in the market-specific set of workers Ξ_a.

Our final step involves taking expectations with respect to industries $l = 1, ..., L$. For this reason, we replace the market index a with its industry/year representation, (l, t), and we use the notation D to represent an individual's industrial affiliation. Given the independence between X and δ_a conditional on $\boldsymbol{\eta}_a$ for all $a = 1, ..., A$, and the normalizations of ξ and

$\eta_{a,m}$, it follows from the Law of Total Probability that

$$\begin{aligned}
&\sum_{l=1}^{L}\left\{\int\limits_{\boldsymbol{\eta}_a\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\xi \leq \xi_{jt}|X = X_{jt}, \boldsymbol{\eta}_a, a)\mathbf{dG}_a(\boldsymbol{\eta}_a)\right\} \Pr(l|X = X_{jt}, t) \\
&= \sum_{l=1}^{L}\left\{\int\limits_{\boldsymbol{\eta}_{l,t}\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\xi \leq \xi_{jt}|X = X_{jt}, \boldsymbol{\eta}_{l,t}, D = l, t) \Pr(D = l|X = X_{jt}, t)\mathbf{dG}_{l,t}(\boldsymbol{\eta}_{l,t})\right\} \\
&= \sum_{l=1}^{L}\left\{\int\limits_{\boldsymbol{\eta}_{l,t}\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\xi \leq \xi_{jt}, D = l|X = X_{jt}, \boldsymbol{\eta}_{l,t}, t)\mathbf{dG}_{l,t}(\boldsymbol{\eta}_{l,t})\right\} \\
&= \sum_{l=1}^{L}\left\{\int\limits_{\boldsymbol{\eta}_{l,t}\in[\mathbf{0},\mathbf{1}]^{M_0}} \Pr(\xi \leq \xi_{jt}, D = l|\boldsymbol{\eta}_{l,t}, t)\mathbf{dG}_{l,t}(\boldsymbol{\eta}_{l,t})\right\} \\
&= \sum_{l=1}^{L} \Pr(\xi \leq \xi_{jt}, D = l|t) \\
&= F_{\xi,t}(\xi_{jt}) \\
&= \xi_{jt}.
\end{aligned}$$

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

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

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

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

[13] 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.

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

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

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

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

[18] 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.

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

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

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

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

[23] 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.

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

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

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

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

[28] 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.

[29] 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.

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

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

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

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

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

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

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

[37] Halton, J. (1960), "On the Efficiency of Evaluating Certain Quasi-Random Sequences of Points in Evaluating Multi-dimensional Integrals," *Numerische Mathematik* 2: 84–90.

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

[39] 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.

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

[41] Judd, K. (1998), *Numerical Methods in Economics*. MIT Press, Cambridge MA.

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

[43] 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.

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

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

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

[47] Lise, J. and F. Postel-Vinay (2019), "Multidimensional Skills, Sorting, and Human Capital Accumulation," forthcoming, *American Economic Review.*

[48] Mardia, J., J. Kent and M. Bibby (1979), *Multivariate Analysis.* Academic Press, New York.

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

[50] 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

[51] 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.

[52] 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.

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

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

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

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

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

[58] 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.

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

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

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

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

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