Semi-Nonparametric Models of Multidimensional Matching: An Optimal Transport Approach

Dongwoo Kim[1] Young Jun Lee[2]

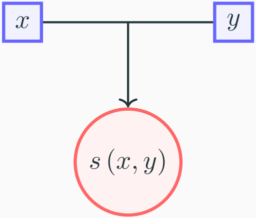
[1]Simon Fraser University

[2]Korea Institute for International Economic Policy

ASSA San Francisco, 2025

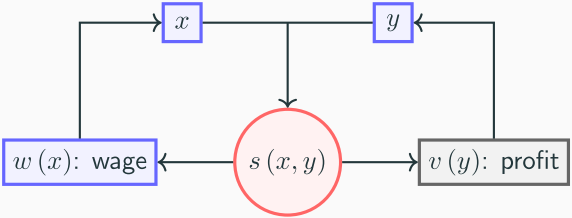
Worker-Job Matching

- Matching of worker x and job y produces a joint surplus $s(x, y)$



Worker-Job Matching

- Matching of worker x and job y produces a joint surplus $s(x, y)$
- Matching with transfers: surplus of a pair is split between worker and job



- We want to estimate $s(x, y)$ given data on matched pair $\{(x_i, y_i)\}_{i=1}^n$ and wage $\{w_i\}_{i=1}^n$ in order to understand:
  - How does attributes of x and y affect surplus?
  - Assortativeness or complementarities between workers and jobs
  - Evolution of wage distribution along with technological progress

Human types are continuous

This Paper

- **Multi-dimensional framework**: structural models of matching between workers and jobs
  - Multivariate and continuous $x \in \mathcal{X} \in \mathbb{R}^d$ and $y \in \mathcal{Y} \in \mathbb{R}^d$ represent worker and job types
  - Optimal transport literature $\Rightarrow$ equilibrium assignment $T^* : \mathcal{X} \to \mathcal{Y}$ and wage $w^* : \mathcal{X} \to \mathbb{R}_+$ depend on $s(x, y)$ and marginal distributions of x and y
- **Semi-nonparametric models**: sieve simultaneous estimation
  - No distributional assumption on x and y
    $\Rightarrow$ finite-dimensional θ for $s(x, y; \theta)$ and infinite-dimensional function $w(x)$
- **Empirical application**: U.S. matched pair data in 1990 and 2010
  - Cognitive and manual skills (worker) and those skill requirements (job)
  - Technological progress in favour of cognitive skills and its effect on *wage polarization*

Related literature

- Assortative matching in the context of couple matching: Becker (1991), Grossbard-Schechtman (1993), Pencavel (1998), Choo and Siow (2006)
- Optimal transport theory: Monge (1781), Villani (2003, 2008), De Philippis and Figalli (2014), Brenier (1991), Caffarelli (1992a,b, 1996), Cordero-Erausquin and Figalli (2019)
- Economic applications of optimal transport: Galichon (2017), Ekeland (2010), Chiappori *et al.* (2010), Chiong *et al.* (2016), Lindenlaub (2017), Galichon *et al.* (2019), Galichon and Salanié (2022), Gualdani and Sinha (2023)
- Sieve method for nonparametric models: Chen (2007), Shen (1997), Newey and Powell (2003), Ai and Chen (2003, 2007), Chen and Pouzo (2009)

Agents, Technology, Preferences

- Workers: $x = (x_1, \ldots, x_d) \in \mathcal{X} \subset \mathbb{R}^d$, with $x \sim P$
- Jobs: $y = (y_1, \ldots, y_d) \in \mathcal{Y} \subset \mathbb{R}^d$, with $y \sim Q$
- Twist condition: for any fixed x and $y_1 \neq y_2$, $\partial s\left(x, y_1\right) / \partial x \neq \partial s\left(x, y_2\right) / \partial x$
  - Supermodularity in one-dimension: $y \leq y' \Rightarrow \partial s\left(x, y\right) / \partial x \leq \partial s\left(x, y'\right) / \partial x$

Agents, Technology, Preferences

- Workers: $x = (x_1, \ldots, x_d) \in \mathcal{X} \subset \mathbb{R}^d$, with $x \sim P$
- Jobs: $y = (y_1, \ldots, y_d) \in \mathcal{Y} \subset \mathbb{R}^d$, with $y \sim Q$
- Twist condition: for any fixed x and $y_1 \neq y_2$, $\partial s\left(x, y_1\right) / \partial x \neq \partial s\left(x, y_2\right) / \partial x$
    - Supermodularity in one-dimension: $y \leq y' \Rightarrow \partial s\left(x, y\right) / \partial x \leq \partial s\left(x, y'\right) / \partial x$
- Preferences: risk-neutrality across workers and jobs
- Frictionless matching: perfect and costless information about potential matches
- Matching with transfers: allows compensating transfers to take place between workers and jobs

Division of Surplus between Worker and Job

- Workers and jobs have many potential partners:

$$\max_{y \in \mathcal{Y}} \{s(x,y) - v(y)\} \ \forall x \in \mathcal{X}, \quad \max_{x \in \mathcal{X}} \{s(x,y) - w(x)\} \ \forall y \in \mathcal{Y}.$$

- $w(x)$: "price" the firm offering job y must "pay" to attract worker x

Division of Surplus between Worker and Job

- Workers and jobs have many potential partners:

$$\max_{y\in\mathcal{Y}}\{s(x,y)-v(y)\}\ \forall x\in\mathcal{X},\quad \max_{x\in\mathcal{X}}\{s(x,y)-w(x)\}\ \forall y\in\mathcal{Y}.$$

- $w(x)$: "price" the firm offering job y must "pay" to attract worker x

- Stability (as a property of w and v): if $w(x)+v(y)<s(x,y)$ for some $(x,y)\in\mathcal{X}\times\mathcal{Y}$, the matching is not stable.

- Cost-minimization problem subject to stability constraints:

$$\inf_{w,v}\{\mathbb{E}_P[w(x)]+\mathbb{E}_Q[v(y)]\}\ \textit{s.t.}\ w(x)+v(y)\geq s(x,y),\ \forall(x,y)\in\mathcal{X}\times\mathcal{Y}. \quad (1)$$

- Solution to (1), (w^*,v^*): equilibrium wage function for x and equilibrium profit function from job y

Maximization of Aggregate Surplus: 2 Workers and 2 Jobs

- $x \in \mathcal{X} = \{H, L\}$ and $y \in \mathcal{Y} = \{C, M\}$
- $\pi_0(x, y) \in \{0, 1\}$: 1 if worker x and job y are matched, 0 else.
- Feasible matching on $\mathcal{X} \times \mathcal{Y}$: $\pi(x, y) = \pi_0(x, y)/2$, where π_0 satisfies

$$\pi_0(x, C) + \pi_0(x, M) = 1, \ \forall x \in \mathcal{X} \quad \& \quad \pi_0(H, y) + \pi_0(L, y) = 1, \ \forall y \in \mathcal{Y}.$$

s	C	M
H	9	3
L	7	5

Joint surplus

π_0	C	M
H	1	0
L	1	0

Nonfeasible

π_0	C	M
H	0	1
L	1	0

Feasible

Maximization of Aggregate Surplus: 2 Workers and 2 Jobs

- $x \in \mathcal{X} = \{H, L\}$ and $y \in \mathcal{Y} = \{C, M\}$
- $\pi_0(x, y) \in \{0, 1\}$: 1 if worker x and job y are matched, 0 else.
- Feasible matching on $\mathcal{X} \times \mathcal{Y}$: $\pi(x, y) = \pi_0(x, y)/2$, where π_0 satisfies

$$\pi_0(x, C) + \pi_0(x, M) = 1, \ \forall x \in \mathcal{X} \quad \& \quad \pi_0(H, y) + \pi_0(L, y) = 1, \ \forall y \in \mathcal{Y}.$$

s	C	M
H	9	3
L	7	5

Joint surplus

π_0	C	M
H	1	0
L	1	0

Nonfeasible

π_0	C	M
H	0	1
L	1	0

Feasible

(π, s)	C	M
H	(.5,9)	(0,3)
L	(0,7)	(.5,5)

Optimal

- Feasible matching, π^*, is optimal for $(\mathcal{X}, \mathcal{Y}, s)$ if, for all feasible matchings π',

$$\sum_{x,y} s(x, y)\, \pi^*(x, y) \geq \sum_{x,y} s(x, y)\, \pi'(x, y).$$

$\Rightarrow$ $\pi^* = \arg\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi [s(x, y)]$, where $\mathcal{M}(P, Q)$: set of feasible matchings.

Matching with a Continuum of Agents as Optimal Transport Problem

- Social planner's problem: For

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_{\pi}\left[s\left(x,y\right)\right]$$

  - $\mathcal{M}\left(P,Q\right)$: set of feasible probability distributions, π, over $\mathcal{X}\times\mathcal{Y}$ with

$$\int_{\mathcal{Y}} d\pi\left(x,y\right) = P\left(x\right) \text{ and } \int_{\mathcal{X}} d\pi\left(x,y\right) = Q\left(y\right)$$

Matching with a Continuum of Agents as Optimal Transport Problem

- Social planner's problem: For

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_{\pi}\left[s\left(x,y\right)\right]$$

  - $\mathcal{M}\left(P,Q\right)$: set of feasible probability distributions, π, over $\mathcal{X} \times \mathcal{Y}$ with

$$\int_{\mathcal{Y}} d\pi\left(x,y\right) = P\left(x\right) \text{ and } \int_{\mathcal{X}} d\pi\left(x,y\right) = Q\left(y\right)$$

  - This is exactly the Kantorovich formulation of optimal transport.

Matching with a Continuum of Agents as Optimal Transport Problem

- Social planner's problem: For

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_{\pi}\left[s\left(x,y\right)\right]$$

  - $\mathcal{M}\left(P,Q\right)$: set of feasible probability distributions, π, over $\mathcal{X} \times \mathcal{Y}$ with

$$\int_{\mathcal{Y}} d\pi\left(x,y\right) = P\left(x\right) \text{ and } \int_{\mathcal{X}} d\pi\left(x,y\right) = Q\left(y\right)$$

  - This is exactly the Kantorovich formulation of optimal transport.

- This is a linear programming problem $\Rightarrow$ Monge-Kantorovich duality

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_{\pi}\left[s\left(x,y\right)\right] = \min_{w(x)+v(y)\geq s(x,y)} \left\{\mathbb{E}_{P}\left[w\left(x\right)\right] + \mathbb{E}_{Q}\left[v\left(y\right)\right]\right\}$$

$$= \max_{T} \mathbb{E}_{P}\left[s\left(x,T\left(x\right)\right)\right] \quad \text{s.t. } T\left(P\right) = Q$$

Some Properties of Optimal Transport

- Additional noninteraction terms, $u(x)$ and/or $t(y)$, do not affect T^*, but added to w^* and v^*, respectively: for $s(x,y) = s^o(x,y) + u(x) + t(y)$,

$$\begin{aligned} &\min_{w(x)+v(y)\geq s(x,y)} \{\mathbb{E}_P[w(x)] + \mathbb{E}_Q[v(y)]\} \\ &= \max_{\pi\in\mathcal{M}(P,Q)} \mathbb{E}_\pi[s(x,y)] = \max_{\pi\in\mathcal{M}(P,Q)} \mathbb{E}_\pi[s^o(x,y)] + \mathbb{E}_P[u(x)] + \mathbb{E}_Q[t(y)] \\ &= \min_{w^o(x)+v^o(y)\geq s^o(x,y)} \{\mathbb{E}_P[w^o(x)] + \mathbb{E}_Q[v^o(x)]\} + \mathbb{E}_P[u(x)] + \mathbb{E}_Q[t(y)]. \end{aligned}$$

  - If $w^{o*}(x)$ is a solution to the OT problem with $s^o(x,y)$, then $w^{o*}(x) + u(x)$ is a solution to the OT problem with $s(x,y)$

Some Properties of Optimal Transport

- Additional noninteraction terms, $u(x)$ and/or $t(y)$, do not affect T^*, but added to w^* and v^*, respectively: for $s(x,y) = s^o(x,y) + u(x) + t(y)$,

$$\begin{aligned}
&\min_{w(x)+v(y)\geq s(x,y)} \left\{\mathbb{E}_P\left[w(x)\right] + \mathbb{E}_Q\left[v(y)\right]\right\} \\
&= \max_{\pi\in\mathcal{M}(P,Q)} \mathbb{E}_\pi\left[s(x,y)\right] = \max_{\pi\in\mathcal{M}(P,Q)} \mathbb{E}_\pi\left[s^o(x,y)\right] + \mathbb{E}_P\left[u(x)\right] + \mathbb{E}_Q\left[t(y)\right] \\
&= \min_{w^o(x)+v^o(y)\geq s^o(x,y)} \left\{\mathbb{E}_P\left[w^o(x)\right] + \mathbb{E}_Q\left[v^o(x)\right]\right\} + \mathbb{E}_P\left[u(x)\right] + \mathbb{E}_Q\left[t(y)\right].
\end{aligned}$$

  - If $w^{o*}(x)$ is a solution to the OT problem with $s^o(x,y)$, then $w^{o*}(x) + u(x)$ is a solution to the OT problem with $s(x,y)$

- Brenier (1991): If $s^o(x,y) = x'y$, there exists the unique (up to constant) convex solution w^{o*} such that $T^*(x) = dw^{o*}(x)/dx$.
  - Convexity of $w^{o*}(x) \Rightarrow$ monotonicity of $dw^{o*}(x)/dx$: if $\mathcal{X} = [0,1]$ and $x \sim \mathcal{U}([0,1])$, $T^* : [0,1] \to \mathcal{Y}$ is a quantile.

Bilinear $s(x, y)$

- Two-dimensional model for simplicity
  - Worker's cognitive and manual skills: $x = (x_C, x_M)$
  - Job's cognitive and manual skill demands: $y = (y_C, y_M)$

Bilinear $s\left(x, y\right)$

- Two-dimensional model for simplicity
  - Worker's cognitive and manual skills: $x = \left(x_C, x_M\right)$
  - Job's cognitive and manual skill demands: $y = \left(y_C, y_M\right)$

- Bilinear technology: with $\theta = \operatorname{vec}\left(A, b\right)$,

$$s\left(x, y; \theta\right) = x'Ay + x'b = \begin{pmatrix} x_C & x_M \end{pmatrix} \begin{pmatrix} \alpha_{CC} & \alpha_{CM} \\ \alpha_{MC} & \alpha_{MM} \end{pmatrix} \begin{pmatrix} y_C \\ y_M \end{pmatrix} + \begin{pmatrix} x_C & x_M \end{pmatrix} \begin{pmatrix} \beta_C \\ \beta_M \end{pmatrix}$$

- $x'Ay$: value of interaction
  - α_{CC} and α_{MM}: within-task complementarity or substitutability
  - α_{CM} and α_{MC}: between-task complementarity or substitutability
- $x'b$: value of non-interaction

Unique Existence of Equilibrium Wage and Matching Functions

Application of Brenier's theorem to the assignment problem from $\mathcal{X}$ to $A\mathcal{Y}$:

Assumption 1

P and Q have finite second moments, and P is absolutely continuous.

Proposition 1

If Assumption 1 holds, then there exists the unique (up to constant) convex solution, $w^{o*}(x)$, to the dual problem with $s^o(x,y) = x'Ay$, and the equilibrium wage function $w^*(x)$ is given by

$$w^*(x) = w^{o*}(x) + \beta_C x_C + \beta_M x_M + c$$

with a constant of normalization c, and the equilibrium assignment is given by

$$A\begin{pmatrix} y_C^*(x) & y_M^*(x) \end{pmatrix}' = \begin{pmatrix} \partial w^{o*}(x)/\partial x_C & \partial w^{o*}(x)/\partial x_M \end{pmatrix}'.$$

Empirical Model

- Bridging the equilibrium with real-world data
  1. Unobserved worker's characteristics affecting surplus: $s\left(x, y^{*}; \theta\right) = x' A y^{*} + x' b + \varepsilon_{w}$
  2. Accounting for measurement errors, ε_{y}: $y = y^{*} + \varepsilon_{y}$

- Observations are denoted by $z_i = \left(w_i, y_i, x_i\right)$, where $i = 1, \ldots, n$, arising from the model with moment conditions:

$$\begin{aligned} w_i &= w^{*}\left(x_i\right) + \varepsilon_{wi} = \underbrace{w^{o*}\left(x_i\right) + c}_{:= w\left(x_i\right)} + x_i' b + \varepsilon_{wi}, \\ y_i &= y_i^{*} + \varepsilon_{yi} = A^{-1} dw\left(x_i\right)/dx + \varepsilon_{yi}. \end{aligned} \tag{2}$$

- Parameters of interests: $\theta = \mathrm{vec}\left(A^{-1}, b\right)$ and $w\left(x\right)$

Comparison between This Paper and Lindenlaub (2017)

- Common settings
  - Bi-linear function for joint surplus from matches: $s\left(x, y^{*}; \theta\right) = x' A y^{*} + x' b$
  - System of $(1+d)$ regressions: (i) $w_i = w^{*}\left(x_i\right) + \varepsilon_{wi} = w\left(x_i\right) + x_i' b + \varepsilon_{wi}$, (ii) $y_i = y_i^{*} + \varepsilon_{yi} = A^{-1} dw\left(x_i\right)/dx + \varepsilon_{yi}$

Comparison between This Paper and Lindenlaub (2017)

- Common settings
  - Bi-linear function for joint surplus from matches: $s(x, y^*; \theta) = x' A y^* + x' b$
  - System of $(1 + d)$ regressions: (i) $w_i = w^*(x_i) + \varepsilon_{wi} = w(x_i) + x_i' b + \varepsilon_{wi}$, (ii) $y_i = y_i^* + \varepsilon_{yi} = A^{-1} dw(x_i)/dx + \varepsilon_{yi}$

- Distributional assumption on x and y^*
  - Lindenlaub (2017): $x \sim N(0, \Sigma_x)$, $y^* \sim N(0, \Sigma_{y^*})$
    $\Rightarrow w(x; A, \Sigma_x, \Sigma_{y^*}) = x' J(A, \Sigma_x, \Sigma_{y^*}) x$
  - This paper: no distributional specification $\Rightarrow w(x)$

Comparison between This Paper and Lindenlaub (2017)

- Common settings
  - Bi-linear function for joint surplus from matches: $s(x, y^*; \theta) = x' A y^* + x' b$
  - System of $(1+d)$ regressions: (i) $w_i = w^*(x_i) + \varepsilon_{wi} = w(x_i) + x_i' b + \varepsilon_{wi}$, (ii) $y_i = y_i^* + \varepsilon_{yi} = A^{-1} dw(x_i)/dx + \varepsilon_{yi}$

- Distributional assumption on x and y^*
  - Lindenlaub (2017): $x \sim N(0, \Sigma_x)$, $y^* \sim N(0, \Sigma_{y^*})$ $\Rightarrow w(x; A, \Sigma_x, \Sigma_{y^*}) = x' J(A, \Sigma_x, \Sigma_{y^*}) x$
  - This paper: no distributional specification $\Rightarrow w(x)$

- Distributional assumption on $\varepsilon = (\varepsilon_w, \varepsilon_y')'$
  - Lindenlaub (2017): normally distributed, uncorrelated, and homoskedastic ε $\Rightarrow$ parametric nonlinear MLE
  - This paper allows for nonnormally distributed, correlated, and heteroskedastic ε $\Rightarrow$ semi-nonparametric sieve GLSE

Conditional Moment Restrictions

- $\mathbb{E}\left[\varepsilon_i | x_i\right] = 0$ implies that $\mathbb{E}\left[\rho\left(z_i;\theta,w\right)|x_i\right] = 0$ at a true parameter (θ_0, w_0), where

$$\rho\left(z_i;\theta,w\right) = \begin{pmatrix} \rho_w\left(w_i, x_i;\theta,w\right) \\ \rho_y\left(y_i, x_i;\theta,w\right) \end{pmatrix} = \begin{pmatrix} w_i - \left(w\left(x_i\right) + x_i' b\right) \\ y_i - A^{-1} dw\left(x_i\right)/dx \end{pmatrix}.$$

- **Theorem 1:** If $\mathbb{E}\left[\varepsilon_i | x_i\right] = 0$, then $\theta_0 = \mathrm{vec}\left(A_0^{-1}, b_0\right)$ and $w_0\left(x\right)$ in the model (2) are identified. details

Conditional Moment Restrictions

- $\mathbb{E}[\varepsilon_i|x_i] = 0$ implies that $\mathbb{E}[\rho(z_i;\theta,w)|x_i] = 0$ at a true parameter (θ_0, w_0), where

$$\rho(z_i;\theta,w) = \begin{pmatrix} \rho_w(w_i, x_i;\theta,w) \\ \rho_y(y_i, x_i;\theta,w) \end{pmatrix} = \begin{pmatrix} w_i - (w(x_i) + x_i'b) \\ y_i - A^{-1}dw(x_i)/dx \end{pmatrix}.$$

- **Theorem 1:** If $\mathbb{E}[\varepsilon_i|x_i] = 0$, then $\theta_0 = \mathrm{vec}\left(A_0^{-1}, b_0\right)$ and $w_0(x)$ in the model (2) are identified. ▸ details

- Empirical criterion for GLS estimation:

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

  - $\hat{\Sigma}(x_i)$: positive definite weighting matrix.
  - Minimization over $\Theta \times \mathcal{W}$, but $\mathcal{W}$ is an infinite dimensional function space

Sieve GLS Estimation

- Sieve method: Minimize Q_n over a sequence of approximating spaces

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

where $\{p_{j_C}(x_C)\}_{j_C=0}^{k_{C_n}}$ and $\{p_{j_M}(x_M)\}_{j_M=0}^{k_{M_n}}$: known basis functions

Sieve GLS Estimation

- Sieve method: Minimize Q_n over a sequence of approximating spaces

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

where $\{p_{j_C}(x_C)\}_{j_C=0}^{k_{C_n}}$ and $\{p_{j_M}(x_M)\}_{j_M=0}^{k_{M_n}}$: known basis functions

---

Algorithm: Computing the Sieve GLS Estimator of θ and w

---

1. Obtain an initial consistent sieve LS estimator $\left(\widetilde{\theta}_n, \widetilde{w}_n(x) = w_n(x;\tilde{\gamma}_n)\right)$ by $\min_{(\theta,w)\in\Theta\times\mathcal{W}_n} \sum_{i=1}^n \rho(z_i;\theta,w)' \rho(z_i;\theta,w)$;
2. Obtain a consistent estimator $\hat{\Sigma}_n(x)$ of $\Sigma_0(x) = \mathrm{Var}[\rho(Z;\theta,w)|\, x]$ using $\left(\widetilde{\theta}_n, \widetilde{w}_n\right)$ from sieve LS estimation;
3. Obtain the optimally weighted sieve GLS estimator $\left(\hat{\theta}_n, \hat{w}_n(x) = w_n(x;\hat{\gamma}_n)\right)$ by $\min_{(\theta,w)\in\Theta\times\mathcal{W}_n} \sum_{i=1}^n \rho(z_i;\theta,w)' \left[\hat{\Sigma}_n(x_i)\right]^{-1} \rho(z_i;\theta,w)$.

- **Theorem 2 and Theorem 3**: asymptotic normality and semiparametric efficiency of $\hat{\theta}_n$

Comparing Estimators

Table 1: Comparison of key assumptions in estimation procedures

Assumptions	Lindenlaub	Kim and Lee		
	ML	SML	SLS	SGLS
Joint normal (X, Y)	✓			
Normal errors	✓	✓		
Uncorrelated errors	✓	✓		
Homoskedasticity	✓	✓	✓	

Simulation 1: Lindenlaub (2017)'s Gaussian Model

- $s\left(x, y^*; \alpha_{CC}, \alpha_{MM}, \beta_C, \beta_M\right) = \alpha_{CC} x_C y_C^* + \alpha_{MM} x_M y_M^* + \beta_C x_C + \beta_M x_M$
- Distributions for x and y^*:

$$\begin{pmatrix} x_C \\ x_M \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_x \\ \rho_x & 1 \end{bmatrix}\right), \quad \begin{pmatrix} y_C^* \\ y_M^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{y^*} \\ \rho_{y^*} & 1 \end{bmatrix}\right)$$

- Measurement errors:

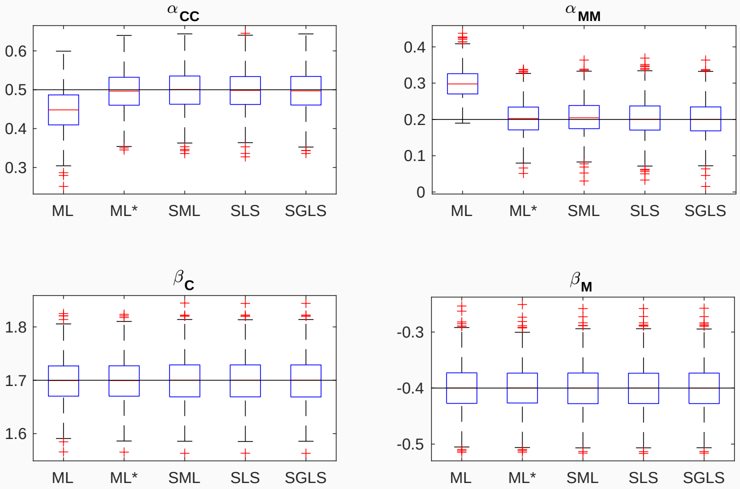
$$\varepsilon_w \sim N\left(0, \sigma_w^2\right), \quad \varepsilon_C \sim N\left(0, \sigma_C^2\right), \quad \varepsilon_M \sim N\left(0, \sigma_M^2\right),$$

- Parameter values in simulations:

$$(\alpha_{CC}, \alpha_{MM}, \beta_C, \beta_M, w_0, \rho_x, \rho_{y^*}) = (0.5, 0.2, 1.7, -0.4, 30, -0.4, -0.5).$$

- Closed-form solution can be used to obtain a sample of wage and matching.
- Sample size: $n = 3000$; # of Monte Carlo simulations: 1000

Simulation 1: Estimation Performance



ML vs ML* ML: $w(x) = x' J(A, \Sigma_x, \Sigma_y)\, x$; ML*: $w(x) = x' J(A, \Sigma_x, \Sigma_{y^*})\, x$

Simulation 2: Non-Gaussian Model

- x and y^*: Gaussian mixture distributions from two Gaussian components K_1 and K_2 ($\rho_x = 0.4$ and $\rho_y = 0.5$) with one-half weight for each, where

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

- ε: Gaussian mixture distribution from M_1 and M_2, where

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

- Parameter values, sample size, # of simulations are same as in Simulation 1.
- No closed-form solution for equilibrium wage and assignment
$\Rightarrow$ GUROBI OPTIMIZER solves the equilibrium numerically through linear programming for each Monte Carlo sample.

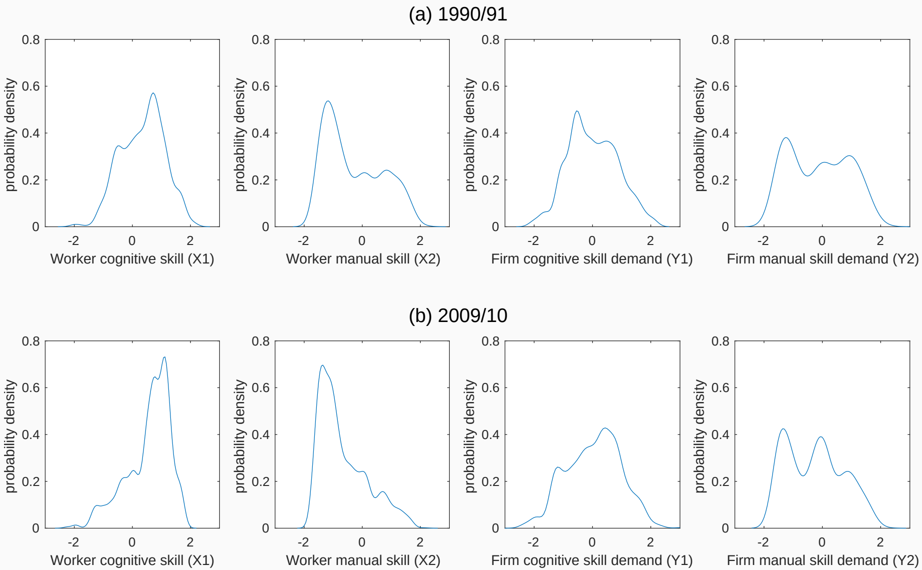
Simulation 2: Estimation Performance

		ML*	SML	SLS	SGLS
α_{CC}	Bias	0.2896	-0.0014	-0.0016	-0.0017
	SE	0.2228	0.0429	0.0425	0.0385
	RMSE	0.3653	0.0429	0.0425	0.0385
α_{MM}	Bias	0.2446	-0.0017	0.0006	-0.0027
	SE	0.2621	0.0454	0.0431	0.0336
	RMSE	0.3584	0.0454	0.0431	0.0337
β_C	Bias	0.7123	-0.0007	-0.0013	0.0008
	SE	0.1081	0.0428	0.0426	0.0364
	RMSE	0.7204	0.0428	0.0426	0.0364
β_M	Bias	-0.1288	-0.0001	0.0002	-0.0011
	SE	0.0937	0.0421	0.0419	0.0305
	RMSE	0.1592	0.0421	0.0419	0.0305

Empirical Model with the U.S. Data

- Data constructed by Lindenlaub (2017)
- Job's skill demands: $y = (y_C, y_M)$ (cognitive and manual)
  - Obtain (y_C, y_M) using principal component analysis from O*NET
  - (1,0.11) for 'Physicist', (0.34, 1.00) for 'Dancer'
  - (0.62,0.03) for 'Economist'
- Worker's skills: $x = (x_C, x_M)$ (cognitive and manual)
  - Match data from NLSY on education and training of workers to (y_C, y_M)
  - Selected groups: employed workers aged between 27 and 29 years during years (1) 1990/91 (NLSY79, $n_1 = 2984$) and (2) 2009/10 (NLSY97, $n_2 = 4495$)
- $s\left(x, y^*; \alpha_C, \alpha_M, \beta_C, \beta_M\right) = \alpha_C x_C y_C^* + \alpha_M x_M y_M^* + \beta_C x_C + \beta_M x_M$
  - α_C and α_M: worker-job complementarity
  - β_C and β_M: productivity of skills independent of a job's skill demands

Distributions of Skill Supply and Demand



Test for Normality of Transformed Data

$$
\begin{tikzcd}
\begin{pmatrix} x_C \\ x_M \end{pmatrix} \sim P_{\mathcal{X}} \arrow[r] \arrow[d, "T^*(x)"] & \begin{array}{l} \widetilde{x}_C = \Phi^{-1}\left(F_{x_C}\left(x_C\right)\right) \sim N(0,1) \\ \widetilde{x}_M = \Phi^{-1}\left(F_{x_M}\left(x_M\right)\right) \sim N(0,1) \end{array} \arrow[r, "?"] & \begin{pmatrix} \widetilde{x}_C \\ \widetilde{x}_M \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho_{\widetilde{x}} \\ \rho_{\widetilde{x}} & 1 \end{pmatrix}\right) \arrow[d, "\widetilde{T}^*(\widetilde{x})"'] \\
\begin{pmatrix} y_C \\ y_M \end{pmatrix} \sim P_{\mathcal{Y}} \arrow[r] & \begin{array}{l} \widetilde{y}_C = \Phi^{-1}\left(F_{y_C}\left(y_C\right)\right) \sim N(0,1) \\ \widetilde{y}_M = \Phi^{-1}\left(F_{y_M}\left(y_M\right)\right) \sim N(0,1) \end{array} \arrow[r, "?"] & \begin{pmatrix} \widetilde{y}_C \\ \widetilde{y}_M \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho_{\widetilde{y}} \\ \rho_{\widetilde{y}} & 1 \end{pmatrix}\right)
\end{tikzcd}
$$

Figure 1: Lindenlaub (2017)'s transformation: Φ denotes the standard normal c.d.f.

Test for Normality of Transformed Data

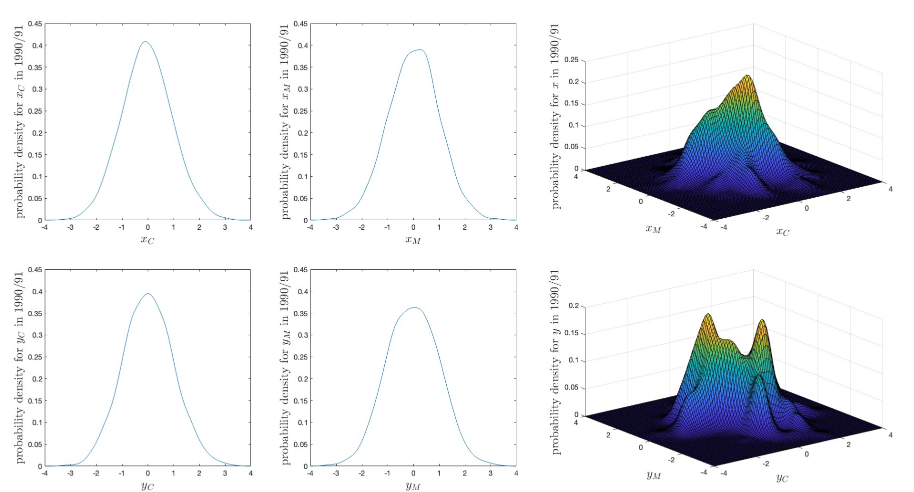
$$
\begin{array}{ccccc}
\begin{pmatrix} x_C \\ x_M \end{pmatrix} \sim P_{\mathcal{X}} & \longrightarrow & \begin{array}{l} \widetilde{x}_C = \Phi^{-1}\left(F_{x_C}\left(x_C\right)\right) \sim N\left(0,1\right) \\ \widetilde{x}_M = \Phi^{-1}\left(F_{x_M}\left(x_M\right)\right) \sim N\left(0,1\right) \end{array} & \xrightarrow{?} & \begin{pmatrix} \widetilde{x}_C \\ \widetilde{x}_M \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho_{\widetilde{x}} \\ \rho_{\widetilde{x}} & 1 \end{pmatrix}\right) \\
\Big\downarrow T^*\left(x\right) & & & & \widetilde{T}^*\left(\widetilde{x}\right) \Big\downarrow \\
\begin{pmatrix} y_C \\ y_M \end{pmatrix} \sim P_{\mathcal{Y}} & \longrightarrow & \begin{array}{l} \widetilde{y}_C = \Phi^{-1}\left(F_{y_C}\left(y_C\right)\right) \sim N\left(0,1\right) \\ \widetilde{y}_M = \Phi^{-1}\left(F_{y_M}\left(y_M\right)\right) \sim N\left(0,1\right) \end{array} & \xrightarrow{?} & \begin{pmatrix} \widetilde{y}_C \\ \widetilde{y}_M \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho_{\widetilde{y}} \\ \rho_{\widetilde{y}} & 1 \end{pmatrix}\right)
\end{array}
$$

Figure 1: Lindenlaub (2017)'s transformation: Φ denotes the standard normal c.d.f.

Table 2: Mardia's multivariate normality test statistics: p-values in parentheses

	1990/91 ($n_1 = 2984$)		2009/10 ($n_2 = 4495$)	
	$\widetilde{x}$	$\widetilde{y}$	$\widetilde{x}$	$\widetilde{y}$
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)

Nonnormality of Transformed Data



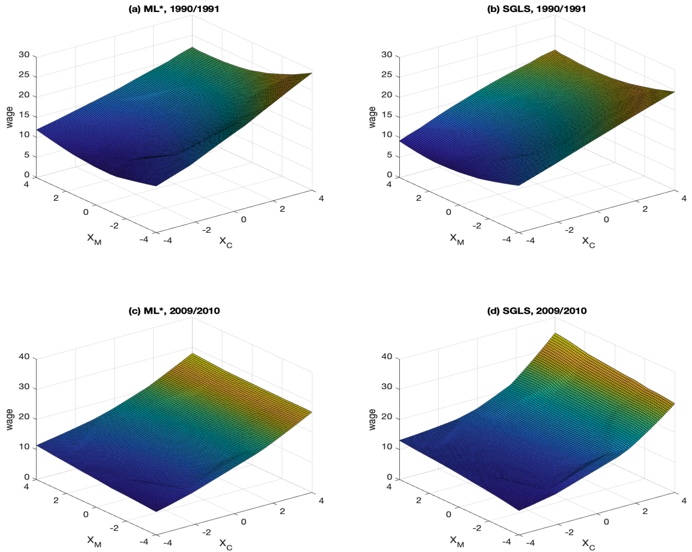
Estimation Results

Table 3: Parameter estimates of worker-job surplus on transformed data

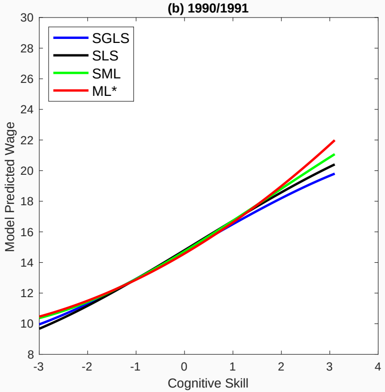
	1990/91		2009/10	
	ML*	SGLS	ML*	SGLS
α_C	0.765	0.000	1.119	2.290
	(0.574)	(0.000)	(0.408)	(0.265)
α_M	1.270	0.856	0.486	0.291
	(0.148)	(0.098)	(0.633)	(0.059)
β_C	1.711	1.585	2.208	2.198
	(0.589)	(0.416)	(0.540)	(0.534)
β_M	-0.392	-0.388	0.243	0.327
	(0.406)	(0.309)	(0.731)	(0.532)

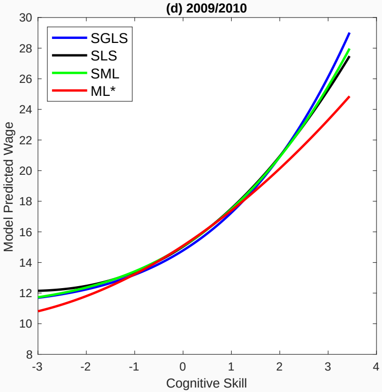
- Changes in complementarities across periods: cognitive attributes $\uparrow$, but manual inputs $\downarrow$
- Normality of x and y for ML* $\Rightarrow$ difference between $\hat{\alpha}_{ML^*}$ and $\hat{\alpha}_{SML}$
- Normality, uncorrelatedness, and homoskedasiticity of ε for SML $\Rightarrow$ difference between $\hat{\alpha}_{SML}$ and $\hat{\alpha}_{SGLS}$

Estimated Wage Functions



Estimated Wage Functions w.r.t. Cognitive Skill





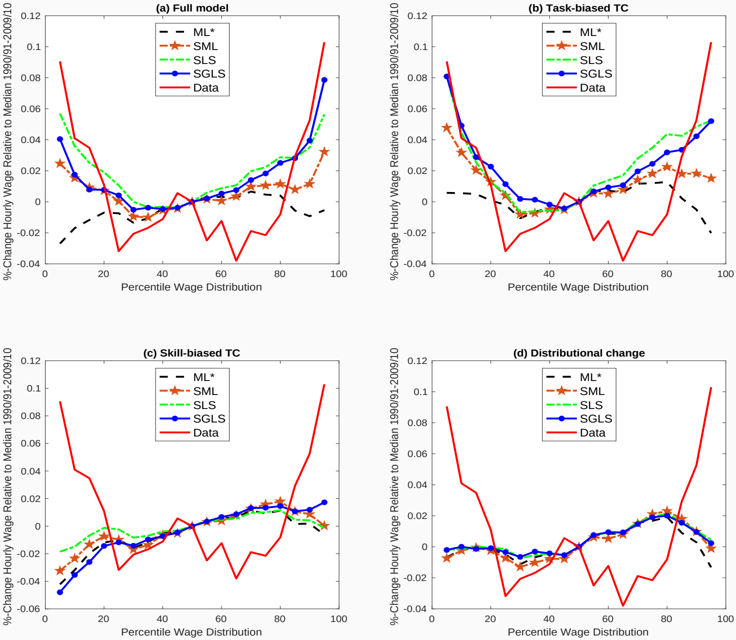
Counterfactual Study

- 'Wage polarization' in the U.S.: stronger wage growth in the lower and upper tails of the distribution compared to the median

- Analyzing whether wage polarization can be triggered by the estimated surplus or distributional changes when fed into the model

- Baseline: $\hat{w}_i^{1990} = \hat{w}^{1990}\left(x_i^{1990}\right) + \hat{\beta}_C^{1990} x_{Ci}^{1990} + \hat{\beta}_M^{1990} x_{Mi}^{1990}$

- Changes in wage distribution of $\hat{w}_i^{1990}$ to
  (a) $\hat{w}_{1i}^{2010} = \hat{w}^{2010}\left(x_i^{2010}\right) + \hat{\beta}_C^{2010} x_{Ci}^{2010} + \hat{\beta}_M^{2010} x_{Mi}^{2010}$: Full model
  (b) $\hat{w}_{2i}^{2010} = \hat{w}^{2010}\left(x_i^{1990}\right) + \hat{\beta}_C^{1990} x_{Ci}^{1990} + \hat{\beta}_M^{1990} x_{Mi}^{1990}$: Task-biased TC
  (c) $\hat{w}_{3i}^{2010} = \hat{w}^{1990}\left(x_i^{1990}\right) + \hat{\beta}_C^{2010} x_{Ci}^{1990} + \hat{\beta}_M^{2010} x_{Mi}^{1990}$: Skill-biased TC
  (d) $\hat{w}_{4i}^{2010} = \hat{w}^{1990}\left(x_i^{2010}\right) + \hat{\beta}_C^{1990} x_{Ci}^{2010} + \hat{\beta}_M^{1990} x_{Mi}^{2010}$: Distributional Change



(a) Full model
(b) Task-biased TC
(c) Skill-biased TC
(d) Distributional change
ML*
SML
SLS
SGLS
Data
%-Change Hourly Wage Relative to Median 1990/91-2009/10
Percentile Wage Distribution

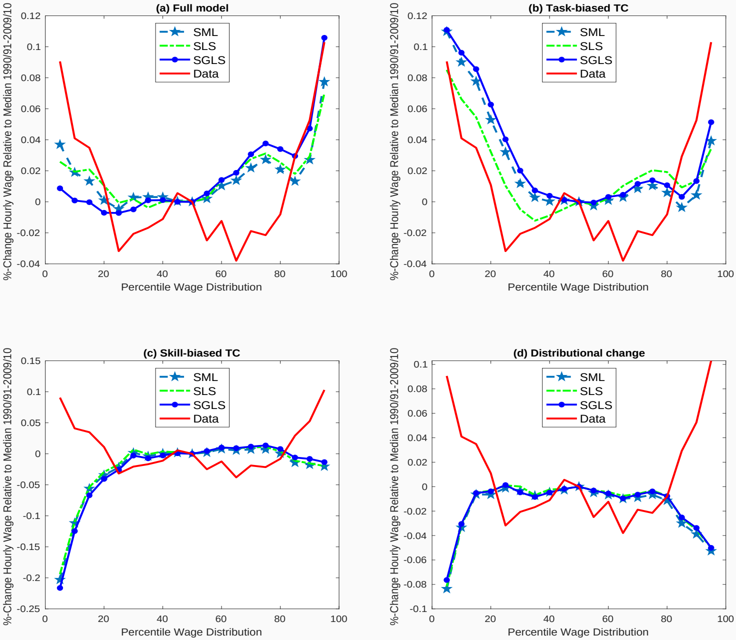
Estimation on Original Data

Our method works for any arbitrary underlying distributions of (x, y), Lindenlaub's model doesn't.

Table 4: 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	6.032	5.784	6.471
	(0.000)	(0.000)	(0.000)	(0.110)	(0.184)	(0.196)
α_{MM}	1.947	2.292	2.234	0.000	1.020	0.003
	(0.060)	(0.081)	(0.079)	(0.000)	(0.024)	(0.000)
β_C	2.238	2.246	2.108	3.395	3.299	3.707
	(0.083)	(0.225)	(0.220)	(0.101)	(0.396)	(0.292)
β_M	-0.341	-0.441	-0.317	0.254	0.384	0.440
	(0.073)	(0.211)	(0.213)	(0.097)	(0.333)	(0.239)

Standard errors in the parentheses


(a) Full model
(b) Task-biased TC
(c) Skill-biased TC
(d) Distributional change
SML
SLS
SGLS
Data
%-Change Hourly Wage Relative to Median 1990/91-2009/10
Percentile Wage Distribution

What we have done:

- Empirical model of multidimensional matching in line with theory
- Identification and estimation of parametric surplus function and nonparametric equilibrium wage and matching functions
- More explanation power in the evolution of wage inequality in the U.S. between 1990 and 2010

Next steps:

- Inclusion of more skills, e.g., interpersonal and digital skills: Lindenlaub and Postel-Vinay (2021)
- Multidimensional measurement error on x in $w(x)$: Schennach (2013), Ben-Moshe (2021)
- Search frictions and unobserved heterogeneity: Eeckhout and Kircher (2011), Bojilov and Galichon (2016)

Thank you!

AI, C. and CHEN, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, **71** (6), 1795–1843.

— and — (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, **141** (1), 5–43.

BECKER, G. S. (1991). *A treatise on the family: Enlarged edition*. Harvard university press.

BEN-MOSHE, D. (2021). Identification of linear regressions with errors in all variables. *Econometric Theory*, **37** (4), 633–663.

BOJILOV, R. and GALICHON, A. (2016). Matching in closed-form: equilibrium, identification, and comparative statics. *Economic Theory*, **61**, 587–609.

BRENIER, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, **44** (4), 375–417.

CAFFARELLI, L. A. (1992a). Boundary regularity of maps with convex potentials. *Communications on Pure and Applied Mathematics*, **45** (9), 1141–1151.

— (1992b). The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, **5** (1), 99–104.

— (1996). Boundary regularity of maps with convex potentials–ii. *Annals of mathematics*, **144** (3), 453–496.

CHEN, X. (2007). Chapter 76 large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, vol. 6, Elsevier, pp. 5549–5632.

— and POUZO, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics*, **152** (1), 46–60.

CHIAPPORI, P.-A., MCCANN, R. J. and NESHEIM, L. P. (2010). Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. *Economic Theory*, **42**, 317–354.

CHIONG, K. X., GALICHON, A. and SHUM, M. (2016). Duality in dynamic discrete-choice models. *Quantitative Economics*, **7** (1), 83–115.

CHOO, E. and SIOW, A. (2006). Who marries whom and why. *Journal of political Economy*, **114** (1), 175–201.

CORDERO-ERAUSQUIN, D. and FIGALLI, A. (2019). Regularity of monotone transport maps between unbounded domains. *Discrete and Continuous Dynamical Systems*, **39** (12), 7101–7112.

DE PHILIPPIS, G. and FIGALLI, A. (2014). The Monge-Ampère equation and its link to optimal transportation. *Bull. Amer. Math. Soc. (N.S.)*, **51** (4), 527–580.

EECKHOUT, J. and KIRCHER, P. (2011). Identifying sorting-in theory. *The Review of Economic Studies*, **78** (3), 872–906.

EKELAND, I. (2010). Notes on optimal transportation. *Economic Theory*, pp. 437–459.

GALICHON, A. (2017). A survey of some recent applications of optimal transport methods to econometrics. *The Econometrics Journal*, **20** (2), C1–C11.

—, KOMINERS, S. D. and WEBER, S. (2019). Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Political Economy*, **127** (6), 2875–2925.

— and SALANIÉ, B. (2022). Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, **89** (5), 2600–2629.

GROSSBARD-SCHECHTMAN, S. (1993). A theory of marriage, labor and divorce.

GUALDANI, C. and SINHA, S. (2023). Partial identification in matching models for the marriage market. *Journal of Political Economy*, **131** (5), 1109–1171.

LINDENLAUB, I. (2017). Sorting multidimensional types: Theory and application. *The Review of Economic Studies*, **84** (2), 718–789.

— and POSTEL-VINAY, F. (2021). *The Worker-Job Surplus*. Working Paper 28402, National Bureau of Economic Research.

MONGE, G. (1781). Histoire de l'académie royale des sciences de paris. *De L'imprimerie Royale*.

NEWEY, W. K. and POWELL, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, **71** (5), 1565–1578.

PENCAVEL, J. (1998). Assortative mating by schooling and the work behavior of wives and husbands. *The American Economic Review*, **88** (2), 326–329.

Identification: Assumptions and Proof ◂ main

Assumptions for Theorem 1

(i) P and Q have finite second moments, and P is absolutely continuous;
(ii) A is invertible;
(iii) There exist $y_1, y_2, y_3 \in \mathcal{Y}$ s.t. $y_1 - y_2$ and $y_2 - y_3$ are linearly independent;
(iv) $\mathbb{E}\left[\varepsilon|x\right] = 0$.

Sketch of proof of Theorem 1:

1. $w_0^*\left(x\right) = \mathbb{E}\left[w|x\right]$ is identified and $\mathbb{E}\left[y|x\right] = A_0^{-1}\left(\nabla w_0^*\left(x\right) - b_0\right)$.
2. $\exists x_1, x_2, x_3 \in \mathcal{X}$ such that $\nabla w_0\left(x_1\right) - \nabla w_0\left(x_2\right)$ and $\nabla w_0\left(x_2\right) - \nabla w_0\left(x_3\right)$ are linearly independent.
3. A_0 is identified from that $\nabla w_0\left(x\right) - \nabla w_0\left(x'\right) = \nabla w_0^*\left(x\right) - \nabla w_0^*\left(x'\right)$ and

$$\begin{aligned}&\begin{pmatrix}\mathbb{E}\left[y|x_1\right] - \mathbb{E}\left[y|x_2\right] & \mathbb{E}\left[y|x_2\right] - \mathbb{E}\left[y|x_3\right]\end{pmatrix}\\ &= A^{-1}\begin{pmatrix}\nabla w_0^*\left(x_1\right) - \nabla w_0^*\left(x_2\right) & \nabla w_0^*\left(x_2\right) - \nabla w_0^*\left(x_3\right)\end{pmatrix}.\end{aligned}$$

4. In turn, b_0 and $w_0\left(x\right)\left(= w_0^*\left(x\right) - x'b_0\right)$ are also identified.

Estimation Bias in Lindenlaub (2017) main

Estimation with $\{z_i = (w_i, x_i, y_i)\}_{i=1}^{n}$ from

$$w_i = w^*(x_i) + \varepsilon_{wi}, \quad y_i = y_i^* + \varepsilon_{yi} = T^*(x_i) + \varepsilon_{yi}.$$

- If $x \sim N(0, \Sigma_x)$ and $y^* \sim N(0, \Sigma_{y^*})$, then closed-form equilibrium wage and matching functions should be

$$w^*(x; A, b, c, \Sigma_x, \Sigma_{y^*}) = \frac{1}{2} x' J(A, \Sigma_x, \Sigma_{y^*}) x + x'b + c,$$
$$T^*(x; A, \Sigma_x, \Sigma_{y^*}) = J(A, \Sigma_y, \Sigma_{y^*}) x,$$

where $J(A, \Sigma_x, \Sigma_{y^*}) = \begin{pmatrix} J_{11}(A, \Sigma_x, \Sigma_{y^*}) & J_{12}(A, \Sigma_x, \Sigma_{y^*}) \\ J_{21}(A, \Sigma_x, \Sigma_{y^*}) & J_{22}(A, \Sigma_x, \Sigma_{y^*}) \end{pmatrix}$.

Estimation Bias in Lindenlaub (2017) main

Estimation with $\{z_i = (w_i, x_i, y_i)\}_{i=1}^{n}$ from

$$w_i = w^*(x_i) + \varepsilon_{wi}, \quad y_i = y_i^* + \varepsilon_{yi} = T^*(x_i) + \varepsilon_{yi}.$$

- If $x \sim N(0, \Sigma_x)$ and $y^* \sim N(0, \Sigma_{y^*})$, then closed-form equilibrium wage and matching functions should be

$$w^*(x; A, b, c, \Sigma_x, \Sigma_{y^*}) = \frac{1}{2} x' J(A, \Sigma_x, \Sigma_{y^*}) x + x'b + c,$$
$$T^*(x; A, \Sigma_x, \Sigma_{y^*}) = J(A, \Sigma_y, \Sigma_{y^*}) x,$$

  where $J(A, \Sigma_x, \Sigma_{y^*}) = \begin{pmatrix} J_{11}(A, \Sigma_x, \Sigma_{y^*}) & J_{12}(A, \Sigma_x, \Sigma_{y^*}) \\ J_{21}(A, \Sigma_x, \Sigma_{y^*}) & J_{22}(A, \Sigma_x, \Sigma_{y^*}) \end{pmatrix}$.

- However, Lindenlaub (2017)'s solution uses $\Sigma_y = \Sigma_{y^*} + \Sigma_{\varepsilon_y}$ instead of Σ_{y^*}, where $\varepsilon_y \sim N\left(0, \Sigma_{\varepsilon_y}\right)$.