Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector

Eric A. Verhoogen

Quarterly Journal of Economics, 123 (2), May 2008

Appendix II: Online Appendix

This appendix describes the processing of the various datasets. For variable definitions, refer to Appendix I. Additional tables and figures referred to in the main text appear at the end of this appendix.

II.A *Encuesta Industrial Anual (EIA)*

The *Encuesta Industrial Anual (EIA)* [Annual Industrial Survey] is carried out yearly, in the spring, with data referring to the previous calendar year. It contains information on employment, hours, wages, expenditures, sales, other revenues, inventories, and capital assets and investment. A companion survey, the *Encuesta Industrial Mensual (EIM)* [Monthly Industrial Survey], is carried out monthly using a less extensive survey at the same set of plants; the EIM is the source of information on wages and employment for the two occupational categories, *obreros* (production, or blue-collar, workers) and *empleados* (non-production, or white-collar, workers). INEGI has two separate datafiles, one covering the 1993-2001 period and an earlier on covering the 1984-1994 period.[1] On the basis of these datasets, I created two balanced panels, which I refer to simply as the EIA 1993-2001 Panel and the EIA 1984-2001 Panel, and one unbalanced panel, which I refer to as the EIA 1993-2001 Unbalanced Panel.

II.A.1 Sampling Design

In 1993, 205 of the 309 6-digit industries (*clases*) in the Mexican industrial classification system were chosen to be included in the EIA.[2] From a list of the universe of manufacturing plants in Mexico (generated in preparation for the upcoming 1994 industrial census), non-*maquiladora* plants within each *clase* were ranked in decreasing order of the value of production (*valor de producción*), the value of the output of the plant priced at the "factory"

---

[1] The earlier dataset is also described in the data appendix of Grether [1996].

[2] The Mexican industrial classification on which the surveys used in this paper are based is the *Clasificacíon Mexicana de Actividades y Productos (CMAP)* [Mexican Classification of Activities and Products]. It is organized in 6-digit industries called *clases* (classes), 4-digit industries called *ramas* (branches), and 2-digit industries called *divisiones* (divisions). The classification was revised in 1993. The initial sample for the 1993-2001 EIA and EIM panels was drawn in 1993 using the updated classification, usually referred to as CMAP-94. (It was published in 1994.) In the manufacturing sector in the CMAP-94, there are 309 *clases*, 54 *ramas*, and 9 *divisiones*.

price (*venta de fabrica*). Plants were then selected in decreasing order of value of output until the set of selected plants made up 85% of the total value of output (not including *maquiladoras*) of the *clase*. This rule was subject to the following qualifications:

1. If a plant employed 100 or more workers, it was added to the sample, regardless of whether the 85% level had been reached.

2. If more than 100 plants were required to cover 85% of output within the *clase*, the number of plants was limited to 100. (There were no cases in which this rule conflicted with (1).)

3. If four or fewer plants made up 85% or more of total output of the *clase*, then to preserve the confidentiality of those plants, *all* plants in the *clase* were included, superseding rule (2).

These criteria generated an initial sample of 7,157 plants in 1993. These plants were subsequently followed over time. Specific analysts in the INEGI offices in Aguascalientes, Mexico, are assigned to follow particular plants over time, and to double-check inconsistencies or sudden changes in the plant, in many cases by calling the establishment on the phone. The same is true for the monthly EIM survey of the same plants. As a result, the quality of the data in the EIA and EIM tends to be better than in surveys with less regular contact between the INEGI analysts and the survey respondents.

A small number of plants were added to the survey after 1993, but they were not added in a systematic way, and I ignore the new establishments. The questionnaire used in the EIA changed over time prior to 1997. In 1997, a consistent format for the questionnaire was settled upon, and has since remained in effect. Variables collected prior to 1997 that are no longer included in the survey have been discarded from the dataset. An exception is foreign ownership, which was collected in 1994 but not thereafter, which has been preserved and can still be linked to the ongoing panel.

An important advantage of the EIA and the EIM is that the analysts tracking each establishment keep track of why some establishments fail to respond or respond in an irregular way. I classified plants that went bankrupt or suspended operations as non-random exiters.[3] I classified plants that switched to industries not covered by the survey, that switched from manufacturing to wholesale or retail sales, that merged with other establishments, or that failed to provide data for an unexplained reason as random exiters. I assume that these random exiters are ignorable in the sense of Griliches [1986].

The 1984-2001 panel used in this paper was formed by linking the 1993-2001 panel with the earlier 1984-1994 panel. The design of the 1984-1994 sample was similar to that of the 1993-2001 sample. In 1984, under an earlier industrial classification system, 129 *clases* were selected for the panel. Establishments within each *clase* were chosen following the same criteria described for the 1993-2001 panel above. The principal difference was that establishments were included within each industry until 80% of the total value of production in the industry was covered. The initial sample consisted of 3,199 plants. The EIA surveys prior to 1992 did not elicit information on exports. Information on exports is available from supplementary surveys funded by the World Bank for the same sample of plants, but

---

[3]The set of plants that suspended operations includes plants that suspended operations due to strikes or to shortages of materials.

only for the years 1986-1990. Data on exports are thus not available for 1984-1985 or 1991. Many plants cannot be linked from the 1984-1994 to the 1993-2001 samples because industry definitions and the set of industries surveyed differ between the two panels. Because the EIA focuses on large plants, however, it is still possible to link a significant number of establishments.

Prior to 2000, the instructions for the EIA requested that respondents report the average number of workers, either permanent or casual, paid or unpaid, who worked in the plant and were under the control of the plant's management. In 2000, the instructions were clarified to specify the inclusion of workers under subcontract for other firms. No information is available on whether the workers are formal, in the sense of being officially registered with the Mexican social security agency, or informal. Anecdotal evidence suggests that many workers not registered with the social security agency are still reported in the EIA, and hence that the EIA employment totals include some informal workers.

II.A.2 Cleaning Procedure

The cleaning procedure for the 1993-2001 data was the following:

1. I removed establishments that report data for more than one establishment or that have their data reported by another establishment. In some multi-establishment firms, survey respondents are unable or unwilling to report information separately for each establishment. In these cases, respondents report joint data for the establishments on a single survey. Only in 1998 did INEGI begin keeping track systematically of the reporting patterns. Establishments that have their information reported elsewhere are easy to identify; their records appears in zeros. Establishments that report information for more other establishments in the same firm are more difficult to identify. I discarded any establishment that had its information reported elsewhere in any year or that reported information from another establishment in 1998 or later.

2. I removed establishments owned in whole or in part by government entities.

3. I removed establishments that appear in any year to be *maquiladoras*. Although the EIA sample was designed to exclude *maquiladoras*, it is possible that some were included by mistake, either because of errors in sampling or because they entered the government's *maquiladora* program subsequent to the drawing of the EIA sample. I classify a plant as a *maquiladora* if either of the following conditions is satisfied:

   (a) The plant derives 95% or more of its sales from exports.

   (b) The plant derives 95% or more of its total revenues[4] from income from sub-contracting services (*ingresos por servicios de maquila*), which are not reported under sales.[5]

4. I removed establishments that exited randomly, by the above definition, in any year during the 1993-2001 period.

---

[4]Total revenues are defined as (total sales + income from subcontracting for other plants - expenditures on subcontracting by other plants), in thousands of 1994 pesos, deflated by producer price index.

[5]While it is possible that the income from subcontracting services is for domestic subcontracting, the more likely explanation is that such income reflects assembly for export.

5. I set to missing the values of key variables that changed by more than a factor of five from one year to the next. This procedure was carried out for: total employment, blue-collar employment, white-collar employment, total hours, blue-collar hours, white-collar hours, real total remuneration, real blue-collar remuneration, real white-collar remuneration, real domestic sales.

6. I imputed new values for missing values using a variant of the sequential regression multivariate imputation technique implemented by Abowd and Woodcock [2001]. The steps were the following:

   (a) I classified variables into five groups:

   Group 1: total employment, total hours, total remuneration, total sales.

   Group 2: materials costs, electricity costs, revenues, capital stock.

   Group 3: blue-collar employment, white-collar employment, blue-collar hours, white-collar hours, blue-collar remuneration, white-collar remuneration.

   Group 4: initial capital stock in machinery and equipment, initial capital stock in land and buildings, initial capital stock in transportation equipment and other assets, all in 1993.

   Group 5: foreign ownership, investment in machinery and equipment, investment in land and buildings, investment in transportation equipment and other assets, export share, imported inputs share.

   (b) For Group 1, I regressed each variable (in the order given) on the other variables in the group, a lead and lag of the same variable, and year dummies, using the plants remaining in the sample after the steps above. (All variables in logs, deflated as described in section II.A.3 below.) The predicted values from this regression were imputed for the missing values.[6]

   (c) For Group 2, I regressed each variable on the other variables in Groups 1-2, a lead and a lag of the same variable, and year dummies, and imputed with the predicted values. For Group 3, I did the same, including variables from Groups 1-3.

   (d) For Group 4, I regressed each variable on the variables in Groups 1-3 and a lead of the same variable, and imputed with the predicted values.

   (e) The Group 5 variables had to be dealt with specially, due to their discreteness or censoring. For foreign ownership, I estimated a probit on Group 1-4 variables in 1994 (the only year during 1993-2001 that foreign ownership is observed), then drew randomly from a Bernoulli distribution with probability of success equal to the predicted probability from the probit. For each of the investment variables, I first imputed whether investment was zero or positive, following the same procedure as for foreign ownership; I then took predicted values from a regression of investment on Group 1-4 variables to replace the missing values. Schafer and Olsen [2000] argue that this "two-part" method is preferable to a tobit model in cases such as this one where zero values are economically meaningful, rather

---

[6]If a lead and/or lag was missing, I used the predicted value from an analogous regression without a lead and/or lag.

than being stand-ins for latent negative values.[7] For export share, I first imputed domestic sales using Group 1-4 variables as in step 6c above, then imputed export sales using the same method as for investment in this step, then calculated export share and used it to rescale domestic and export sales to sum to total sales from step 6b. For imported inputs share, I followed the same procedure as for export share.

7. I re-checked the imputed values from the previous step to see whether values changed by more than a factor of five from one year to the next, and, if so, set them again to missing.

8. I removed establishments that were missing data on key variables after steps 1-7. Establishments classified as non-exiters were removed if they were missing data in any year over the 1993-2001 period. Establishments classified as non-random exiters were removed if they were missing data in any year prior to the year of exit. The key variables were blue-collar employment, white-collar employment, total employment, blue-collar hours, white-collar hours, total hours, blue-collar remuneration, white-collar remuneration, total remuneration, domestic sales, export sales total sales, the investment variables, the initial capital stock variables, foreign ownership, geographical location (state), materials costs, and electricity costs.

Appendix Table V summarizes the numbers of plants lost in each step. In the end I am left with a balanced panel of 3,263 plants that have complete data over the entire 1993-2001 period, and an unbalanced panel that includes 1,139 additional plants that went out of business during the period.

Following a suggestion of Angrist and Krueger [1999], I "winsorized" the key variables at the tails, replacing values in the lower or upper 1% tails with values at the 1st or 99th percentiles, respectively, within each year. The following variables were winsorized: total employment, blue-collar employment, white-collar employment, total hours, blue-collar hours, white-collar hours, real average wage, real blue-collar wage, real white-collar wage, white-collar/blue-collar wage ratio, real domestic sales, real revenues, real capital-labor ratio.

The cleaning procedure for the 1984-1994 EIA dataset was similar to that for the 1993-2001 dataset. The 1984-1994 dataset indicates neither whether multi-establishment firms have consolidated information under one establishment nor whether plants have public ownership, so steps 1 and 2 were skipped. Also, the 1984-1994 EIA dataset does not contain information on the reason for exit, so step 4 was skipped as well. I took the information from 1984-1992 from the 1984-1994 dataset, and information for 1993-2001 from the 1993-2001 dataset.[8]

The facts that we do not know the reason for exit in 1984-1994 and that the sampling design of the EIA changed for the 1993-2001 dataset would make the analysis of an unbalanced 1984-2001 panel difficult. Instead, I focus on the balanced 1984-2001 panel. After

---

[7]Note that the two-part method is equivalent a Heckman [1976] selection model with the correlation between the errors in the two stages set to zero.

[8]As mentioned above, the 1984-1994 dataset does not include information on export vs. domestic sales or imported vs. domestic inputs in 1984, 1985 or 1991. These variables were not included in the cleaning procedure for those years.

the cleaning procedure, there are 1,114 establishments that appear with complete data over the entire 1984-2001 period.

II.A.3 Construction of Capital Stock Variable

Capital stock was constructed using the perpetual-inventory method. I classified capital 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 were assumed to be 10%, 5% and 20% respectively. Total capital stock is the sum of the three types of capital. I took the book value of capital stock in the initial year (1984 or 1993 depending on the panel) as the initial value. Both the initial book value and subsequent investments are measured in thousands of 1994 pesos, deflated by the producer price index.[9]

II.B *Encuesta Nacional de Empleo, Salarios, Tecnología y Capacitacíon (ENESTyC)*

The *Encuesta Nacional de Empleo, Salarios, Tecnología y Capacitacíon* (ENESTyC) [National Survey of Employment, Wages, Technology and Training] is a special survey that includes qualitative questions regarding training, turnover, technology use and a variety of workplace practices. It was carried out in 1992, 1995, 1999, and 2001, with many of the questions referring to the previous calendar year. The survey covered 5,071 establishments in 1992, 5,242 in 1995, 7,429 in 1999, and 8,856 in 2001. Unlike the EIA, the ENESTyC is based on a representative sample of plants and includes *maquiladoras.* The sampling frame in each year was stratified by total employment, with plants with 100 or more employees sampled with certainty, and a sample of plants with fewer than 100 employees drawn at random. In 1995, two samples were drawn. One was a probabilistic sample similar to the samples in 1992, 1999, and 2001, with the difference that *maquiladoras* were excluded. The second sample was a follow-up sample for the 1992 sample. All respondents to the 1992 survey that could be located were included. A few *maquiladoras* were included in the 1995 survey through this follow-up sample.

The surveys were designed as separate cross-sections, not as a panel, but because large establishments have been sampled with certainty, it is possible to link a fair number of plants over time. The different waves of the survey did not employ the same identification codes, and many of the establishments had to be linked across years by establishment name and street address. Although many questions changed in the questionnaires between waves, several key variables are comparable across waves.

My judgment after working extensively with these data is that the EIA is the more reliable source because plants are followed more closely by INEGI personnel. For this reason, I focus on plants that appear in both the ENESTyC and the EIA, and use information (i.e. on domestic sales) from the EIA when possible. To construct the EIA-ENESTyC Panel, I

---

[9]The raw data contains some negative values for fixed assets and investment. I set these to zero before calculating capital stock.

limited the sample to plants in the balanced EIA 1993-2001 Panel that could be linked to the ENESTyC and had non-missing data for ISO 9000 certification — the most important ENESTyC variable for testing the quality-upgrading hypothesis — in the three waves in which ISO 9000 is available: 1995, 1999, and 2001. This yielded a panel of 844 plants. Limiting the panel to plants with complete data on all variables included in Table IV would have made the panel prohibitively small. Instead, I allow the sample size to change when focusing on the different dependent variables in Table IV.

The ENESTyC asked a number of different questions about employee training. It appears from the pattern of responses, however, that respondents misunderstood many of the specific questions, or used different rules of thumb to guide their answers. The most reliable measure of training appears to be simply whether or not a formal training program exists at the plant, rather than how many workers received training.

The ENESTyC reports wage and employment data by 4 occupational categories: unskilled blue-collar workers, skilled blue-collar workers, administrative and technical workers, and managers. The definitions of the categories in the official documentation, however, are imprecise, especially on the distinction between unskilled and skilled blue-collar workers, and this seems to have led to a significant amount of noise in the data, with some plants reporting all their blue-collar workers under the skilled blue-collar category, and others under the unskilled blue-collar category. The coarser distinction between blue-collar and white-collar workers is more familiar to survey respondents and appears to be more reliable.

The information on white-collar and blue-collar schooling is available only in the 1992, 1999 and 2001 waves. I used information from the 1992 wave rather than the 1995 wave as the pre-peso-crisis information.

I censored outliers for the turnover rate, the accident rate, and the absentee rate following the same procedure as for the EIA, assigning all variables in the bottom and top 1% to the values at the 1st and 99th percentiles respectively.

II.C *Estadísticas Mensuales de la Industria Maquiladora de Exportación (EMIME)*

In order to remain in the Mexican government export-promotion program, *maquiladora* plants must report monthly production statistics. These administrative records are collected in a datafile referred to by INEGI as the *Estadísticas Mensuales de la Industria Maquiladora de Exportación (EMIME)* [Monthly Statistics on Maquiladora Export Industry], which represents a census of participants in the program each month. The EMIME contains information on employment and wages by occupational categories and spending on inputs. Hours worked by occupational category is available beginning in 1997. The EMIME does not contain a measure of total sales, nor does it include measures of the share of inputs that are imported or the share of output that is exported. The industrial classification used by the EMIME is *grupo económico* [economic group], of which there are 12.[10] Data are available for 1990-2004. To maximize comparability with the EIA 1993-2001 Panel, I draw the set of plants with complete data in every year over the 1993-2001 period. I refer to

---

[10]The 12 groups are: 1. Selection, preparation, packaging and canning of food; 2. Assembly of apparel and other textile products; 3. Shoes and leather goods; 4. Furniture and accessories and other metal and wood products; 5. Chemical products; 6. Construction, reconstruction and assembly of transportation equipment and accessories; 7. Assembly and repair of tools and machinery (not electric); 8. Assembly and repair of machinery (electric and electronic); 9. Materials and accessories (electric and electronic); 10. Assembly of toys and sporting goods; 11. Other manufacturing; 12. Services.

these 1,088 plants as the EMIME panel.

II.D *Encuesta Nacional de Empleo Urbano (ENEU)*

The *Encuesta Nacional de Empleo Urbano (ENEU)* [National Urban Employment Survey] is a household survey modeled on the Current Population Survey (CPS) in the United States. Households are interviewed quarterly for five quarters, and then rotate out of the sample. The original ENEU sample focused on the 16 largest Mexican cities.[11] Over time, the coverage of cities expanded but I focus on the original 16 cities in order to maximize comparability across years.[12]

The ENEU sample used in this paper consists of men, ages 16-64, who worked 35 or more hours in previous week. Self-employed workers are excluded.

The hourly wage figures were constructed as follows:

1. I recovered monthly wages for the job worked last week as converted from weekly or bi-weekly basis by INEGI enumerators. Top-coded reports were assigned 1.5 times the top-code value. Individuals who reported not working in previous week were dropped.

2. Monthly hours were calculated as 4.3 times hours worked in the previous week. Responses of "irregular hours, less than 35", "irregular hours, between 35 and 48" and "irregular hours, more than 48" were assigned values of 20, 42 and 60 hours per week, respectively.

3. Hourly wage was calculated as monthly wage/monthly hours. The wage was deflated to constant 1994 pesos using the main consumer price index (INPC) from *Banco de Mexico*, the Mexican central bank.

4. I winsorized the wage data at the 1st and 99th percentiles, within each year.

All calculations use the sampling weights reported by INEGI.

II.E ADMINISTRATIVE RECORDS OF THE *Instituto Mexicano del Seguro Social (IMSS)*

All private, formal-sector Mexican employers are required to report wages for their employees, and pay social-security taxes on the basis of their reports to the Mexican social security agency, *Instituto Mexicano del Seguro Social (IMSS)*. The raw data can be considered a census of private, formal-sector establishments and their workforces for 1985-2001. The number of workers in the raw data at a given point in time ranges from approximately 5 million in 1985 to approximately 11 million in 2001. These data are the basis for Kaplan and Verhoogen [2006], and readers are referred there for further details. To mimic the EIA sample, we drew balanced panels of plants with 50 or more workers in all years 1993-2001 in two broad sectors, tradables (manufacturing) and non-tradables (construction, retail trade, transportation, and services). The panels have 3628 and 3672 establishments respectively.

The IMSS data contain information on the daily wage of individuals. The figures are based on a measure of total compensation, called the *salario base de cotización*, which

---

[11]The 16 cities are: Mexico City, Guadalajara, Monterrey, Puebla, Leon, Torreon, San Luis Potosi, Merida, Chihuahua, Tampico, Orizaba, Veracruz, Ciudad Juarez, Tijuana, Matamoros and Nuevo Laredo.

[12]Mexico also had a nationally representative survey that covered rural areas, the *Encuesta Nacional de Empleo (ENE)*, but until it was combined with the ENEU in 2000 it was carried out at less regular intervals.

includes both wages and benefits, including payments made in cash, bonuses, commissions, room and board, overtime payments, and in-kind benefits. We extracted data for September 30 for each year. At the establishment level, the data contain information only on location and industry (using the IMSS's own 4-digit industrial categories).

The wage variable used in Appendix Table IV is log of average real daily wage (over all individuals in plant) on Sept. 30 of the given year.

REFERENCES

Abowd, John M. and Simon Woodcock, "Disclosure Limitation in Longitudinal Linked Data," in Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies ed. by P. Doyle, J. Lane, L. Zayatz and J. Theeuwes (Amsterdam: North Holland, 2001).

Angrist, Joshua and Alan Krueger, "Empirical Strategies in Labor Economics," in Handbook of Labor Economics vol. 3A ed. by Orley Ashenfelter and David Card (Amsterdam: Elsevier Science, 1999).

Grether, Jean-Marie, "Mexico, 1985-1990: Trade Liberalization, Market Structure and Manufacturing Performance," in Mark J. Roberts and James R. Tybout, eds., Industrial Evolution in Developing Countries, New York: Oxford University Press for the World Bank, 1996, chapter 11.

Griliches, Z., "Economic Data Issues," in Handbook of Econometrics vol. 3 ed. by Z. Griliches and M. Intrilligator (Amsterdam: North-Holland, 1986).

Heckman, James, "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, V (1976), 475–492.

Kaplan, David S. and Eric A. Verhoogen, "Exporting and Individual Wage Premia: Evidence from Mexican Employer-Employee Data," Unpub. paper, ITAM and Columbia University, 2006.

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

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

Schafer, Joseph L. and Maren K. Olsen, "Modeling and Imputation of Semi-continuous Survey Variables," Unpub. Paper, The Methodology Center Technical Paper 00-39, Pennsylvania State University, 2000.

Table A1
Correlations among Alternative Proxies, EIA Balanced Panel, 1993

	log (domestic sales)	log (employment)	export share index	ISO 9000 index	1st prin. component	TFP (fixed effect)	log (dom sales/worker)	export share
log (domestic sales)	1.00							
log (employment)	0.80	1.00						
export share index	0.80	0.86	1.00					
ISO 9000 index	0.40	0.43	0.64	1.00				
1st principal component	0.87	0.86	0.96	0.63	1.00			
TFP (fixed effect)	0.81	0.64	0.66	0.40	0.74	1.00		
log (domestic sales per worker)	0.65	0.06	0.25	0.12	0.36	0.54	1.00	
export share		0.19	0.23	0.29	0.32	0.11		1.00

Notes: Table reports bivariate correlation coefficients for indicated variables in 1993. All variables have been deviated from industry means. Correlations between actual export share and log domestic sales, log domestic sales per worker omitted because measurement error in sales generates severe and misleading negative bias; see discussion in Section 5. N=3263. Definitions of proxies in Appendix I. Further details on dataset in Section IV of text and Appendix II.

Table A2
Selection-Correction Model, EIA 1993-2001 Unbalanced Panel

		survive to end of period (1)	Δ (export share of sales) (2)	Δ log(white-collar wage) (3)	Δ log(blue-collar wage) (4)	Δ log(wage ratio) (5)	Δ log(K/L ratio) (6)	Δ white-coll. emp. share (7)
1993-1997	log domestic sales, 1993	0.331***	0.013***	0.050***	0.034***	0.013	0.061***	0.004
		[0.025]	[0.004]	[0.015]	[0.010]	[0.017]	[0.020]	[0.003]
1997-2001	log domestic sales, 1997	0.357***	0.005*	0.010	-0.001	0.014	0.022	0.003
		[0.028]	[0.003]	[0.010]	[0.008]	[0.012]	[0.014]	[0.002]
Difference in coefficients			0.008	0.040**	0.035***	-0.001	0.039*	0.001
			[0.005]	[0.018]	[0.013]	[0.021]	[0.024]	[0.004]

Notes: Table reports coefficients on log domestic sales for 12 separate two-step selection-correction models, without instruments (Heckman, 1976). (Co-variate at left; dependent variables at top, with changes in Columns 2-7 over period at left.) Column 1 reports first stage, which is common to second stages in Columns 2-7. All regressions include 205 industry (6-digit) and 32 state dummies. N=4402 in 1993, of which 3682 survive until 1997. N=3682 in 1997, of which 3263 survive until 2001. Variable definitions in Appendix I. Further details on dataset in Section IV of text and Appendix II. Standard errors in brackets. *** indicates significance at 1% level, ** at 5% level, * at 10% level.

Table A3

Summary Statistics on Debt Portfolios, Publicly Listed Mexican Firms, 1993

	All sectors			Mfg. Only		
	non-exps.	exporters	all	non-exps.	exporters	all
foreign share of short-term debt	0.17 [.02]	0.42 [.02]	0.32 [.02]	0.20 [.05]	0.38 [.03]	0.33 [.02]
foreign share of long-term debt	0.19 [.03]	0.47 [.04]	0.35 [.03]	0.13 [.05]	0.36 [.05]	0.30 [.04]
foreign share of total debt	0.20 [.03]	0.46 [.03]	0.35 [.02]	0.20 [.05]	0.39 [.03]	0.34 [.03]
short-term share of total debt	0.65 [.03]	0.65 [.02]	0.65 [.02]	0.64 [.04]	0.68 [.03]	0.67 [.02]
debt/assets ratio	0.43 [.02]	0.41 [.02]	0.42 [.01]	0.43 [.04]	0.41 [.02]	0.41 [.02]
cost of capital (%)	13.23 [1.13]	12.31 [.67]	12.69 [.61]	17.80 [2.83]	12.53 [.86]	13.90 [.99]
N	93	132	225	29	83	112

Notes: Data for 1993 for publicly listed firms on Mexican stock market (Bolsa Mexicana de Valores). Exporter means export sales > 0. Cost of capital defined as 100*(interest paid)/(total debt). For further details on dataset, see Pratap and Urrutia (2004) and Aguiar (2005). Standard errors of means in brackets.

Table A4
Comparing Manufacturing and Non-Tradable Sectors, Data from Social Security Records

	Dependent variable: Dataset:	Δ log(hourly wage) EIA 1993-2001 Panel (Mfg.) (1)	Δ log(daily wage)		Difference (IMSS Mfg. vs. IMSS Services) (4)
			IMSS Mfg. Panel (2)	IMSS Services Panel (3)	
1993-1997	log employment 1993	0.072*** [0.007]	0.056*** [0.004]	0.015*** [0.004]	0.041*** [0.006]
	R-squared	0.157	0.185	0.159	
1997-2001	log employment 1997	0.025*** [0.006]	0.022*** [0.004]	-0.001 [0.006]	0.023*** [0.007]
	R-squared	0.109	0.117	0.087	
Difference (1993-1997 vs. 1997-2001)		0.047*** [0.009]	0.034*** [0.006]	0.015** [0.008]	0.018** [0.009]
N		3263	3628	3672	

Notes: Table reports coefficients on initial log employment for 6 separate regressions. (Co-variate at left; dependent variables at top, with changes over period at left.) Column (1) uses the EIA 1993-2001 balanced panel and includes 205 industry (6-digit) and 32 state dummies. The wage variable is change in log real hourly wage for white- and blue-collar workers together (total real wage bill/total hours worked). Columns (2) and (3) use balanced establishment-level panels for 1993-2001 from administrative earnings records of the Instituto Mexicano del Seguro Social (IMSS), the Mexican social security agency, and include 122 industry dummies (using an IMSS-specific industrial classification) and 32 state dummies. Wage variable log of average real individual daily wage for permanent (non-casual) workers on Sept. 30 of each year. Robust standard errors in brackets. Standard errors on differences in coefficients between 1993-1997 and 1997-2001 allow for cross-equation correlation. Further details on datasets in Section IV of text and Appendix II. *** indicates significance at 1% level, ** at 5% level, * at 10% level.

Table A5
Cleaning Procedure for EIA 1993-2001 Panel

	Non-exiters	Exiters		All
		Non-random	Random	
Raw sample (# of establishments)	4831	1902	424	7157
Subtractions				
1. Remove establishments whose information is reported in another establishment in the same firm.	-418	-186	-85	-689
2. Remove establishments that report information from other establishments in the same firm.	-230	-21	-6	-257
3. Remove establishments owned in whole or in part by government.	-39	-22	0	-61
4. Remove establishments that appear to be maquilas.	-374	-278	-51	-703
5. Remove plants that exit randomly.	0	0	-282	-282
6. Remove establishments with incomplete information (after imputation procedure described in Appendix B).	-507	-256	0	-763
Sample after cleaning	3263	1139	0	4402

Notes: Plants are assumed to be maquilas either if exports make up 95% or more of total sales, or if subcontrating income makes up 95% or more of total revenues. Non-random exiters defined as plants that go bankrupt or that suspend operations (possibly temporarily), including suspensions due to lack of materials or strikes. Random exiters are defined as plants that switch to industries not covered by the survey, that switch from manufacturing to primarily wholesale or retail sales, that merge with other establishments, or that fail to provide data for unexplained reasons. Random exiters are assumed to be ignorable in the sense of Griliches (1986). The EIA 1993-2001 panel includes non-exiters only. The EIA 1993-2001 unbalanced panel includes non-exiters and non-random exiters. Further details on dataset in Section IV of text and Appendix II.

Figure A1: Shift Toward Less-Skill-Intensive Sectors, 1988-1998
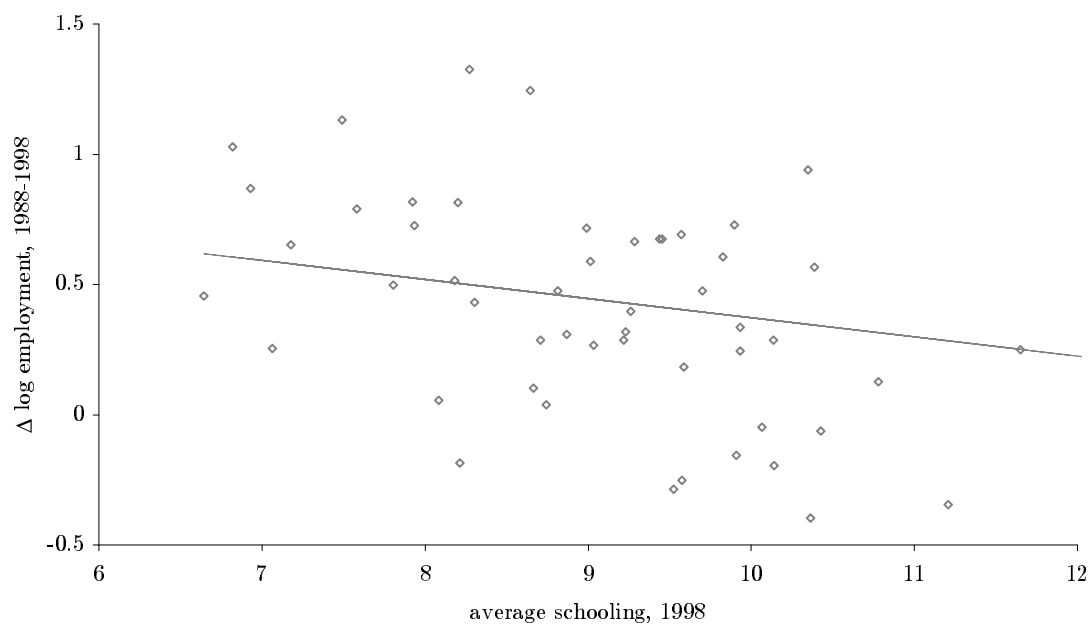

Notes: Data on employment in 1988 and 1998 by 4-digit manufacturing industry (including *maquiladoras*) from the Mexican *Censos Industriales* (Industrial Censuses). Data on schooling by 4-digit industry from *Encuesta Nacional de Empleo Urbano (ENEU);* further details in Appendix II. Regression weighted by employment in 1988.

Figure A2: Shift Toward Less-Capital-Intensive Sectors, 1988-1998
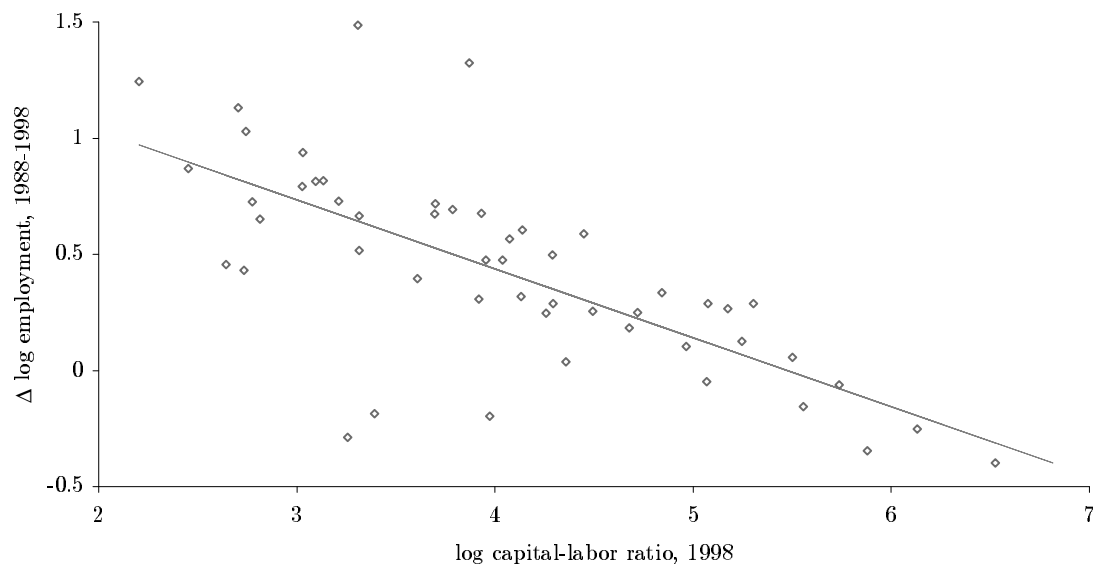
Notes: Data on employment in 1988 and 1998 and capital-labor ratio in 1998 by 4-digit manufacturing industry (including *maquiladoras*) from the Mexican *Censos Industriales* (Industrial Censuses). Regression weighted by employment in 1988.

Figure A3: Black Market and Nominal Exchange Rates, 1994-1995
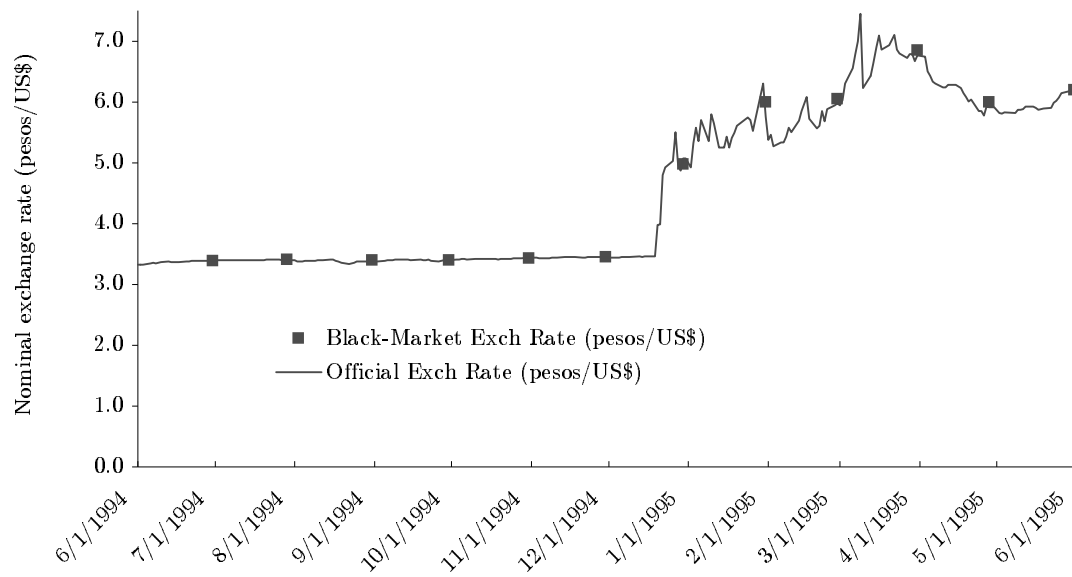
Figure A4: Non-Parametric Regressions, Levels of Key Variables vs. log Domestic Sales, 1993 and 1997

Fig. A4a: Export share of sales
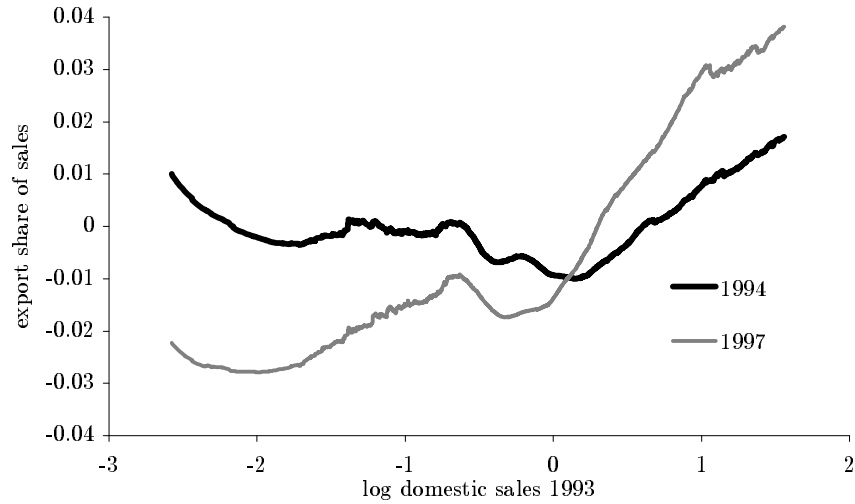
Fig. A4b: Log white-collar wage
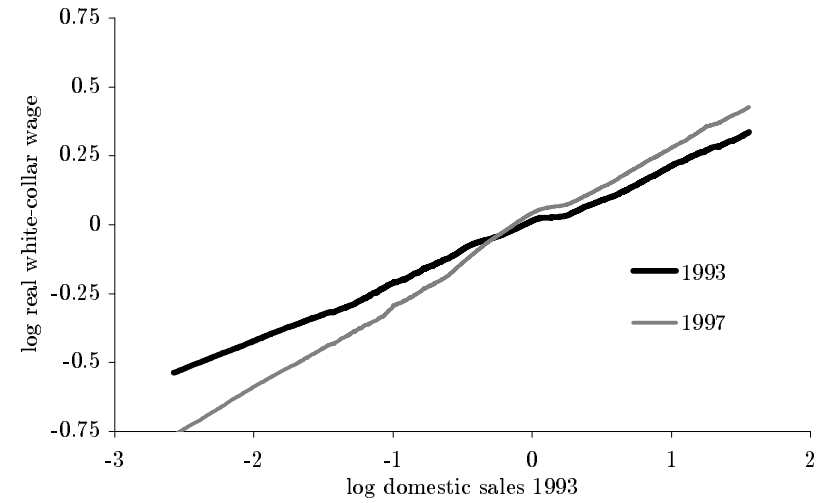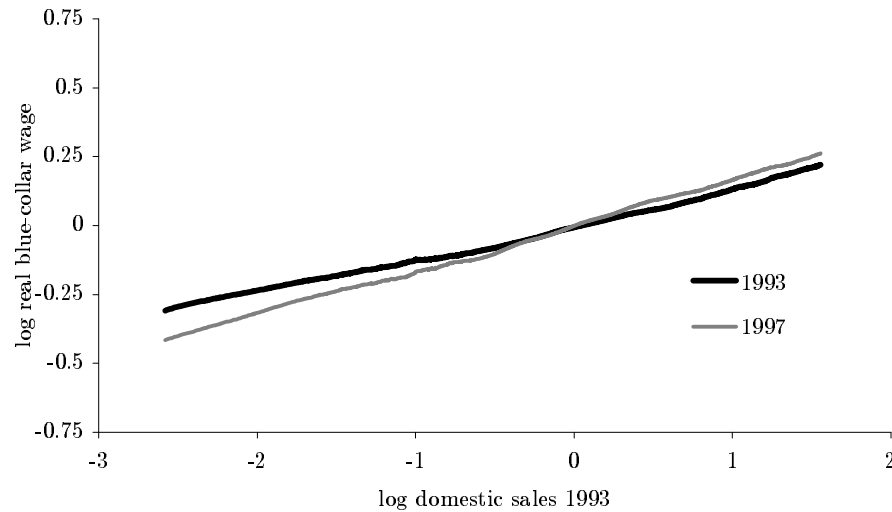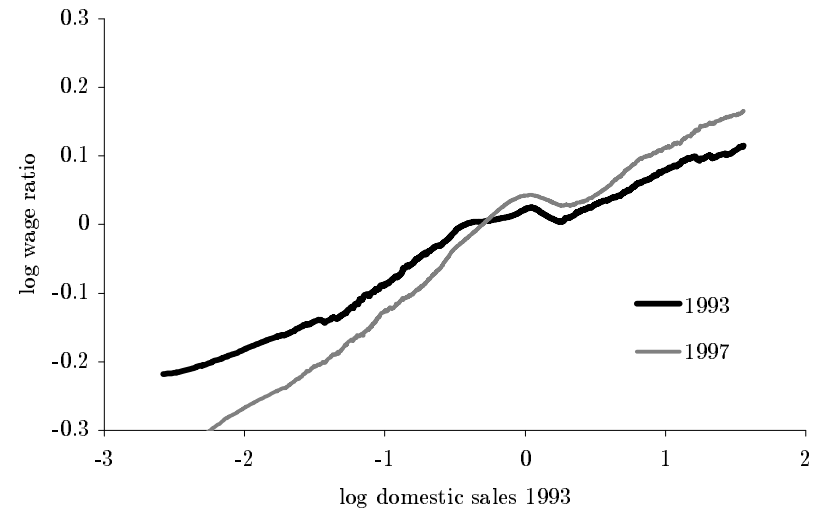
Fig. A4c: Log blue-collar wage

Fig. A4d: Log wage ratio

Notes: All variables deviated from industry-year means. Graphs are non-parametric regressions (bandwidth = .5) of levels of indicated variables in indicated year on deviated log domestic sales in 1993, using EIA 1993-2001 Panel. Plants below 1st or above 99th percentile in deviated log domestic sales in 1993 omitted from graph (but not regression procedure.) App. Fig. 4a uses 1994 as pre-crisis year to avoid the mechanical negative bias in the relationship between domestic sales in 1993 and export share in 1993. Variable definitions in Appendix I. Further details on dataset in Section IV of text and Appendix II.

Figure A5: Non-Parametric Regressions, Changes 1993-1997 and 1997-2001
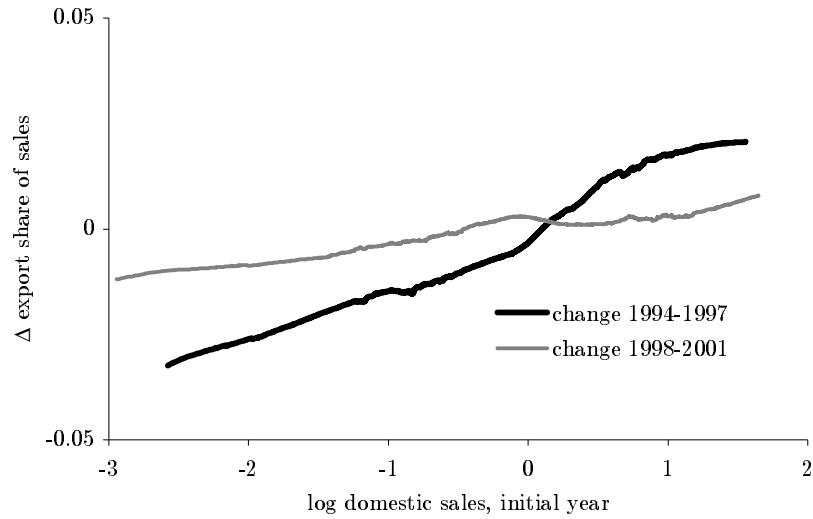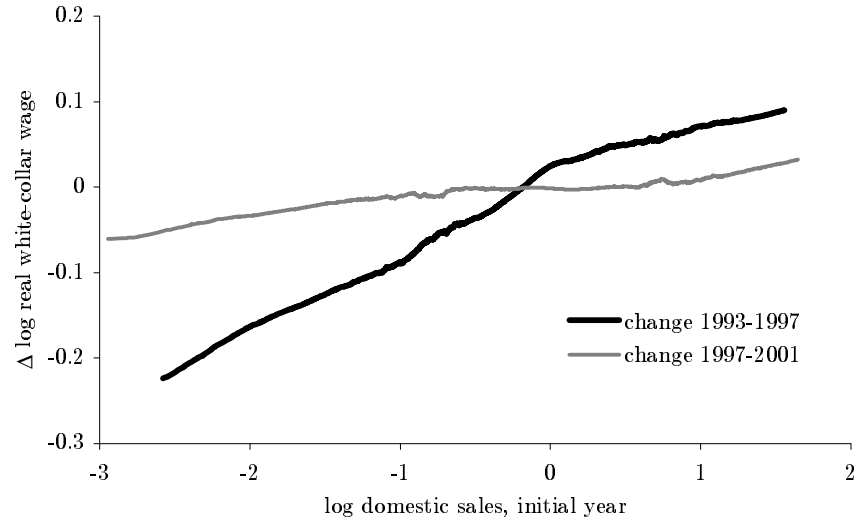
Fig. A5a: Changes in export share of sales

change 1994-1997
change 1998-2001

Fig. A5b: Changes in log white-collar wage

change 1993-1997
change 1997-2001

Fig. A5c: Changes in log blue-collar wage

change 1993-1997
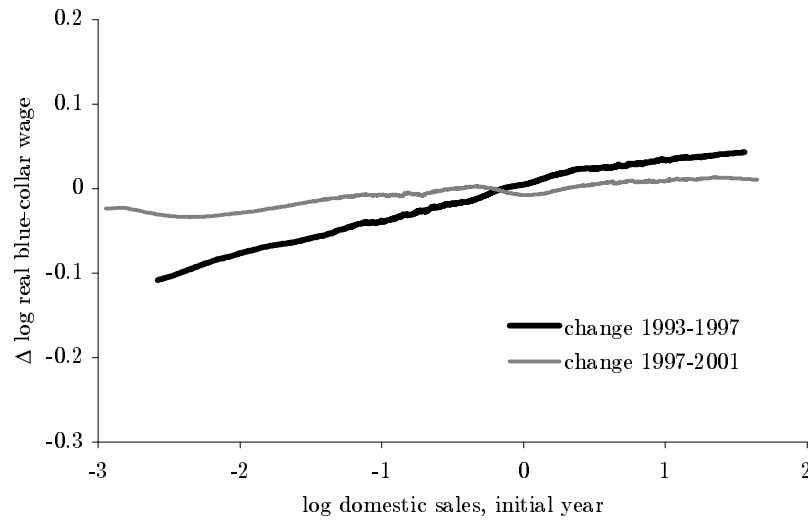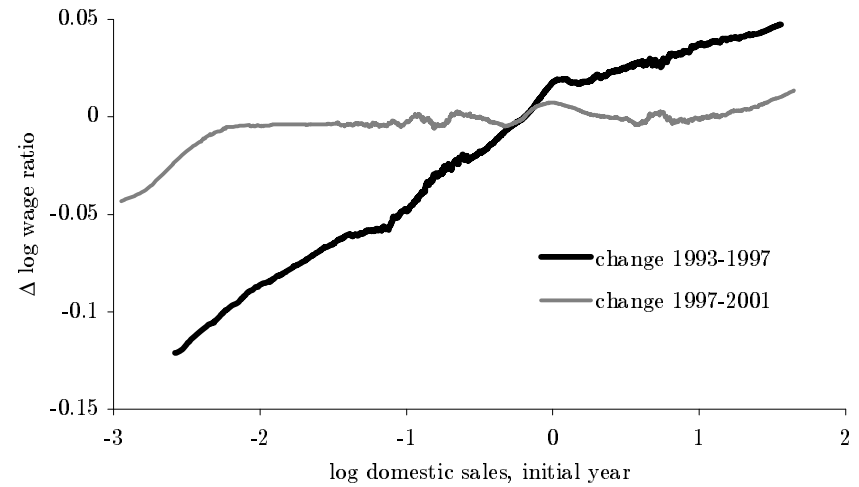change 1997-2001

Fig. A5d: Changes in log wage ratio

change 1993-1997
change 1997-2001

Notes: All variables deviated from industry means. Graphs are non-parametric regressions (bandwidth = .5), of changes of indicated variables over indicated periods on log domestic sales in initial year (1993 or 1997), using EIA 1993-2001 Panel. App. Fig. Va changes omit initial year to avoid bias from mean reversion. Variable definitions in Appendix I. Further details on dataset in Section IV of text and Appendix II.