Sorting Workers Between and Within Industries

John M. Abowd (Cornell, US Census, CREST, and NBER)

Francis Kramarz (CREST-INSEE,CEPR, IZA)

Paul Lengermann (US Census)

Sébastien Pérez-Duarte (CREST-INSEE)

Outline of the paper

- Estimate a wage equation with both firm specific and worker specific effects, for France and the US
- Correlate these two effects within and between industries
- Observe the differences both within and between countries
- Posit a model of wages and worker allocation, simple yet sufficiently flexible to accommodate for patterns in the correlation of the effects
- Estimate this model and ponder

Some separate strands of the literature

- Inter-industry wage differentials
  - ✓Krueger and Summers (1987,1988: firms matter),Murphy and Topel (1987: persons matter), Dickens and Katz (1987), Gibbons and Katz (1987), Brown and Medoff (1991) firm size
- Optimal sorting
  - ✓Becker (1973 marriage), Shimer and Smith (2000 with frictions), Shimer (2002)

Basic Statistical Model

$$y_{it} - \mu_y = \theta_i + \psi_{J(i,t)} + (x_{it} - \mu_x)\beta + \varepsilon_{it}$$

- The dependent variable is compensation
  - ✓ France: employee + employer cost of employment
  - ✓ United States: employer reported gross employee earnings
- The function J(*i,t*) indicates the employer of *i* at date *t*.
- The first component is the person effect.
- The second component is the firm effect.
- The third component is the time-varying measured characteristics effect.
- The fourth component is the statistical residual, orthogonal to all other effects in the model.

Summary of Data Sources

- French Employer Payroll Reports
  - ✓ 1/25 sample of French workforce
  - ✓ Individual and employing firm identified
  - ✓ annual 1976-1996
  - ✓ 1.5 million individuals, more than 13 million observations used
- United States UI Records
  - ✓ Sample of 9 States employment (California, Florida, Illinois, Maryland, Minnesota, New Jersey, North Carolina, Pennsylvania,Texas) from the Census LEHD program
  - ✓ Individual and taxable employing entity identified
  - ✓ 200 million observations, 46 million individuals and 2 million firms

Estimation

- Identification of both firm and person effects possible by job mobility, through a *connected* group of workers and firms
- Exact least squares solution, see Abowd, Creecy, and Kramarz (2002) for France and for the US

Simple textbook model

- Positive assortive mating (Becker 1973)
- Say $Y_{ij}=f(q_i,q_j)$ output of the match, if Y_{ij} supermodular then q_i positively related to q_j
- This is the case if $f(q_i,q_j) = A(q_i)^{\boldsymbol{a}} (q_j)^{\boldsymbol{b}}$
- If rent sharing rule,

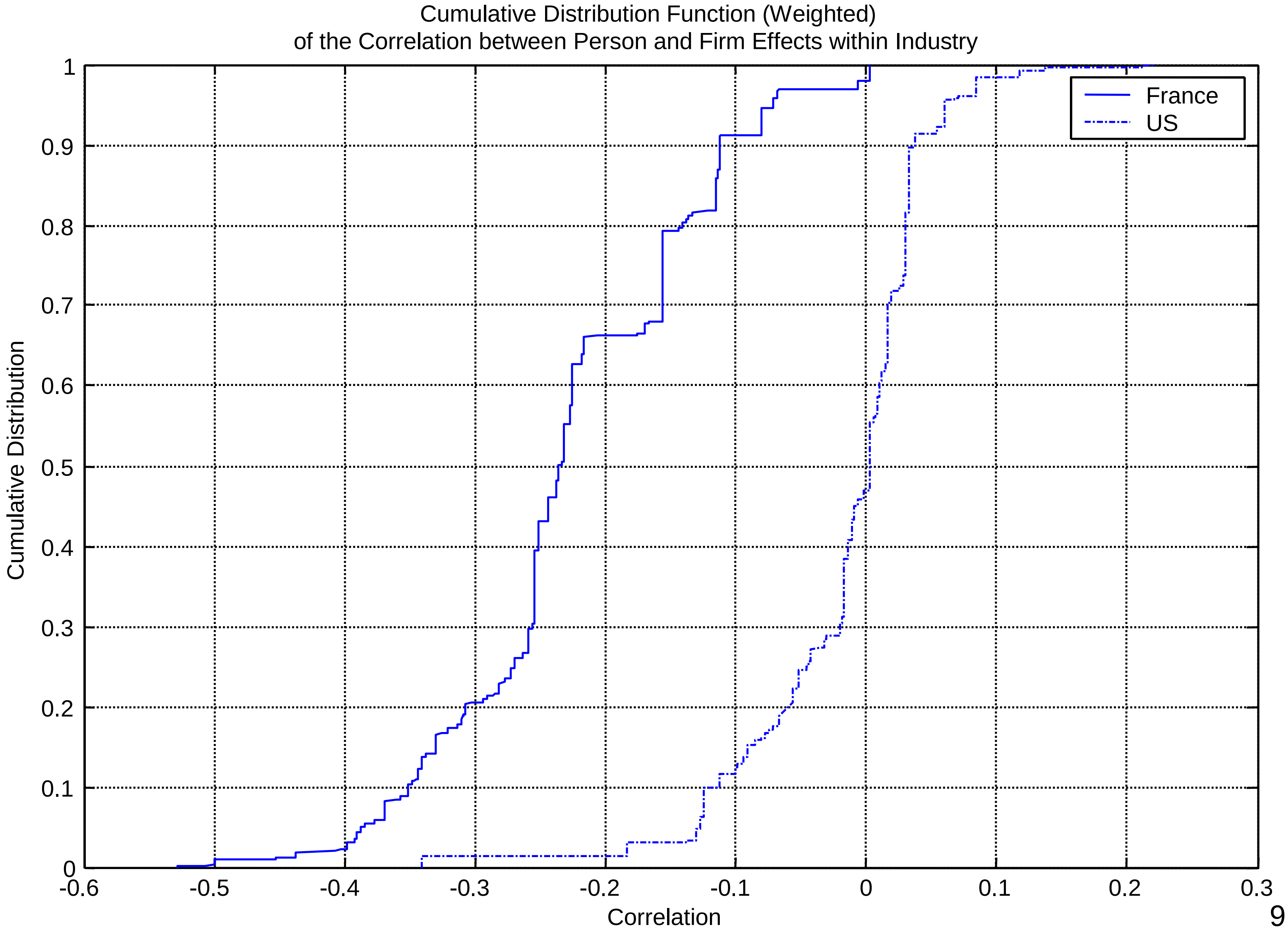
$$\log w_{ij} = \text{Constant} + \boldsymbol{a} \log q_i + \boldsymbol{b} \log q_j$$

A Puzzle

- Correlation pattern between and within industries

Table 1					
Decomposition of Person and Firm Effects					
		Variance of Person Effect	Variance of Firm Effect	Covariance of Person and Firm Effects	Correlation of Person and Firm Effects
France	*Total*	0.2490	0.2176	-0.0562	-0.2415
	Within	0.2450	0.2029	-0.0600	**-0.2690**
	Between	0.0040	0.0147	0.0038	**0.4893**
US	*Total*	0.6292	0.1304	0.0058	0.0203
	Within	0.6149	0.0945	0.0006	**0.0025**
	Between	0.0143	0.0359	0.0052	**0.2293**

Distribution of Within Industry Correlations



Within Industry Correlations, France

France, 2-digit industries	Correlation
Miscellaneous repair services	-0.5289
Bakery products	-0.5077
Lumber mills	-0.5006
Furniture and fixtures manufacture	-0.4992
Automobile dealers, auto parts and repair	-0.4787
Leather products except footwear	-0.4527
Construction	-0.4402
Waste product management	-0.4368
Retail specialty non food trade	-0.4092
Tobacco products manufacture	-0.0805
Gas Distribution	-0.0802
Health Services	-0.0799
Office and accounting machines	-0.0713
Iron and steel foundries	-0.0685
Advertising and consulting services	-0.0668
Telecommunications and postal	-0.0601
Coal mining	-0.0061
Mining and Quarrying of Non-Metallic Minerals	0.0025
Railroad transportation	0.0029

Within Industry Correlations, US

US 2-digit SIC Industries	Correlation
Private households	-0.3401
Pipelines, except natural gas	-0.2295
Miscellaneous services	-0.1838
Environmental quality and housing	-0.1723
Communication	-0.1574
Bituminous coal mining	-0.1519
Agriculture-livestock	-0.1384
Miscellaneous repair	-0.1351
Trucking and warehousing	-0.1339
Local and interurban passenger transport	0.0683
Holding and other investments	0.0704
Motion pictures	0.0851
Agricultural services	0.1171
Tobacco products	0.1268
Apparel	0.1371
Electric and electronic equipment	0.1698
Business services	0.2108
Railroad transport	0.2226

Comparison of Correlation Patterns France—US

Correlation across comparable industries in France and in the US

French Weights

	Corr US	Corr FR
Corr US	1	0.658
Corr FR		1

US Weights

	Corr US	Corr FR
Corr US	1	0.540
Corr FR		1

48 groups of industries

Correlation across comparable industries x size cat. in France and in the US

French Weights

	Corr US	Corr FR
Corr US	1	0.448
Corr FR		1

US Weights

	Corr US	Corr FR
Corr US	1	0.469
Corr FR		1

450 groups in the US (industry x size category), 405 in France

Understanding the Negative Correlation

- “Negative sorting” (France) or “No sorting” (US) within industries ? Contradicts standard Becker-type assignment model (assortive mating)
- Possible explanation in a model with heterogeneous firms and two types of contracts, with self selection of workers across firms
- Another possible explanation: bias in the estimated correlation of person and firm effects

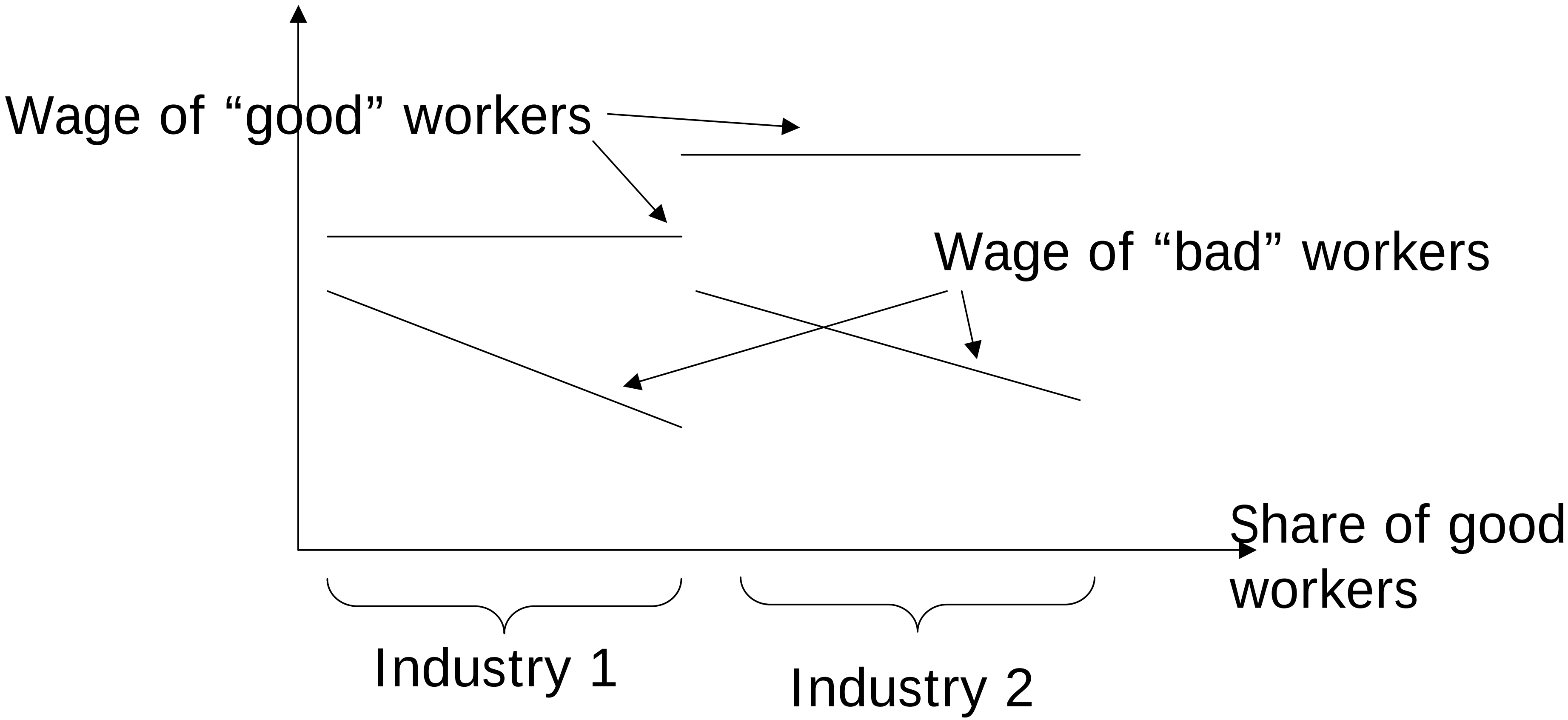
The Theoretical Model

- Two types of workers (good and bad e.g.), multiple sectors, multiple firms per sector (continuum)
- Firms are indexed by *x,* proportion of "good" jobs they must fill (say, because of structure of capital in firm), and *t,* size of firm (exogenous)
- The distribution of *(x,t)* is given in the economy and can be different across industries
- Firms post wage profiles for each job type, and workers apply to a particular job type in each firm
- Firms prefer good workers (low cost of effort) to be in type 1 jobs, and bad workers to be in type 2 jobs
- Fixed wages will depend on *t* (monitoring costs) and *x* (bargaining power)

A Simpler Model

- Assume that for a given t there is a unique x, and such that x is increasing in t : then we can index firms by x only.
- If sectors are ranked along the x dimension, if good workers have higher wages in high x firms, if bad workers receive higher pay in the low-x firm than the high-x firm of the same sector, then:
  - ✓ Total correlation of the estimated person and firm effects can be close to zero
  - ✓ Inter-industry correlation is positive
  - ✓ Within-industry correlation is negative

A Simpler Model (2)



Induced Covariance Decomposition
Structural Model

- The statistical model to be estimated
  - ✓ The distribution of (x,t) is

    $G_{\mathrm{k}}(x,t) = \mathrm{Beta}_{\mathrm{k}}(x|t)\ \mathrm{LogNormal}_{\mathrm{k}}(t)$

    Beta_k is $\mathrm{Beta}(p_0+p_1 t, q_0+q_1 t)$,

    $\mathrm{LogNormal}_k$ is the observed distribution of sizes
  - ✓ Good workers are paid $w_{1xt} = w_{1k}$ specific to each industry k
  - ✓ Bad workers are paid

    $w_{2xt} = u_{0k} + u_{1k}\,x + u_{2k}\log(\mathrm{t}) + u_{3k}\,x\log(\mathrm{t})$

What is Sorting in this Model?

- Between contracts: type 1 (resp. type 2) workers self-select to incentive contracts (resp. efficiency wage or rent-sharing contracts)
- Between sectors: through the fraction of type 1 jobs. Question: how are w_1 and w_2 related with the proportion of type 1 jobs?
- Within sectors: how do the distribution of type 2 jobs and the wages of type 2 jobs vary with firm size?

Worker and Firm Effects Formulas

- Worker effects

$$\begin{cases} \hat{\theta}_1 = \dfrac{\sum_k \int x(1-x)tw_{1k}\, dG_k(x,t)}{\sum_k \int x(1-x)t\, dG_k(x,t)} \\ \hat{\theta}_2 = \dfrac{\sum_k \int x(1-x)tw_{2k}(x,t)\, dG_k(x,t)}{\sum_k \int x(1-x)t\, dG_k(x,t)} \end{cases}$$

- Firm effects $\hat{\psi}_k(x,t) = x(w_{1k} - \hat{\theta}_1) + (1-x)(w_{2k}(x,t) - \hat{\theta}_2)$
- Industry effects for instance

$$E(\hat{\theta}|k) = \frac{\int_k (xt\hat{\theta}_1 + (1-x)t\hat{\theta}_2)\, dG(x,t)}{\int_k t\, dG(x,t)}$$

- Variance and covariance ...

Estimation Procedure

- Data available: moments by industry and size of the firms of θ and ψ (variance and covariance): 95 in France (resp. 81 in the US) industries, 10 size categories, 893x3 (resp. 825x3) usable moments.
- Fit by ALS on least squares criterion

$$\sum_{k,t} \text{weight}_{kt} \times \left(\left(V_{kt}\theta - V_{kt}\hat{\theta} \right)^2 + \left(V_{kt}\varphi - V_{kt}\hat{\varphi} \right)^2 + \left(\text{cov}_{kt}(\theta, \varphi) - \text{cov}_{kt}(\hat{\theta}, \hat{\varphi}) \right)^2 \right)$$

Estimation Procedure (2)

- Number of parameters: 855 (resp. 729), non separable
- Each function evaluation (for the French case)
  - ✓ 14250 numerical integrations.
  - ✓ One iteration in Gauss-Newton: 12,198,000 n.i.
  - ✓ Parallel implementation of Matlab (CMTM) with 6 processors.
- One week for convergence, 6 months to understand the results...

Distribution of Moments and Estimated Parameters across Industries --- France

	5	10	25	50	75	90	95
var(worker)	0.1209	0.1236	0.1566	0.2118	0.2871	0.3844	0.4281
var(firm)	0.0111	0.0177	0.0494	0.1151	0.2354	0.3665	0.5509
cov	-0.1338	-0.1062	-0.0762	-0.0440	-0.0186	-0.0060	-0.0018
corr	-0.4722	-0.4066	-0.3465	-0.2627	-0.1698	-0.0805	-0.0672
mu	0.3272	0.4155	0.7635	1.3521	1.8643	2.3382	2.7106
sigma	0.6872	0.8326	1.1512	1.4808	1.7837	2.1407	2.5375
w	-1.9266	-1.2973	-0.6244	-0.2050	0.5863	1.4288	2.1730
u0	-2.1244	-1.4059	-0.5557	0.2422	0.6479	1.4960	2.0511
u1	-2.2799	0.2470	2.3270	4.6181	6.4519	9.1281	10.1577
u2	-1.5278	-1.1286	-0.0944	0.0724	0.1115	0.2338	0.5211
u3	-2.1324	-0.9641	-0.5243	-0.1637	0.3157	1.0318	1.4746
p0	-0.3489	-0.0162	0.2458	0.5714	1.4824	3.7917	4.7016
p1	0.0013	0.0021	0.0048	0.0154	0.1249	0.4670	0.7379
q0	-0.2873	-0.1584	0.0318	1.5044	2.9610	4.5854	6.3167
q1	0.0017	0.0072	0.0362	0.0750	0.1619	0.4642	0.7662
umean	-3.5729	-2.5746	0.4030	1.3404	2.0040	3.2849	4.0211
xmean	0.0666	0.0744	0.1071	0.1600	0.6752	0.7923	0.8225

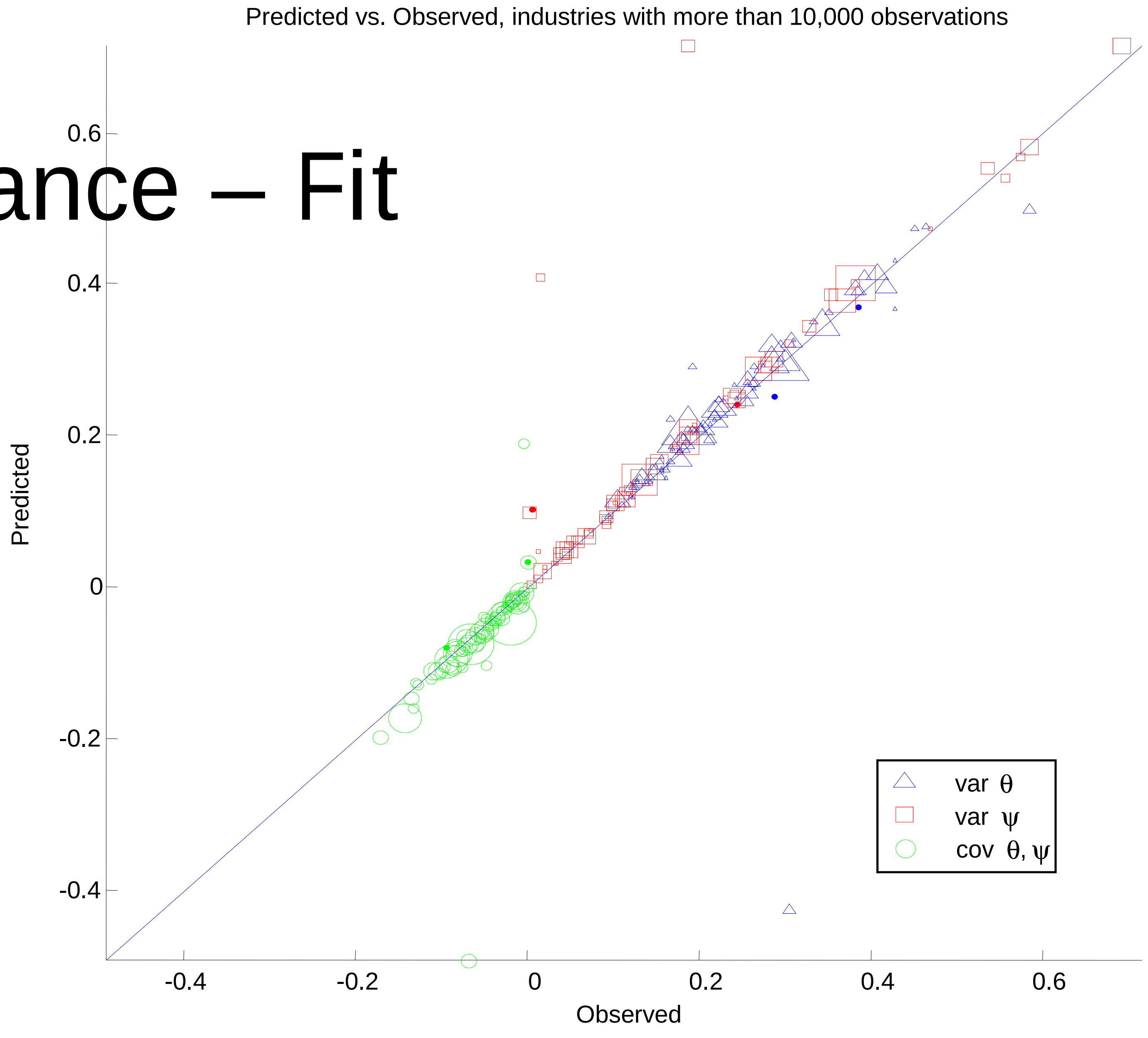
Notes: distribution of parameters across 3-digit industries, 95 observations. var(worker) is variance of worker effect within each industry, var(firm) is variance of firm effect within each industry,
cov is the within-industry covariance of worker and firm effects, corr is the within-industry covariance of worker and firm effects; mu, mean of log wage in industry; sigma, the variance; w, wage of type 1 workers;
u, wage of type 2 workers = u0 + u1*x + u2*log(t) + u3*x*log(t), where x is the share of type 1 workers and t is the size of the firm; Beta(p,q), law of x conditional on t, p = p0 + p1*t, q = q0 + q1*t. umean and xmean are the means of u and x by industry

Distribution of Moments and Estimated Parameters across Industries --- US

	5	10	25	50	75	90	95
var(worker)	0.3550	0.4094	0.4800	0.5735	0.6794	0.7836	0.8167
var(firm)	0.0277	0.0372	0.0502	0.0741	0.1275	0.1803	0.2787
cov	-0.0414	-0.0282	-0.0156	-0.0040	0.0037	0.0184	0.0353
corr	-0.1641	-0.1343	-0.0865	-0.0189	0.0168	0.0691	0.1315
mu	1.0068	1.1365	1.3460	1.7438	2.2785	2.6163	2.7356
sigma	1.0188	1.0672	1.2240	1.4189	1.7225	1.8132	2.0736
w	-4.5149	-4.1178	-2.9114	-0.3029	1.7964	3.3906	6.2562
u0	-5.7070	-3.3737	-1.7297	0.6216	3.0555	4.2577	4.4851
u1	-22.2332	-20.2782	-16.5279	-6.4376	-0.8993	6.9752	17.1130
u2	-2.3100	-1.1792	-0.7068	0.0040	0.7661	2.0524	2.7908
u3	-9.2132	-6.6776	-4.4677	0.0062	1.1365	2.9054	3.3464
p0	-0.0017	0.0512	1.4918	5.6510	14.1194	44.0079	80.9299
p1	-0.0008	0.0000	0.0031	0.2035	1.2667	3.5913	7.2754
q0	-4.1133	-1.2626	0.1969	3.1457	10.0295	60.4979	92.8351
q1	0.0000	0.0003	0.0118	0.1332	4.4323	13.1642	26.5358
umean	-14.6648	-11.8218	-8.0924	-4.2634	-1.3799	3.5764	7.1882
xmean	0.1184	0.1460	0.1874	0.4185	0.7414	0.7818	0.8178

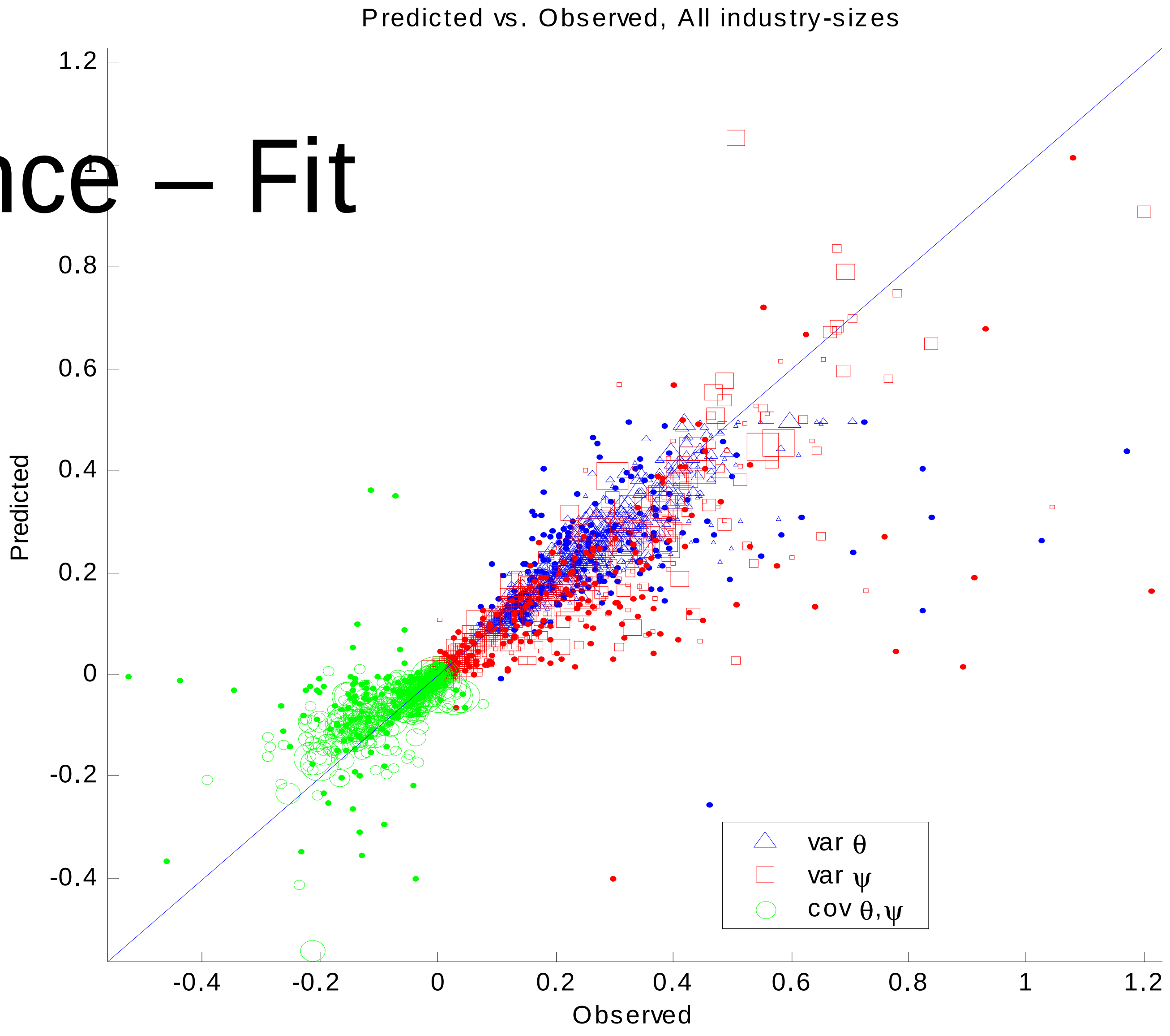
Notes: distribution of parameters across 3-digit industries, 81 observations. var(worker) is variance of worker effect within each industry, var(firm) is variance of firm effect within each industry,
cov is the within-industry covariance of worker and firm effects, corr is the within-industry covariance of worker and firm effects; mu, mean of log wage in industry; sigma, the variance; w, wage of type 1 workers;
u, wage of type 2 workers = u0 + u1*x + u2*log(t) + u3*x*log(t), where x is the share of type 1 workers and t is the size of the firm; Beta(p,q), law of x conditional on t, p = p0 + p1*t, q = q0 + q1*t. umean and xmean are the means of u and x by industry

France – Fit

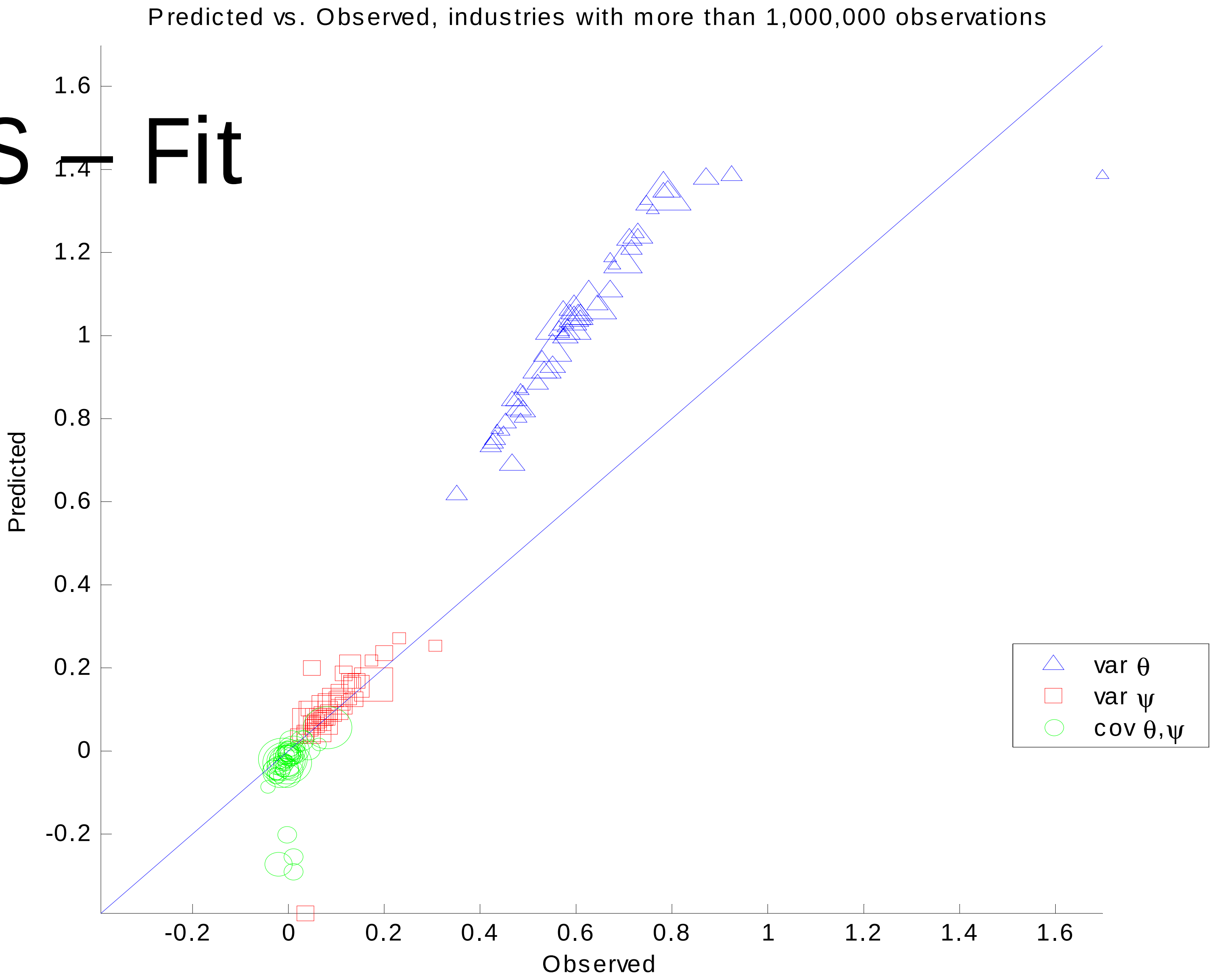


Predicted vs. Observed, industries with more than 10,000 observations
Predicted
Observed
0.6
0.4
0.2
0
-0.2
-0.4
-0.4
-0.2
0
0.2
0.4
0.6
var θ
var ψ
cov θ, ψ

France – Fit


Predicted vs. Observed, All industry-sizes
1.2
0.8
0.6
0.4
0.2
0
-0.2
-0.4
Predicted
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
Observed
var θ
var ψ
cov θ,ψ

US – Fit



Predicted vs. Observed, industries with more than 1,000,000 observations
Predicted
Observed
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
var θ
var ψ
cov θ,ψ

Estimation Results for France
(within industries)

	mean u		mean u		mean x	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
mean x	-18.4550	1.2550	-17.6058	1.3243		
log(size)	-0.2962	0.0325			-0.00192	0.00098
R-square	0.8550		0.8375		0.9672	

Notes: 783 industry-size observations. Each regression includes a full set a industry indicators and is weighted by the number of observations in the cell.

Estimation Results for France
(between industries)

	mean u		mean x		mean w	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
mean x	-4.8167	0.5413			2.1041	0.4976
log(size)	0.0065	0.0600	-0.02000	0.00390		
R-square	0.0954		0.0326		0.1657	

Notes: 783 industry-size observations for the first two regressions and 92 industry observations for the last. Regressions are weighted by the number of observations in the cell.

Estimation Results for US
(within industries)

	mean u		mean u		mean x	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
mean x	8.1279	1.8398	8.5352	1.8891		
log(size)	0.1933	0.0315			0.00062	0.00067
R-square	0.9640		0.9619		0.9866	

Notes: 742 industry-size observations. Each regression includes a full set a industry indicators and is weighted by the number of observations in the cell.

Estimation Results for US
(between industries)

	mean u		**mean x**		**mean w**	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
mean x	-3.6640	1.0293			-5.5831	1.0058
log(size)	0.6994	0.1273	0.00225	0.00455		
R-square	0.0557		0.0003		0.2832	

Notes: 742 industry-size observations for the first two regressions and 80 industry observations for the last. Regressions are weighted by the number of observations in the cell.

Tentative Explanation

- Main parameters explaining within-industry correlation: distribution of the share of good workers in the firm
- Similarly for the between-industry effects
- Possibly sorting for good workers.
- Results (therefore, sorting) appear to be very different in France and in the US
- For instance, the within-industry results support a rent-sharing interpretation in France where bad workers have more bargaining power (through unions) when their fraction in the firm is high

Other Explanation

- The covariance estimate is biased downwards:

$$w_{it} = \alpha_i + \varphi_{J(i,t)} + \varepsilon_{it} = \alpha_i + \sum_{j=1}^{J} \varphi_j \delta_{it}^j + \varepsilon_{it}$$

$$\bar{w}_{it} = w_{it} - w_{i\cdot} = \sum_{j=1}^{J} \varphi_j \left(\delta_{it}^j - \delta_{i\cdot}^j \right) + \varepsilon_{it} - \varepsilon_{i\cdot}$$

$$\begin{aligned} cov(\hat{\alpha}_i, \hat{\varphi}_{J(i,t)}) &= cov(\alpha_i, \varphi_{J(i,t)}) + cov(\hat{\alpha}_i - \alpha_i, \hat{\varphi}_{J(i,t)} - \varphi_{J(i,t)}) \\ &\quad + cov(\hat{\alpha}_i - \alpha_i, \varphi_{J(i,t)}) + cov(\alpha_i, \hat{\varphi}_{J(i,t)} - \varphi_{J(i,t)}) \end{aligned}$$

$$cov(\hat{\alpha}_i - \alpha_i, \hat{\varphi}_{j_0} - \varphi_{j_0}) = -\delta_{i\cdot}^{j_0} var(\hat{\varphi}_{j_0} - \varphi_{j_0}) - \sum_{j \neq j_0} \delta_{i\cdot}^j cov\left(\hat{\varphi}_j - \varphi_j, \hat{\varphi}_{j_0} - \varphi_{j_0} \right)$$

Other Explanation (continued)

- Simulations: Show that the bias is indeed related to mobility
- When mobility is very low (lower than in the data), the bias exists in the direction of the observed covariance (negative)
- When mobility is low (even lower than in the data), the bias seems to disappear. Maybe, the second (positive) part compensates
- More simulations using the data-generated mobility patterns are needed to be sure