

Semi-nonparametric models of multidimensional matching: an optimal transport approach

Dongwoo Kim
Young Jun Lee

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP09/26



SEMI-NONPARAMETRIC MODELS OF MULTIDIMENSIONAL MATCHING: AN OPTIMAL TRANSPORT APPROACH

DONGWOO KIM

Simon Fraser University and Korea University

YOUNG JUN LEE

Hanyang University

This paper develops a set of empirically tractable and flexible sieve estimators for semi-nonparametric multidimensional matching models with transferable utility, focusing on worker-job matching. We generalize the parametric quadratic-Gaussian framework employed by Bojilov and Galichon (2016) and Lindenlaub (2017), which relies on joint normality of observed characteristics. We allow unrestricted distributions of characteristics and show identification of the production technology and the equilibrium wage and matching functions using optimal transport theory. Given identification, we propose efficient, consistent, and asymptotically normal sieve estimators. We revisit Lindenlaub's empirical application and show that, between 1990 and 2010, the U.S. economy experienced much larger technological progress favoring cognitive abilities than the original findings suggest. Furthermore, our flexible model specifications provide a significantly better fit for patterns in the evolution of wage inequality.

Dongwoo Kim: dongwook@sfu.ca, dongwoo_kim@korea.ac.kr

Young Jun Lee: youngjunlee@hanyang.ac.kr

This paper is based on the third chapter of Lee's doctoral dissertation at University College London. We thank Dennis Kristensen, Krishna Pendakur, Minchul Shin, Martin Weidner, and Daniel Wilhelm for their helpful suggestions. We also benefited greatly from comments by seminar and conference participants at UCL, Bocconi University, KIEP, University of Calgary, Northwestern University, Kyung Hee University, University of Seoul, Hanyang University, KAEA virtual seminar, IAAE 2024, KERIC 2024, ESWC 2025, and EcoSta 2025. Kim gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada under the Insight Grant (435-2024-0322) and Korea University Grants (K2528561, K2528721, K2614081, K2613941). Lee gratefully acknowledges support from the Italian Ministry of University and Research under PRIN 2017 (2017TMFPSH). All errors are our own.

Keywords: Multidimensional matching, transferable utility, optimal transport, sieve extremum estimation, technological progress, wage polarization.

1. INTRODUCTION

In two-sided markets, agents form optimal matches based on their preferences and characteristics, generating a shareable joint surplus. Matching models are widely used to analyze these dynamics in settings such as the labor market (workers and jobs) and the marriage market (spouses). To keep the analysis tractable, empirical work often relies on low-dimensional heterogeneity, e.g., focusing on one salient attribute or aggregating multiple characteristics into a scalar index.¹ While such scalar-index approaches implicitly encode multidimensional characteristics, they restrict the matching mechanism to operate along a single dimension. In the labor market, for instance, workers develop specialized skills (cognitive for mathematicians and manual for gymnasts), and the single index model cannot capture this specialization. These considerations motivate matching models that accommodate multidimensional heterogeneity more directly, a goal pursued in a growing literature.²

Following the seminal work of [Tinbergen \(1956\)](#), recent contributions have advanced the analysis of matching with multiple continuous attributes. These frameworks typically achieve tractability by imposing parametric restrictions on the distribution of observable characteristics and/or on the structure of the surplus. In particular, [Bojilov and Galichon \(2016\)](#) and [Lindenlaub \(2017\)](#) develop parametric matching models that assume joint normality of attributes and a bilinear surplus, which yields closed-form expressions for the equilibrium assignment. However, the normality assumption can be consequential. If the true joint distribution deviates from Gaussianity, the implied assignment patterns may be distorted. In empirical applications, this approach often requires additional steps to align the data with the normality assumption, for instance, transforming each marginal to stan-

¹ Assortative spousal matching on income, wages, education, risk aversion, and preference for childbearing are investigated by [Becker \(1991\)](#), [Grossbard-Schechtman \(1993\)](#), [Pencavel \(1998\)](#), [Choo and Siow \(2006\)](#), [Chiappori and Reny \(2016\)](#), [Legros and Newman \(2007\)](#) and [Chiappori and Orefice \(2008\)](#) among many others. [Becker \(1973\)](#) and [Chiappori et al. \(2012\)](#) investigate spousal matching that hinges on “ability indices”.

² Studies such as [Willis and Rosen \(1979\)](#) and [Papageorgiou \(2014\)](#) favor the multidimensional setup over a single index model in the labor market context. Spousal choices are also based on a variety of attributes as shown in [Becker \(1991\)](#), [Weiss and Willis \(1997\)](#), [Qian \(1998\)](#), [Silventoinen et al. \(2003\)](#), [Hitsch et al. \(2010\)](#), and [Orefice and Quintana-Domeque \(2010\)](#).

dard normal. Such transformations can alter the interpretation of the underlying attributes, and the resulting joint distribution need not be well approximated by a multivariate normal. Consequently, the estimated assignment mechanism may not accurately reflect the true matching process.

In this paper, we address these challenges by developing a set of empirically tractable and flexible sieve estimators for semi-nonparametric multidimensional matching models with transferable utility in the context of worker-job matching. We generalize the quadratic-Gaussian structure employed in [Bojilov and Galichon \(2016\)](#) and [Lindenlaub \(2017\)](#) by accommodating arbitrary distributions of attributes while maintaining the bilinear surplus. In this context, each worker possesses distinct skills, and each job requires specific skills to produce output according to the production technology. The social planner's problem is to optimally assign workers to jobs to maximize total output in the economy, which can be formulated as a Monge-Kantorovich optimal transport problem. Unique solutions for the equilibrium assignment and wage functions are derived using an optimal transport approach ([Villani, 2003, 2008](#), [De Philippis and Figalli, 2014](#)).³ We then propose an econometric framework to apply the theoretical model to real-world data. Following [Lindenlaub \(2017\)](#), we introduce error terms to bridge the gap between the deterministic equilibrium predictions and observed data. Unlike [Lindenlaub \(2017\)](#), however, we do not require the errors to be normally distributed or uncorrelated.

We estimate the production technology, equilibrium assignment, and wage functions using the semiparametric M-estimation techniques proposed by [Ai and Chen \(2003\)](#) and [Chen \(2007\)](#). To our knowledge, this paper is the first in the literature to introduce semiparametric M-estimators to multidimensional matching models. Depending on the assumptions on the error terms, the model is estimated by sieve maximum likelihood (SML), least squares (SLS), or generalized least squares (SGLS). These estimators are efficient, asymptotically normal, and easy to implement. Our estimators perform well in extensive simulation ex-

³Applications of optimal transport have proven successful in multiple fields of economics (e.g., [Ekeland \(2010\)](#), [Chiappori et al. \(2010\)](#), [Chiong et al. \(2016\)](#), [Lindenlaub \(2017\)](#), [Galichon and Salanié \(2022\)](#), among others). [Galichon \(2017\)](#) offers a comprehensive survey of the literature.

periments for a wide class of data-generating processes. Although we focus on worker-job matching, our method can be broadly applied to other matching problems, such as couple matching in the marriage market.

We revisit [Lindenlaub \(2017\)](#)'s worker-job data from the U.S. and estimate production technology, equilibrium assignment, and wage functions to investigate the technological shift and its effects on wage inequality between 1990 and 2010. Our results show greater technological progress favoring cognitive skills than Lindenlaub's estimates suggest. Furthermore, the greater flexibility of our models provides much greater explanatory power for the evolution of wage inequality, particularly the '*wage polarization*' phenomenon featuring stronger wage growth in the bottom and upper tails of the wage distribution relative to the median. The Gaussian model fails to predict wage polarization because, on top of misspecification bias, it restricts the equilibrium wage function to a quadratic form.

This paper is organized as follows. The remainder of this section discusses the related literature. Section 2 introduces the multidimensional matching model with a bilinear surplus and characterizes the equilibrium using optimal transport theory. Section 3 proposes the econometric matching models and establishes identification. Section 4 presents the sieve estimators. Section 5 derives the asymptotic properties of our sieve GLS estimator. Section 6 conducts simulation experiments. Section 7 revisits [Lindenlaub \(2017\)](#)'s empirical analysis. Section 8 concludes. Technical proofs and additional theoretical details are provided in the appendix.

1.1. *Related literature*

[Choo and Siow \(2006\)](#) (CS henceforth) introduced an empirical transferable utility (TU) model for matching with discrete types and multidimensional unobserved heterogeneity.⁴ Their framework assumes that i.i.d. unobserved heterogeneity follows the extreme value type I (Gumbel) distribution, under which the systemic match surplus is identified by logit formulae. Several extensions remain within this discrete-type framework: [Galichon and Salanié \(2022\)](#) generalize the distributional assumption on unobservables beyond the logit

⁴See [Galichon et al. \(2019\)](#) for the imperfectly transferable utility model with unobserved heterogeneity.

specification and establish identification of the social surplus, while [Gualdani and Sinha \(2023\)](#) show partial identification of the systemic surplus under nonparametric assumptions on the error distribution.

[Dupuy and Galichon \(2014\)](#) extend the CS framework to continuous types while maintaining the i.i.d. Gumbel structure for unobserved heterogeneity, formulating the matching problem as an entropy-regularized optimal transport problem. [Bojilov and Galichon \(2016\)](#) specialize the [Dupuy and Galichon](#) framework by imposing the quadratic-Gaussian structure—a bilinear surplus function combined with jointly normal attributes—and derive closed-form solutions for the equilibrium matching distribution. Taking a fundamentally different approach, [Lindenlaub \(2017\)](#) formulates a deterministic matching model without unobserved heterogeneity in the surplus, so the equilibrium assignment is governed by a standard Monge-Kantorovich optimal transport problem.⁵ Although [Lindenlaub](#) also employs the quadratic-Gaussian structure to obtain closed-form solutions, the absence of unobserved heterogeneity yields a purely deterministic assignment rather than the stochastic matching patterns in the CS–[Dupuy and Galichon](#)–[Bojilov and Galichon](#) tradition.

All these approaches share the separability assumption that unobserved heterogeneity (or, in our case, measurement error) does not interact with observed characteristics in generating the surplus. However, the nature of unobservables differs fundamentally across frameworks. In the CS tradition (including [Dupuy and Galichon](#) and [Bojilov and Galichon](#)), unobserved heterogeneity enters the surplus function directly and determines stochastic matching patterns; the systemic surplus is identified from matching patterns—i.e., the joint distribution of matched characteristics—via logit inversion or its generalizations, and individual-level wage data are not required. In contrast, our framework and [Lindenlaub](#)'s introduce measurement errors that are additive and separable, capturing discrepancies between the deterministic equilibrium predictions and observed data. Our econometric model explicitly exploits observed wages alongside matching patterns: the equilibrium wage func-

⁵Alternatively, [Lise and Postel-Vinay \(2020\)](#) consider a search-theoretic model in which workers are matched to firms in a dynamic setup.

tion enters as a directly estimable object, which strengthens identification and estimation of the technology parameters.

Our paper extends [Lindenlaub](#)'s deterministic matching framework by dispensing with distributional assumptions on the characteristics, thereby offering a more flexible and robust econometric framework for multidimensional matching. We propose efficient econometric procedures that jointly estimate both finite-dimensional parameters and infinite-dimensional functions using conditional moments implied by the model equilibrium, leveraging the sieve M-estimation literature.⁶ In particular, we employ the sieve GLS estimator proposed in [Chen \(2007\)](#) for our most flexible model specification. This estimator is efficient and computationally simpler than the SMD estimator.

We establish the convergence, efficiency, and asymptotic normality of our sieve estimators relying on the smoothness of unknown nonparametric components (optimal transport maps). This smoothness condition can be verified by applying the results in the mathematical literature on optimal transport maps. [Caffarelli \(1992a,b, 1996\)](#) show the smoothness of transport maps when the distributions of characteristics on both sides are compactly supported. [Cordero-Erausquin and Figalli \(2019\)](#) further extend the earlier results to cases where the distributions may have unbounded supports. The degree of smoothness of a transport map depends on how smooth the densities are.

2. MULTIDIMENSIONAL MATCHING WITH A BILINEAR SURPLUS

We consider an environment where every worker with a bundle of skills sorts into a job demanding specific combinations of those skills. Let $\mathcal{X} \subset \mathbb{R}^d$ and $\mathcal{Y} \subset \mathbb{R}^d$ be spaces of worker and job characteristics endowed with probability measures P and Q , respectively. Workers and jobs are described by the corresponding vectors of characteristics $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Every matched pair produces output according to production technology $s(x, y)$,

⁶[Shen \(1997\)](#) establishes asymptotic properties of smooth functionals of sieve MLE. [Newey and Powell \(2003\)](#), [Ai and Chen \(2003, 2007\)](#), and [Blundell et al. \(2007\)](#) propose efficient sieve IV and sieve minimum distance (SMD) estimators. [Chen and Pouzo \(2009\)](#) further show that the SMD estimator under proper penalization is consistent and efficient when residuals are potentially nonsmooth. [Chen \(2007\)](#) provides an extensive overview of sieve estimation of semi-nonparametric models.

and the surplus is shared through a competitive process that determines equilibrium wages $w^*(x)$ and profits $v^*(y)$. The equilibrium is characterized by the Walrasian stability condition: no worker-job pair can do better by deviating. As is well known in the optimal transport literature, under regularity conditions the stable matching can be formulated as a Monge-Kantorovich optimal transport problem whose solution determines a unique equilibrium assignment map $T^* : \mathcal{X} \rightarrow \mathcal{Y}$, mapping each worker x to her equilibrium job $y = T^*(x)$, and whose dual yields the equilibrium wage and profit functions.⁷

We specify a bilinear surplus function, as in [Tinbergen \(1956\)](#), [Bojilov and Galichon \(2016\)](#), and [Lindenlaub \(2017\)](#):

$$s(x, y) := s(x, y; A, b) = x' Ay + x' b, \quad (2.1)$$

where A is a $d \times d$ matrix and b is a $d \times 1$ vector. The diagonal elements of A capture within-task complementarities and the off-diagonal elements indicate between-task complementarities. $x' b$ represents non-interaction skill terms.⁸ We keep the parametric surplus specification for two primary reasons: i) interpretability and comparability; and ii) identification power. The quadratic surplus provides a direct economic interpretation of the technology parameters (in A) as measures of production complementarity. As [Bojilov and Galichon \(2016\)](#) advocated, the quadratic surplus is widely used in applied studies and provides simple, intuitive, and meaningful interpretations of worker-job interactions.⁹

⁷For the general theory, see [Villani \(2003, 2008\)](#), [Chiappori et al. \(2010\)](#), and [Galichon \(2017\)](#). For reference on matching with transferable utility, see [Browning et al. \(2014\)](#) and [Chiappori \(2017\)](#). The bilinear surplus specification used in this paper dates back to at least [Tinbergen \(1956\)](#), and has been employed in [Bojilov and Galichon \(2016\)](#), [Lindenlaub \(2017\)](#), and the recent working paper by [Boerma et al. \(2025\)](#).

⁸In the transferable utility framework with one-to-one matching, a linear term in y can be absorbed into the firm-side profit function without affecting equilibrium matching or wages, so it is normalized to zero for identification and parsimony.

⁹[Lindenlaub and Postel-Vinay \(2025\)](#) also employ a similar bilinear surplus specification when they empirically implement their theoretical model with a general surplus, referring to the bilinear specification as a second-order Taylor series approximation.

Matching assortativity depends crucially on the properties of the surplus function. To fix ideas, consider [Becker \(1973\)](#)'s spousal matching model where men and women are endowed with "ability indices" x and y , respectively. If $\partial^2 s(x, y) / \partial x \partial y \geq 0$, then $T^*(x) = F_y^{-1}(F_x(x))$ where F_x and F_y are the cumulative distribution functions of x and y . T^* having this property is defined as positive assortative matching (PAM) in the sense that high-type males match high-type females. Negative assortative matching (NAM) is the opposite.

The properties of A are pivotal to the assortativity of the equilibrium assignment in our specification. T^* satisfies PAM if A is diagonal with all positive principal minors (Proposition 2 of [Lindenlaub \(2017\)](#)). The assignment is unaffected by non-interaction terms because $\mathbb{E}_\pi [X'AY + X'b] = \mathbb{E}_\pi [X'AY] + \mathbb{E}_P [X'b]$ and the latter does not depend on π . Therefore, the dual optimal transport problem with $s(x, y)$ can be rewritten in terms of $s^o(x, y) = x'Ay$ as follows:

$$\begin{aligned} & \inf_{w \in \mathcal{W}} \left\{ \mathbb{E}_P [w(X)] + \mathbb{E}_Q \left[\sup_{x \in \mathcal{X}} \{s(x, Y) - w(x)\} \right] \right\} \\ &= \inf_{w^o \in \mathcal{W}} \left\{ \mathbb{E}_P [w^o(X)] + \mathbb{E}_Q \left[\sup_{x \in \mathcal{X}} \{s^o(x, Y) - w^o(x)\} \right] \right\} + \mathbb{E}_P [X]'b. \end{aligned} \quad (2.2)$$

Note that the solution w^* to the problem with the original s is obtained by $w^*(x) = w^{o*}(x) + x'b + c$ with any constant c where w^{o*} is the solution to the problem with s^o .

We now consider the case where the attributes of workers and firms are two-dimensional ($d = 2$), tailoring the model to our data. Every worker is endowed with a bundle of cognitive and manual skills, $x = (x_C, x_M)$. In turn, each firm is endowed with both cognitive and manual skill requirements, $y = (y_C, y_M)$. With $A = ((\alpha_{CC}, \alpha_{MC})', (\alpha_{CM}, \alpha_{MM})')$ and $b = (\beta_C, \beta_M)'$, define $\delta := \frac{\alpha_{MM}}{\alpha_{CC}}$, which represents the relative level of complementarities across cognitive and manual tasks. When both α_{CC} and α_{MM} are positive, $\delta < 1$ indicates that worker-job complementarity in the cognitive task is stronger than in the manual task.

Under joint normality of x and y , [Lindenlaub \(2017\)](#) and [Bojilov and Galichon \(2016\)](#) derive closed-form solutions for T^* and w^* and estimate the production technology parameters. In practice, however, x and y are non-normal as natural skills tend to have skewed

distributions. To align the data with the model, [Lindenlaub \(2017\)](#) converts each element of x and y into standard normal. Their dependence is then modeled using a Gaussian copula. Figure 1 illustrates how she derives the equilibrium assignment and wage function from transformed data. If the transformed data (\tilde{x} and \tilde{y}) are bivariate normal, this transformation provides a way to study properties of matching in terms of technology parameters. However, as the equilibrium assignments are based on the original data, not the transformed data, $T^*(x)$ (optimal assignment map from the original data) and $\tilde{T}^*(\tilde{x})$ (optimal assignment from transformed data) in Figure 1 may not deliver the same empirical implications. Moreover, the joint distribution of \tilde{x} and \tilde{y} is not in general normal, and hence, the model can be misspecified even after transforming the data. To avoid such misspecification, we allow x and y to have any arbitrary distributions in our model.

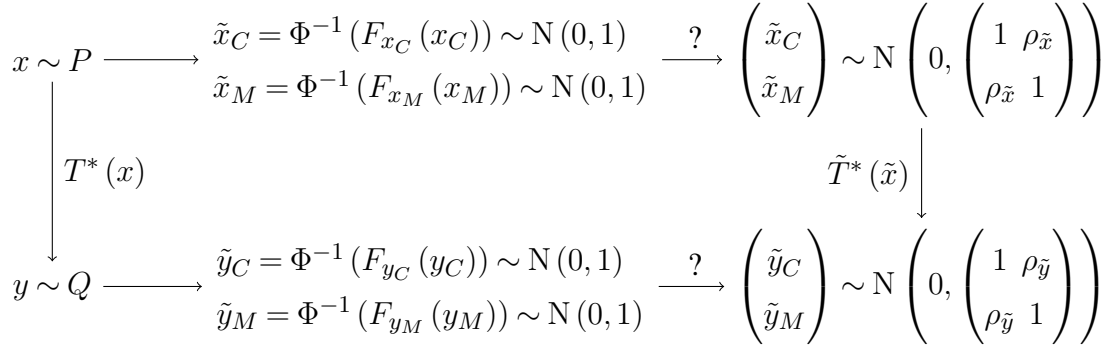


FIGURE 1.—[Lindenlaub \(2017\)](#)'s transformation. This figure should be read from left to right: starting from arbitrary marginals for x and y , transformations are applied to obtain standard normal marginals, which are then associated by the Gaussian copula. The true dependence structure after transformations can differ from the Gaussian copula. $T^*(x)$ and $\tilde{T}^*(\tilde{x})$ denote the optimal assignment maps in the original problem and in the transformed data, respectively. Φ denotes the standard normal c.d.f. F_{x_C} and F_{x_M} denote the c.d.f. for x_C and x_M , respectively.

We further impose conditions on the probability measures P and Q , as well as on A .

ASSUMPTION 1: (i) P and Q have finite second moments, and (ii) P is absolutely continuous with respect to the Lebesgue measure.

ASSUMPTION 2: The matrix A in the production technology (2.1) is invertible.

The following proposition derives the equilibrium assignment and wage as the solution to the dual optimal transport problem (2.2). This result builds on the characterization of optimal transport maps in Villani (2003, 2008). In the special case of jointly normally distributed characteristics, closed-form solutions are derived in Lindenlaub (2017) and (with entropy regularization) in Bojilov and Galichon (2016). Our contribution here is to state the result for general distributions under the bilinear surplus specification and to show that the equilibrium assignment is uniquely recovered from the gradient of the convex wage function.

PROPOSITION 1: *Let Assumption 1 hold. Then, there exists a unique (up to a constant) convex solution, $w^{o*}(x)$, to the second dual problem in (2.2), and the equilibrium wage (w^*) and assignment ($y^*(x) = T^*(x)$ where $y^*(x) = (y_1^*(x), \dots, y_d^*(x))'$) are given by*

$$w^*(x) = w^{o*}(x) + x'b + c, \quad Ay^*(x) = \nabla w^{o*}(x),$$

where c is the constant of integration. In addition, if Assumption 2 holds, the unique equilibrium assignment function is given by $y^*(x) = A^{-1}\nabla w^{o*}(x)$.

We can interpret this problem as assigning from \mathcal{X} to $A\mathcal{Y} := \{Ay : y \in \mathcal{Y}\}$. Denoting $z = Ay$, the surplus function can be written as $s(x, z) = x'z$, which satisfies the twist condition, regardless of the invertibility of A . Hence, Assumption 1 alone guarantees the existence of the convex solution, $w^{o*}(x)$, which implicitly depends on A . Assumption 2 is required to ensure the unique one-to-one mapping between x and y .

3. ECONOMETRIC MODEL AND IDENTIFICATION

This section describes the econometric model using the theoretical results in the previous section. Let $\{(w_i, x_i', y_i')\}_{i=1}^n$ represent an i.i.d. sequence of n matched observations on worker i 's wage w_i , her bundle of skills x_i , and the matched job's skill demands y_i . As Proposition 1 provides the equilibrium mapping from x to w and y that is interpreted as Walrasian equilibrium, we can write down a system of equations for w_i and y_i in terms of x_i . Pure Monge-Kantorovich optimal transport models (without unobserved heterogeneity) typically predict deterministic assignments between agent characteristics. To apply the

model in empirical analysis, the models are typically regularized by introducing unobserved heterogeneity, search frictions, or measurement errors. Here we introduce measurement errors in the equilibrium wage and assignment functions to keep the econometric model in line with optimal transport theory. Then, we show that the equilibrium functions and technology parameters in A are identified from the following nonparametric regression system.

$$\begin{aligned} w_i &= w^*(x_i) + \varepsilon_{wi} = w^{o*}(x_i) + x_i' b + c + \varepsilon_{wi}, \\ y_i &= y_i^*(x_i) + \varepsilon_{yi} = A^{-1} \nabla w^{o*}(x_i) + \varepsilon_{yi}. \end{aligned} \tag{3.1}$$

Here, ε_{wi} is a scalar measurement error in the observed wage. ε_{yi} is a $d \times 1$ vector of measurement errors in the firm's skill demands.

In the empirical application, the introduction of measurement errors can be motivated by the construction of skill measures from data. [Sanders \(2014\)](#), [Lindenlaub \(2017\)](#), and [Lise and Postel-Vinay \(2020\)](#) use the U.S. Department of Labor Occupational Characteristics Database (O*NET) to determine the levels of skills required for each categorical task. O*NET data provide rich information (more than 270 descriptors) on skill requirements for a large number of occupations. Measurement errors could arise in three ways. First, researchers conventionally classify the descriptors into predetermined skill categories, e.g., “cognitive” and “manual”. However, this decision may be far from clear-cut for many descriptors. Second, the descriptors are aggregated within each category using principal component analysis. This procedure inevitably produces measurement errors even if the descriptors are correctly classified. Lastly, there may be unobserved factors not included in O*NET. As discussed in the Related Literature, unobserved heterogeneity also arises in other matching frameworks, but our additive and separable measurement errors are fundamentally different from the surplus-entering unobservables in the CS tradition.

The closed-form solutions in the Gaussian model involve the productivity correlation $\rho_{y^*} = \text{corr}(y_C^*, y_M^*)$, which is the correlation of the latent equilibrium skill demands. However, the econometrician only observes $y = y^* + \varepsilon_y$, so the observed correlation $\text{corr}(y_C, y_M)$ generally differs from ρ_{y^*} due to attenuation from measurement error. In

the Gaussian ML implementation of [Lindenlaub \(2017\)](#), the observed correlation is used in place of the productivity correlation, which introduces a discrepancy in the likelihood function. Our sieve-based methods do not rely on the closed-form solution under bivariate normality, thereby avoiding this difficulty.

Unlike [Lindenlaub \(2017\)](#)'s approach, we do not need to impose distributional assumptions on measurement errors—assumptions that are vulnerable to misspecification. The moment conditions (3.2) below correspond to what is usually referred to as *weakly classical measurement error* in the terminology of [Schennach \(2016\)](#). We impose the following conditions assuming the exogeneity of x_i :

$$\mathbb{E}[\varepsilon_{wi}|x_i] = 0, \quad \mathbb{E}[\varepsilon_{yi}|x_i] = 0. \quad (3.2)$$

Let $\theta = (\text{vec}(A^{-1})', b')'$ denote a vector of unknown finite-dimensional parameters and $\theta \in \Theta$ where Θ is a compact subset of \mathbb{R}^{d^2+d} . The normalizing constant is not in our parameters of interest, and henceforth, we refer to $w(x) := w^{o*}(x) + c$ as the constant added infinite-dimensional parameter. We denote $z_i = (w_i, x_i', y_i')'$ and $\rho(z_i; \theta, w) = (\rho_w(w_i, x_i; \theta, w), \rho_y(y_i, x_i; \theta, w))'$, where

$$\rho_w(w_i, x_i; \theta, w) = w_i - (w(x_i) + x_i' b), \quad \rho_y(y_i, x_i; \theta, w) = y_i - A^{-1} \nabla w(x_i).$$

For each observation i , the model (3.1) satisfies the moment conditions (3.2). This implies that the following conditional moments hold:

$$\mathbb{E}[\rho(z_i; \theta, w) | x_i] = 0, \quad (3.3)$$

at the true parameter (θ_0, w_0) . Then (θ_0, w_0) are identified via the model (3.3) by [Proposition 1](#) and the exogeneity of x_i as well as the following assumption on \mathcal{Y} , which requires that the support of job attributes is not contained in a lower-dimensional affine subspace.

ASSUMPTION 3: *There exist $y_1, \dots, y_d, y_{d+1} \in \mathcal{Y}$ such that $\{y_1 - y_2, \dots, y_d - y_{d+1}\}$ is linearly independent.*

[Proposition 1](#) implies the existence of a deterministic equilibrium characterized by a unique convex function w_0 . When there is no non-interaction term ($b_0 = 0$), it follows that

$\nabla w_0^* = \nabla w_0$. The strict convexity of w_0 further implies that $\mathbb{E} [\nabla w_0(x_i) \nabla w_0(x_i)']$ has a full rank, thus identifying A_0 . Additionally, Assumption 3 is sufficient to identify the nonzero vector b_0 , as stated in the following theorem.

THEOREM 1: *Let Assumptions 1-3 hold and the moment conditions (3.3) be satisfied. Then, θ_0 and $w_0 = w_0^{o*} + c_0$ are identified.*

We can identify w_0^{o*} and c_0 separately under a normalization such as $w_0^{o*}(x_0) = 0$ for some $x_0 \in \mathcal{X}$ or $\int_{\mathcal{X}} w_0^{o*}(x) dx = 0$ when \mathcal{X} is bounded. With the former constraint, c_0 and $w_0^{o*}(x)$ are identified with $w_0^*(x) = w_0^{o*}(x) + x'b$ since $c_0 = w_0^{o*}(x_0) + c_0 = w_0^*(x_0) - x_0'b$.

REMARK 1: Our framework can be extended to allow for endogenous x_i using non-parametric instrumental-variables (NPIV) methods. Specifically, if one has instruments z_i satisfying $E[\varepsilon_{w_i}|z_i] = E[\varepsilon_{y_i}|z_i] = 0$, the equilibrium wage and assignment functions are identified from moment conditions of the form

$$E[(w_i - w^*(x_i))\phi(z_i)] = 0, \quad E[(y_i - y^*(x_i))\phi(z_i)] = 0$$

for a family of test functions $\phi(\cdot)$, under standard completeness conditions. In this case, the technology parameters and the equilibrium wage function can be estimated by NPIV estimators, e.g., [Blundell et al. \(2007\)](#). NPIV estimation is in general an ill-posed problem. However, in our model, the equilibrium wage function is convex, so the finite sample properties can be improved by imposing a shape restriction as shown in [Chetverikov and Wilhelm \(2017\)](#).

REMARK 2: Our framework extends to the case where the two sides of the market have different numbers of observable characteristics ($d_x \neq d_y$). Under the bilinear surplus, the matching problem is equivalent to optimally assigning x to $Ay \in \mathbb{R}^{d_x}$, regardless of the dimension of y . This allows identification of the equilibrium wage function and the technology parameters without imposing $d_x = d_y$; see [Chiappori et al. \(2017, 2020\)](#), [McCann and Pass \(2020\)](#), and [Nenna and Pass \(2023\)](#) for general unequal-dimension optimal transport formulations, and Appendix B for the formal identification result under our bilinear

surplus specification. When $d_x \geq d_y$ and A has rank d_y , the equilibrium assignment function $y^*(x) = (A'A)^{-1}A'\nabla w^{o*}(x)$ is also identified, and the matching between x and y is pure.

4. SIEVE-BASED SEMIPARAMETRIC ESTIMATION

The model parameters are identified by the semiparametric conditional moment restrictions (3.3). If the function w is parametrically specified, these moment conditions lead to standard GMM estimation. As w is infinite-dimensional in our specification, we approximate it using sieves. The unknown function $w \in \mathcal{W}$ is approximated by $w_n \in \mathcal{W}_n$ where \mathcal{W}_n is an approximating multivariate function space becoming dense in \mathcal{W} as $n \rightarrow \infty$. We generate \mathcal{W}_n via tensor-product construction. One could consider a kernel-based approximation for the equilibrium wage and assignment functions. However, using sieves (e.g., polynomials or splines) is particularly advantageous in our setting. First, the equilibrium assignment is determined by $y = \nabla w(x)$. Sieves allow us to compute this gradient analytically, so that the functional form information is easily incorporated in estimation. Second, the sieve objective functions are linear in parameters, so that obtaining the score and Hessian is relatively straightforward. Finally, our theory implies that w is convex. The sieve framework accommodates global convexity constraints in a straightforward way, improving stability and computational efficiency.

From now on, we assume that there are sets of firms and workers with $d = 2$. Every worker is endowed with a bundle of cognitive and manual skills, $x = (x_C, x_M)$. In turn, each firm is endowed with both cognitive and manual skill demands, $y = (y_C, y_M)$. y_C (y_M) corresponds to the productivity or skill requirement of cognitive task C (manual task M). In our case,

$$\mathcal{W}_n = \left\{ w_n : \mathcal{X} \rightarrow \mathbb{R}, w_n(x; \gamma) = \sum_{j_C=0}^{k_{Cn}} \sum_{j_M=0}^{k_{Mn}} \gamma_{j_C j_M} p_{j_C}(x_C) p_{j_M}(x_M), \gamma_{j_C j_M} \in \mathbb{R} \right\}, \quad (4.1)$$

where $\{p_{j_C}(x_C)\}_{j_C=0}^{k_{Cn}}$ and $\{p_{j_M}(x_M)\}_{j_M=0}^{k_{Mn}}$ are basis functions of x_C and x_M . The tensor-product space is simple to extend to higher dimensions and easy to implement. For our second and third conditional moment restrictions, we approximate $\partial w_0(x) / \partial x_C$ and

$\partial w_0(x)/\partial x_M$ with the same parameter values $\{\gamma_{j_C j_M}\}$ used to approximate $w_0(x)$ in \mathcal{W}_n :

$$\begin{aligned}\partial w_n(x; \gamma)/\partial x_C &= \sum_{j_C=0}^{k_{Cn}} \sum_{j_M=0}^{k_{Mn}} \gamma_{j_C j_M} (\partial p_{j_C}(x_C)/\partial x_C) p_{j_M}(x_M), \\ \partial w_n(x; \gamma)/\partial x_M &= \sum_{j_C=0}^{k_{Cn}} \sum_{j_M=0}^{k_{Mn}} \gamma_{j_C j_M} p_{j_C}(x_C) (\partial p_{j_M}(x_M)/\partial x_M).\end{aligned}$$

We first consider the model (3.1) with normally distributed, mean-zero measurement errors that are mutually uncorrelated. Then, we can estimate the parameters using sieve maximum likelihood (SML). Assuming $\varepsilon_i \sim N(0, \Sigma)$, we write the log-likelihood function of model (3.1) as

$$L^*(\theta, \Sigma, w(\cdot)) = -\frac{n}{2} \log \det(\Sigma) + \sum_{i=1}^n \log |\det(\partial \rho_i / \partial (w_i, y_{C_i}, y_{M_i}))| - \frac{1}{2} \sum_{i=1}^n \rho_i' \Sigma^{-1} \rho_i,$$

where $\rho_i = \rho(z_i; \theta, w)$. Solving $\partial L^* / \partial \Sigma = 0$ for Σ , we get $\Sigma = \frac{1}{n} \sum_{i=1}^n \rho_i \rho_i'$, which yields the concentrated log-likelihood function of our model

$$L(\theta, w(\cdot)) = -\frac{n}{2} \log \det \left(\frac{1}{n} \sum_{i=1}^n \rho_i \rho_i' \right). \quad (4.2)$$

The value of (θ, w) maximizing (4.2) is the sieve nonlinear full information maximum likelihood estimator of (θ, w) .

The normality assumption on measurement errors has no theoretical or empirical basis. Without any distributional assumptions on measurement errors, we can still estimate (θ, w) using several sieve M-estimators. As $\rho(z; \theta, w) - \rho(z; \theta_0, w_0)$ does not depend on y under Assumption 2, we can apply the sieve generalized least squares (GLS) procedure (Chen, 2007) that minimizes the following objective function with respect to (θ, w) :

$$\min_{(\theta, w)} \sum_{i=1}^n \rho(z_i; \theta, w)' \left[\hat{\Sigma}_0(x_i) \right]^{-1} \rho(z_i; \theta, w),$$

where $\hat{\Sigma}_0(x)$ is a consistent estimator of the optimal weighting matrix $\Sigma_0(x) := \text{Var}[\rho(z_i; \theta, w) | x_i = x]$. In addition, if A is diagonal, we can rewrite the last two moment conditions as $\rho_C(y_C, x; \kappa_C, w) = y_C - \kappa_C \nabla_C w(x)$ and $\rho_M(y_M, x; \kappa_M, w) = y_M -$

$\kappa_M \nabla_M w(x)$, where $\kappa_C = \alpha_{CC}^{-1}$ and $\kappa_M = \alpha_{MM}^{-1}$. Table I outlines the three-step procedure to compute the SGLS estimator.

TABLE I

THREE-STEP PROCEDURE FOR SIEVE GLS ESTIMATION (CHEN, 2007)

ALGORITHM: Computing the Sieve GLS Estimator of θ and w

1. Obtain an initial consistent sieve LS estimator $(\tilde{\theta}_n, \tilde{w}_n)$ by

$$\min_{(\theta, w)} \sum_{i=1}^n \rho(z_i; \theta, w)' \rho(z_i; \theta, w),$$
2. Obtain a consistent estimator $\hat{\Sigma}_0(x)$ of $\Sigma_0(x) = \text{Var}[\rho(z_i; \theta, w) | x_i = x]$ using $(\tilde{\theta}_n, \tilde{w}_n)$ and sieve LS estimation.
3. Obtain the optimally weighted sieve GLS estimator $(\hat{\theta}_n, \hat{w}_n)$ by

$$\min_{(\theta, w)} \sum_{i=1}^n \rho(z_i; \theta, w)' [\hat{\Sigma}_0(x_i)]^{-1} \rho(z_i; \theta, w).$$

The SGLS estimator allows for arbitrary correlation between measurement errors and heteroskedasticity. We can impose homoskedasticity by assuming $\Sigma_0(x) = \Sigma_0$ so that the optimal weighting matrix does not vary with x . If we further assume that the measurement errors are uncorrelated, i.e., Σ_0 is diagonal, we can use the sieve least squares (SLS) estimator from step 1 of the three-step procedure. We summarize the key differences in assumptions imposed in different estimation procedures in Table II.

TABLE II

COMPARISON OF KEY ASSUMPTIONS IN ESTIMATION PROCEDURES

| Assumptions | Lindenlaub (2017) | Sieve Estimators | | |
|--|-------------------|------------------|-----|------|
| | ML | SML | SLS | SGLS |
| Joint normality of x and y | ✓ | | | |
| Normality of measurement errors | ✓ | ✓ | | |
| Uncorrelated measurement errors | ✓ | ✓ | | |
| Homoskedasticity of measurement errors | ✓ | ✓ | ✓ | |

We implement the sieve estimators using finite-dimensional Bernstein polynomials to construct the approximating space \mathcal{W}_n of \mathcal{W} on $[0, 1]^2$. The basis functions are $p_{j_C}(x_C) = \binom{k_{C_n}}{j_C} (x_C)^{j_C} (1 - x_C)^{k_{C_n} - j_C}$ and $p_{j_M}(x_M) = \binom{k_{M_n}}{j_M} (x_M)^{j_M} (1 - x_M)^{k_{M_n} - j_M}$, where $j_C = 0, 1, \dots, k_{C_n}$, $j_M = 0, 1, \dots, k_{M_n}$, and $\binom{k}{j}$ is a binomial coefficient.¹⁰

If $\gamma_{j_C j_M} = w(j_C/k_{C_n}, j_M/k_{M_n})$, the Bernstein polynomial $w_n(x; \gamma)$ converges uniformly to $w(x)$ by the Stone-Weierstrass approximation theorem (see, e.g., [Lorentz \(1986\)](#)). This provides an approach to imposing shape restrictions on the sieve estimator with a linear constraint which can be solved easily.¹¹ Without any constraint, the equilibrium wage function, $w(x) + x'b$, is unique and convex. To obtain a more stable estimator, without loss of generality, we impose linear constraints on the Bernstein polynomials, which are necessary for the function to be convex. A detailed description of implementing this convexity constraint in the estimation procedures is provided in [Appendix C](#).

To understand technological changes in the production function, the parametric components of the model are of primary interest. The SGLS estimator is ideal in this case because for θ (i) it is easy to use, \sqrt{n} -consistent, and asymptotically normal; (ii) it is semiparametrically efficient; and (iii) the asymptotic variance estimator of $\hat{\theta}$ is consistent and easy-to-compute.¹² We formally derive its asymptotic properties in the following section.

5. ASYMPTOTIC THEORY FOR THE SGLS ESTIMATOR

We establish consistency, convergence rate, asymptotic normality, and semiparametric efficiency of our SGLS estimator using results in [Chen and Shen \(1998\)](#), [Ai and Chen \(2003\)](#), and [Chen \(2007\)](#). Define $\lambda := (\theta, w(\cdot))$. Let $\hat{\lambda}_n$ and λ_0 denote our sieve GLS es-

¹⁰ x does not lie in $[0, 1]^2$ in many applications. To satisfy the domain restriction for our simulation studies and empirical application, we use the following linear transformation when $(x_C, x_M) \in [\underline{x}_C, \bar{x}_C] \times [\underline{x}_M, \bar{x}_M]$: $p_{j_C}(x_C) = \binom{k_{C_n}}{j_C} x_1^{j_C} (1 - x_1)^{k_{C_n} - j_C}$ and $p_{j_M}(x_M) = \binom{k_{M_n}}{j_M} x_2^{j_M} (1 - x_2)^{k_{M_n} - j_M}$, where $x_1 = (x_C - \underline{x}_C)/(\bar{x}_C - \underline{x}_C)$ and $x_2 = (x_M - \underline{x}_M)/(\bar{x}_M - \underline{x}_M)$.

¹¹[Compiani \(2022\)](#) uses linear constraints for the function w to impose monotonicity restrictions and a so-called “diagonal dominance” constraint.

¹²The sieve minimum distance estimator can be considered. However, when $\rho(z; \theta, w) - \rho(z; \theta_0, w_0)$ does not depend on y , the SGLS estimator is simpler to implement and computationally faster.

estimator and the true parameter values, respectively. We first show that $\hat{\lambda}_n$ converges to λ_0 at a rate faster than $n^{-1/4}$ under a pseudo norm $\|\cdot\|$. For any $\lambda_1 = (\theta_1, w_1(\cdot))$, $\lambda_2 = (\theta_2, w_2(\cdot)) \in \Lambda$, $\|\cdot\|$ is defined as

$$\|\lambda_1 - \lambda_2\|^2 = \mathbb{E} \left[\left(\frac{d\rho(z_i; \lambda_0)}{d\lambda} [\lambda_1 - \lambda_2] \right)' \Sigma(x_i)^{-1} \left(\frac{d\rho(z_i; \lambda_0)}{d\lambda} [\lambda_1 - \lambda_2] \right) \right],$$

where

$$\frac{d\rho(z; \lambda_0)}{d\lambda} [\lambda_1 - \lambda_2] = \begin{pmatrix} w_1(x) - w_2(x) + x'(b_1 - b_2) \\ \nabla_C w_0(x) (\kappa_{C1} - \kappa_{C2}) + \kappa_{C0} (\nabla_C w_1(x) - \nabla_C w_2(x)) \\ \nabla_M w_0(x) (\kappa_{M1} - \kappa_{M2}) + \kappa_{M0} (\nabla_M w_1(x) - \nabla_M w_2(x)) \end{pmatrix}.$$

The pseudo metric is comparatively weaker than the standard sup or L_2 metric, wherein convergence of $\hat{\lambda}_n$ to λ_0 under the standard metric implies convergence under the pseudo metric. [Ai and Chen \(2003\)](#) show that $\hat{\lambda}_n$ converging at a rate faster than $n^{-1/4}$ under the weaker metric $\|\cdot\|$ suffices to derive the \sqrt{n} -asymptotic normality of the parametric component, $\hat{\theta}_n$.

Let $\Lambda = \Theta \times \mathcal{W}$ be equipped with a norm $\|\lambda\|_s = |\theta|_e + \|w\|_\infty + \|\nabla_C w\|_\infty + \|\nabla_M w\|_\infty$, where $|\cdot|_e$ denotes the Euclidean norm and $\|w\|_\infty = \sup_{x \in \mathcal{X}} |w(x)|$ is the supremum norm. We introduce the Hölder class of functions. Let $[m]$ be the largest nonnegative integer such that $[m] < m$. A real-valued function w on \mathcal{X} is said to be in Hölder space $\Lambda^m(\mathcal{X})$ if it is $[m]$ times continuously differentiable on \mathcal{X} and

$$\max_{\ell_1 + \ell_2 \leq [m]} \sup_x \left| \frac{\partial^{\ell_1 + \ell_2} w(x)}{\partial x_C^{\ell_1} \partial x_M^{\ell_2}} \right| + \sup_{m_1 + m_2 = [m]} \sup_{x, x'} \left| \frac{\partial^{[m]} w(x)}{\partial x_C^{m_1} \partial x_M^{m_2}} - \frac{\partial^{[m]} w(x')}{\partial x_C^{m_1} \partial x_M^{m_2}} \right| / |x - x'|_e^{m - [m]}$$

is finite. We provide the following assumptions for convergence.

ASSUMPTION 4: (i) $\{w_i, y_i', x_i'\}_{i=1}^n$ are i.i.d.; (ii) \mathcal{X} is compact and a Cartesian product of compact intervals \mathcal{X}_C and \mathcal{X}_M .

ASSUMPTION 5: $\Sigma(x)$ and $\Sigma_0(x) \equiv \text{Var}(\rho(z_i, \lambda_0) | x_i = x)$ are positive definite and bounded uniformly over $x \in \mathcal{X}$.

ASSUMPTION 6: $\Lambda \equiv \Theta \times \mathcal{W}$ is compact under $\|\cdot\|_s$.

ASSUMPTION 7: (i) $w \in \Lambda^m(\mathcal{X})$ with $m > 2$; (ii) $\forall w \in \Lambda^m(\mathcal{X}), \exists w_n(x; \gamma) \in \mathcal{W}_n$ such that $\|w_n - w\|_\infty = O((k_{Cn}k_{Mn})^{-m/2})$ with $k_{Cn}, k_{Mn} = O(n^{1/2(m+1)})$.

Assumption 4 requires the data to be i.i.d. and the support of x to be compact. Assumption 5 ensures the existence and invertibility of the weight matrix for GLS estimation. Assumption 6 requires the parameter space to be compact under the sup norm. We do not explicitly require identification of λ here as Assumptions 1–3 guarantee it by Theorem 1. Assumption 7 quantifies the sieve approximation error. Most papers in the literature require $m > d_{\mathcal{X}}/2$, where $d_{\mathcal{X}}$ is the dimension of \mathcal{X} . However, our objective function involves $\nabla_C w(x)$ and $\nabla_M w(x)$, so we need a higher order of m . Note that the smoothness of w can be verified using the theory of optimal transport. The degree of the smoothness of the solution function w depends on how smooth the densities of x and y^* are.¹³ The following proposition establishes the convergence rate of $\hat{\lambda}_n$.

PROPOSITION 2: If Assumptions 1–7 hold, then $\|\hat{\lambda}_n - \lambda_0\| = o_p(n^{-1/4})$.

We now derive the asymptotic normality of the parametric components of the SGLS estimator, $\hat{\theta}_n$. Define $D_v(x) := (D_{v_1}(x), D_{v_2}(x), D_{v_3}(x), D_{v_4}(x))$ where

$$D_{v_1}(x) = \begin{pmatrix} v_1(x) \\ \kappa_{C0} \nabla_C v_1(x) - \nabla_C w_0(x) \\ \kappa_{M0} \nabla_M v_1(x) \end{pmatrix}, \quad D_{v_2}(x) = \begin{pmatrix} v_2(x) \\ \kappa_{C0} \nabla_C v_2(x) \\ \kappa_{M0} \nabla_M v_2(x) - \nabla_M w_0(x) \end{pmatrix},$$

$$D_{v_3}(x) = \begin{pmatrix} v_3(x) - x_C \\ \kappa_{C0} \nabla_C v_3(x) \\ \kappa_{M0} \nabla_M v_3(x) \end{pmatrix}, \quad D_{v_4}(x) = \begin{pmatrix} v_4(x) - x_M \\ \kappa_{C0} \nabla_C v_4(x) \\ \kappa_{M0} \nabla_M v_4(x) \end{pmatrix}.$$

¹³Assuming the densities are bounded away from zero and infinity, if the densities of variables x and y^* belong to the space Λ^{m-2} , the function w_0 is a member of $\Lambda^m(\mathcal{X})$. For a more comprehensive understanding, refer to Caffarelli (1992a,b, 1996) which cover the case of compactly supported \mathcal{X} and \mathcal{Y}^* . Cordero-Erausquin and Figalli (2019) provides an extended result for distributions with unbounded supports.

Let $v^* = (v_1^*, v_2^*, v_3^*, v_4^*)$, where v_j^* solves

$$\inf_{v_j} \mathbb{E} \left[D_{v_j}(x_i)' \Sigma(x_i)^{-1} D_{v_j}(x_i) \right]. \quad (5.1)$$

ASSUMPTION 8: (i) $\mathbb{E} [D_{v^*}(x_i)' D_{v^*}(x_i)]$ is positive definite; (ii) Each element of v^* belongs to the Hölder space $\Lambda^m(\mathcal{X})$ with $m > 2$.

ASSUMPTION 9: $\theta_0 \in \text{int}(\Theta)$.

Assumptions 8–9 are standard conditions for M-estimation that guarantee the uniqueness of the solution within the parameter space. Under Assumptions 1–8, it follows from Lemma B.1 in Ai and Chen (2003) that $|\hat{\theta}_n - \theta_0|_e = o_p(n^{-1/4})$, $\|\hat{w}_n - w_0\|_2 = \left(\mathbb{E} \left[(\hat{w}_n(x_i) - w_0(x_i))^2 \right] \right)^{1/2} = o_p(n^{-1/3})$, and $\|\nabla_C \hat{w}_n - \nabla_C w_0\|_2, \|\nabla_M \hat{w}_n - \nabla_M w_0\|_2 = o_p(n^{-1/4})$. Now the following theorem provides the asymptotic normality of $\hat{\theta}_n$.

THEOREM 2: Let Assumptions 1–9 hold. Then, $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, V_1^{-1} V_2 V_1^{-1})$, where

$$V_1 = \mathbb{E} \left[D_{v^*}(x_i)' \Sigma(x_i)^{-1} D_{v^*}(x_i) \right],$$

$$V_2 = \mathbb{E} \left[D_{v^*}(x_i)' \Sigma(x_i)^{-1} \Sigma_0(x_i) \Sigma(x_i)^{-1} D_{v^*}(x_i) \right].$$

The asymptotic variance $V_1^{-1} V_2 V_1^{-1}$ can be consistently estimated (see Remark 4.2 in Chen (2007)) and the standard errors of $(\hat{\alpha}_{CC}, \hat{\alpha}_{MM}) = (1/\hat{\kappa}_C, 1/\hat{\kappa}_M)$ are obtained by using the delta method. Furthermore, if all conditions of Theorem 2 are satisfied with $\Sigma(x) = \Sigma_0(x)$, $\hat{\theta}_n$ achieves semiparametric efficiency with a consistent estimator $\hat{\Sigma}_0(x)$ of $\Sigma_0(x)$. The estimation of $\Sigma_0(x)$ is straightforward through series least squares estimation, using the initial consistent SLS estimator $(\tilde{\theta}_n, \tilde{w}_n)$. To ensure the efficiency of the SGLS estimator, $\hat{\Sigma}_0(x)$ is required to exhibit the following uniform convergence rate.

ASSUMPTION 10: $\hat{\Sigma}_0(x) = \Sigma_0(x) + o_p(n^{-1/4})$ uniformly over $x \in \mathcal{X}$.

Let $v_0 = (v_{01}, v_{02}, v_{03}, v_{04})$, where v_{0j} solves (5.1) with $\Sigma(x)$ replaced by $\Sigma_0(x)$. Now the following theorem establishes the semiparametric efficiency of $\hat{\theta}_n$.

THEOREM 3: *Suppose that all conditions of Theorem 2 with $\Sigma(x) = \Sigma_0(x)$ and $v^* = v_0$ hold, and Assumption 10 is satisfied. Then, $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d N(0, V_0^{-1})$, with $V_0 = \mathbb{E} \left[D_{v_0}(x_i)' \Sigma_0(x_i)^{-1} D_{v_0}(x_i) \right]$.*

Note that the assumptions introduced in this section are standard regularity conditions for sieve M-estimation. They do not impose additional restrictive structures compared to the assumptions required for Lindenlaub (2017)'s maximum likelihood (ML) estimation. Rather, they are the nonparametric generalizations of the standard regularity conditions for ML estimators.¹⁴

6. MONTE CARLO SIMULATIONS

This section evaluates the finite sample performances of our sieve estimators using known data-generating processes (DGPs). We first generate Monte Carlo samples from Lindenlaub (2017)'s quadratic-Gaussian model. Workers' skill bundle, x , and occupations' skill requirements, y , follow joint Gaussian distributions:

$$\begin{pmatrix} x_C \\ x_M \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_x \\ \rho_x & 1 \end{pmatrix} \right), \quad \begin{pmatrix} y_C \\ y_M \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_y \\ \rho_y & 1 \end{pmatrix} \right),$$

from which $\{x_i\}_{i=1}^n$ are drawn with sample size $n = 3000$. Following Lindenlaub (2017), we rule out between-task complementarities in the production technology (2.1) so that A is a diagonal matrix ($A = \text{diag}(\alpha_{CC}, \alpha_{MM})$). Then, the equilibrium assignment y^* and wage w^* have closed-form solutions as shown in Lindenlaub (2017). Both depend on ρ_x, ρ_y , and technology parameters. We generate $\{y_i^*\}_{i=1}^n$ and the equilibrium wages $\{w_i^*\}_{i=1}^n$ using closed-form solutions. Lastly, we draw measurement errors from mean-zero Gaussian distributions:

$$\varepsilon_w \sim N(0, \sigma_w^2), \quad \varepsilon_C \sim N(0, \sigma_C^2), \quad \varepsilon_M \sim N(0, \sigma_M^2),$$

¹⁴For example, Assumptions 4, 5, 8, and 9 have exact counterparts in standard finite-dimensional MLE theory. Assumptions 6, 7, and 10 are merely the technical conditions required to allow the underlying distributions of x and y to be "unknown smooth distributions".

and add them to w_i^* , y_{Ci}^* , and y_{Mi}^* , respectively, to generate the observable data $(w_i, y_i, x_i)_{i=1}^n$ following (3.1). The true parameter values used in simulations are:

$$(\alpha_{CC}, \alpha_{MM}, \beta_C, \beta_M, c, \rho_x, \rho_y, \sigma_w, \sigma_C, \sigma_M) = (0.5, 0.2, 1.7, -0.4, 30, -0.4, -0.5, 2, 1, 1),$$

which are set close to the empirical ML estimates in [Lindenlaub \(2017\)](#).

We estimate the production technology parameters using [Lindenlaub \(2017\)](#)'s parametric ML estimator and our sieve estimators (SML, SLS, and SGLS) across 1000 Monte Carlo samples. As we discussed earlier, the original ML estimator suffers from bias because the solution uses the measurement-error-contaminated productivity correlation $\tilde{\rho}_y = \text{corr}(y)$, which is different from the true productivity correlation $\rho_y = \text{corr}(y^*)$. If the measurement errors in y are negligible, e.g., (σ_C, σ_M) are close to 0, the ML estimator works well for this DGP. However, given the current parameter specification, the measurement errors are substantial, so the ML estimator can be inconsistent. To address this issue, we define a corrected ML estimator (referred to as ‘ML*’) that uses the corrected productivity correlation $\rho_y = \tilde{\rho}_y \sqrt{\text{var}(y_1)\text{var}(y_2)} / \sqrt{(\text{var}(y_1) - \sigma_C^2)(\text{var}(y_2) - \sigma_M^2)}$. This correction in turn yields much more precise estimates than the original ML estimator. The sieve estimators do not share this problem.

The box plots in [Figure 2](#) summarize the distributions of parameter estimates delivered by the 5 estimators we consider. For (β_C, β_M) , all the estimators work well. On the other hand, for $(\alpha_{CC}, \alpha_{MM})$, the original ‘‘ML’’ estimator shows substantial bias due to the use of the observed (rather than latent) productivity correlation in the likelihood. Our corrected ‘‘ML*’’ estimator, on the contrary, accurately recovers the true parameters and performs the best under the Gaussian DGP as it relies on the true specification. In this regard, we focus on ‘‘ML*’’ as the appropriate parametric benchmark and omit the ‘‘ML’’ results in subsequent analyses. The sieve estimators also perform very well for $(\alpha_{CC}, \alpha_{MM})$ as the distributions of their parameter estimates are centered around the true parameter values. We document the estimators’ bias and root-mean-squared errors (RMSE) in [Table III](#). Notably, the sieve estimators, while being more robust, tend to be no less efficient than the parametric estimators. The SML estimator’s root-mean-squared errors (RMSE) are slightly larger than

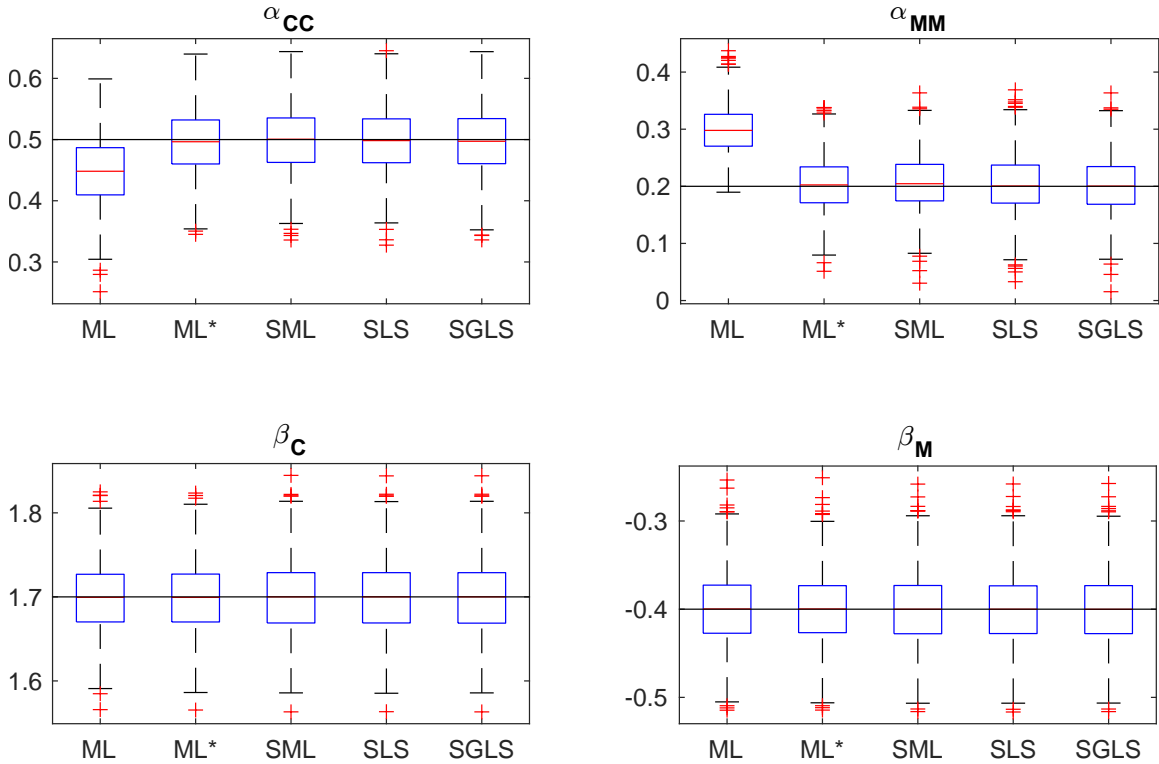


FIGURE 2.—Box plots of parameter estimates (Gaussian DGP)

those of the ML^* estimator for most parameters. The SLS and SGLS estimators—which perform very similarly because the measurement errors are uncorrelated—are slightly less efficient than the SML estimator, as expected.

Now we consider DGPs for which the Gaussian model is moderately misspecified. We first generate $\{x_{1i}, x_{2i}\}_{i=1}^n$ and $\{y_{1j}, y_{2j}\}_{j=1}^n$ separately from the Gumbel copula with shape parameter values 1.3 and 1.4, respectively. These variables are uniformly distributed in $[0, 1]$ by construction. Then we transform them into standard normally distributed variables:

$$x_{Ci} = \Phi^{-1}(x_{1i}), \quad x_{Mi} = \Phi^{-1}(1 - x_{2i}), \quad y_{Cj} = \Phi^{-1}(y_{1j}), \quad y_{Mj} = \Phi^{-1}(1 - y_{2j}),$$

TABLE III

FINITE SAMPLE PERFORMANCES OF THE ESTIMATORS (GAUSSIAN DGP)

| | | ML | ML* | SML | SLS | SGLS |
|---------------|------|---------|---------|---------|---------|---------|
| α_{CC} | Bias | -0.0536 | -0.0041 | -0.0007 | -0.0018 | -0.0027 |
| | RMSE | 0.0775 | 0.0513 | 0.0520 | 0.0523 | 0.0523 |
| α_{MM} | Bias | 0.0992 | 0.0044 | 0.0065 | 0.0031 | 0.0022 |
| | RMSE | 0.1077 | 0.0473 | 0.0491 | 0.0502 | 0.0489 |
| β_C | Bias | -0.0006 | -0.0007 | -0.0009 | -0.0009 | -0.0009 |
| | RMSE | 0.0416 | 0.0414 | 0.0427 | 0.0427 | 0.0427 |
| β_M | Bias | -0.0000 | -0.0001 | -0.0000 | -0.0000 | -0.0000 |
| | RMSE | 0.0398 | 0.0395 | 0.0397 | 0.0397 | 0.0397 |

so that (x_{Ci}, x_{Mi}) and (y_{Cj}, y_{Mj}) are negatively correlated. Define matrices x and y by

$$x := \begin{bmatrix} x_{C1} & x_{M1} \\ \vdots & \vdots \\ x_{Cn} & x_{Mn} \end{bmatrix}, \quad y := \begin{bmatrix} y_{C1} & y_{M1} \\ \vdots & \vdots \\ y_{Cn} & y_{Mn} \end{bmatrix}.$$

The skill demand and supply bundles are standard normally distributed in this DGP so the ML* estimator is directly applicable without transformation, although it is misspecified because their joint distributions are not Gaussian. The Gumbel copula exhibits asymmetric tail dependence (the upper tail has stronger dependence than the lower tail), whereas the Gaussian copula has symmetric dependence. However, given the current parameter setup, the Gumbel copula does not drastically differ from the Gaussian copula. The production technology is specified as before.

There exists no closed-form solution for the equilibrium assignment in this case. We therefore numerically solve the equilibrium matching through linear programming for each Monte Carlo sample. To do so, we first compute the pairwise surplus of each possible match between x and y and construct the surplus matrix S whose ij entry is the surplus generated by worker i and firm j . Let \mathbb{I}_n denote an $n \times n$ identity matrix and $\mathbf{1}_n$ be an $n \times 1$ vector of ones. Let f be the vector obtained by stacking the columns of S . Then the solution (q^*) to

the following linear programming problem provides the equilibrium assignment:

$$\max_q f'q, \quad \text{s.t.} \quad \begin{bmatrix} \underbrace{A_1}_{n \times n^2} \\ \underbrace{A_2}_{n \times n^2} \end{bmatrix} q \leq \underbrace{b}_{2n \times 1}, \quad (6.1)$$

where $A_1 := \mathbb{I}_n \otimes \mathbf{1}'_n$, $A_2 := \mathbf{1}'_n \otimes \mathbb{I}_n$, and $b := \mathbf{1}_{2n}$. Reshaping q^* into an $n \times n$ matrix gives the optimal transportation matrix, T . Then the optimal matching for x is given by $y^* := Ty$. The wage w^* is computed from the solution to the dual problem (6.1).¹⁵

For measurement errors, we consider two different specifications. In the first case, the errors are independently drawn from a gamma distribution, $\Gamma(a, b)$, with $a = 1$, $b = 2$. They are demeaned and scaled to have the same means and variances specified in the Gaussian DGP. Under this specification, the ML^* and SML estimators are further misspecified as measurement errors are non-normal. In the second case, we generate the errors from a joint normal distribution in which the errors are correlated as follows:

$$\begin{pmatrix} \varepsilon_w \\ \varepsilon_C \\ \varepsilon_M \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 0.5 \\ 1 & 0.5 & 1 \end{pmatrix} \right).$$

The ML^* and SML estimators are still misspecified as measurement errors are correlated. The SLS estimator is consistent but not as efficient as the SGLS estimator, since it does not take the correlation structure of errors into account. The observable data $(y_i, x_i, w_i)_{i=1}^n$ are generated by adding the measurement errors to w_i^* , y_{Ci}^* , and y_{Mi}^* respectively.

The estimation results are provided in Table IV. In both cases, the ML^* estimator is misspecified for the distributions of X , Y , and the measurement errors so that it performs the worst for all the parameters. It exhibits especially large biases and RMSE for complementarity parameters. Even for the linear productivity parameters, the ML^* estimator shows

¹⁵Even with the moderate sample size $n = 3000$, the constraint matrix is enormous ($6000 \times 9,000,000$). Solving the linear program (6.1) over many Monte Carlo samples is computationally demanding. We employ GUROBI OPTIMIZER 10.0 to solve it efficiently.

TABLE IV

FINITE SAMPLE PERFORMANCES OF THE ESTIMATORS (GUMBEL DGP)

| | | Gamma errors | | | | Joint Gaussian errors | | | |
|---------------|------|--------------|---------|---------|---------|-----------------------|---------|---------|---------|
| | | ML* | SML | SLS | SGLS | ML* | SML | SLS | SGLS |
| α_{CC} | Bias | -0.0608 | -0.0021 | -0.0015 | -0.0020 | -0.2032 | -0.0023 | 0.0008 | 0.0112 |
| | RMSE | 0.3780 | 0.0538 | 0.0535 | 0.0538 | 0.2971 | 0.0914 | 0.0925 | 0.0809 |
| α_{MM} | Bias | -0.0550 | 0.0019 | 0.0024 | 0.0020 | -0.3339 | 0.0018 | 0.0034 | -0.0123 |
| | RMSE | 0.4209 | 0.0512 | 0.0527 | 0.0513 | 1.6626 | 0.0886 | 0.0898 | 0.0827 |
| β_C | Bias | 0.0055 | 0.0003 | 0.0003 | 0.0003 | -0.0015 | -0.0054 | -0.0054 | -0.0024 |
| | RMSE | 0.1138 | 0.0406 | 0.0405 | 0.0406 | 0.1013 | 0.0762 | 0.0757 | 0.0760 |
| β_M | Bias | -0.0056 | 0.0011 | 0.0011 | 0.0011 | -0.0327 | -0.0058 | -0.0056 | -0.0032 |
| | RMSE | 0.1033 | 0.0398 | 0.0398 | 0.0399 | 0.1137 | 0.0735 | 0.0728 | 0.0677 |

much larger RMSEs than the sieve estimators. In contrast, all the sieve estimators work equally well for the linear coefficients. The SML estimator is misspecified for the distributions of measurement errors but it produces accurate estimates for the complementarity parameters $(\alpha_{CC}, \alpha_{MM})$ in both cases. The SLS and SGLS estimators perform similarly to the SML estimator when the measurement errors are drawn from the Gamma distributions. In the case of correlated errors, the SLS estimator performs similarly to the SML estimator. The SGLS estimator outperforms the other estimators as it takes into account the correlations between measurement errors, resulting in more efficient estimation.

Lastly, we consider a DGP in which [Lindenlaub \(2017\)](#)'s model is more severely misspecified. Specifically, we draw x and y from finite Gaussian mixture distributions. Each mixture distribution has two Gaussian components with one-half weight for each. For both x and y , the Gaussian components, K_1 and K_2 , are specified as follows.

$$K_1 \sim N \left(\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \quad K_2 \sim N \left(\left(\begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \right).$$

We set ρ equal to 0.4 for x and 0.5 for y . The equilibrium assignments and wages are solved via linear programming as before. We also generate the measurement errors from a joint

Gaussian mixture distribution which has two components:

$$M_1 \sim N \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0.3 \\ 0.7 & 0.3 & 1 \end{pmatrix} \right), \quad M_2 \sim N \left(\begin{pmatrix} -3 \\ -3 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 & 0.7 \\ 0.7 & 1 & 0.3 \\ 0.7 & 0.3 & 1 \end{pmatrix} \right).$$

In this case, both (x, y) and measurement errors have bi-modal distributions that are far from a normal distribution.

TABLE V
FINITE SAMPLE PERFORMANCES OF THE ESTIMATORS (GAUSSIAN MIXTURE DGP)

| | | ML* | SML | SLS | SGLS |
|---------------|------|---------|---------|---------|---------|
| α_{CC} | Bias | 0.2896 | -0.0014 | -0.0016 | -0.0017 |
| | RMSE | 0.3653 | 0.0429 | 0.0425 | 0.0385 |
| α_{MM} | Bias | 0.2446 | -0.0017 | 0.0006 | -0.0027 |
| | RMSE | 0.3584 | 0.0454 | 0.0431 | 0.0337 |
| β_C | Bias | 0.7123 | -0.0007 | -0.0013 | 0.0008 |
| | RMSE | 0.7204 | 0.0428 | 0.0426 | 0.0364 |
| β_M | Bias | -0.1288 | -0.0001 | 0.0002 | -0.0011 |
| | RMSE | 0.1592 | 0.0421 | 0.0419 | 0.0305 |

As the margins of x and y are not standard normal, the ML* estimator is not directly applicable. Therefore, we use the inverse transform method to convert x and y to standard normal variables for the ML* estimator. As the joint distribution after transformation is non-normal, ML* is misspecified. Our sieve estimators can be applied without this transformation so we use untransformed data for the sieve-based estimators. The estimates of technology parameters are reported in Table V. Not surprisingly, the ML* estimator delivers parameter estimates that are very different from the true values. On top of misspecification, the transformation procedure introduces additional bias as the actual assignments are determined on the original data. By contrast, the sieve estimators still perform well in this case. Estimators relying on fewer assumptions deliver more accurate estimates. The SML esti-

mator produces the least precise estimates among the sieve estimators. The SGLS estimator incorporates the correlation structure among measurement errors so it possesses substantial efficiency gains compared to the SLS estimator.

TABLE VI

AVERAGE COMPUTATION TIME (IN SECONDS) AND RMSE (IN PARENTHESES)

| | | Time (RMSE) | | | |
|-----------------------|---|-------------------------------|---------------|-------------------------------|---------------|
| | | <i>2 characteristics (2D)</i> | | <i>3 characteristics (3D)</i> | |
| Sieve Order (k_n) | | $n = 5,000$ | $n = 10,000$ | $n = 5,000$ | $n = 10,000$ |
| ML* | – | 0.010 (0.034) | 0.016 (0.024) | 0.100 (0.047) | 0.157 (0.033) |
| SML | 2 | 0.034 (0.035) | 0.024 (0.024) | 0.169 (0.046) | 0.281 (0.032) |
| | 3 | 0.066 (0.035) | 0.106 (0.025) | 0.459 (0.046) | 0.675 (0.032) |
| SLS | 2 | 0.058 (0.035) | 0.101 (0.025) | 0.375 (0.047) | 0.636 (0.033) |
| | 3 | 0.133 (0.035) | 0.237 (0.025) | 1.568 (0.048) | 2.539 (0.033) |
| SGLS | 2 | 0.171 (0.035) | 0.368 (0.024) | 0.738 (0.046) | 1.254 (0.032) |
| | 3 | 0.320 (0.035) | 0.631 (0.025) | 3.221 (0.046) | 5.196 (0.032) |

Note: Times are measured in MATLAB on a MacBook Pro with Apple Silicon (M1, 8-core CPU).

Lastly, we investigate the computational scalability of our sieve estimators in settings where worker and job attributes have more than two dimensions. In additional simulations, we compare computation times of the parametric ML* estimator and sieve-based estimators in Gaussian designs with two (2D) and three characteristics (3D). The results are summarized in Table VI. As expected, the ML* estimator is the fastest in this Gaussian benchmark. One more characteristic increases the computing time for ML* roughly tenfold. The sieve-based methods are only moderately slower in absolute terms. For instance, with three characteristics, $n = 10,000$, and the Bernstein polynomial degree of 2 ($k_n = 2$) for each attribute, SML requires 0.281 seconds versus 0.157 seconds for ML*, while delivering virtually identical RMSE. In all cases reported, computation times remain well below one second for SML and below a few seconds for the more demanding SGLS specifications. The additional dimension tends to increase the computation time for the sieve estimators no more than tenfold, and often much less. Furthermore, relatively low sieve or-

ders are sufficient to match the performance of the correct parametric model. With $k_n = 2$, all sieve estimators attain RMSEs that are essentially indistinguishable from those of ML*. This suggests that, in practice, choosing modest sieve orders keeps the effective parameter dimension small, while retaining flexibility.

Regarding scalability beyond 3D, our sieve estimators can be computationally manageable for moderate numbers of characteristics (e.g., 3–5) and low sieve orders, as the simulations illustrate in the most demanding SGLS configurations. In applications with richer covariate sets, it is natural to combine our method with standard dimension-reduction tools (e.g., economically motivated indices or principal components) and to keep sieve orders small. The sieve estimators also offer a great “implementation” advantage. Computational burden in practice is not only about CPU time but also about implementation cost. In ML estimation, the log-likelihood depends on the parameters in a complex nonlinear way, so deriving analytic expressions of the gradient and Hessian becomes cumbersome and error-prone for higher dimensions. By contrast, for our sieve-based estimators, the objective functions are much simpler in the sieve coefficients: their gradients and Hessians have straightforward closed-form expressions that can be implemented with relatively little coding effort. This substantially reduces the “programmer time” required to obtain a stable and efficient implementation in higher-dimensional specifications.

Our simulation exercises provide evidence that the Gaussian model can be misleading when the model is misspecified. Even with moderate misspecification, the ML estimator does not produce reliable estimates. Furthermore, the actual impact of technological shifts on wage distribution may differ from the prediction of the Gaussian model. [Lindenlaub \(2017\)](#) shows that (i) wage distributions are positively skewed for any pairs of α_{CC} and α_{MM} , (ii) the variance of the wage distribution increases as cognitive or manual skill complementarity increases, and (iii) wage skewness is minimized when $\alpha_{CC} = \alpha_{MM}$. However, in our simulations using non-normal distributions (detailed in Appendix D), the resulting wage distribution’s skewness does not reach its minimum when $\alpha_{CC} = \alpha_{MM}$. On the contrary, our sieve estimators do not suffer from these problems. Therefore, our estimators can be more generally applicable regardless of underlying distributions of x , y , and

measurement errors with no need for transformation. They also show excellent finite sample performances and remain computationally feasible for large samples and for settings with more than two attributes.

7. EMPIRICAL APPLICATION TO U.S. WORKER-JOB MATCHING

We revisit the dataset constructed by [Lindenlaub \(2017\)](#) to learn how production technology in the U.S. has evolved. We estimate the production technology parameters in the model using the dataset and sieve estimators. The National Longitudinal Survey of Youth (NLSY) data and U.S. Department of Labor Occupational Characteristics Database (O*NET) are used to construct workers' cognitive and manual skills as well as the occupational skill requirements of firms. To assess the effect of technological changes on wage inequality, we compare estimation results based on two cohorts: the first cohort commencing in 1979 (referred to as NLSY79) and the second commencing in 1997 (referred to as NLSY97). Following Lindenlaub's main specification, we focus on employed workers aged 27 to 29 during the years 1990-91 and 2009-10, sourced from the NLSY79 and NLSY97 cohorts, respectively.¹⁶ The wage, w , is defined as the CPI-adjusted hourly rate.

Firms' skill demands, (y_C, y_M) , are constructed from the O*NET, which contains information on skill requirements for each occupation. [Sanders \(2014\)](#) classifies occupational skill requirements into cognitive and manual categories and constructs two task scores for over 400 occupations. These scores are employed to obtain skill demands at the occupation level.¹⁷ The skill bundle (x_C, x_M) is constructed using survey responses in the NLSY on education and training that are predetermined prior to employment. Given college education, apprenticeships, government training degrees, and occupational training history, workers are qualified for specific occupations from which their skill bundles are deter-

¹⁶The dataset excludes military samples and oversamples of special demographic/racial groups to give primary focus on the core sample of the NLSY.

¹⁷For instance, the occupation 'dancer' has a normalized cognitive score (y_C) of 0.34 and a normalized manual score (y_M) of 1. By contrast, 'physicist' has a skill demand bundle of $(y_C = 1, y_M = 0.11)$. We use the normalized task scores for illustration purposes following [Lindenlaub \(2017\)](#). The original supports of worker skills and firms' skill demands are provided in Table [VII](#).

TABLE VII

SUMMARY STATISTICS OF WORKER SKILLS (x) AND FIRMS' SKILL DEMANDS (y)

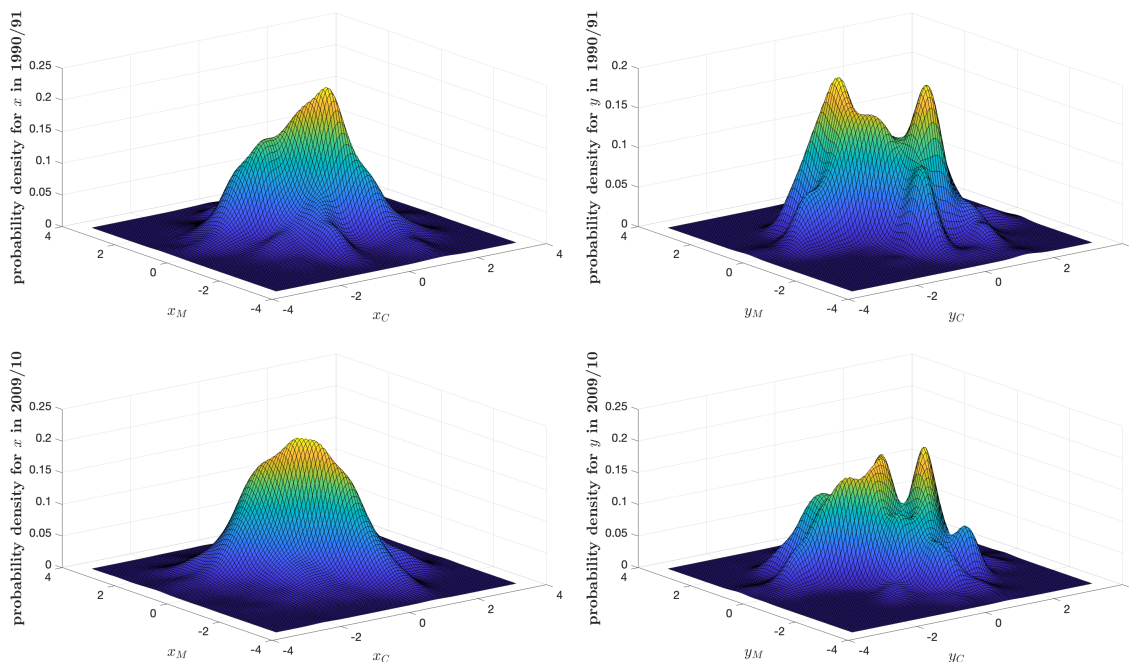
| | 1990/91 ($n = 2984$) | | | | 2009/10 ($n = 4495$) | | | |
|------|------------------------|---------|---------|---------|------------------------|---------|---------|---------|
| | x_C | x_M | y_C | y_M | x_C | x_M | y_C | y_M |
| Mean | 0.3596 | -0.2912 | 0.0135 | -0.1189 | 0.5667 | -0.6601 | 0.0468 | -0.2509 |
| SD | 0.7423 | 0.9923 | 0.8490 | 1.0240 | 0.7556 | 0.8358 | 0.9280 | 0.9656 |
| Min | -2.0595 | -1.7004 | -2.0622 | -1.6949 | -2.3019 | -1.8116 | -2.5200 | -1.6597 |
| Max | 2.1649 | 2.1855 | 2.0925 | 2.1895 | 1.9160 | 2.1838 | 3.0504 | 2.1351 |

mined.¹⁸ We argue that measurement errors in y_i arise primarily from coarsening and aggregation of O*NET descriptors and survey noise. As the workers' skills are independent of their current (realized) occupation and skill requirements of each occupation are not tailored to individual workers, the exogeneity of measurement errors we impose for estimation is an empirically reasonable approximation in our setting.

Table VII presents summary statistics of workers' skills and firms' skill demands. In 1990/91, workers had higher cognitive skills on average than manual skills, and firms also required more cognitive skills than manual skills. Two decades later, workers had increased cognitive skills and decreased manual skills on average compared to 1990/91, with firms also showing a similar trend. The skill correlation (ρ_x) shifted from -0.40 to -0.52 , indicating increased worker specialization. In contrast, the productivity correlation (ρ_y) remained stable at -0.49 . Initially, jobs were more specialized than workers, but skill supply caught up, resulting in slightly greater worker specialization in 2009/10.

Lindenlaub (2017) transforms each element of x and y into a standard Gaussian variable, and their dependence is modeled using the Gaussian copula. However, while their marginal distributions are standard normal, the transformed variables are not guaranteed to have a joint normal distribution. Let \tilde{x} and \tilde{y} be Gaussian-transformed x and y respectively. As

¹⁸For example, a worker who studied economics at a university is qualified for the 'economist' job. Then the worker possesses a normalized skill bundle of ($x_C = 0.615, x_M = 0.034$), that is required to be an economist.

FIGURE 3.—Joint densities of \tilde{x} and \tilde{y} (transformed data)

shown in Figure 3, the joint distributions of \tilde{x} and \tilde{y} are not normal. In particular, the joint density of \tilde{y} is multi-modal in both periods.

We employ Mardia's test (Mardia, 1970) to formally test the joint normality of \tilde{x} and \tilde{y} . We first create an $n \times n$ matrix for \tilde{x} :

$$C = (c_{ij}) = x^* S^{-1} (x^*)',$$

where the i -th row of x^* is $x_i^* = \tilde{x}_i - \sum_{i=1}^n \tilde{x}_i/n$, and define multivariate measures of skewness and kurtosis as follows:

$$b_1 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n c_{ij}^3, \quad b_2 = \frac{1}{n} \sum_{i=1}^n c_{ii}^2.$$

Under bivariate normality, the limiting distribution of $\frac{nb_1}{6}$ is a chi-square distribution with $d(d+1)(d+2)/6$ degrees of freedom and the limiting distribution of $\frac{\sqrt{n}(b_2 - d(d+2))}{\sqrt{8d(d+2)}}$ is the standard normal distribution where d is the dimensionality of \tilde{x} . We conduct the same procedure for \tilde{y} . Table VIII shows the test statistics. The test strongly rejects the bivariate

normality of \tilde{x} and \tilde{y} in both periods. The normality assumption imposed in Lindenlaub’s model is not satisfied even after transforming x and y . Therefore, our semiparametric approach is more appropriate in this case.

TABLE VIII

MARDIA’S MULTIVARIATE NORMALITY TEST STATISTICS (P-VALUES IN PARENTHESES)

| | 1990/91 ($n = 2984$) | | 2009/10 ($n = 4495$) | |
|----------|------------------------|----------------|------------------------|----------------|
| | \tilde{x} | \tilde{y} | \tilde{x} | \tilde{y} |
| Skewness | 4.58 (0.333) | 100.09 (0.000) | 16.34 (0.003) | 145.14 (0.000) |
| Kurtosis | 4.44 (0.000) | 0.29 (0.774) | 14.42 (0.000) | 1.98 (0.048) |

We estimate the production technology in each period separately. As in the simulations, we shut down between-task complementarities (A is diagonal) following Lindenlaub (2017). As described in the simulation section, the corrected ML procedure (ML^*) modifies Lindenlaub (2017)’s MLE to use the latent productivity correlation in place of the observed one. Here we further generalize ML^* to accommodate the case where the true skill requirements \tilde{y}^* have variances not equal to 1 (since the inverse transform method converts the measurement error contaminated \tilde{y} , not \tilde{y}^*) and allow the Gaussian measurement errors to be correlated with each other. We then estimate the technology parameters using our sieve estimators (SML, SLS, and SGLS) with Bernstein polynomial basis functions of degree 3, which perform the best in terms of information criteria and model fit. We compare our semiparametric estimates of the technology parameters to Lindenlaub’s original estimates and the ML^* estimates.

The estimation results are provided in Table IX. All the models clearly show a substantial shift in the relative importance between manual and cognitive tasks over the two decades. Our results are qualitatively consistent with Lindenlaub’s but quantitatively very different. In 1990/91, the estimated complementarity in manual tasks was much larger than in cognitive tasks. The Gaussian model (ML^*) indicates that the complementarity in manual tasks is roughly 1.7 times as large as that of cognitive tasks. When we dispense with the

TABLE IX

ESTIMATES OF PRODUCTION TECHNOLOGY PARAMETERS ON TRANSFORMED DATA

| | 1990/91 | | | | | 2009/10 | | | | |
|---------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------------|------------------|------------------|------------------|------------------|
| | ML | ML* | SML | SLS | SGLS | ML | ML* | SML | SLS | SGLS |
| α_{CC} | 0.203 (0.342) | 0.765 (0.574) | 0.454 (0.009) | 0.000 (0.000) | 0.000 (0.000) | 0.739 (0.198) | 1.119 (0.408) | 2.048 (0.036) | 2.293 (0.256) | 2.290 (0.265) |
| α_{MM} | 0.479 (0.175) | 1.270 (0.148) | 1.422 (0.031) | 1.084 (0.113) | 0.856 (0.098) | 0.055 (0.154) | 0.486 (0.633) | 0.237 (0.006) | 1.033 (0.215) | 0.291 (0.059) |
| β_C | 1.686 (0.143) | 1.711 (0.589) | 1.692 (0.068) | 1.719 (0.434) | 1.585 (0.416) | 2.203 (0.152) | 2.208 (0.540) | 2.115 (0.068) | 2.063 (0.589) | 2.198 (0.534) |
| β_M | -0.421 (0.141) | -0.392 (0.406) | -0.374 (0.068) | -0.382 (0.329) | -0.388 (0.309) | 0.210 (0.152) | 0.243 (0.731) | 0.198 (0.076) | 0.180 (0.545) | 0.327 (0.532) |

Standard errors in parentheses. ML indicates the original estimates in [Lindenlaub \(2017\)](#).

normality assumption on skill demand and supply, the ratio becomes larger than 3. When we further generalize the model by removing the Gaussian assumption on measurement errors, the complementarity in cognitive tasks shrinks close to 0 (but significantly larger than 0), whereas that of manual tasks is still close to 1. The estimates of linear productivity coefficients β_C and β_M are similar across all the specifications as shown in the simulations.

This pattern reverses over the following 20 years. All the complementarity estimates for 2009/10 indicate a substantial increase in the complementarity between cognitive worker and job attributes, whereas the complementarity in manual tasks sharply decreased. In the Gaussian model (ML*), the ratio of estimated complementarities in manual tasks to cognitive tasks ($\frac{\alpha_{MM}}{\alpha_{CC}}$) is around 0.43, which is similar to the estimated value by SLS. However, SML and SGLS deliver much smaller values close to 0.1. The linear coefficients are similar across specifications. These patterns imply substantial changes in the relative complementarities across tasks because of technological advances. Lindenlaub describes this phenomenon as “task-biased technological change in favor of cognitive tasks”. The cognitive dimension became much more important in labor market sorting. Our semiparametric models suggest that the “task-biased technological change” favoring cognitive tasks in the

last two decades may have been much larger than previously found. The increases in β_C and β_M indicate that both cognitive and manual skill productivity have risen. However, the estimated manual skill productivity in both periods is insignificant in most specifications. Therefore, following Lindenlaub’s description, we can conclude that the U.S. economy has also experienced “skill-biased technological change” in favor of cognitive skills.

We now investigate the effect of estimated technological changes on wage inequality. Wage inequality in the U.S. labor market until the late 2000s is well characterized by *wage*

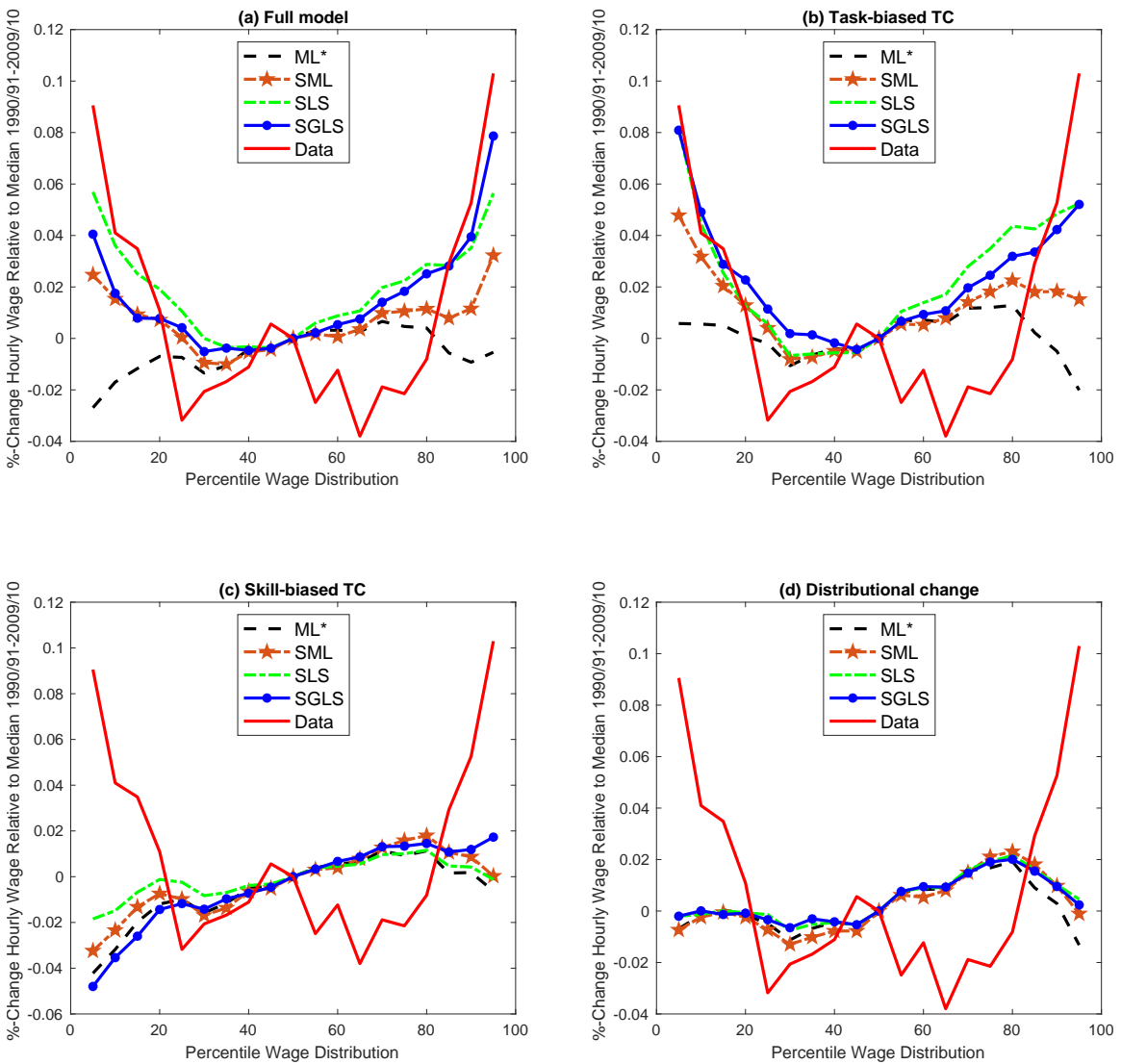


FIGURE 4.—Actual and model predicted wage polarization (transformed data)

polarization, which is defined as stronger wage growth in the bottom and upper tails relative to the median. The red solid line in Figure 4 plots how wages changed relative to the median wage between 1990/91 and 2009/10 by wage percentile, implying that the U.S. labor market experienced a spike in the upper-tail wage inequality, while the lower-tail inequality declined. This phenomenon is surprisingly well predicted by our models, as shown in Figure 4 (a). All the semiparametric models predict substantial wage polarization once the estimated parameter values are fed in. By contrast, the Gaussian model fails to account for wage polarization in both tails. We can also observe that the model fit improves as the model becomes more flexible.

To further explore why the matching model requires greater flexibility to account for wage polarization, we compare the curvature of estimated wage functions across different models in Figures 5–6. In 1990/91, all models produce almost linear wage functions, indicating a relatively uniform relationship between wages and cognitive skills. However, in 2009/10, our semiparametric models predict a significantly steeper curvature, particularly at high cognitive skill levels, whereas the Gaussian model still generates a more linear wage function. The Gaussian model constrains the wage function to a quadratic form of standard normal variables, limiting its shape to a low-degree polynomial. In contrast, our models do not impose such constraints, allowing for greater flexibility in curvature that better fits the data. Notably, our most flexible model predicts a sharply increasing slope in the wage function for 2009/10, which is relatively flat at low cognitive skill levels and very steep at high skill levels, generating substantial wage polarization.

To understand the driving forces behind wage polarization, we isolate the effects of technological and distributional changes in Figure 4. We only keep task-biased technological change (shutting down changes in linear productivity coefficients) in panel (b), skill-biased technological change (shutting down changes in complementarity parameters) in (c), and distributional change (shutting down both changes in linear productivity and complementarity parameters) in (d). We find that task-biased technological change explains wage polarization remarkably well. In particular, all three semiparametric models exhibit an excellent fit in the lower tail in panel (b), while the Gaussian model shows only a slight decline

in lower tail inequality. In contrast, skill-biased technological change exacerbates wage inequality in the lower tail as shown in panel (c). The distributional change has a negligible impact on wage inequality. In summary, despite their parsimony, our matching models effectively account for the evolution of wage inequality over the past 20 years in the U.S., with task-biased technological change being the primary driver of the observed pattern.

Lastly, we estimate our semiparametric models on the original data. Unlike the Gaussian model, our models can be applied directly to the data without any transformation. As illustrated in Figure 1, the matching occurs based on the original distributions. Hence,

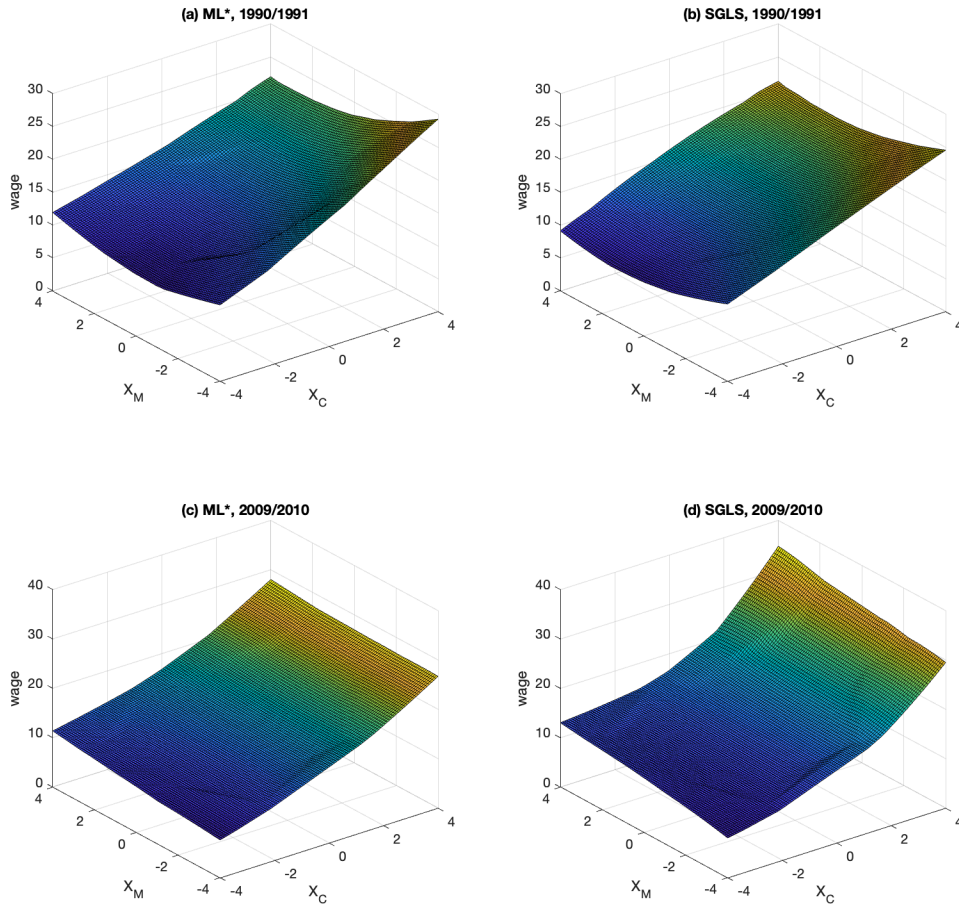


FIGURE 5.—Estimated wage functions (transformed data)

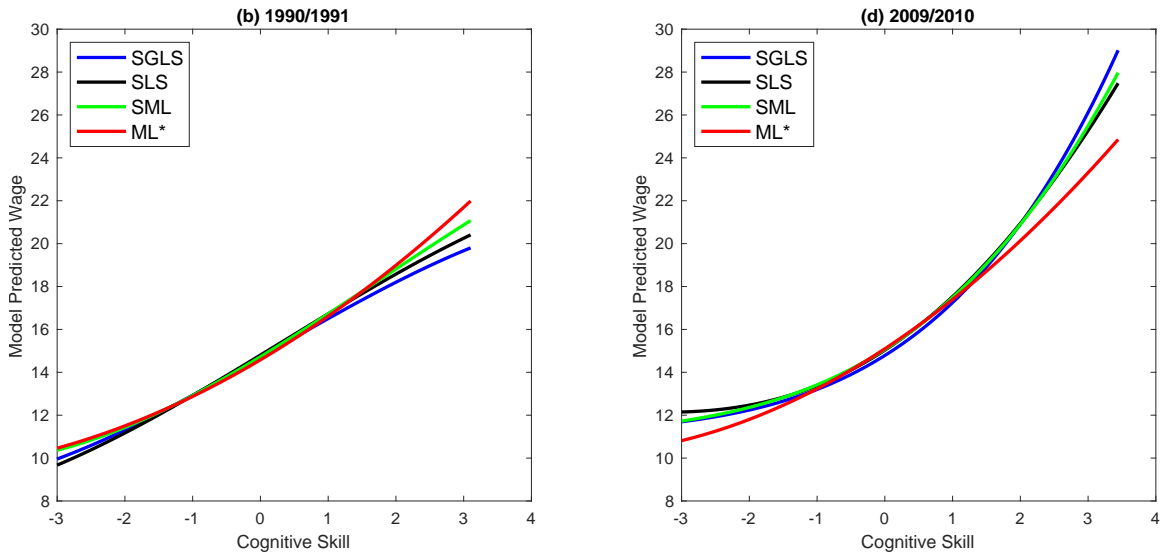


FIGURE 6.—Predicted wage with respect to cognitive skill (transformed data)

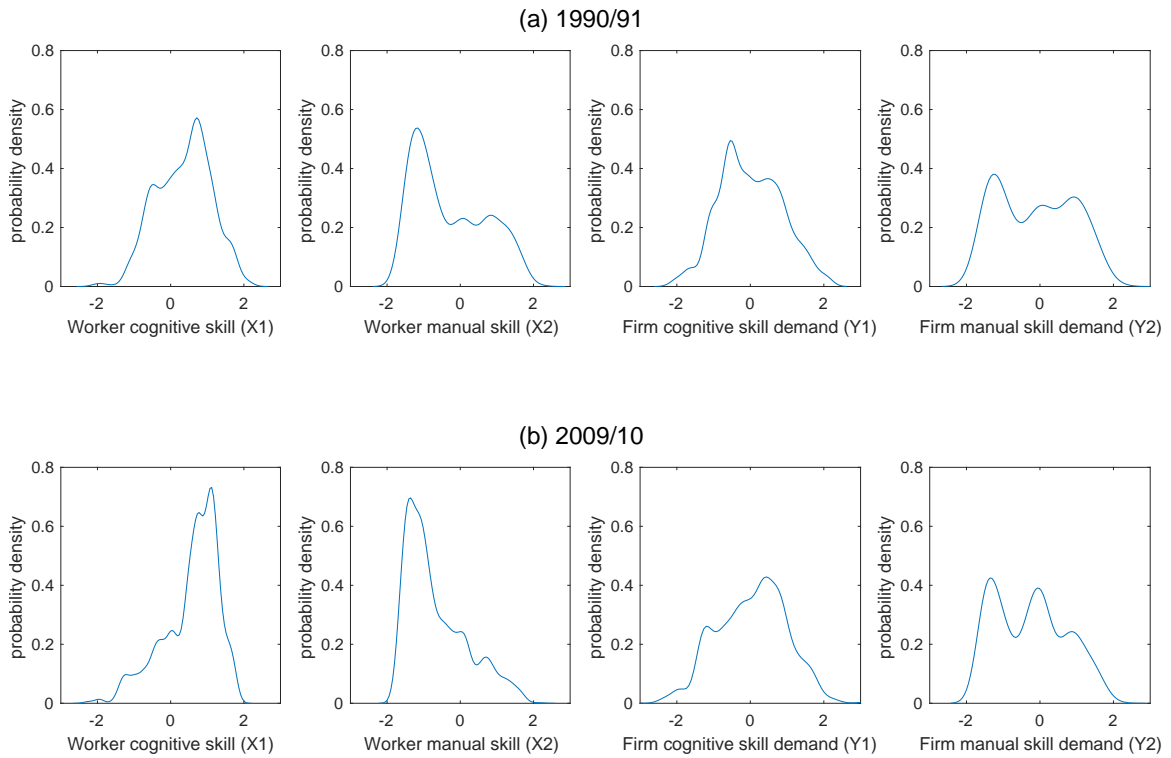


FIGURE 7.—Marginal distributions of skill supply and demand

transforming marginal distributions to standard normal can lead to a different solution from $T(x)$. The marginal distributions of workers' skill supply x and firms' skill demand y in Figure 7 are skewed and multi-modal, quite different from standard normal. Therefore, it is crucial to investigate the robustness of the Gaussian model using the original data. We report the estimated parameters in Table X, using Bernstein polynomial basis functions of degree 4 to accommodate the less well-behaved original data. The estimated parameters reveal similar patterns across our models. Notably, the complementarity in cognitive tasks increased significantly from near 0 in 1990/91 to around 6 in 2009/10, while the complementarity in manual tasks decreased from near 2 to almost 0. Additionally, linear skill productivity improved, with a larger increase in cognitive skill productivity. These findings confirm that the U.S. economy experienced substantial task-biased and skill-biased technological changes favoring cognitive skills over the two decades.

TABLE X

ESTIMATES OF PRODUCTION TECHNOLOGY PARAMETERS ON ORIGINAL DATA

| | 1990/91 | | | 2009/10 | | |
|---------------|-------------------|-------------------|-------------------|------------------|------------------|------------------|
| | SML | SLS | SGLS | SML | SLS | SGLS |
| α_{CC} | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 6.032 (0.110) | 5.784 (0.184) | 6.471 (0.196) |
| α_{MM} | 1.947 (0.060) | 2.292 (0.081) | 2.234 (0.079) | 0.000 (0.000) | 1.020 (0.024) | 0.003 (0.000) |
| β_C | 2.238 (0.083) | 2.246 (0.225) | 2.108 (0.220) | 3.395 (0.101) | 3.299 (0.396) | 3.707 (0.292) |
| β_M | -0.341 (0.073) | -0.441 (0.211) | -0.317 (0.213) | 0.254 (0.097) | 0.384 (0.333) | 0.440 (0.239) |

Standard errors in parentheses

The estimated models on the original data effectively capture the patterns of wage polarization, particularly in the upper tail, as shown in Figure 8. While the models slightly under-predict wage polarization in the lower tail, they confirm that task-biased technological change was the primary driver of wage polarization. In the absence of skill-biased

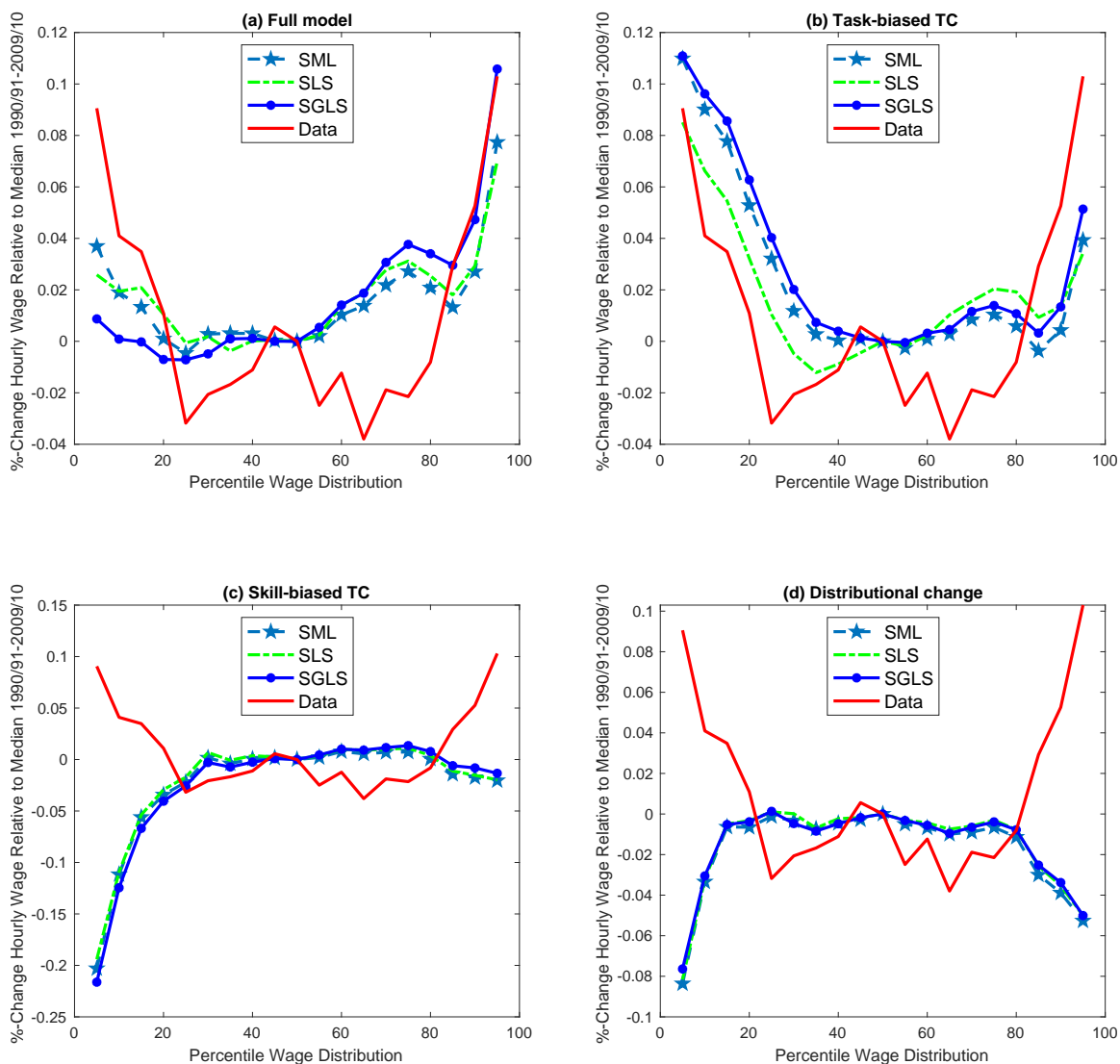


FIGURE 8.—Actual and model predicted wage polarization (original data)

technological change, the model shows a significant relative wage increase in the lower tail. However, skill-biased technological change had a negative impact on wage inequality, exacerbating lower tail inequality. Distributional change improved upper tail inequality but worsened lower tail inequality.

In summary, the estimated semiparametric models on the original data exhibit similar patterns to those on the Gaussian transformed data, with task- and skill-biased technological changes being more pronounced. Our models demonstrate a remarkable fit for wage

polarization, highlighting the substantial changes in production technology in the U.S. over the past two decades. This exercise showcases the versatility and effectiveness of our semi-parametric models and sieve-based estimators, which can accommodate any underlying joint distributions of skill supply and demand without requiring data transformation or distributional assumptions. Moreover, our approach builds upon standard sieve-based estimators, which have been shown to achieve semiparametric efficiency and are easy to implement in practice.

8. CONCLUSION

Theoretical matching models often face empirical challenges due to discrepancies between model assumptions and real-world data. [Lindenlaub \(2017\)](#) presents a tractable theoretical model suitable for comparative statics and qualitative analysis of multidimensional matching. However, its empirical application is limited by restrictive distributional assumptions on observed characteristics and measurement errors. We generalize this model by relaxing these key distributional restrictions, enabling our models to accommodate datasets with matched pairs, regardless of the underlying characteristic and error distributions. Our simulation results demonstrate the accuracy of our semi-nonparametric estimators across various data-generating processes. Moreover, our flexible models generate significant wage polarization, aligning with U.S. data patterns, whereas the parametric Gaussian model falls short. Our estimated models indicate that task-biased technological progress, favoring cognitive abilities over manual skills, is the primary driver of wage polarization.

Our empirical results are consistent with task-biased technical change raising the relative productivity (and returns) to cognitive skill bundles and contributing to wage polarization through equilibrium sorting. Looking ahead, the diffusion of generative artificial intelligence may further reshape the relative productivity of cognitive tasks and—where AI complements higher-order cognition and coordination—could amplify these forces, although the direction and magnitude remain an empirical question. A first policy implication is to broaden access to cognitive and hybrid skill formation through higher education, STEM pathways, and mid-career retraining targeted to task bundles whose returns are rising. Second, policies that foster broad diffusion of productivity-enhancing technologies and expand

access to high-quality job opportunities—across firms, industries, and regions—may mitigate inequality by reducing barriers to adoption and complementarity-driven sorting. Finally, redistributive instruments such as wage insurance and earnings-smoothing policies can help insure workers against technology-driven wage risk during transitions.

Our study opens up promising research avenues. First, incorporating additional dimensions such as interpersonal and digital skills into our model holds great potential. By employing advanced techniques such as artificial neural networks ([Chen et al., 2023](#)) to approximate equilibrium functions in high-dimensional spaces, one could enhance the model’s accuracy in elucidating intricate matching patterns and their impact on wage inequality. Second, addressing measurement errors in assessing worker skills is crucial. Overcoming this challenge requires innovative econometric approaches that can accurately estimate models amidst multidimensional measurement errors. Finally, our framework’s application extends beyond the worker-job matching problem, offering insights into matching problems in diverse contexts such as the marriage market.

While our approach offers valuable insights, it is essential to acknowledge its limitations. As in [Lindenlaub \(2017\)](#), our frictionless matching model, grounded in optimal transport, assumes that all workers are employed and all jobs are filled. The lack of treatment of unemployed workers and job vacancies (unmatched individuals on both sides of the market) is a limitation of the balanced optimal transport formulation employed here and in [Lindenlaub \(2017\)](#). Incorporating unemployment and vacancies would require extending the framework to partial or unbalanced optimal transport, or to models with search frictions. Introducing randomness in assignment through factors such as search frictions and unobserved heterogeneity poses a fruitful challenge, especially when extending existing one-dimensional theories, e.g., those in [Eeckhout and Kircher \(2011\)](#), to multidimensional settings.

REFERENCES

- AI, CHUNRONG AND XIAOHONG CHEN (2003): “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions,” *Econometrica*, 71 (6), 1795–1843. [[4](#), [7](#), [18](#), [19](#), [21](#)]

- (2007): “Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables,” *Journal of Econometrics*, 141 (1), 5–43. [7]
- BECKER, GARY S. (1973): “A Theory of Marriage: Part I,” *Journal of Political Economy*, 81 (4), 813–846. [3, 9]
- BECKER, GARY S (1991): *A treatise on the family: Enlarged edition*, Harvard university press. [3]
- BLUNDELL, RICHARD, XIAOHONG CHEN, AND DENNIS KRISTENSEN (2007): “Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves,” *Econometrica*, 75 (6), 1613–1669. [7, 14]
- BOERMA, JOB, ANDREA OTTOLINI, AND ALEH TSYVINSKI (2025): “Multidimensional sorting: comparative Statics,” Working Paper. [8]
- BOJILOV, RAICHO AND ALFRED GALICHON (2016): “Matching in closed-form: equilibrium, identification, and comparative statics,” *Economic Theory*, 61, 587–609. [3, 4, 6, 8, 9, 11]
- BROWNING, MARTIN, PIERRE-ANDRÉ CHIAPPORI, AND YORAM WEISS (2014): *Economics of the Family*, Cambridge University Press. [8]
- CAFFARELLI, LUIS A. (1992a): “Boundary regularity of maps with convex potentials,” *Communications on Pure and Applied Mathematics*, 45 (9), 1141–1151. [7, 20]
- CAFFARELLI, LUIS A (1992b): “The regularity of mappings with a convex potential,” *Journal of the American Mathematical Society*, 5 (1), 99–104. [7, 20]
- (1996): “Boundary regularity of maps with convex potentials–II,” *Annals of mathematics*, 144 (3), 453–496. [7, 20]
- CHEN, JIAFENG, XIAOHONG CHEN, AND ELIE TAMER (2023): “Efficient estimation of average derivatives in NPIV models: Simulation comparisons of neural network estimators,” *Journal of Econometrics*, 235 (2), 1848–1875. [43]
- CHEN, XIAOHONG (2007): “Chapter 76 Large Sample Sieve Estimation of Semi-Nonparametric Models,” Elsevier, vol. 6 of *Handbook of Econometrics*, 5549–5632. [4, 7, 16, 17, 18, 21, 1, 2, 3]
- CHEN, XIAOHONG AND DEMIAN POUZO (2009): “Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals,” *Journal of Econometrics*, 152 (1), 46–60. [7]
- CHEN, XIAOHONG AND XIAOTONG SHEN (1998): “Sieve Extremum Estimates for Weakly Dependent Data,” *Econometrica*, 66 (2), 289–314. [18, 2]
- CHETVERIKOV, DENIS AND DANIEL WILHELM (2017): “Nonparametric instrumental variable estimation under monotonicity,” *Econometrica*, 85 (4), 1303–1320. [14]
- CHIAPPORI, PIERRE-ANDRÉ (2017): *Matching with Transfers: The Economics of Love and Marriage*, Princeton University Press. [8]
- CHIAPPORI, PIERRE-ANDRÉ, ROBERT MCCANN, AND BRENDAN PASS (2020): “Multidimensional matching: theory and empirics,” Working Paper. [14]
- CHIAPPORI, PIERRE-ANDRÉ, ROBERT J MCCANN, AND LARS P NESHEIM (2010): “Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness,” *Economic Theory*, 42, 317–354. [4, 8]

- CHIAPPORI, PIERRE-ANDRÉ, ROBERT J MCCANN, AND BRENDAN PASS (2017): “Multi-to One-Dimensional Optimal Transport,” *Communications on Pure and Applied Mathematics*, 70 (12), 2405–2444. [14]
- CHIAPPORI, PIERRE-ANDRÉ AND SONIA OREFFICE (2008): “Birth control and female empowerment: An equilibrium analysis,” *Journal of Political Economy*, 116 (1), 113–140. [3]
- CHIAPPORI, PIERRE-ANDRÉ, SONIA OREFFICE, AND CLIMENT QUINTANA-DOMEQUE (2012): “Fatter attraction: anthropometric and socioeconomic matching on the marriage market,” *Journal of Political Economy*, 120 (4), 659–695. [3]
- CHIAPPORI, PIERRE-ANDRÉ AND PHILIP J. RENY (2016): “Matching to share risk,” *Theoretical Economics*, 11 (1), 227–251. [3]
- CHIONG, KHAI XIANG, ALFRED GALICHON, AND MATT SHUM (2016): “Duality in dynamic discrete-choice models,” *Quantitative Economics*, 7 (1), 83–115. [4]
- CHOO, EUGENE AND ALOYSIUS SIOW (2006): “Who marries whom and why,” *Journal of political Economy*, 114 (1), 175–201. [3, 5]
- COMPIANI, GIOVANNI (2022): “Market counterfactuals and the specification of multiproduct demand: A non-parametric approach,” *Quantitative Economics*, 13 (2), 545–591. [18]
- CORDERO-ERAUSQUIN, DARIO AND ALESSIO FIGALLI (2019): “Regularity of monotone transport maps between unbounded domains,” *Discrete and Continuous Dynamical Systems*, 39 (12), 7101–7112. [7, 20]
- DE PHILIPPIS, GUIDO AND ALESSIO FIGALLI (2014): “The Monge-Ampère equation and its link to optimal transportation,” *Bull. Amer. Math. Soc. (N.S.)*, 51 (4), 527–580. [4]
- DUPUY, ARNAUD AND ALFRED GALICHON (2014): “Personality Traits and the Marriage Market,” *Journal of Political Economy*, 122 (6), 1271–1319. [6]
- EECKHOUT, JAN AND PHILIPP KIRCHER (2011): “Identifying Sorting-In Theory,” *The Review of Economic Studies*, 78 (3), 872–906. [43]
- EKELAND, IVAR (2010): “Notes on optimal transportation,” *Economic Theory*, 41, 437–459. [4]
- FLOATER, MICHAEL S (1994): “A weak condition for the convexity of tensor-product Bézier and B-spline surfaces,” *Advances in Computational Mathematics*, 2, 67–80. [11]
- GALICHON, ALFRED (2017): “A survey of some recent applications of optimal transport methods to econometrics,” *The Econometrics Journal*, 20 (2), C1–C11. [4, 8]
- GALICHON, ALFRED, SCOTT DUKE KOMINERS, AND SIMON WEBER (2019): “Costly Concessions: An Empirical Framework for Matching with Imperfectly Transferable Utility,” *Journal of Political Economy*, 127 (6), 2875–2925. [5]
- GALICHON, ALFRED AND BERNARD SALANIÉ (2022): “Cupid’s invisible hand: Social surplus and identification in matching models,” *The Review of Economic Studies*, 89 (5), 2600–2629. [4, 5]
- GROSSBARD-SCHECHTMAN, SHOSHANA (1993): “A Theory of Marriage, Labor and Divorce,” . [3]
- GUALDANI, CRISTINA AND SHRUTI SINHA (2023): “Partial identification in matching models for the marriage market,” *Journal of Political Economy*, 131 (5), 1109–1171. [6]

- HITSCH, GÜNTER J, ALI HORTAÇSU, AND DAN ARIELY (2010): “Matching and sorting in online dating,” *American Economic Review*, 100 (1), 130–163. [3]
- LEGROS, PATRICK AND ANDREW F NEWMAN (2007): “Beauty is a beast, frog is a prince: Assortative matching with nontransferabilities,” *Econometrica*, 75 (4), 1073–1102. [3]
- LINDENLAUB, ILSE (2017): “Sorting Multidimensional Types: Theory and Application,” *The Review of Economic Studies*, 84 (2), 718–789. [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 17, 22, 23, 27, 30, 31, 32, 34, 35, 42, 43]
- LINDENLAUB, ILSE AND FABIEN POSTEL-VINAY (2025): “The Worker-Job Surplus,” Working Paper. [8]
- LISE, JEREMY AND FABIEN POSTEL-VINAY (2020): “Multidimensional Skills, Sorting, and Human Capital Accumulation,” *American Economic Review*, 110 (8), 2328–76. [6, 12]
- LORENTZ, GEORGE G (1986): *Bernstein polynomials*, American Mathematical Soc. [18]
- MARDIA, K. V. (1970): “Measures of multivariate skewness and kurtosis with applications,” *Biometrika*, 57 (3), 519–530. [33]
- MCCANN, ROBERT J AND BRENDAN PASS (2020): “Optimal transportation between unequal dimensions,” *Archive for Rational Mechanics and Analysis*, 238 (3), 1475–1520. [14]
- NENNA, LUCA AND BRENDAN PASS (2023): “A note on Cournot-Nash equilibria and Optimal Transport between unequal dimensions,” in *Optimal Transport Statistics for Economics and Related Topics*, Springer, 117–130. [14]
- NEWKEY, WHITNEY K. AND JAMES L. POWELL (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71 (5), 1565–1578. [7]
- OREFFICE, SONIA AND CLIMENT QUINTANA-DOMEQUE (2010): “Anthropometry and socioeconomics among couples: Evidence in the United States,” *Economics & Human Biology*, 8 (3), 373–384. [3]
- PAPAGEORGIOU, THEODORE (2014): “Learning your comparative advantages,” *Review of Economic Studies*, 81 (3), 1263–1295. [3]
- PENCANEL, JOHN (1998): “Assortative mating by schooling and the work behavior of wives and husbands,” *The American Economic Review*, 88 (2), 326–329. [3]
- QIAN, ZHENCHAO (1998): “Changes in assortative mating: The impact of age and education, 1970–1890,” *Demography*, 35 (3), 279–292. [3]
- SANDERS, CARL (2014): “Skill accumulation, skill uncertainty, and occupational choice,” Working Paper. [12, 31]
- SCHENNACH, SUSANNE M. (2016): “Recent Advances in the Measurement Error Literature,” *Annual Review of Economics*, 8, 341–377. [13]
- SHEN, XIAOTONG (1997): “On methods of sieves and penalization,” *The Annals of Statistics*, 25 (6), 2555 – 2591. [7]
- SILVENTOINEN, KARRI, JAAKKO KAPRIO, EERO LAHELMA, RICHARD J VIKEN, AND RICHARD J ROSE (2003): “Assortative mating by body height and BMI: Finnish twins and their spouses,” *American Journal of Human Biology*, 15 (5), 620–627. [3]

- TINBERGEN, JAN (1956): “On the theory of income distribution,” *Weltwirtschaftliches archiv*, 77, 155–175. [3, 8]
- VILLANI, CÉDRIC (2003): *Topics in optimal transportation*, vol. 58, American Mathematical Soc. [4, 8, 11, 1]
- VILLANI, CÉDRIC (2008): *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Springer. [4, 8, 11]
- WANG, J. AND S.K. GHOSH (2012): “Shape restricted nonparametric regression with Bernstein polynomials,” *Computational Statistics & Data Analysis*, 56 (9), 2729–2741. [10]
- WEISS, YORAM AND ROBERT J WILLIS (1997): “Match quality, new information, and marital dissolution,” *Journal of labor Economics*, 15 (1, Part 2), S293–S329. [3]
- WILLIS, ROBERT J AND SHERWIN ROSEN (1979): “Education and self-selection,” *Journal of political Economy*, 87 (5, Part 2), S7–S36. [3]

APPENDIX A: TECHNICAL PROOFS

PROOF OF PROPOSITION 1: In equilibrium, the firm maximizes its profit, so the first-order condition of the firm's maximization problem is satisfied:

$$\nabla w^*(x) - b = \nabla_x x' \tilde{y} \Big|_{\tilde{y}=T(x)}.$$

By Theorem 2.12 in Villani (2003), $\nabla_x x' \tilde{y} \Big|_{\tilde{y}=T(x)} = \nabla w^{o*}(x)$ and therefore,

$$\nabla w^*(x) = \nabla w^{o*}(x) + b.$$

This implies that $w^*(x) = w^{o*}(x) + x'b + c$ where c is the constant of integration. *Q.E.D.*

PROOF OF THEOREM 1: We first note from $\mathbb{E}[\varepsilon_{wi}|x_i] = 0$ that the convex function $w_0^*(x) = \mathbb{E}[w_i|x_i = x] = w_0(x) + x'b_0$ is identified. Since $\nabla w_0^*(x)$ is also identified and $\mathbb{E}[y_i|x_i = x] = A_0^{-1}(\nabla w_0^*(x) - b_0)$, the invertibility of A_0 implies that A_0 and b_0 are identified: We consider $y_1, \dots, y_{d+1} \in \mathcal{Y}$ satisfying Assumption 3. Then, there are corresponding $d + 1$ distinct points $x_1, \dots, x_{d+1} \in \mathcal{X}$ such that $A_0 y_1 = \nabla w_0(x_1), \dots, A_0 y_{d+1} = \nabla w_0(x_{d+1})$. It follows from the invertibility of A_0 that

$$\nabla w_0(x_1) - \nabla w_0(x_2) = A_0(y_1 - y_2), \dots, \nabla w_0(x_d) - \nabla w_0(x_{d+1}) = A_0(y_d - y_{d+1})$$

are independent. Since $\nabla w_0^*(x) = \nabla w_0(x) + b_0$, the differences $\nabla w_0^*(x_1) - \nabla w_0^*(x_2), \dots, \nabla w_0^*(x_d) - \nabla w_0^*(x_{d+1})$ are also linearly independent. Then, A_0 is identified from

$$\begin{pmatrix} \mathbb{E}[y_i|x_i = x_1] - \mathbb{E}[y_i|x_i = x_2] \\ \vdots \\ \mathbb{E}[y_i|x_i = x_d] - \mathbb{E}[y_i|x_i = x_{d+1}] \end{pmatrix}' = A_0^{-1} \begin{pmatrix} \nabla w_0^*(x_1) - \nabla w_0^*(x_2) \\ \vdots \\ \nabla w_0^*(x_d) - \nabla w_0^*(x_{d+1}) \end{pmatrix}'$$

In turn, b_0 and $w_0(x) = w_0^*(x) - x'b_0$ are also identified. *Q.E.D.*

PROOF OF PROPOSITION 2: We obtain the result by applying Theorem 3.2 in Chen (2007). We note that

$$\begin{aligned} & \rho'(z; \lambda) \Sigma(x)^{-1} \rho(z; \lambda) - \rho'(z; \lambda_0) \Sigma(x)^{-1} \rho(z; \lambda_0) \\ &= [\rho(z; \lambda) - \rho(z; \lambda_0)]' \Sigma(x)^{-1} [\rho(z; \lambda) - \rho(z; \lambda_0)] - 2\varepsilon' \Sigma(x)^{-1} [\rho(z; \lambda) - \rho(z; \lambda_0)]. \end{aligned}$$

2

and

$$\rho(z; \lambda) - \rho(z; \lambda_0) = \begin{pmatrix} w(x) - w_0(x) + x'(b - b_0) \\ \nabla_C w_0(x) (\kappa_C - \kappa_{C0}) + \kappa_C \nabla_C \{w(x) - w_0(x)\} \\ \nabla_M w_0(x) (\kappa_M - \kappa_{M0}) + \kappa_M \nabla_M \{w(x) - w_0(x)\} \end{pmatrix}.$$

Then, it follows from Assumption 5 that

$$\|\lambda - \lambda_0\|^2 \asymp \mathbb{E} \left[\rho'(z_i; \lambda) \Sigma(x_i)^{-1} \rho(z_i; \lambda) - \rho'(z_i; \lambda_0) \Sigma(x_i)^{-1} \rho(z_i; \lambda_0) \right],$$

i.e., there exists a finite $C_1 > 0$ such that

$$C_1^{-1} \|\lambda - \lambda_0\|^2 \leq \mathbb{E} \left[\rho'(z_i; \lambda) \Sigma(x_i)^{-1} \rho(z_i; \lambda) - \rho'(z_i; \lambda_0) \Sigma(x_i)^{-1} \rho(z_i; \lambda_0) \right] \leq C_1 \|\lambda - \lambda_0\|^2.$$

Condition 3.6 in Chen (2007) is assumed with Assumption 4. Now we check Conditions 3.7 and 3.8 in Chen (2007). Again, Assumption 5 implies that there exists a $C_2 > 0$ such that

$$\begin{aligned} & \mathbb{E} \left[\left(\rho'(z_i; \lambda_0) \Sigma(x_i)^{-1} \rho(z_i; \lambda_0) - \rho'(z_i; \lambda) \Sigma(x_i)^{-1} \rho(z_i; \lambda) \right)^2 \right] \\ & \leq C_2 \mathbb{E} \left[|\rho(z_i; \lambda_0) - \rho(z_i; \lambda)|_e^4 \right]. \end{aligned}$$

By Lemma 2 in Chen and Shen (1998), we have $\|w - w_0\|_\infty \leq c \|w - w_0\|_2^{m/(m+1)}$ and $\|\nabla_C \{w - w_0\}\|_\infty, \|\nabla_M \{w - w_0\}\|_\infty \leq c \|w - w_0\|_2^{(m-1)/m}$ for some finite $c > 0$, where $\|w\|_2^2 = \mathbb{E} [w^2(x_i)]$. Since $\|\cdot\|_2 \asymp \|\cdot\|$,

$$\mathbb{E} \left[\left(\rho'(z_i; \lambda_0) \Sigma(x_i)^{-1} \rho(z_i; \lambda_0) - \rho'(z_i; \lambda) \Sigma(x_i)^{-1} \rho(z_i; \lambda) \right)^2 \right] \leq C_3 \|\lambda - \lambda_0\|^{2[1+(m-1)/m]}.$$

So Condition 3.7 is satisfied for all $\varepsilon \leq 1$. On the other hand,

$$\begin{aligned} & \left| \rho'(z_i; \lambda_0) \Sigma(x_i)^{-1} \rho(z_i; \lambda_0) - \rho'(z_i; \lambda) \Sigma(x_i)^{-1} \rho(z_i; \lambda) \right| \\ & \leq \|\lambda - \lambda_0\|_\infty |\Sigma(x_i)|^{-1} (2|\varepsilon|_e + \|\lambda\|_\infty + \|\lambda_0\|_\infty), \end{aligned}$$

almost surely. Using Lemma 2 in Chen and Shen (1998) again, Condition 3.8 is satisfied.

To apply Theorem 3.2 in Chen (2007), it remains to compute the deterministic approximation error rate $\inf_{\lambda \in \Theta \times \mathcal{W}_n} \|\lambda - \lambda_0\|$ and the metric entropy with bracketing. By the same proof as that for Proposition 3.3 in Chen (2007), they are also computed, and then the result follows. *Q.E.D.*

PROOF OF THEOREM 2: We obtain the limiting distribution of $\hat{\theta}_n$ by verifying that Assumptions 4.1 and 4.2 of Proposition 4.4 in Chen (2007) are satisfied. It is easy to see that Assumptions 4.2.(ii) and (iv) are satisfied with the expression for ρ . Assumptions 4.1.(i) and 4.2.(i) are our assumption 9 and 5. Assumption 4.1.(ii) is implied by our assumption 5 and 8. Assumption 4.1.(iii) is implied by our Proposition 2 and Assumption 8: there is $\pi_n u^* \in \mathcal{W}_n$ such that $\|\pi_n u^* - u^*\| \times \|\hat{\lambda}_n - \lambda_0\| = o_p(n^{-1/2})$. Let $u_\theta^* = (u_{\theta_1}^*, u_{\theta_2}^*, u_{\theta_3}^*, u_{\theta_4}^*)' = \left(\mathbb{E} \left[D_{v^*}(x_i)' \Sigma(x_i)^{-1} D_{v^*}(x_i) \right] \right)^{-1} \eta$, $u_w^* = -v^* u_\theta^*$ and $u^* = (u_\theta^*, u_w^*)$, where $\eta \in \mathbb{R}^4$ is an arbitrary unit vector. It remains to show that Assumption 4.2.(iii) (Conditions 4.2' and 4.3') in Chen (2007) are satisfied with

$$\begin{aligned} \frac{\partial \ell(z; \bar{\lambda})}{\partial \lambda} [\pi_n u^*] &= \begin{pmatrix} \bar{w}(x) + x' \bar{b} - w \\ \bar{\kappa}_C \nabla_C \bar{w}(x) - y_C \\ \bar{\kappa}_M \nabla_M \bar{w}(x) - y_M \end{pmatrix}' \Sigma(x)^{-1} \begin{pmatrix} \pi_n u_w^* + u_{\theta_3}^* x_C + u_{\theta_4}^* x_M \\ u_{\theta_1}^* \nabla_C \bar{w}(x) + \bar{\kappa}_C \nabla_C (\pi_n u_w^*(x)) \\ u_{\theta_2}^* \nabla_M \bar{w}(x) + \bar{\kappa}_M \nabla_M (\pi_n u_w^*(x)) \end{pmatrix} \\ &= \begin{pmatrix} \bar{w}(x) - w_0(x) + x' (\bar{b} - b_0) - \varepsilon_w \\ \bar{\kappa}_C \nabla_C \bar{w}(x) - \kappa_{C0} \nabla_C w_0(x) - \varepsilon_C \\ \bar{\kappa}_M \nabla_M \bar{w}(x) - \kappa_{M0} \nabla_M w_0(x) - \varepsilon_M \end{pmatrix}' \Sigma(x)^{-1} \begin{pmatrix} \pi_n u_w^* + u_{\theta_3}^* x_C + u_{\theta_4}^* x_M \\ u_{\theta_1}^* \nabla_C \bar{w}(x) + \bar{\kappa}_C \nabla_C (\pi_n u_w^*(x)) \\ u_{\theta_2}^* \nabla_M \bar{w}(x) + \bar{\kappa}_M \nabla_M (\pi_n u_w^*(x)) \end{pmatrix} \end{aligned}$$

for all $\bar{\lambda} \in \Theta \times \mathcal{W}_n$ with $\|\bar{\lambda} - \lambda_0\| = o(1)$. Since

$$\begin{pmatrix} \hat{w}(x) - w_0(x) + x' (\hat{b} - b_0) \\ \hat{\kappa}_C \nabla_C \hat{w}(x) - \kappa_{C0} \nabla_C w_0(x) \\ \hat{\kappa}_M \nabla_M \hat{w}(x) - \kappa_{M0} \nabla_M w_0(x) \end{pmatrix} = \frac{d\rho(z; \lambda_0)}{d\lambda} [\hat{\lambda}_n - \lambda_0] + \begin{pmatrix} 0 \\ (\hat{\kappa}_C - \kappa_{C0}) (\nabla_C \hat{w}(x) - \nabla_C w_0(x)) \\ (\hat{\kappa}_M - \kappa_{M0}) (\nabla_M \hat{w}(x) - \nabla_M w_0(x)) \end{pmatrix},$$

and

$$\begin{aligned} &\begin{pmatrix} \pi_n u_w^* + u_{\theta_3}^* x_C + u_{\theta_4}^* x_M \\ u_{\theta_1}^* \nabla_C \hat{w}(x) + \hat{\kappa}_C \nabla_C (\pi_n u_w^*(x)) \\ u_{\theta_2}^* \nabla_M \hat{w}(x) + \hat{\kappa}_M \nabla_M (\pi_n u_w^*(x)) \end{pmatrix} \\ &= \frac{d\rho(Z; \lambda_0)}{d\lambda} [\pi_n u^*] + \begin{pmatrix} 0 \\ u_{\theta_1}^* \nabla_C (\hat{w}(x) - w_0(x)) + (\hat{\kappa}_C - \kappa_{C0}) \nabla_C (\pi_n u_w^*(x)) \\ u_{\theta_2}^* \nabla_M (\hat{w}(x) - w_0(x)) + (\hat{\kappa}_M - \kappa_{M0}) \nabla_M (\pi_n u_w^*(x)) \end{pmatrix}, \end{aligned}$$

Condition 4.3' is satisfied given the definition of $\|\cdot\|$ and

$$\langle \hat{\lambda}_n - \lambda_0, \pi_n u^* \rangle = \mathbb{E} \left[\left(\frac{d\rho(z_i; \lambda_0)}{d\lambda} [\hat{\lambda}_n - \lambda_0] \right)' \Sigma(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [\pi_n u^*] \right].$$

Condition 4.2' can be verified by applying Lemma 4.2 in [Chen \(2007\)](#). The condition on the metric entropy with bracketing for Lemma 4.2 is satisfied with Assumption 7. *Q.E.D.*

PROOF OF THEOREM 3: We follow the proofs of Theorem 4.1 and 6.2 in [Ai and Chen \(2003\)](#). Let $\mathcal{N}_{on} \equiv \left\{ \lambda \in \Theta \times \mathcal{W}_n : \|\lambda - \lambda_0\|_s = o(1), \|\lambda - \lambda_0\| = o\left(n^{-1/4}\right) \right\}$. By Proposition 2, the sieve estimator $\tilde{\lambda}_n$ in Step 1 satisfies $\|\tilde{\lambda}_n - \lambda_0\|_s = o_p(1)$ and $\|\tilde{\lambda}_n - \lambda_0\| = o_p\left(n^{-1/4}\right)$. Hence $\tilde{\lambda}_n \in \mathcal{N}_{on}$. Using the proof similar to those of Proposition 2, we can also show that $\hat{\lambda}_n \in \mathcal{N}_{on}$. Let $u_{0\theta} = (u_{0\theta 1}, u_{0\theta 2}, u_{0\theta 3}, u_{0\theta 4})' = \left(\mathbb{E} \left[D_{v_0}(x_i)' \Sigma(x_i)^{-1} D_{v_0}(x_i) \right] \right)^{-1} \eta$, $u_{0w} = -v_0 u_{0\theta}$ and $u_0 = (u_{0\theta}, u_{0w})$, where $\eta \in \mathbb{R}^4$ is an arbitrary unit vector. Then, we have

$$(\theta - \theta_0)' \eta = \langle \lambda - \lambda_0, u_0 \rangle = \mathbb{E} \left[\left(\frac{d\rho(z_i; \lambda_0)}{d\lambda} [\lambda - \lambda_0] \right)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] \right]$$

for all $\lambda \in \Lambda$. Let $\varepsilon_n = o\left(n^{-1/2}\right) > 0$. Denote $u_{n0} := (u_{n0\theta}, u_{n0w}) = \pi_n u_0$ to simplify notation. We take a continuous path $\lambda(t) = \hat{\lambda}_n \pm t\varepsilon_n u_{n0}$. Then $\{\lambda(t) : t \in [0, 1]\}$ in \mathcal{N}_{on} . Let

$$\hat{Q}_n(\lambda(t)) = -\frac{1}{n} \sum_{i=1}^n \rho(z_i; \lambda(t))' \hat{\Sigma}_0(x_i)^{-1} \rho(z_i; \lambda(t)).$$

By definition of $\hat{\lambda}_n$, and a Taylor expansion around $t = 0$ up to second order, we obtain

$$0 \leq \hat{Q}_n(\hat{\lambda}_n) - \hat{Q}_n(\hat{\lambda}_n \pm \varepsilon_n u_{n0}) = - \left. \frac{d\hat{Q}_n(\lambda(t))}{dt} \right|_{t=0} - \frac{1}{2} \left. \frac{d^2\hat{Q}_n(\lambda(t))}{dt^2} \right|_{t=s},$$

for some $s \in [0, 1]$, where

$$\begin{aligned} - \left. \frac{d\hat{Q}_n(\lambda(t))}{dt} \right|_{t=0} &= \frac{\pm 2\varepsilon_n}{n} \sum_{i=1}^n \rho(z_i; \hat{\lambda}_n)' \hat{\Sigma}_0(x_i)^{-1} \frac{d\rho(z_i; \hat{\lambda}_n)}{d\lambda} [u_{n0}] \\ - \left. \frac{d^2\hat{Q}_n(\lambda(t))}{dt^2} \right|_{t=s} &= \underbrace{\frac{2\varepsilon_n^2}{n} \sum_{i=1}^n \rho(z_i; \lambda(s))' \hat{\Sigma}_0(x_i)^{-1} \frac{d^2\rho(z_i; \lambda(s))}{d\lambda d\lambda} [u_{n0}, u_{n0}]}_{:=A_1} \\ &\quad + \underbrace{\frac{2\varepsilon_n^2}{n} \sum_{i=1}^n \left(\frac{d\rho(z_i; \lambda(s))}{d\lambda} [u_{n0}] \right)' \hat{\Sigma}_0(x_i)^{-1} \frac{d\rho(z_i; \lambda(s))}{d\lambda} [u_{n0}]}_{:=A_2}, \end{aligned}$$

with

$$\frac{d\rho(z; \lambda(s))}{d\lambda} [\varepsilon_n u_{n0}] \equiv \frac{d\rho(z; \lambda(t))}{dt} \Big|_{t=s}, \quad \frac{d^2\rho(z; \lambda(s))}{d\lambda d\lambda} [\varepsilon_n u_{n0}, \varepsilon_n u_{n0}] \equiv \frac{d^2\rho(z; \lambda(t))}{dt^2} \Big|_{t=s}.$$

We note from Assumption 8.(ii) that each element of

$$\frac{d^2\rho(z; \lambda(s))}{d\lambda d\lambda} [u_{n0}, u_{n0}] = \frac{d^2\rho(z; \lambda + tu_{n0})}{dt^2} \Big|_{t=s} = \begin{pmatrix} 0 \\ -2u_{0\theta 1} \nabla_C u_{n0w}(x) \\ -2u_{0\theta 2} \nabla_M u_{n0w}(x) \end{pmatrix}$$

is uniformly bounded over $x \in \mathcal{X}$. For A_1 , we write

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \rho(z_i; \lambda(s))' \hat{\Sigma}_0(x_i)^{-1} \frac{d^2\rho(z_i; \lambda(s))}{d\lambda d\lambda} [u_{n0}, u_{n0}] \\ &= \frac{1}{n} \sum_{i=1}^n (\rho(z_i; \lambda(s)) - \rho(z_i; \lambda_0))' \hat{\Sigma}_0(x_i)^{-1} \frac{d^2\rho(z_i; \lambda(s))}{d\lambda d\lambda} [u_{n0}, u_{n0}] \\ & \quad + \frac{1}{n} \sum_{i=1}^n \rho(z_i; \lambda_0)' (\hat{\Sigma}_0(x_i)^{-1} - \Sigma_0(x_i)^{-1}) \frac{d^2\rho(z_i; \lambda(s))}{d\lambda d\lambda} [u_{n0}, u_{n0}] \\ & \quad + \frac{1}{n} \sum_{i=1}^n \rho(z_i; \lambda_0)' \Sigma_0(x_i)^{-1} \frac{d^2\rho(z_i; \lambda(s))}{d\lambda d\lambda} [u_{n0}, u_{n0}]. \end{aligned}$$

Since

$$\begin{aligned} \rho(z; \lambda(s)) - \rho(z; \lambda_0) &= \frac{d\rho(z; \lambda_0)}{d\lambda} [\hat{\lambda}_n - \lambda_0] + \begin{pmatrix} 0 \\ (\hat{\kappa}_C - \kappa_{C0}) (\nabla_C \hat{w}(x) - \nabla_C w_0(x)) \\ (\hat{\kappa}_M - \kappa_{M0}) (\nabla_M \hat{w}(x) - \nabla_M w_0(x)) \end{pmatrix} \\ & \quad \mp s\varepsilon_n \begin{pmatrix} u_{n0w}(x) + x_C u_{0\theta 1} + x_M u_{0\theta 2} \\ \hat{\kappa}_C \nabla_C u_{n0w}(x) + u_{0\theta 3} \nabla_C \hat{w}(x) + u_{0\theta 3} \nabla_C u_{n0w}(x) \\ \hat{\kappa}_M \nabla_M u_{n0w}(x) + u_{0\theta 4} \nabla_M \hat{w}(x) + u_{0\theta 4} \nabla_M u_{n0w}(x) \end{pmatrix}, \end{aligned}$$

the first term of the right-hand side is $o_p\left(n^{-1/4}\right)$ uniformly over $\lambda(s) \in \mathcal{N}_{on}$, which implies that A_1 is $o_p\left(\varepsilon_n^2\right)$. For A_2 , we write

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \left(\frac{d\rho(z_i; \lambda(s))}{d\lambda} [u_{n0}] \right)' \hat{\Sigma}_0(x_i)^{-1} \frac{d\rho(z_i; \lambda(s))}{d\lambda} [u_{n0}] \\
&= \frac{1}{n} \sum_{i=1}^n \left(\frac{d\rho(z_i; \lambda(s))}{d\lambda} [u_{n0}] - \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_{n0}] \right)' \hat{\Sigma}_0(x_i)^{-1} \frac{d\rho(z_i; \lambda(s))}{d\lambda} [u_{n0}] \\
&+ \frac{1}{n} \sum_{i=1}^n \left(\frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_{n0}] \right)' \hat{\Sigma}_0(x_i)^{-1} \left(\frac{d\rho(z_i; \lambda(s))}{d\lambda} [u_{n0}] - \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_{n0}] \right) \\
&+ \frac{1}{n} \sum_{i=1}^n \left(\frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_{n0}] \right)' \left(\hat{\Sigma}_0(x_i)^{-1} - \Sigma_0(x_i)^{-1} \right) \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_{n0}] \\
&+ \frac{1}{n} \sum_{i=1}^n \left(\frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_{n0}] \right)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_{n0}]
\end{aligned}$$

Since

$$\begin{aligned}
\frac{d\rho(z; \lambda(s))}{d\lambda} [u_{n0}] - \frac{d\rho(z; \lambda_0)}{d\lambda} [u_{n0}] &= \begin{pmatrix} 0 \\ u_{0\theta 1} \nabla_C (\hat{w}(x) - w_0(x)) + (\hat{\kappa}_C - \kappa_{C0}) \nabla_C u_{n0w}(x) \\ u_{0\theta 2} \nabla_M (\hat{w}(x) - w_0(x)) + (\hat{\kappa}_M - \kappa_{M0}) \nabla_M u_{n0w}(x) \end{pmatrix} \\
&\pm s\varepsilon_n \begin{pmatrix} 0 \\ u_{0\theta 1} \nabla_C u_{n0w}(x) + u_{0\theta 1} \nabla_C (\pi_n u_w^*(x)) \\ u_{0\theta 2} \nabla_M u_{n0w}(x) + u_{0\theta 2} \nabla_M (\pi_n u_w^*(x)) \end{pmatrix},
\end{aligned}$$

the first two terms on the right-hand side are $o_p\left(n^{-1/4}\right)$ and the third term is $o_p(1)$ uniformly over $\lambda(s) \in \mathcal{N}_{on}$, which implies that A_2 is $O_p\left(\varepsilon_n^2\right)$. Moreover, since $\varepsilon_n = o\left(n^{-1/2}\right) > 0$, we obtain uniformly over $\lambda(s) \in \mathcal{N}_{on}$:

$$\frac{1}{n} \sum_{i=1}^n \rho\left(z_i; \hat{\lambda}_n\right)' \hat{\Sigma}_0(x_i)^{-1} \frac{d\rho\left(z_i; \hat{\lambda}_n\right)}{d\lambda} [u_{n0}] = o_p\left(n^{-1/2}\right).$$

Write

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \rho(z_i; \hat{\lambda}_n)' \hat{\Sigma}_0(x_i)^{-1} \frac{d\rho(z_i; \hat{\lambda}_n)}{d\lambda} [u_{n0}] \\
&= \frac{1}{n} \sum_{i=1}^n \rho(z_i; \hat{\lambda}_n)' \hat{\Sigma}_0(x_i)^{-1} \left(\frac{d\rho(z_i; \hat{\lambda}_n)}{d\lambda} [u_{n0}] - \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] \right) \\
& \quad + \frac{1}{n} \sum_{i=1}^n \rho(z_i; \hat{\lambda}_n)' \left[\hat{\Sigma}_0(x_i)^{-1} - \Sigma_0(x_i)^{-1} \right] \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] \\
& \quad + \frac{1}{n} \sum_{i=1}^n \rho(z_i; \hat{\lambda}_n)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0].
\end{aligned}$$

Using proofs for second derivative terms together with Assumption 10, the first two terms on the right-hand side are $o_p(n^{-1/2})$. Since $\left\{ \rho(z_i; \lambda)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] : \lambda \in \mathcal{N}_{on} \right\}$ is a Donsker class,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \left(\rho(z_i; \hat{\lambda}_n) - \rho(z_i; \lambda_0) \right)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] \\
&= \mathbb{E} \left[\left(\rho(z_i; \hat{\lambda}_n) - \rho(z_i; \lambda_0) \right)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] \right] + o_p(n^{-1/2}).
\end{aligned}$$

With $\rho(z_i; \hat{\lambda}_n) - \rho(z_i; \lambda_0) - \frac{d\rho(z_i; \lambda_0)}{d\lambda} [\hat{\lambda}_n - \lambda_0] = o_p(n^{-1/2})$ uniformly over $x_i \in \mathcal{X}$,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \rho(z_i; \hat{\lambda}_n)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] \\
&= \langle \hat{\lambda}_n - \lambda_0, u_0 \rangle + \frac{1}{n} \sum_{i=1}^n \rho(z_i; \lambda_0)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] + o_p(n^{-1/2}).
\end{aligned}$$

Then,

$$\sqrt{n}(\theta - \theta_0)' \eta = \sqrt{n} \langle \hat{\lambda}_n - \lambda_0, u_0 \rangle = -\frac{1}{\sqrt{n}} \sum_{i=1}^n \rho(z_i; \lambda_0)' \Sigma_0(x_i)^{-1} \frac{d\rho(z_i; \lambda_0)}{d\lambda} [u_0] + o_p(1),$$

and Theorem 3 follows from applying a standard CLT for i.i.d. data.

Q.E.D.

APPENDIX B: EXTENSION TO UNEQUAL DIMENSIONS

We extend the identification results in Section 3 to the case where worker and job characteristics have different dimensions, $d_x \neq d_y$. Without Assumption 2 (the invertibility of

A), we can express the equilibrium as

$$w^*(x) = w^{o*}(x) + x'b + c, \quad A \left(y_1^*(x), \dots, y_{d_y}^*(x) \right)' = \nabla w^{o*}(x).$$

Then, the econometric model with measurement errors can be defined by:

$$\begin{aligned} w_i &= w^*(x_i) + \varepsilon_{wi} = w^{o*}(x_i) + x_i'b + c + \varepsilon_{wi}, \\ Ay_i &= Ay_i^*(x_i) + \varepsilon_{yi} = \nabla w^{o*}(x_i) + \varepsilon_{yi}. \end{aligned}$$

The matrix $A_0 \in \mathbb{R}^{d_x \times d_y}$, the vector $b_0 \in \mathbb{R}^{d_x}$, and the convex function $w_0(\cdot) = w^{o*}(\cdot) + c$ are identified by the following theorem.

THEOREM 4: *Let Assumption 1 hold. Assume that there exist $d_y + 1$ values of x such that the $d_y \times d_y$ matrix formed by differences in the conditional mean of job characteristics,*

$$M_Y := \left(\mathbb{E}[y_i | x_i = x_1] - \mathbb{E}[y_i | x_i = x_{d_y+1}], \dots, \mathbb{E}[y_i | x_i = x_{d_y}] - \mathbb{E}[y_i | x_i = x_{d_y+1}] \right)$$

is invertible. Then, A_0, b_0 , and $w_0(\cdot)$ are identified.

Intuitively, the additional identification condition (the invertibility of M_Y) only requires that the conditional mean of job characteristics varies in d_y linearly independent directions across worker types. In the proof of Theorem 1 for the equal-dimension case, Assumption 3 is used to establish such linear independence of the conditional means. In the unequal-dimension case, we impose this rank condition directly by assuming that M_Y is invertible for $d_y + 1$ worker types. Hence, the invertibility of M_Y provides a sufficient condition for identification, which can be viewed as a conditional-mean analogue of Assumption 3.

Note that the equilibrium assignment function is also identified when we have $d_x \geq d_y$ and A has rank d_y , as it is recovered by $y^*(x) = (A'A)^{-1}A'\nabla w^{o*}(x)$. In this case, the twist condition is satisfied and the matching between x and y is pure, which implies the uniqueness of stable matching. When $d_y > d_x$, the twist condition is not met, so the purity and uniqueness of matching between x and y are not guaranteed.

PROOF OF THEOREM 4: Similar to the proof of Theorem 1, it follows from the exogeneity condition $\mathbb{E}[\varepsilon_{wi} | x_i] = 0$ that the convex function $w_0^*(\cdot) = w_0(x) + x'b_0$ is identified

from the conditional expectation of wages. Since w_0^* is convex and continuously differentiable, its gradient $\nabla w_0^*(x)$ is also identified. Moreover,

$$\nabla w_0^*(x) = \nabla w_0(x) + b_0 \quad \Rightarrow \quad \nabla w_0(x) = A_0 \mathbb{E}[y_i | x_i = x] = \nabla w_0^*(x) - b_0.$$

Now consider the points x_1, \dots, x_{d_y+1} satisfying the new assumption. For each $k = 1, \dots, d_y$,

$$A_0 \left(\mathbb{E}[y_i | x_i = x_k] - \mathbb{E}[y_i | x_i = x_{d_y+1}] \right) = \nabla w_0^*(x_k) - \nabla w_0^*(x_{d_y+1}).$$

Define the $d_x \times d_y$ matrix

$$M_X := \left(\nabla w_0^*(x_1) - \nabla w_0^*(x_{d_y+1}) \cdots \nabla w_0^*(x_{d_y}) - \nabla w_0^*(x_{d_y+1}) \right),$$

and the $d_y \times d_y$ matrix

$$M_Y := \left(\mathbb{E}[y_i | x_i = x_1] - \mathbb{E}[y_i | x_i = x_{d_y+1}] \cdots \mathbb{E}[y_i | x_i = x_{d_y}] - \mathbb{E}[y_i | x_i = x_{d_y+1}] \right).$$

Stacking the above d_y equalities gives

$$A_0 M_Y = M_X.$$

By the invertibility of M_Y , we can solve uniquely for A_0 :

$$A_0 = M_X M_Y^{-1}.$$

This shows that $A_0 \in \mathbb{R}^{d_x \times d_y}$ is identified, without any restriction on the relative sizes of d_x and d_y .

Having identified A_0 , we can recover b_0 . For any x in the support of x_i ,

$$b_0 = \nabla w_0^*(x) - A_0 \mathbb{E}[y_i | x_i = x],$$

and the right-hand side is identified from the data. Thus b_0 is identified. This completes the proof. *Q.E.D.*

APPENDIX C: BERNSTEIN POLYNOMIALS WITH CONVEX CONSTRAINTS

The equilibrium wage function, $w(x)$, obtained from optimal transport theory is unique and convex (up to a constant). However, finite sample estimators of $w(x)$, $\hat{w}_n(x; \gamma)$, might be nonconvex at the values close to the boundary of \mathcal{X} . To obtain a more stable estimator for $w(x)$, we impose a convexity restriction in the sieve-based estimation procedures without loss of generality. Among many possible linear approximating spaces, we particularly consider the following Bernstein polynomial sieve space:

$$\mathcal{W}_n = \left\{ w_n : \mathcal{X} \rightarrow \mathbb{R} : w_n(x; \gamma) = \sum_{j_1, \dots, j_d=0}^{k_n} \gamma_{j_1 \dots j_d} \left[\prod_{\ell=1}^d p_{j_\ell}(x_\ell) \right] : \right. \\ \left. p_{j_\ell}(x_\ell) = \binom{k_n}{j_\ell} \left(\frac{x_\ell - \underline{x}_\ell}{\bar{x}_\ell - \underline{x}_\ell} \right)^{j_\ell} \left(\frac{\bar{x}_\ell - x_\ell}{\bar{x}_\ell - \underline{x}_\ell} \right)^{k_n - j_\ell} \right\},$$

for $j_\ell = 0, 1, 2, \dots, k_n$ where p_{j_ℓ} is the Bernstein basis polynomial.

Let \mathcal{W}^{cvx} be the set of midpoint convex functions:

$$\mathcal{W}^{cvx} = \left\{ w \in C(\mathcal{X}) : 2w\left(\frac{x_1 + x_2}{2}\right) \leq w(x_1) + w(x_2), \forall x_1, x_2 \in \mathcal{X} \right\},$$

where $C(\mathcal{X})$ is the class of all continuous functions on \mathcal{X} . We do not assume that the true function $w(x)$ has derivatives of any order. In fact, \mathcal{W}^{cvx} is the class of all continuous convex functions because a continuous function that is midpoint convex is convex.

First, we consider the one-dimensional ($d = 1$) constrained Bernstein polynomial sieve space, $\mathcal{W}_n^{cvx} = \{w_n(x; \gamma) \in \mathcal{W}_n : A\gamma \geq 0\}$, where

$$A\gamma \equiv \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ & & & \ddots & & \\ 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix}_{(k_n-1) \times (k_n+1)} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{k_n} \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Since the second derivatives of $w_n(x; \gamma)$ can be written as

$$w_n^{(2)}(x; \gamma) = \frac{k_n(k_n-1)}{(\bar{x} - \underline{x})^2} \sum_{j=0}^{k_n-2} (\gamma_{j+2} - 2\gamma_{j+1} + \gamma_j) \binom{k_n-2}{j} \left(\frac{x - \underline{x}}{\bar{x} - \underline{x}} \right)^j \left(\frac{\bar{x} - x}{\bar{x} - \underline{x}} \right)^{k_n-2-j},$$

the above restriction ensures $w_n^{(2)}(\cdot) \geq 0$ for all n . Wang and Ghosh (2012) show that $\{\mathcal{W}_n^{cvx}\}$ is nested and dense in \mathcal{W}^{cvx} with respect to sup-norm.

For the two-dimensional sieve in eq.(4.1), we consider the following linear constraints:

$$\begin{aligned} \gamma_{j_C+2, j_M} - 2\gamma_{j_C+1, j_M} + \gamma_{j_C, j_M} &\geq 0, \quad \forall j_C = 0, \dots, k_{C_n} - 2, j_M = 0, \dots, k_{M_n}, \\ \gamma_{j_C, j_M+2} - 2\gamma_{j_C, j_M+1} + \gamma_{j_C, j_M} &\geq 0, \quad \forall j_C = 0, \dots, k_{C_n}, j_M = 0, \dots, k_{M_n} - 2. \end{aligned} \quad (\text{C.1})$$

Then, the Bernstein polynomial sieve space with linear constraints (C.1) is nested and dense in

$$\begin{aligned} \widetilde{\mathcal{W}}^{cvx} = \left\{ w \in C(\mathcal{X}) : 2w\left(\frac{x_{C1} + x_{C2}}{2}, x_M\right) \leq w(x_{C1}, x_M) + w(x_{C2}, x_M) \ \& \right. \\ \left. 2w\left(x_C, \frac{x_{M1} + x_{M2}}{2}\right) \leq w(x_C, x_{M1}) + w(x_C, x_{M2}), \right. \\ \left. \forall (x_{C1}, x_M), (x_{C2}, x_M), (x_C, x_{M1}), (x_C, x_{M2}) \in \mathcal{X} \right\}, \end{aligned}$$

which is larger than

$$\begin{aligned} \mathcal{W}^{cvx} = \left\{ w \in C(\mathcal{X}) : 2w\left(\frac{x_{C1} + x_{C2}}{2}, \frac{x_{M1} + x_{M2}}{2}\right) \leq w(x_{C1}, x_{M1}) + w(x_{C2}, x_{M2}), \right. \\ \left. \forall (x_{C1}, x_{M1}), (x_{C2}, x_{M2}) \in \mathcal{X} \right\}. \end{aligned}$$

Note that [Floater \(1994\)](#) provides sufficient conditions for the two-dimensional Bernstein polynomial $w_n(x; \gamma)$ to be convex, which include linear inequalities (C.1) as well as additional nonlinear constraints. We use (C.1) for our estimation because (i) they are easy to impose in the optimization procedure and to extend to higher-dimensional functions, and (ii) $w(x) \in \mathcal{W}^{cvx} \subset \widetilde{\mathcal{W}}^{cvx}$.

APPENDIX D: EFFECTS OF PRODUCTION TECHNOLOGY ON WAGE DISTRIBUTION

In this section, we check whether the [Lindenlaub \(2017\)](#) model's predictions on the effects of technological changes on wage distribution hold for the transformed data and other distributions. We consider three DGPs with the same quadratic production function, $s(x, y) = \alpha_{CC}x_Cy_C + \alpha_{MM}x_My_M$, and three different skill distributions, respectively: (1) bivariate normal distribution ([Lindenlaub, 2017](#)), (2) transformed Gumbel copula, and (3) untransformed Gumbel copula.

For the first DGP, we set workers' skill bundle, x , and jobs' skill requirements, y , to follow standard joint normal distributions with $\rho_x = -0.2$ and $\rho_y = -0.6$, respectively. For both nonnormal transformed and untransformed DGPs, we set $x = (x_1, x_2)$ and $y = (y_1, y_2)$ to follow the Gumbel copula with the shape parameter values 1.25 and 2.5, respectively. Then Kendall's correlation coefficients of x and y are 0.2 and 0.6. For the transformed data, we convert x and y into standard normally distributed variables:

$$x_{Ci} = \Phi^{-1}(x_{1i}), x_{Mi} = \Phi^{-1}(1 - x_{2i}), y_{Cj} = \Phi^{-1}(y_{1j}), y_{Mj} = \Phi^{-1}(1 - y_{2j}),$$

so that (x_{Ci}, x_{Mi}) and (y_{Cj}, y_{Mj}) are negatively correlated. For the untransformed data,

$$x_{Ci} = x_{1i}, x_{Mi} = 1 - x_{2i}, y_{Cj} = y_{1j}, y_{Mj} = 1 - y_{2j}.$$

As mentioned in the main text, there is no guarantee that the transformed data follow a joint normal distribution, implying that no closed-form solution exists for the equilibrium wage and assignment function. Hence, we numerically solve the equilibrium matching through linear programming for each Monte Carlo sample and different parameter values of $(\alpha_{CC}, \alpha_{MM})$.

Figure D.1 plots the skewness and variance of the distribution of wages for each DGP. As [Lindenlaub \(2017\)](#) derived, for all three DGPs, wage distributions are positively skewed for different pairs of α_{CC} and α_{MM} , and the wage dispersion increases as cognitive or manual skill complementarity increases. However, for the transformed Gumbel copula, the skewness decreases as α_{CC} decreases and α_{MM} increases. This simulation result is inconsistent with the theoretical result for normally distributed X and Y , in which the skewness is minimized when $\alpha_{CC} = \alpha_{MM}$. It implies that estimating the Gaussian model in [Lindenlaub \(2017\)](#) with transformed data can mislead the effects of technological changes on wages and inequality.

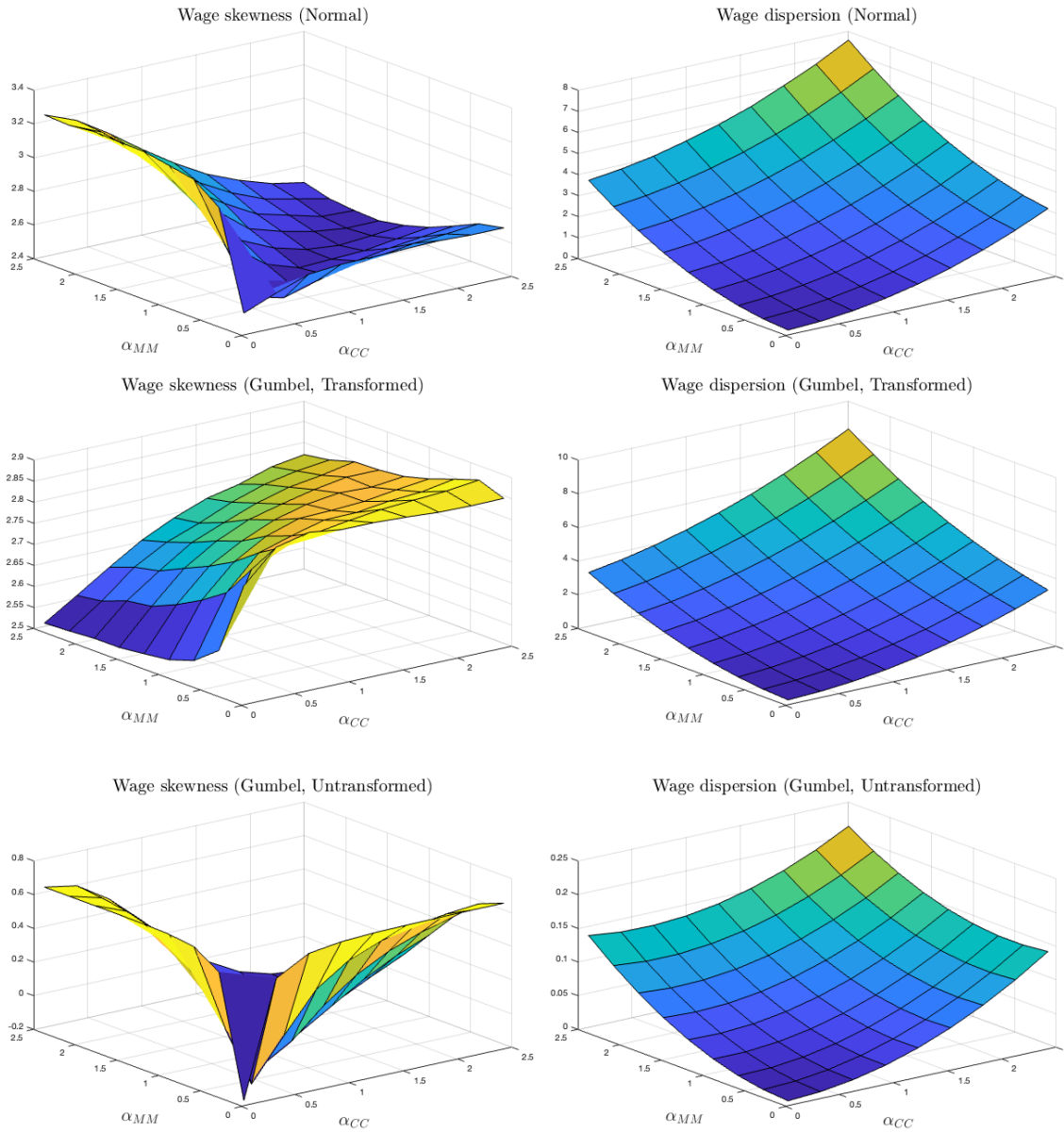


FIGURE D.1.—Effects of changes in production technology on wages and inequality