Assortative Matching in Skill: Evidence from Sweden

Rikard Forslid (SU) Erik Lindqvist (SSE & IFN) Jonas Vlachos (SU & IFN)

January 8, 2017

Two views

- Why do firms differ with respect to the skill level of their workforces?
- Two (broad) views:
  1. Sorting reflect complementarities in worker skills (Kremer 1993; Kremer and Maskin, 1996; Garicano and Rossi-Hansbergs, 2006)
  2. Workers sort into different type of jobs (Acemoglu 1999; Caselli 1999)

Data and sample

- We combine three data sets:
  1. Employer-employee matched data for the universe of Swedish firms and workers
  2. Data on workers' occupation
  3. Data on cognitive and non-cognitive skills for men measured at the Swedish military draft (age 18-19)
- We use the draft skill measures to construct a composite skill measure, s, which we normalize to mean zero and unit variance
- We focus on a sample of 30-55 year-old men from whom we observe both occupation and draft-level skills

Decomposing the between-firm variance in skill

Let s_j denote the average skill level in firm j and N_j the number of workers in firm j. We decompose the between-firm variance in skills, $Var\left(s_j - \overline{s}\right)^2$, using

$$\widehat{s}_j = \frac{1}{N_j} \sum_h N_{hj} \overline{s}_h$$

where $\overline{s}_h$ is the average skill level in occupation h and N_{hj} the number of workers in firm j whose occupation is h. Hence

$$Var\left(s_j - \overline{s}\right)^2$$

$$= \underbrace{Var\left(\widehat{s}_j - \overline{s}\right)}_{\text{between occupations}} + \underbrace{Var\left(s_j - \widehat{s}_j\right)}_{\text{within occupations}} + \underbrace{2Cov\left(\left(\widehat{s}_j - \overline{s}\right), \left(s_j - \widehat{s}_j\right)\right)}_{\text{covariance}}$$

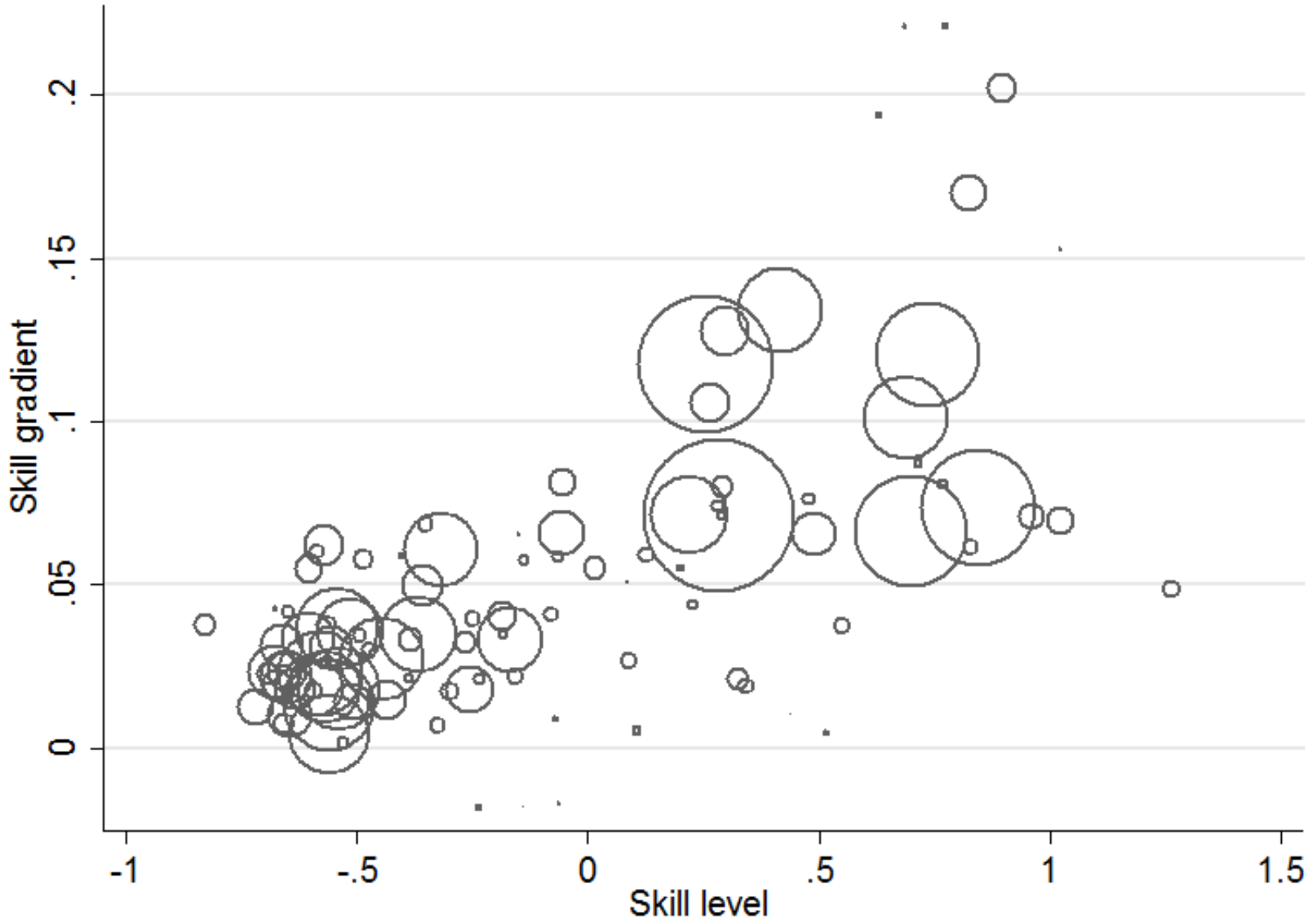
Sorting into occupations

- Consider an economy with two non-substitutable occupations, engineers (E) and mechanics (M)
- Workers differ according to a unidimensional measure of skill, s.
- Assume the return to skill, β, is higher for engineers than for mechanics, i.e., $\beta_E > \beta_M$
- In this economy, there is a cutoff $\widetilde{s}$ such that workers for whom $s > \widetilde{s}$ become engineers and those for whom $s < \widetilde{s}$ become mechanics. We thus have $\overline{s}_E > \overline{s}_M$.

Preliminaries

- A basic prediction of this models is that there should be a positive correlation between the average skill in an occupation $(\overline{s}_h)$ and the occupation-specific return to skill (β_h)
- We test this by regressing log wages on s separately for all 80 occupations in our data

Figure 1. The Return to Skill By Occupation



Decomposing the between-firm variance

- Let share of engineers in firm j be denoited by α_j. The predicted average skill level of firm j based on its occupational structure is thus $\widehat{s}_j = \alpha_j \overline{s}_E + (1 - \alpha_j)\, \overline{s}_M$.
- We use $\widehat{s}_j$ to decompose the between-firm variance of skill

$$\underbrace{Var\left(\widehat{s}_j - \overline{s}\right)}_{\text{between occupations}} + \underbrace{Var\left(s_j - \widehat{s}_j\right)}_{\text{within occupations}} + \underbrace{2Cov\left(\left(\widehat{s}_j - \overline{s}\right), \left(s_j - \widehat{s}_j\right)\right)}_{\text{covariance}}$$

- The between-occupation component reflects the extent to which firms differ in the shares of engineers (α_j) in their workforce
- The within-occupation component reflects the extent to which the best engineers and the best mechanics work in the same firms
- The covariance term is positive if the best engineers and mechanics work in firms with a high fraction of engineers

Sorting into firms

- Case 1: α_j differ across firms. No complementarities.
- The firm-level predicted average skill level is $\widehat{s}_j = \alpha_j \overline{s}_E + (1 - \alpha_j)\, \overline{s}_M$. We have

$$Var\left(\widehat{s}_j - \overline{s}\right) > 0$$

due to variation in occupational structure (α_j). Also,

$$Var\left(s_j - \widehat{s}_j\right) > 0,$$

since workers are randomized within occupations, whereas

$$Cov\left(\left(\widehat{s}_j - \overline{s}\right), \left(s_j - \widehat{s}_j\right)\right) = 0$$

in expectation.

Sorting into firms

- Case 2: α_j is the same across firms, but there is a positive complementarity in worker skills
- We have

$$Var\left(\widehat{s}_j - \overline{s}\right) = 0 \text{ and } Cov\left(\left(\widehat{s}_j - \overline{s}\right), \left(s_j - \widehat{s}_j\right)\right) = 0$$

since $\widehat{s}_j = \alpha_j \overline{s}_E + (1 - \alpha_j)\, \overline{s}_M = \overline{s}$ is the same for all firms by construction. However,

$$Var\left(s_j - \widehat{s}_j\right) > 0$$

and larger than predicted by random allocation of workers within each occupation to firms.

Sorting into firms

- Depending on the type of complementarities, several types of assortative matching within occupations are possible:
  - The best engineers may work together
  - The best mechanics may work together
  - The best engineers may work with the best mechanics (as in the knowledge-hierarchy model by Garicano and Rossi-Hansberg)

Decomposing the between-firm variance

	Actual	Simulated	Predicted	
Between-firm variance	0.201	0.139		
between occupations	0.127	0.127	0	+
within occupations	0.037	0.012	+	+
2*covariance	0.037	0.000	0	+
Occupational structure		X		X
Complementarities			X	X

Simulations based on random assignment of workers within occupations to firms, holding firm occupational structure fixed.

Decomposing the sub-components

- We now turn to decomposing each of the three sub-components of the between-firm variance in skill
- In doing so, we divide workers into five broad occupational groups based on their qualification level. This division is based on Statistics Sweden's classification of occupations, i.e., not our skill measure, s_i. We also consider a finer division into ten groups, also based on the type of work.

Qualification level	Share	$\overline{s}_h$
Managers	0.096	0.60
High	0.171	0.74
Medium	0.233	0.27
Low	0.467	-0.49
Unqualified	0.033	-0.63

Decomposing the between-occupation component

Qualifications	**Type**	**Share**	$\Delta \left[\frac{1}{N} \sum_j N_j \left(\widehat{s}_j - \overline{s} \right)^2 \right]$
Managers		0.097	0.024
High	Technical	0.123	-0.032
	Business	0.048	-0.004
Medium	Technical	0.137	0.028
	Sales	0.096	0.017
Low	Support	0.046	0.006
	Services	0.046	0.005
	Craftsmen	0.153	-0.001
	Operators	0.221	-0.014
Unqualified		0.033	-0.001

Decomposing the between-occupation component

Qualifications	**Type**	**Share**	$\Delta \left[\frac{1}{N} \sum_j N_j \left(\widehat{s}_j - \overline{s} \right)^2 \right]$
Managers		0.097	0.024
High	**Technical**	**0.123**	**-0.032**
	Business	0.048	-0.004
Medium	Technical	0.137	0.028
	Sales	0.096	0.017
Low	Support	0.046	0.006
	Services	0.046	0.005
	Craftsmen	0.153	-0.001
	Operators	0.221	-0.014
Unqualified		0.033	-0.001

Decomposing the within-occupation component

By worker qualification level

	Managers	High	Medium	Low	Unqualified
Managers	0.158				
	(0.048)				
High	0.035	0.063			
		(0.013)			
Medium	0.046	0.024	0.069		
			(0.013)		
Low	0.012	0.005	0.008	0.017	
				(0.004)	
Unqualified	0.014	0.002	0.011	0.003	0.025
					(0.009)

Note: We scale up small occupational groups in order to get a comparable measure of the degree of sorting.

Decomposing the within-occupation component

	M	HQ1	HQ2	MQ1	MQ2	LQ1	LQ2	LQ3	LQ4
Managers	.158								
HQ1: Technical	.033	.064							
HQ2: Business	.030	.023	.089						
MQ1: Technical	.033	.018	.019	.065					
MQ2: Sales	.056	.023	.019	.025	.100				
LQ1: Support	.007	.007	.000	.010	.014	.042			
LQ2: Services	.012	-.001	.004	.004	.010	.006	.033		
LQ3: Craftsmen	.007	.003	.002	.002	.004	.000	.001	.022	
LQ4: Operators	.014	.009	-.001	.006	.005	.006	.001	.004	.019

Decomposing the covariance

Qualifications	$\frac{1}{N_1}\sum_j N_{j,1}\left(2\overline{\varepsilon}_{1j}\left(\widehat{s}_j-\overline{s}\right)\right)$
Managers	0.124
High	0.034
Medium	0.060
Low	0.008
Unqualified	0.003

Decomposing the covariance

Qualifications	**Type**	$\frac{1}{N_1}\sum_j N_{j,1}\left(2\overline{\varepsilon}_{1j}\left(\widehat{s}_j-\overline{s}\right)\right)$
High	Technical	0.023
	Business	0.061
Medium	Technical	0.048
	Sales	0.076
Low	Support	0.005
	Services	0.004
	Craftsmen	0.005
	Operators	0.012

Summary

- Occupational structure can explain 70% of the between-firm variance in skill
  - Key role for high-qualified technical occupations
    - Broadly in line with theories of that emphasize the role of technology for sorting (Acemoglu 1999; Caselli 1999)
- Positive assortative matching between and within managers and workers in high- and medium qualified occupations
- Close to zero assortative matching between and within workers in low- and unqualified occupations

Caveats

- Focus on *firms* not *teams*
  - Most assortative matching in the economy might take place within rather than between firms
    - The relative importance of assortative matching increases slightly when restricting the sample to firms with $<$150 employees
- Occupational data are imperfect
  - Basic assumption that occupation captures "task" and not "skill level"
  - Coarse coding of occupations + missclassifications