Exports and Wage Premia: Evidence from Mexican Employer-Employee Data*

Judith A. Frías

David S. Kaplan

Eric A. Verhoogen

David Alfaro-Serrano

Sept. 2018

Appendix

For Online Publication

A Data Appendix

A.1 IMSS individual-level data

All private Mexican employers are legally required to report wages for their employees to the Mexican social security agency, *Instituto Mexicano del Seguro Social (IMSS)*. Not all employers comply; those that do not are commonly considered to be in the informal sector. The raw IMSS data can thus be considered a census of private, formal-sector establishments and their workforces for 1985-2005. (Most public-sector workers and employees of the state-run oil company are covered by other insurance programs.)

The IMSS data contain information on the daily wage of individuals. The wages are a measure of total compensation, called the *salario base de cotización*, which includes both earnings and benefits, including payments made in cash, bonuses, commissions, room and board, overtime payments, and in-kind benefits. The data are reported as a sequence of spells for each worker, with beginning and end dates. In principle it is possible to recover a wage for every individual for every day of every year. We extracted data for September 30 for each year. At the level of individuals, the data also contain information on age, sex, and state and year of the individual's first registration with IMSS. At the establishment level, the data contain information only on location and industry (using the IMSS's own 4-digit industrial categories, of which there are 276.)

We impose the following criteria in cleaning the data. (1) In its internal records, IMSS classifies wage records by types referred to as *modalidades*. We use only *modalidades* corresponding to permanent workers and for which consistent, reliable wage figures are available.[1] (2) We require that an individual have a positive wage. (3) We require that municipality and industry are reported for establishment. (4) We winsorize wages within year, assigning wages above the 90th percentile to the 90th percentile and wages below the 10th percentile to the 10th percentile, for the reasons discussed in Section 3. (5) If wages for more than one establishment are observed simultaneously for a given individual, we keep only the highest-wage observation. (6) We require that individuals be 14 years or older and 64 years or younger. (7) We require that workers be employed in an establishment in the largest connected set of establishments, as described in Section 2 above.

The total number of workers with wage data in the "raw" IMSS files (i.e. the sample size after step 3 of the cleaning procedure described in the previous paragraph) ranges from approximately 4 million in 1985 to approximately 10 million in 2005. The numbers of individuals in the cleaned data, after step (6) above but before limiting to the largest connected sets, are in Appendix Table A.2. The numbers after limiting to the largest connected sets are in Table 1. Additional details on the IMSS data are available in Castellanos et al. (2004) and Kaplan et al. (2005, 2007).

A.2 EIA plant-level data

The cleaning procedure for the plant-level data from the *Encuesta Industrial Anual (EIA)* [Annual Industrial Survey] is the same as described in Appendix II of Verhoogen (2008), and rather than repeat the entire description we focus here on key points.

For the reasons discussed in Section 3, we focus in this paper on the EIA data from 1993-2003. The sample was drawn in 1993, to include the largest plants in 205 of the 309 6-digit industries (*clases*) in the Mexican industrial classification system, covering 85% of the value of production in each industry. These plants were followed over time, with minimal refreshing of the sample.

Capital stock was constructed using the perpetual-inventory method. Capital was classified into three types: machinery and equipment, land and buildings, and transportation equipment and other fixed assets. Following Olley and Pakes (1996), each type of capital was assumed to evolve according to $K_{jt} = (1 - \delta_j)K_{jt-1} + i_{jt-1}$, where j indexes the three types of capital. Following Levinsohn and Petrin (2003), the depreciation rates, δ_j for machinery and equipment, land and buildings, and transportation equipment

[1] In the IMSS internal classification system, we use *modalidades* 10, 13 and 17. This excludes rural casual laborers, self-employed individuals who are insured through IMSS, employees of rural agricultural cooperatives and credit unions, freelance workers, taxi drivers, domestic workers, miscellaneous public-sector workers insured through IMSS, and a number of smaller categories.

were assumed to be 10%, 5% and 20% respectively. Total capital stock is the sum of the three types of capital. The book value of capital stock in 1993 was taken as the initial value.

The following cleaning procedures were implemented. (1) Plants in multi-plant firms for which complete information was not reported separately by plant were dropped. (2) Plants owned in whole or in part by government entities were dropped. (3) Establishments that appeared to be *maquiladoras*, because they derived more than 95% of their income from exports or subcontracting, were dropped. (4) Variables that changed within a plant by more than a factor of 10 from one year to the next were set to missing. (5) Missing values of variables were imputed following the procedure described in Appendix II of Verhoogen (2008). (6) After imputation, plants with incomplete information on any key variable (employment, hours, wage bill, total costs, domestic sales, total sales, capital stock) were dropped. (7) The key variables listed in the previous point were "winsorized" at the 1st and 99th percentiles, following a suggestion Angrist and Krueger (1999).

We then selected a balanced panel of plants with complete data in all years 1993-2003, which we refer to as the EIA panel. 3,529 plants are included in this balanced panel. We then linked the EIA panel to the IMSS data and collapsed to the period level (period 1 is 1992-1994, period 2 is 1996-1998, period 3 is 2000-2002; see Section 3 for justification), averaging variables within period.[2] We then selected plants with estimated plant and average person components for all three periods. 2,625 plants satisfied this requirement. We refer to this balanced plant-period-level panel as the EIA-IMSS panel.

References

Angrist, Joshua D. and Alan B. Krueger, "Empirical Strategies in Labor Economics," in Orley C. Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Elsevier Science, 1999.

Castellanos, Sara, Rodrigo Garcia-Verdu, and David Kaplan, "Wage Rigidities in Mexico: Evidence from Social Security Records," *Journal of Development Economics*, 2004, *75* (2), 507–33.

Kaplan, David S., Gabriel Martinez Gonzalez, and Raymond Robertson, "What Happens to Wages After Displacement?," *Economía*, 2005, *5* (2), 197–242.

_ , _ , and _ , "Mexican Employment Dynamics: Evidence from Matched Firm-Worker Data," 2007. World Bank Policy Research working paper No. 4433.

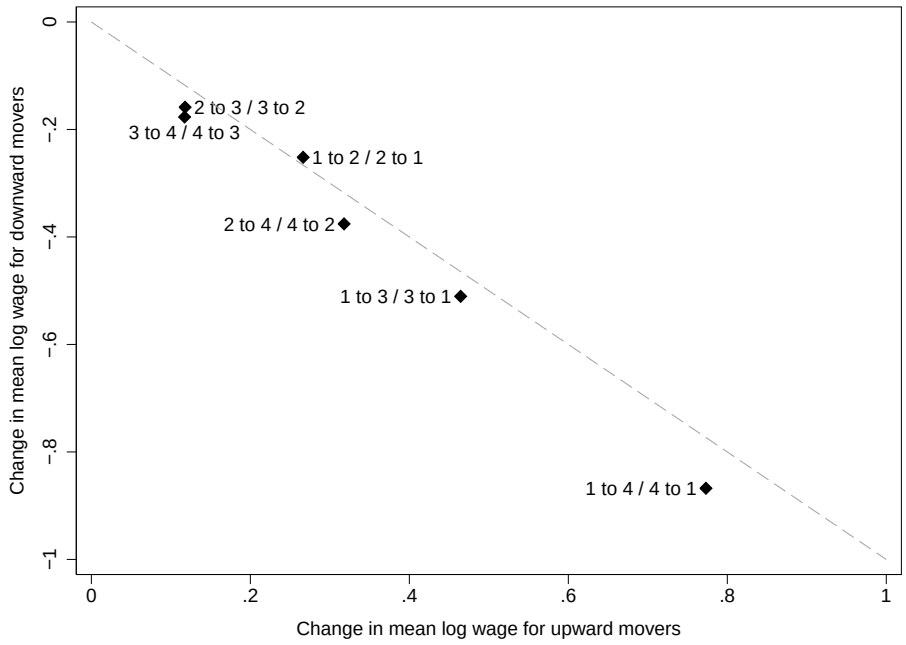
Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, April 2003, *70*, 317–341.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Verhoogen, Eric, "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 2008, *123* (2), 489–530.

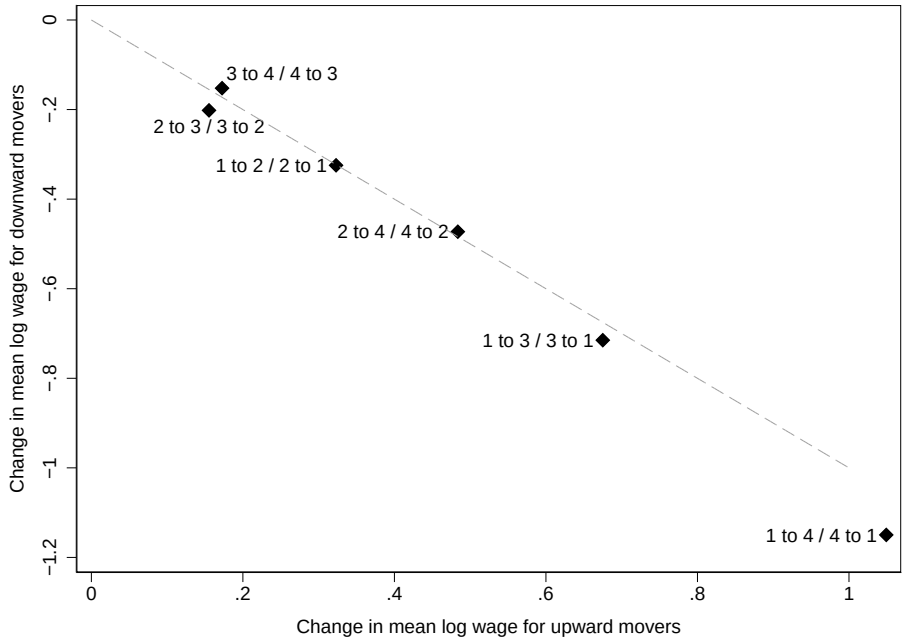
[2]For period 1, we averaged EIA variables for 1993-1994, since the EIA variables are not available in 1992.

Figure A.1. Comparing upward and downward moves, IMSS data, 1992-1995



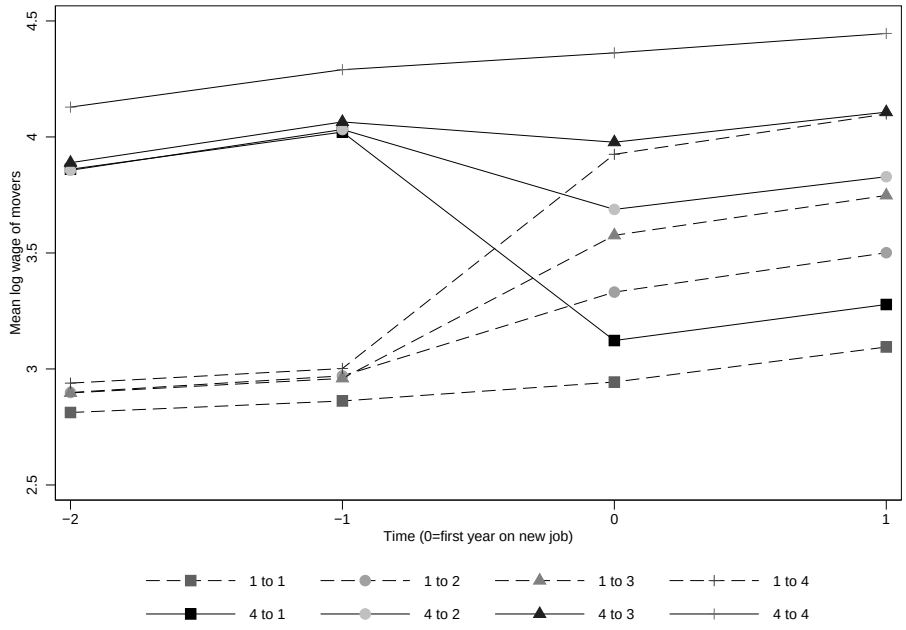
Notes: Sample is all workers observed in 1992-1995 in the IMSS database (after cleaning steps 1-6 described in Appendix A.1) who changed job between 1993 and 1994 and held both the preceding and new job for at least two years. The dashed line is at −45 degrees. Each dot plots upward and downward transitions between two types of firms, classified according to quartiles of average coworkers' wage. Wage changes are changes in log real wage, averaged over workers making same transition, between 1993 and 1994.

Figure A.2. Comparing upward and downward moves, IMSS data, 2000-2003



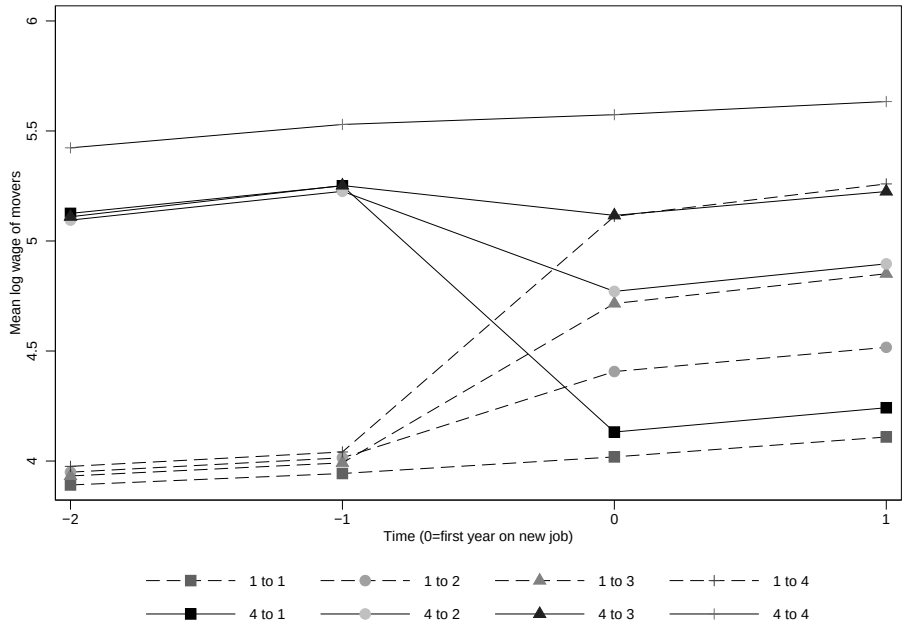
Notes: Sample is all workers observed in 2000-2003 in the IMSS database (after cleaning steps 1-6 described in Appendix A.1) who changed job between 2000 and 2001 and held both the preceding and new job for at least two years. The dashed line is at −45 degrees. Each dot plots upward and downward transitions between two types of firms, classified according to quartiles of average coworkers' wage. Wage changes are changes in log real wage, averaged over workers making same transition, between 2000 and 2001.

Figure A.3. Movers' mean nominal wages, IMSS data, 1992-1995



Notes: Figure is similar to Figure 4 but shows *nominal* wage changes. Sample is all workers observed in 2000-2003 in the IMSS database (after cleaning steps 1-6 described in Appendix A.1) who changed job between 2001 and 2002 and held both the preceding and new job for at least two years. Each line corresponds to a transition between types of firms classified by quartiles of the average coworkers' wage.

Figure A.4. Movers' mean nominal wages, IMSS data, 2000-2003



Notes: Figure is similar to Figure 5 but shows *nominal* wage changes. Sample is all workers observed in 2000-2003 in the IMSS database (after cleaning steps 1-6 described in Appendix A.1) who changed job between 2001 and 2002 and held both the preceding and new job for at least two years. Each line corresponds to a transition between types of firms classified by quartiles of the average coworkers' wage.

Table A.1. Aggregate labor force statistics

	1990	2000
Total population	81.25	97.48
Eonomically active pop. age > 14	31.23	40.16
Remunerated workers	25.96	32.01
Remunerated workers, private sector	21.27	27.20
Workers registered in IMSS	10.76	15.24
Workers registered in IMSS, permanent	9.53	13.53

Notes: Numbers in millions. Figures drawn from *Anuario Estadístico de los Estados Unidos Mexicanos* [Statistical Yearbook of Mexico], 2005, which draws in turn from the following: decennial population censuses (total population), 1991 *Encuesta Nacional de Empleo* [National Employment Survey] (economically active population age > 14), and INEGI Banco de Información Económica (remunerated employees), IMSS *Memoria Estadística*.

Table A.2. Summary statistics, IMSS individual-level data, before limiting to largest connected sets

					avg. daily wage (raw, 2002 pesos)		avg. daily wage (winsorized, 2002 pesos)	
year	# individuals	# establishments	avg. age	fraction male	mean	std. dev.	mean	std .dev.
1988	5,257,200	426,570	31.76	0.72	146.81	702.83	115.61	62.21
1989	5,993,961	469,018	31.29	0.70	151.86	618.46	125.50	73.70
1990	6,869,806	538,274	31.02	0.69	144.82	417.67	123.23	78.16
1991	7,546,628	596,124	31.01	0.68	153.03	235.67	134.68	86.85
1992	7,756,268	621,246	31.10	0.68	161.18	264.05	142.49	94.67
1993	7,659,363	615,684	31.41	0.68	180.30	249.97	152.09	105.07
1994	7,843,005	619,991	31.58	0.67	190.37	259.41	155.34	109.18
1995	7,413,728	600,015	32.01	0.67	152.28	202.47	122.94	87.77
1996	7,998,174	617,721	31.95	0.67	139.70	187.39	111.67	80.87
1997	8,592,365	640,381	31.92	0.67	140.61	194.91	112.75	83.02
1998	9,001,372	653,151	32.03	0.67	142.99	177.21	115.34	85.46
1999	9,578,857	674,710	32.17	0.66	145.03	176.33	117.56	86.67
2000	10,203,195	711,176	32.32	0.65	153.09	181.65	125.09	91.72
2001	10,103,668	736,849	32.85	0.65	160.87	187.25	132.29	97.03
2002	10,151,601	748,620	33.20	0.65	163.62	189.03	134.94	98.13

Notes: Sample is from IMSS employer-employee records after cleaning steps 1-6 in Appendix A.1 (before restricting to largest connected sets). Winsorization is at 10th and 90th percentiles. Wages are reported both in "raw" (i.e. pre-winsorized) and winsorized form. See Section 3 and Appendix A.1 for further details. Average 2002 exchange rate: 9.60 pesos/US$1.

Table A.3. Summary statistics, EIA panel, 1993

	non-exporters (1)	exporters (2)	all plants (3)
Total revenue	151.88	417.78	226.55
	(7.96)	(47.51)	(14.65)
Employment	184.56	370.31	236.78
	(5.66)	(21.78)	(7.48)
K/L	146.63	194.80	160.17
	(4.82)	(8.96)	(4.30)
Export share of sales		0.15	0.04
		(0.01)	(0.00)
Avg. hourly wage (EIA)	43.64	60.98	48.51
	(0.80)	(1.16)	(0.67)
N	2537	992	3529

Notes: Table reports statistics using 1993 data from EIA panel (before linking to IMSS data). Standard errors of means in parentheses. Exporter defined as export sales > 0. Export share is fraction of total sales derived from exports. Sales are measured in millions of 2002 Mexican pesos, capital-labor ratio in thousands of 2002 pesos, and average daily wage in 2002 pesos. Average 2002 exchange rate: 9.60 pesos/US$1. For further details, refer to Section 3 and Appendix A.2.

Table A.4. Number of EIA plants linked to IMSS data, by connected set status

	EIA panel plants	EIA panel plants linked to IMSS			EIA panel plants not linked to IMSS	EIA-IMSS panel plants
		Total	Connected	Not connected		
Period 1 (1992-1994)	3,529	2,769	2,746	23	760	2,625
Period 2 (1996-1998)	3,529	2,903	2,868	35	626	2,625
Period 3 (2000-2002)	3,529	2,872	2,812	60	657	2,625

Notes: Data from IMSS employer-employee records and EIA plant panel as described in Section 3. "Connected" means contained in the largest connected set, as described in Section 2.

Table A.5. OLS estimates, exports and plant-level outcomes

		(1) △ log K/L	(2) △ log avg. hourly wage (EIA)	(3) △ avg. log daily wage (IMSS)	(4) △ plant component	(5) △ avg. person component
A. Not including initial value of capability proxy						
△ export share		0.043	0.042	0.034	0.058**	-0.024
		(0.082)	(0.039)	(0.023)	(0.026)	(0.024)
B. Including initial value of capability proxy						
△ export share		0.001	0.016	0.019	0.046*	-0.028
		(0.082)	(0.038)	(0.023)	(0.026)	(0.024)
initial log employ.		0.072***	0.045***	0.026***	0.021***	0.005
		(0.009)	(0.004)	(0.003)	(0.004)	(0.003)
6-digit industry × period effects	Y	Y	Y	Y	Y	Y
region (state) × period effects	Y	Y	Y	Y	Y	Y
N (plants)		2625	2625	2625	2625	2625
N (obs)		5250	5250	5250	5250	5250

Notes: Table reports OLS regressions corresponding to equation (4) in the main text. Panel A omits the initial value of the capability proxy, $\widehat{\lambda}_{jp-1}$; Panel B includes it. Export share is fraction of total sales derived from exports. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table A.6. Construction of alternative proxies, TFP and predicted export share

	(1) log VA ind. 31	(2) log VA ind. 32	(3) log VA ind. 33	(4) log VA ind. 34	(5) log VA ind. 35	(6) log VA ind. 36	(7) log VA ind. 37	(8) log VA ind. 38	(9) log VA ind. 39	(10) export share
log w.c. empl.	0.176***	0.122***	0.126***	0.365***	0.311***	0.073**	0.223***	0.156***	0.291**	
	(0.024)	(0.036)	(0.038)	(0.034)	(0.031)	(0.028)	(0.061)	(0.018)	(0.135)	
log b.c. empl.	0.259***	0.295***	0.510***	-0.041	0.038	0.261***	0.141	0.157***	0.014	
	(0.030)	(0.047)	(0.102)	(0.028)	(0.036)	(0.053)	(0.090)	(0.032)	(0.247)	
log capital	0.341***	0.054	0.128	0.145**	0.299***	0.514**	0.165	0.128*	0.247	
	(0.055)	(0.050)	(0.109)	(0.059)	(0.052)	(0.227)	(0.100)	(0.068)	(0.185)	
log empl.										0.053***
										(0.003)
log sales										0.040***
										(0.003)
log K/L										0.027***
										(0.002)
N (plants)	611	575	123	312	743	226	81	814	40	3529
N (obs)	6519	6117	1317	3364	8003	2407	870	8769	429	38819

Notes: Columns 1-9 report coefficients from Levinsohn and Petrin (2003) TFP estimation, with log value-added as outcome, log employment (white-collar and blue-collar separately) and log capital as covariates, and log materials and log electricity as proxies, separately by 2-digit industry. The industries (indicated at top) are: food, beverages, tobacco (31); textiles, apparel, leather goods (32); wood products, including wood furniture (33); paper, papers, products, publishing (34); chemical products (35); non-metallic mineral products (36); basic metal products (37); metal products, machinery, equipment (38); other manufacturing (39). Column 10 reports the coefficients of a tobit model of the export share using log employment, log sales, and log capital-labor ratio as covariates, and including 4-digit sector fixed effects. Observations with negative or zero value-added omitted in Columns 1-9. For all columns, if capital or employment variable has value zero, log is set to zero. Standard errors in parentheses. *10% level, **5% level, ***1% level.

Table A.7. Correlations between proxies, EIA-IMSS panel, pooling periods

	log empl. (hours)	log dom. sales	log TFP (L-P)	pred. exp. share
log employment (hours)	1.0000			
log domestic sales	0.8142	1.0000		
log TFP (Levinsohn-Petrin)	0.5578	0.6979	1.0000	
predicted export share	0.8893	0.9306	0.6247	1.0000

Notes: Table reports bilateral correlation coefficients using the EIA-IMSS panel, pooling periods 1 (1992-1994), 2 (1996-1998), and 3 (2000-2002).