Immigrants and Firms'productivity: Evidence from France

Appendix

Cristina Mitaritonna (CEPII)[*]   Gianluca Orefice (CEPII)[†]

Giovanni Peri (University of California, Davis)[‡]

November 3, 2016

---

[*]CEPII, 113 rue de Grenelle 75007 Paris. Email: cristina.mitaritonna@cepii.fr.

[†]CEPII, 113 rue de Grenelle 75007 Paris. Email: gianluca.orefice@cepii.fr. Corresponding Author.

[‡]Department of Economics UC Davis, One shield Avenue, Davis CA 95616. Email: gperi@ucdavis.edu

Appendix A1: Sample description.

A crucial issue to be clarified is the representativeness of the sample with respect to the initial data source. Our main dataset (containing info on the employment structure of French plants) is the DADS dataset, which is exhaustive and includes all the French establishments (plants) in the period 1995-2005. Notice that one firm may appear many times, depending on the number of establishments she holds. In the period 1995-2005, the DADS contains 13.922.675 observations (triplet plant-sector-year), among these 1.546.871 observations belong to the Manufacturing sector,[1] 12.344.653 to the service sector and the remaining to the primary sector. Because of data availability in balance sheet info, we focus on manufacturing sector only. Thus, when we collapse the 1.546.871 manufacturing observations by counting only once all the multi-sector plants (establishments in more than one sub-sector of the industry sector) we end up with 1.534.582 observations (plant-year).

Since the EAE dataset (containing balance sheet data) does not include the agro-industry sector, we have to exclude from the remaining DADS data also those plants belonging to the agro-industry sector.[2] Then the number of plant-year observations reduces to 1.065.076.

After the merge between the DADS, Custom data and EAE datasets, we hold information for 218.895 plant-year in the manufacturing sector,[3] which are those plants belonging to those firms having more than 20 employees (EAE provides info only for firms bigger than 20 employees). Finally the sample reduces to 160.367 observations if we include the TFP variable, which is available only over the period 1996-2005.

The huge drop in the number of plants (from original DADS to our final sample) is not surprising, because, by data availability constraint (EAE), we get rid of the many individual, micro and small enterprises in France. So it becomes important to clarify the representativeness of our final sample of firms in terms of the share over total employment in France (or number of hours worked). In terms of employment (number of employees), our final dataset covers the 64% of the total French employment (in manufacturing sector in the period 1996-2005); while in terms of total hours worked our final sample represents the 66% of the total.

As a final step, since our main dependent variables (TFP, Capital and export variables) are at firm level, we need to collapse plant level information (SIRET id number) at firm level (SIREN id number). Then we end up with 136244 firm-year combinations.

---

[1] According to the Naf 2-digit classification, the manufacturing sector includes activities from code 10 to 34
[2] Code 10 and code 11 according to the Naf 2 classification
[3] Activities from code 12 to 34 of the Naf 2 classification

Appendix A2: Estimating TFP (total factor productivity)

Let's define a Cobb-Douglas production function as follows:

$$y_{jt} = \beta_0 + \beta_l L_{it} + \beta_k K_{it} + \omega_{it} + \epsilon_{it}$$

where y_{it} is the log of output (value added or revenue) of firm i at time t (year in our data). We use value added to proxy output. As we do not observe physical output, we divide the value added by the Producer price index, 1995 prices at the NAF 2 digit level, and then we take the log (from Insee).

L_{it} and K_{it} are the log of inputs - labor and capital, respectively. The average number of employees during the year is used as a proxy for labour. For capital we used the value of tangible assets at the beginning of the period, deflated by the Real fixed capital stock, 1995 prices (from Euklems, http://www.euklems.net/).

ω_{it} represents unobserved (for the econometrician) inputs that are known to the firm when it decides capital and labour. We refer to ω_{it} as Total Factor Productivity (TFP). ϵ_{it} is the error term. Now if ω_{it} affects the choice of inputs, this leads to a simultaneity problem in the estimations of both β_l and β_k, and thus a biased estimation of TFP.

To solve this problem, Olley and Pakes (1996) propose a semiparametric estimation method, derived from a theoretical model, showing the condition under which an investment proxy controls for correlation between input levels and the unobserved productivity shock. Olley and Pakes (1996) propose a firm-level competition model where firms have idiosyncratic efficiencies and face the same market structure and factor prices. Profits are a function of capital K_{it}, efficiency ω_{it}, factor prices and other firms. ω_{it} follows a first-order Markov process:

$$\omega_{it} = E[\omega_{it}|\omega_{i(t-1)}] + \nu_{it} = h(\omega_{i(t-1)} + \nu_{it})$$

where ν_{it} is uncorrelated with K_{it}, but not necessarily with L_{it}. The model compares for each firm, the value of continuing to produce with the value of liquidation. If firm continues in operation, chose labour and investment, knowing current efficiency ω_{it}). Investment choice at time t, I_{it}, gives the capital stock in the next period:

$$K_{i(t+1)} = (1 - \delta)K_{it} + I_{it}$$

which means that time is needed to build physical capital. Investment is chosen at time t, but it is not productive until period $t+1$. The solution of the model generates two firm decision rules. First, the firm stops producing when its efficiency level falls below a given threshold (which increases monotonically with the capital stock). Second, if the firm does not exit, investment is a function of of current state variables.

$$I_{it} = I_t(K_{it}, \omega_{it}),$$

Assuming monotonicity in the function $I_t(.)$ we can invert and obtain the unobservable productivity as a function of two observed inputs, capital and investment:

$$\omega_{it} = g_t(I_{it}, K_{it}).$$

We can re-express the the Cobb-Douglas production function in logs, in the value added case, as:

$$y_{it} = \beta_l L_{it} + \phi_{it}(K_{it}, I_{it}) + \epsilon_{it}$$

where

$$\phi_{it}(K_{it}, I_{it}) = \beta_0 + \beta_k K_{it} + g_t(I_{it}, K_{it})$$

and

$$E(\epsilon_{it}|L_{it}, K_{it}, I_{it}) = 0$$

The first stage of the Olley and Pakes routine substitutes a third-order polynomial approximation in k_{it} and i_{it} in place of ϕ_{it} and estimates β_l. In the second stage the coefficient β_k is identified as follows. Estimated values for $\widehat{\phi}_{it}$ are computed as

$$\widehat{\phi}_{it} = \widehat{y}_{it} - \widehat{\beta}_l L_{it}.$$

For a candidate value β_k^* we obtain a prediction (upon a constant) of $\widehat{\omega}_{it}$ where

$$\widehat{\omega}_{it} = \widehat{\phi}_{it} - \beta_k^* K_{it}.$$

Assuming that productivity follows a first-order Markov process, $E[\omega_{it}|\omega_{i(t-1)}]$ is given by predicted values from regression:

$$\widehat{\omega}_{it} = \gamma_0 + \gamma_1 \widehat{\omega}_{i(t-1)} + \gamma_2 \widehat{\omega}_{i(t-2)}^2 + \gamma_3 \widehat{\omega}_{i(t-3)}^3 + \epsilon_{it}$$

to which we can refer to $\widehat{E}[\omega_{it}|\omega_{i(t-1)}]$. The estimate of β_k is defined as a solution to the minimization of:

$$min_{\beta_k^*} \sum (y_{it} - \widehat{\beta}_l L_{it} - \beta_k^* K_{it} - \widehat{E}[\omega_{it}|\omega_{i(t-1)}])^2$$

Finally using $\widehat{\beta}_l$ and $\widehat{\beta}$, TFP is estimated as a residual of the Cobb-Douglas production function.

Online Appendix Tables

Table A1

Share of immigrants across French regions in 1995, 2005 and average growth rate over the period 2005-1995

	Share of immigrant over native workers in 1995	*Share of immigrant over native workers in 2005*	*Average yearly growth rate in region immigrant share (2005-1995)*
Region			
Île-de-France	0.099	0.217	0.073
Champagne-Ardenne	0.078	0.077	-0.012
Picardie	0.057	0.076	0.020
Haute-Normandie	0.055	0.083	0.065
Centre	0.076	0.127	0.047
Basse-Normandie	0.023	0.042	0.131
Bourgogne	0.078	0.128	0.069
Nord - Pas-de-Calais	0.063	0.159	0.249
Lorraine	0.075	0.089	0.010
Alsace	0.093	0.143	0.034
Franche-Comté	0.086	0.123	0.018
Pays de la Loire	0.030	0.095	0.144
Bretagne	0.027	0.110	0.727
Poitou-Charentes	0.032	0.142	0.314
Aquitaine	0.048	0.153	0.211
Midi-Pyrénées	0.044	0.139	0.404
Limousin	0.028	0.061	0.135
Rhône-Alpes	0.096	0.187	0.103
Auvergne	0.072	0.101	0.053
Languedoc-Roussillon	0.040	0.110	0.191
Provence-Alpes-Côte d'Azur	0.058	0.215	0.193
Mean	*0.059*	*0.123*	*0.154*

Source: Authors' calculations on DADS data.

Table A2
Share of migrants by education attainment 1995-2005.

	1995	*2005*
Share of immigrants with high school diploma (or lower)	74.49	58.08
Share of immigrants with more than high school diploma	25.51	41.92
Share of immigrants with more than 3-year university diploma	9.11	13.85

Source: INSEE*, Enquete Annuelle.* We thank Ahmed Tritah and Joachim Jarreau for providing us this aggregated statistics.

Table A3
Immigration in the District and other proxies for Firm's productivity

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var :	Per worker Value Added		TFP Ackerberg Caves and Frazer method		Revenues per worker	
Immigrant share in Department	0.523*** (0.146)	0.197 (0.164)	0.230*** (0.030)	0.195*** (0.029)	1.171*** (0.177)	1.045*** (0.162)
(Immigrant share in Department) x (zero initial immi share)	0.063 (0.199)		-0.031 (0.029)		-0.035 (0.172)	
(Immigrant share in Department) x initial TFP below median)		0.774*** (0.256)		0.056 (0.037)		0.333*** (0.166)
Method of Estimation	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Observations	100927	95250	90633	88111	112599	98252
Joint F-stat (Kleibergen-Paap F statistic)	28.43	27.72	28.56	28.84	27.12	27.44

Note: Columns (1)-(6) always include region-by-period, sector-by-period, firm fixed effects and firm level control variables described in the text. The period considered is 1996-2005. The unit of observation is one firm in one year. Standard errors are clustered at the department level. *, **, *** indicate significance at the 10, 5 and 1% confidence level.

Table A4
Immigration in the District and Firm's extensive and intensive margins of trade

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var :	*Log of number of varieties*		*Log of average export value per destination*		*Log of average export value per product*	
Immigrant share in Department	0.321	-0.067	0.408	-0.112	0.483	0.080
	(0.245)	(0.182)	(0.292)	(0.269)	(0.297)	(0.246)
(Immigrant share in Department) x (zero initial immi share)	0.763***		0.893*		0.821*	
	(0.388)		(0.496)		(0.450)	
(Immigrant share in Department) x initial TFP below median)		1.382***		1.731***		1.486***
		(0.316)		(0.576)		(0.512)
Method of Estimation	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Observations	82225	75882	81705	75419	81705	75419
Joint F-stat (Kleibergen-Paap F statistic)	27.04	28.40	27.09	28.43	27.09	28.43

Note: Columns (1)-(6) always include region-by-period, sector-by-period, firm fixed effects and firm level control variables described in the text. The period considered is 1996-2005. The unit of observation is one firm in one year. Standard errors are clustered at the department level *, **, *** indicate significance at the 10, 5 and 1% confidence level.

Table A5
Immigration in the District and Firm's TFP. Robustness using different definition of period

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Dep. Var :				
				TFP calculated using Olley and Pakes method				
Immigrant share in Department	0.126***	0.083***	-0.028	0.011	0.281***	0.239***	0.115***	0.153***
	(0.028)	(0.030)	(0.029)	(0.032)	(0.038)	(0.035)	(0.026)	(0.030)
(Immigrant share in Department) x (zero initial immi share)		0.147***				0.144***		
		(0.041)				(0.050)		
(Immigrant share in Department) x initial TFP below median			0.364***				0.379***	
			(0.058)				(0.062)	
(Immigrant share in Department) x initial Employment below median				0.294***				0.317***
				(0.048)				(0.049)
Method of Estimation	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Sample				All firms				
Observations	100431	98056	95032	98056	100431	98056	95032	98056
Period Definition	3 years	3 years	3 years	3 years	5 years	5 years	5 years	5 years
First stage coefficients :								
Imputed Immi sh.	0.543***	0.516***	0.531***	0.538***	0.693***	0.657***	0.680***	0.689***
Imputed Immi sh. x Zero initial Immi sh		0.767***				0.789***		
Imputed Immi sh. x TFP below median			0.704***				0.715***	
Imputed Immi sh. x Emplo below median				0.702***				0.707***
F-stat of first stage	46.85				53.34			
Joint F-stat (Kleibergen-Paap F statistic)		23.58	22.82	22.94		26.56	24.06	24.75

Note: Columns (1)-(8) always include region-by-period, sector-by-period, firm fixed effects and firm level control variables described in the text. Standard errors are clustered at the department level. The period considered is 1996-2005. The unit of observation is one firm in one year. *, **, *** indicate significance at the 10, 5 and 1% confidence level.

Table A6

Immigration in the District and Firm's TFP. Robustness checks: (i) period 2000-2005, (ii) balanced panel (firms survived over the period 1996-2005) and (iii) CES labor adjusted TFP

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. Var : *TFP calculated using Olley and Pakes method*						*Dep. Var :* *TFP using the CES aggregation of labor*	
Immigrant share in Department	0.073*** (0.020)	0.031** (0.012)	0.053*** (0.016)	0.085*** (0.024)	0.004 (0.022)	0.035 (0.022)	0.237*** (0.063)	0.179*** (0.048)
(Immigrant share in Department) x (zero initial immi share)	0.056** (0.023)			0.167*** (0.043)				
(Immigrant share in Department) x initial TFP below median)		0.123*** (0.033)			0.305*** (0.051)			
(Immigrant share in Department) x initial Employment below median)			0.088*** (0.022)			0.253*** (0.049)		
Sample	Period 2000-2005			Firms survived over period 1996-2005			All firms	
Method of Estimation	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Observations	72387	70176	72382	60123	58400	60123	63150	98606
F-stat of first stage	27.38	27.31	27.52				55.63	59.36
Joint F-stat (Kleibergen-Paap F statistic)				30.60	30.50	31.01		

Note: Columns (1)-(8) always include region-by-period, sector-by-period, firm fixed effects and firm level control variables described in the text. Standard errors are clustered at the department level. The period considered is 1996-2005. The unit of observation is one firm in one year. *, **, *** indicate significance at the 10, 5 and 1% confidence level. TFP in column (7) has been calculated adjusting the number of migrant workers in the firm by their productivity gap with natives (as wage gap).

Results by TFP and Employment quartile

Dep Var :	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	TFP		Domestic mkt sh		Capital		Export	
Immi sh* (initial TFP ≤ 25th pct)	0.527*** (0.036)		0.860*** (0.183)		1.396*** (0.310)		3.860*** (0.977)	
Immi sh* (initial TFP > 25th pct & ≤ 50th pct)	0.250*** (0.038)		0.527*** (0.192)		0.916*** (0.219)		2.069*** (0.762)	
Immi sh* (initial TFP >50th pct & ≤ 75th pct)	0.045 (0.035)		-0.050 (0.191)		0.637*** (0.244)		0.805** (0.392)	
Immi sh* (initial TFP >75th)	-0.034 (0.027)		-0.209 (0.206)		-0.07 (0.188)		-0.765 (0.508)	
Immi sh* (initial Empl. ≤ 25th pct)		0.467*** (0.086)		1.116*** (0.226)		1.046*** (0.276)		3.321*** (0.760)
Immi sh* (initial Empl. > 25th pct & ≤ 50th pct)		0.251*** (0.047)		0.546*** (0.187)		1.343*** (0.214)		2.429*** (0.744)
Immi sh* (initial Empl. > 50th pct & ≤ 75th pct)		0.100** (0.040)		0.004 (0.178)		0.736*** (0.281)		1.792*** (0.589)
Immi sh* (initial Empl. > 75th)		0.007 (0.030)		-0.248 (0.184)		0.214 (0.174)		-0.663 (0.606)
Observations	98056	98056	111609	111609	111928	111928	81901	81901
Sample	All Firms		All Firms		All Firms		All Firms	

Note: All columns include region-by-period, sector-by-period, firm fixed effects and firm level control variables described in the text. The period considered is 1996-2005. The unit of observation is one firm in one year. The standard errors are clustered at the department level. *, **, *** indicate significance at the 10, 5 and 1% confidence level.

Table A8
Immigration in the District and the outcomes of firms with and without immigrants.

	(1)	(2)	(3)	(4)	(5)
Dep Var :	Per capita Value Added	Domestic mkt share	Export	TFP	Capital
Immigrant share* Firm **with** Immi	0.695*** (0.185)	0.708*** (0.165)	2.193*** (0.771)	0.218*** (0.040)	1.104*** (0.225)
Immigrant share* Firm **without** Immi	0.502 (0.353)	0.054 (0.233)	-0.120 (0.668)	0.167*** (0.052)	0.592*** (0.227)
Method of Estimation	2SLS				
Observations	35513	40246	25251	34388	40182
Sample	Firms with zero immigrants in 1996				

Note: All columns include region-by-period, sector-by-period, firm fixed effects and firm level control variables described in the text. The period considered is 1996-2005. The unit of observation is one firm in one year. The standard errors are clustered at the department level. *, **, *** indicate significance at the 10, 5 and 1% confidence level.

Table A9
Region level estimations using Labor Force Survey data 1996-2005.

Dep Var :	(1) Native workers (ln)	(2) Native residents (ln)	(3) Wage for natives	(4) Wage for natives (conditional)
Immigrant share	2.365 (7.184)	2.514 (3.147)	-1.719 (1.707)	-2.712 (1.844)
Method of Estimation	2SLS			
Observations	147	147	147	147
F-stat of first stage	3.16	3.16	3.16	3.16
Coeff imputed immi sh (IV)	0.398*	0.398*	0.398*	0.398*

Note: All columns include region and year fixed effects. In Column (4) the wage is residualized using demographic characteristics (gender, age, education). The period considered is 1996-2005. The unit of observation is the region in one year. *, **, *** indicate significance at the 10, 5 and 1% confidence level.