Increasing the Cost of Informal Employment: Evidence from Mexico

Brenda Samaniego de la Parra Leon Fernandez Bujanda

ONLINE APPENDIX

Identifying Informal Workers at Formal Firms in ENOE

We follow the 17th International Conference of Labor Statisticians resolution for measuring informality.[71] According to this Resolution, informality has two dimensions. The first dimension refers to employers' characteristics: an employer is categorized as informal when it is not a registered business with tax authorities. The second dimension refers to job characteristics. Informal jobs are those that lack the benefits and institutional protection required by the legal framework in the country. Using this definition of informality, an employee can have either a formal or an informal job at a formal firm depending on whether the employer registers the worker with IMSS or not. All jobs are informal at informal firms in Mexico because only formal employers can register their employees.

The first step to identify informal workers at formal firms is to determine which workers in ENOEs household survey are employed at formal firms.[72] INEGI, and previous research[73], uses data on firms' size and industry to determine whether a firm is formal or not. This classification strategy relies on the assumption that larger firms are more likely to be detected by authorities and hence have a higher risk of being informal. Similarly, it assumes that firms in certain industries have more incentives to register with authorities because they either require a larger scale to operate or are more likely to benefit from participating in production networks that require issuing tax deductible sale receipts which are only available to firms registered with the government.

We complement this strategy to identify formal firms using information from the Ministry of Labor and Social Welfare's (STPS's) National Firm Directory (DNE). There are several benefits to this strategy. First, households' reporting of employers' size might be inaccurate. Using size thresholds to identify formal firms can therefore be problematic. Second, Hsieh and Olken (2014) find no evidence in the distribution of firms in Mexico to support size-based sorting into formality. Third, more than half of all employers registered at IMSS have between 2 to 5 employees. Using size to classify employers could therefore lead to misclassifying a large share of registered employers as informal firms. Fourth, since formal firms can hire workers off-the-books and tax authorities do not share information with IMSS, it is not clear whether the relevant measure to determine risk of getting caught is related to aggregate labor force size, share of non-registered workers, a combination of both, or something else entirely, like capital or sales.

[71] International Labour Organization (2003).

[72] All employers and workers in IMSS data are by definition formal.

[73] See, for example, Maloney (1999), Fiess, Fugazza and Maloney (2010), and Alcaraz, Chiquiar and Salcedo (2015) among others.

After identifying formal firms, we classify jobs into formal and informal jobs. We use workers' self-reported data on access to social benefits to determine their job's formality status. When using self-reported access to benefits, time-varying misreporting and misclassification can lead us to overestimate transitions rates across formal and informal jobs. While these errors may cancel in aggregate, stock variables, the estimated flow rates between labor market states may be very sensitive to these spurious transitions (Poterba and Summers (1986)). This concern might be heightened if individuals' incentives to misreport their access to social benefits is correlated with the timing of inspections.

To address this concern, we first point out that if inspections were only changing reporting behavior but not actual access to social security benefits then we would not see any changes in formal jobs in IMSS administrative data. Second, we implement a conservative correction to transitions. We identify sequential back and forth changes in reported access to social security benefits with the same employer and re-code them as misclassifications. If a worker switches between formal and informal status more than once within a three quarter period with the same employer, we consider the "true" formality status as the job in which the worker spent most time with the employer during the 5-quarter period that ENOE tracks the worker. This correction has a negligible effect on the stocks of informal and formal jobs within formal firms. However, it reduces the rate of transitions from formal to informal jobs and vice-versa by 2.5 p.p. and 2.0 p.p., respectively.[74]

Merging Datasets

B1. ENOE and DNE

The National Employment and Occupation Survey (ENOE) interviews 120,260 households every quarter starting in 2005. Among other questions regarding labor market participation, it asks every household member who is employed or involved in any income generating activity the name of the firm, business or institution of employment. ENOE also includes a battery of questions regarding the type of activities performed and goods or services provided by the firm. INEGI then uses the answers provided to these questions to classify the firm into one of 178 NAICS industry codes.

DNE is a list of firms' establishments. Each establishment is identified by the firms' "official name" (*razon social*), the establishments' exact address, and for subset of work-sites in the directory we also observe the firm's tax ID (*Registro Federal de Contribuyentes*). Meanwhile, in ENOE, workers self-report the name

[74] If establishments register their workers after an inspection to avoid being detected in a follow-up visit by IMSS but then un-register them after the verification takes place, observed informal-formal-informal transitions would not be misclassifications but rather real transitions. However, employers have incentives to avoid this "hiding" practice. Registering and unregistering workers within short periods of time can raise flags with authorities making establishments targets of directed inspection visits.

of their employer. Since the dwelling is the unit of observation in ENOE, the survey also includes information on the household's location, but not that of their place of work.[75] These differences generate two challenges when merging DNE and ENOE. First, due to spelling mistakes, abbreviations, and incomplete name reporting by the workers surveyed in ENOE, the name provided by the worker seldom is an exact match to the official name registered by the firm with STPS. Second, if a firm has more than one establishment in the workers' reported location, we need to make a decision about which establishment to match with the worker.

To match ENOE with the DNE and inspections logs, we first perfom basic name cleaning to standardize workers' reported firm names. This includes removing all punctuation, spacing and accents, eliminating articles, spelling out numbers, and replacing common abbreviations and plural forms. We then want to compare firm names in ENOE and find the closest match in the DNE. We define the closest match using a combination of a soundex algorithm and a Levenshtein distance.

Before implementing our matching algorithm, described in more detail below, we must clean ENOE's names further. In ENOE, employers' names are often reported including the type of establishment or sector in which the firm operates. For example, the answer for a worker employed at a 7-Eleven is at times recorded as "*Autoservicio* 7-Eleven" ("Convenience Store 7-Eleven"), "*Tienda* 7-Eleven" ("Store 7-Eleven"), or with the diminutive "*Tiendita* 7-Eleven". Meanwhile the official name (*Razon Social*) for 7-Eleven, as recorded in the DNE, is "*7-Eleven Mexico, S.A. de C.V.*". In this case, the words "*Tienda*" and "*Autoservicio*" are not actual parts of the firms name so we would like to remove them. However, in other cases, these words are useful to distinguish firms with similar names in different sectors, or are part of the official name. To address this issue, we create a word cloud with the most frequently appearing words in workers' reported employer names (see Figure B1). We then reduce these words, and all words with the same root, to the first 5 letters. This procedure reduces the weight given to these words when assessing which name is the closest match in the DNE.

Once we have standardized employers' name in both datasets, we then use a phonetic algorithm, in Spanish, to reduce mismatches from misspelling and typos.[76] Finally, for each employer name reported in ENOE, we identify the closest match in the DNE using the Levenshtein distance.[77] We consider an ENOE-DNE pair to be a match if the Levenshtein similarity ratio is at least 80% and the worker lived in the same state as the firm's location.

On average, we match 43,231 workers per quarter (out of an average of 82,600 wage-earning employees at formal firms). Table B1 presents the distribution of

[75]These two pieces of information, the household's location and their members' employers' names, are collected and recorded by ENOE. However, due to confidentiality requirements, they are only available through INEGI's microdata lab for research purposes.

[76]The algorithm is our own implementation of Amon, Moreno and Echeverri (2012)

[77]For each employer name in ENOE, we actually identified the top 5 matches to conduct manual checks and robustness analysis.

inspections by ENOE wave in the sample of workers that we use for the regression analysis in section III.B. It includes workers employed at a firm that we can match to an establishment in the DNE for at least one of ENOE's waves but did not receive an inspection before the first interview date. For workers employed at establishments that either did not receive an inspection, or that were inspected after the worker exits ENOE's sample, we assign a placebo inspection date so that the distribution of inspections across waves in the control group reflects the distribution of inspections in the treated group.

Figure B1. : ENOE Firm Names Word Cloud

Note: This figure shows a censored version of the word cloud used to standardized employers' names between ENOE and DNE. The goal is to decrease the weight of words frequently used by workers to describe their employers, such as "store," "bank," or "pharmacy," when calculating the Levenshtein distance between ENOE and DNE firm name pairs. Some words were censored to avoid revealing employers' identities.
Source: Own calculations using the ENOE survey panel from 2005 to 2016.

Table B1—: Treatment and Control Groups for Regression Analysis by Timing of Inspection and Formality Status Prior to Inspection

Inspection Wave	Control (Placebo Inspection)		Treated (1st Inspection During ENOE)		Total
	Informal	Formal	Informal	Formal	
1	56,399	169,197	1,881	5,353	232,830
2	43,419	130,256	1,611	4,586	179,872
3	40,860	116,293	1,254	4,446	162,853
4	23,550	78,842	815	2,890	106,097
5	6,143	21,779	252	716	28,890
Total	170,371	516,367	5,813	17,991	710,542

Note: The sample includes all workers employed, for at least one of ENOE's waves, at an establishment matched to the DNE. The control group refers to workers employed at establishments that either did not receive an inspection or whose first inspection arrives after the worker has left ENOE's sample. The treated group includes workers whose establishment of employment receives its first inspection while the worker is included in ENOE's sample. "Inspection Wave" refers to the interview number in ENOE when the inspection (or placebo inspection) occurs.
Source: Own calculations using the ENOE survey panel and the DNE from 2005 to 2016.

B2. IMSS and DNE

IMSS employer-employee administrative data and the DNE share some variables that allow us to identify a firms' establishments in the two datasets: the firm name (*"Razon Social"*) and its tax ID (*RFC*). Due to confidentiality restrictions, we were not allowed to work directly with the non-anonymized data. Instead, that staff at Banxico's EconLab helped us develop and implement a data cleaning and name matching algorithm to match IMSS administrative records with the DNE. The EconLab staff is exceptionally well suited for this task as they are extremely familiar with the data and highly skilled in text analysis and big data.

Merging IMSS records to the DNE involved three steps: 1) tax ID matches, 2) name cleaning and homogeneization, and 3) direct, phonetic, and closest distance) name matching. We start by matching establishments for which we do have tax ID information. If the tax ID matches perfectly, then we consider these employers as matched and keep them in the sample. For the remaining observations in IMSS data, we follow a similar process than the one used to match ENOE and DNE. First, we clean firms' names in both datasets removing acronyms like Corp., Ltd., Inc., etc. We homogenize capitalization and remove accents. Then we identify the employers in IMSS data that have an identical firm name (letter-by-letter match) as a firm in the DNE. We consider these as matches and continue with the rest of the non-matched employers.

We perform a soundex algorithm (in Spanish) converting each firm name in the set of unmatched IMSS employers and in the full set of DNE firms to its phonetic equivalent. We then compare the two datasets and, for each firm in the IMSS data, we find the closest phonetic match in the DNE. If the distance between a firm in IMSS data and its closest phonetic match in the DNE is such that the probability of a true match is at least 95%, we consider it a match. Finally, for the rest of the unmatched employers in IMSS data, we calculate the Levenshtein distance and the Jaro Winkler similarity measure relative to each of the firm's names in the DNE. If the nearest name has a similarity of at least .95, we consider it a match.

Table 1 describes the matched DNE-IMSS sample and compares it to the universe of firms in IMSS data, and the set of employers matched between the DNE and ENOE. Table B2 shows the distribution of new formal hires by employers' inspection status (inspected versus never inspected) and around a 1-year window of the establishment's first inspection.

Even though IMSS records include the entire universe of formal employers in Mexico, we cannot find every firm in the DNE within IMSS records for three main reasons. First, the information we received from both IMSS and STPS only includes tax IDs for a subset of employers. For IMSS records, we had access to tax IDs for employers that were active for at least one month in 2018 and onwards.

Meanwhile, only 28% of all firms in the DNE included tax ID information.[78] Hence, when using tax IDs we are more likely to match younger firms, which are more likely to be active in 2018, and larger firms with longer tenures. For the remaining employers, we rely on fuzzy name and location to do the matching. Second, employers may register their establishments in the DNE but they only appear on IMSS records when they hire at least one wage-earning employee who receives formal, social security benefits. Formal family-owned and operated businesses are not necessarily included in IMSS records since self-registration in social security benefits is not mandatory, but they are included in the DNE.[79] Third, the DNE underrepresents firms with short lifespans who enter and exit the formal sector in between the STPS's annual updating process for the DNE list. This can explain why Construction, which has high firm turnover, is underrepresented within the set of matched firms while Manufacturing is overrepresented. Larger firms that are members of business associations and chambers of commerce, and firms with unions, are arguably more likely to be included in the DNE as these organizations are one of the channels that STPS uses to update the DNE. STPS also updates the list based on the set of firms that participate in its annual worker training programs available for firms with less than 50 workers, so the DNE could overrepresent firms in this size category.

Table B2—: New Formal Jobs at IMSS Employers Matched to DNE Firm in 12-month Window Around Inspection

New Formal Matches by Inspection Status and Time Elapsed Since First Inspection			
Qtrs. to/since Inspection	Inspected	Placebo	Total
-4	106,962	343,195	450,157
-3	118,686	358,751	477,437
-2	122,161	399,449	521,610
-1	117,581	354,471	472,052
0	115,401	323,324	438,725
1	114,665	332,108	446,773
2	111,991	332,134	444,125
3	109,775	324,042	433,817
4	111,670	335,234	446,904
5	117,469	328,030	445,499
Total	1,146,361	3,430,738	4,577,099

Note: The sample includes formal employees at firms that can be matched to an establishment in the DNE where the formal match began within a 12-month window of the establishment's first STPS inspection, for treated employers, or first placebo inspection for the control group.
Source: Own calculations based on data from IMSS administrative records accessed through Banxico's EconLab Convenio No. 45, DNE obtained through official information request numbers 0001400017316 and 0001400017416, and STPS Inspection Logs.

[78]STPS removed the rest of the tax ID's before we received the data in compliance with transparency and data protection regulation.
[79]Small, family operated establishments may have incentives to voluntarily register with STPS due to the agency's offer of free training and management advice.

Verifying Random Selection in Inspections

As shown in table C1, in most cases, inspections are closed without there being any reported violations for the items within STPS's enforcement responsibility. Only 10% of all inspections lead to a fine. Between 2005 and 2016, the average fine was MXN$32,194 (USD$1,740) with a maximum fine of MXN$82,569,000 (USD$4,463,189)[80] and a minimum of MXN$20.57 (USD$1.11).

If auditors attempt to minimize the effort in selecting audit subjects from an ordered list, the selected sample may not be evenly or independently spread throughout the population of eligible firms (see Hall et al. (2012)). Concerns over haphazard sampling are minimized by STPS electronic system to allocate inspection. Nonetheless, we consider two possible orderings for DNE establishments to test for haphazard selection: alphabetical and postal code. Under a null hypothesis of random selection, the order in the alphabetical list or location is uncorrelated with the probability of being inspected. We do this test separately for each year between 2006 and 2015.[81] We fail to reject the null of randomness at a 10% significance level for at least 74% of categories and up to 96%. We do not find evidence of consistent selection of establishments across time based on establishment name or location. Tables C4 and C5 present the haphazard test for 2015.[82]

Table C1—: Distribution of STPS's Inspections by Result (2005-2016)

Result		No. of Inspections	% of All Inspections
Closed without report of violations		266,517	43%
Provided proof of compliance		296,367	48%
Request for time extension granted		184	0%
Fining process started	Fine imposed	23,154	4%
	Fine no yet imposed	34,620	6%

Note: Excludes violations beyond STPS's jurisdiction, including those related to informal employment.
Source: Own calculations using STPS DNE and Inspections logs 2005-2016.

[80]This fine was due to health and hygiene violations in 2013.
[81]We exclude 2005 and 2016 to have complete annual data for all establishments.
[82]We also consider geographical coordinates as an ordering category for haphazard selection. These results, as well as those from other years, are available upon request from the authors.

Table C2—: Testing for Random Probability of Inspections: Employer-level

	Sample: Matched DNE-IMSS Employers			
	Coeff.	s.e.	t-stat	95% CI
Employer Age: Years since hire date for first formal employee				
$[1, 2)$	0.009	0.030	0.288	(-0.050,0.067)
$[2, 3)$	0.022	0.030	0.711	(-0.038,0.081)
$[3, 4)$	-0.007	0.031	-0.217	(-0.067,0.053)
$[4, 5)$	-0.024	0.029	-0.829	(-0.080,0.032)
$[5, 6)$	-0.049	0.029	-1.686	(-0.105,0.008)
$[6, 7)$	-0.042	0.029	-1.440	(-0.099,0.015)
$[7, 8)$	-0.049	0.028	-1.738	(-0.104,0.006)
$[8, 9)$	-0.081	0.028	-2.898	(-0.136,-0.026)
$[9, 10)$	-0.080	0.028	-2.836	(-0.135,-0.025)
$[10, 11)$	-0.087	0.030	-2.934	(-0.145,-0.029)
11 or more	-0.153	0.037	-4.134	(-0.226,-0.081)
Employer Size: Number of Formal Workers				
6-10	0.004	0.015	0.275	(-0.026,0.034)
11-25	0.004	0.013	0.273	(-0.022,0.029)
26-50	0.002	0.014	0.137	(-0.025,0.029)
51-100	-0.001	0.014	-0.065	(-0.029,0.027)
101-250	-0.002	0.015	-0.146	(-0.032,0.027)
251+	0.003	0.017	0.191	(-0.029,0.036)
Employer Sector				
Mining/Extractive	-0.041	0.053	-0.777	(-0.145,0.063)
Manufacturing	-0.003	0.030	-0.106	(-0.062,0.056)
Construction	0.029	0.031	0.915	(-0.033,0.090)
Retail/Wholsesale	0.016	0.044	0.371	(-0.070,0.102)
Restaurant/Lodging	0.015	0.029	0.513	(-0.042,0.072)
Transport/Comms.	0.003	0.032	0.082	(-0.060,0.065)
Prof./Buss. Services	0.019	0.030	0.633	(-0.040,0.077)
Other	0.020	0.032	0.634	(-0.043,0.084)
Share of male workers	-0.001	0.001	-0.616	(-0.002,0.001)
No. of Observations: 119,656				

Note: This table presents evidence of random selection in STPS routine inspections. The baseline estimation sample is employers with an establishment included in the DNE that can be matched to an employer in IMSS records. The dependent variable is a dummy equal to 1 if the employer was inspected by STPS between January 2005 and June 2016. All independent variables are measured the quarter prior to the inspection (or placebo inspection for the control group). Standard errors are robust to heteroskedasticity and we weigh all observations using entropy weights.

Table C3—: Testing for Random Probability of Inspections: Worker-level

	(1) Inspected During ENOE	(2) Informal Prior to Inspection or Placebo
age	0.000	-0.012
	(0.000)	(0.006)
age^2	0.000	0.0001
	(0.000)	(0.000)
male	-0.002	-0.003
	(0.002)	(0.003)
yrs. of schooling	0.001	0.001
	(0.001)	(0.001)
Educational Attainment		
Completed Elementary School	-0.001	-0.031
	(0.004)	(0.013)
Completed 9th grade	-0.002	-0.068
	(0.005)	(0.029)
Completed HS and above	-0.001	-0.071
	(0.007)	(0.034)
Industry		
Manuf.	0.008	-0.032
	(0.011)	(0.046)
Construc.	0.004	-0.055
	(0.011)	(0.033)
Retail	0.004	-0.051
	(0.010)	(0.052)
Restaurants	0.000	0.018
	(0.010)	(0.039)
Transport	0.016	-0.031
	(0.012)	(0.042)
Services	0.000	-0.033
	(0.011)	(0.043)
Establishment size (num. employees)		
2-5	-0.010	0.155
	(0.009)	(0.018)
6-10	-0.002	0.07
	(0.009)	(0.044)
11-15	0.005	0.021
	(0.009)	(0.061)
16-50	0.008	-0.011
	(0.009)	(0.069)
51+	0.017	-0.051
	(0.009)	(0.078)
NA	0.013	-0.008
	(0.010)	(0.066)
Location Size (population)		
15K - 99K	0.002	0.031
	(0.003)	(0.005)
2.5K -14,999	0.001	0.043
	(0.003)	(0.009)
Less than 2.5K	-0.001	0.042
	(0.003)	(0.012)
Adj. R-squared	0.070	0.10
No. of Workers	38,507	289,129

Note: This table presents evidence of random selection in STPS routine inspections. The baseline estimation sample for Column (1) is individuals who are informally employed the quarter prior to an inspection at an establishment that is included in the DNE between 2005 to 2015. The dependent variable is a dummy equal to 1 for individuals employed at inspected establishments. All regressions include worker occupation fixed effects.
Source: Own calculations using ENOE and STPS DNE and Inspections logs.

Table C4—: Haphazard Sampling Test 2015: Alphabetical Order

First Letter in Name	No. of Establishments	% in DNE	No. of Inspections	% of Inspections	test-stat	p_value
A	44,503	10.73%	8,941	10.55%	-1.22	0.222
B	15,690	3.78%	3,026	3.57%	-1.39	0.163
C	56,920	13.73%	11,687	13.80%	0.47	0.641
D	14,563	3.51%	2,965	3.50%	-0.08	0.935
E	18,728	4.51%	3,644	4.30%	-1.42	0.156
F	15,780	3.80%	3,126	3.69%	-0.76	0.447
G	27,110	6.54%	5,751	6.79%	1.67	0.096
H	11,755	2.83%	2,328	2.74%	-0.57	0.569
I	21,143	5.10%	3,953	4.66%	-2.86	0.004
J	5,358	1.29%	1,214	1.43%	0.91	0.362
K	2,880	.69%	517	.61%	-0.55	0.586
L	12,105	2.92%	2,353	2.77%	-0.93	0.353
M	27,912	6.73%	5,848	6.90%	1.14	0.254
N	7,065	1.70%	1,606	1.89%	1.25	0.213
O	12,046	2.90%	2,460	2.90%	-0.01	0.992
OTHER	270	.06%	53	.06%	-0.02	0.987
P	31,598	7.62%	6,362	7.51%	-0.74	0.456
Q	2,308	.55%	406	.47%	-0.50	0.617
R	14,245	3.43%	3,111	3.67%	1.55	0.121
S	36,682	8.85%	8,092	9.55%	4.75	0.000
T	21,710	5.23%	4,139	4.88%	-2.32	0.020
U	2,843	.68%	604	.71%	0.18	0.860
V	6,529	1.57%	1,473	1.73%	1.07	0.287
W	1,625	.39%	331	.39%	-0.01	0.994
X	453	.10%	113	.13%	0.16	0.876
Y	1,157	.27%	233	.27%	-0.03	0.979
Z	1,482	.35%	347	.40%	0.34	0.736
Total	414,460	100%	84,683	100%		

Source: Own calculations using STPS DNE and Inspections logs 2015.

Note: The test-stat column is the test statistic from a χ^2 test of proportions. Under random assignment, the share of establishments in the DNE with names starting with the corresponding row's letter equals the expected share of inspected establishments starting with that letter. The null hypothesis of random selection can be rejected for significance levels greater than the value indicated in the p-value column. "Other" indicates non-alphabetic characters.

Table C5—: Haphazard Sampling Test 2015: Zip Code Order

First Zip Code Digits	No. of Establishments	% in DNE	No. of Inspections	% of Inspections	test-stat	p-value
10	2,175	0.52 %	385	0.45 %	-0.45	0.651
11	11,155	2.69 %	2,062	2.43 %	-1.67	0.094
12	943	0.22 %	195	0.23 %	0.02	0.986
13	1,592	0.38 %	265	0.31 %	-0.46	0.646
14	3,746	0.90 %	523	0.61 %	-1.85	0.064
15	5,010	1.20 %	779	0.91 %	-1.87	0.061
16	1,440	0.34 %	209	0.24 %	-0.65	0.516
17	423	0.10 %	95	0.11 %	0.07	0.948
18	22	0.00 %	3	0.00 %	-0.01	0.991
19	227	0.05 %	62	0.07 %	0.12	0.905
20	7,858	1.89 %	1,371	1.61 %	-1.80	0.072
21	6,046	1.45 %	977	1.15 %	-1.98	0.048
22	4,901	1.18 %	1,242	1.46 %	1.84	0.066
23	8,285	1.99 %	1,328	1.56 %	-2.80	0.005
24	6,754	1.62 %	901	1.06 %	-3.67	0.000
25	6,061	1.46 %	1,404	1.65 %	1.27	0.205
26	5,146	1.24 %	1,286	1.51 %	1.79	0.073
27	3,961	0.95 %	749	0.88 %	-0.46	0.645
28	5,284	1.27 %	1,024	1.20 %	-0.43	0.670
29	5,155	1.24 %	1,356	1.60 %	2.32	0.021
30	2,910	0.70 %	674	0.79 %	0.61	0.545
31	9,386	2.26 %	1,751	2.06 %	-1.28	0.200
32	3,909	0.94 %	608	0.71 %	-1.46	0.145
33	4,906	1.18 %	785	0.92 %	-1.66	0.096
34	6,383	1.54 %	1,715	2.02 %	3.15	0.002
35	2,503	0.60 %	348	0.41 %	-1.25	0.213
36	6,309	1.52 %	1,631	1.92 %	2.62	0.009
37	5,166	1.24 %	1,199	1.41 %	1.10	0.272
38	3,678	0.88 %	809	0.95 %	0.44	0.661
39	6,543	1.57 %	1,197	1.41 %	-1.07	0.284
40	2,154	0.51 %	386	0.45 %	-0.41	0.680
41	646	0.15 %	103	0.12 %	-0.22	0.825
42	6,221	1.50 %	1,217	1.43 %	-0.41	0.679
43	3,701	0.89 %	600	0.70 %	-1.19	0.233
44	12,023	2.90 %	3,636	4.29 %	9.10	0.000
45	6,592	1.59 %	1,640	1.93 %	2.25	0.025
46	453	0.10 %	82	0.09 %	-0.08	0.936
47	323	0.07 %	55	0.06 %	-0.08	0.933
48	2,632	0.63 %	819	0.96 %	2.14	0.032
49	469	0.11 %	77	0.09 %	-0.14	0.886
50	4,622	1.11 %	944	1.11 %	0.00	0.998
51	1,056	0.25 %	135	0.15 %	-0.61	0.539
52	5,196	1.25 %	1,106	1.30 %	0.34	0.734
53	5,026	1.21 %	846	0.99 %	-1.38	0.166
54	12,792	3.08 %	2,202	2.60 %	-3.18	0.001
55	3,956	0.95 %	761	0.89 %	-0.36	0.718

Continues on next page

Table C5—: Continued from previous page

First Zip Code Digits	No. of Establishments	% in DNE	No. of Inspections	% of Inspections	test-stat	p-value
56	2,259	0.54 %	499	0.58 %	0.29	0.775
57	708	0.17 %	203	0.23 %	0.44	0.657
58	7,866	1.89 %	1,918	2.26 %	2.39	0.017
59	700	0.16 %	143	0.16 %	0.00	1.000
60	5,099	1.23 %	1,078	1.27 %	0.28	0.782
61	1,228	0.29 %	266	0.31 %	0.11	0.909
62	7,435	1.79 %	1,329	1.56 %	-1.46	0.145
63	7,031	1.69 %	1,406	1.66 %	-0.23	0.815
64	7,049	1.70 %	1,389	1.64 %	-0.39	0.694
65	759	0.18 %	162	0.19 %	0.05	0.958
66	6,328	1.52 %	1,087	1.28 %	-1.58	0.115
67	6,098	1.47 %	769	0.90 %	-3.65	0.000
68	6,029	1.45 %	1,106	1.30 %	-0.96	0.335
69	222	0.05 %	36	0.04 %	-0.07	0.943
70	2,207	0.53 %	514	0.60 %	0.48	0.631
71	743	0.17 %	160	0.18 %	0.06	0.950
72	10,967	2.64 %	2,472	2.91 %	1.78	0.075
73	1,398	0.33 %	231	0.27 %	-0.42	0.677
74	1,432	0.34 %	211	0.24 %	-0.62	0.534
75	791	0.19 %	142	0.16 %	-0.15	0.881
76	11,428	2.75 %	2,642	3.11 %	2.37	0.018
77	7,556	1.82 %	1,756	2.07 %	1.63	0.104
78	12,452	3.00 %	2,003	2.36 %	-4.18	0.000
79	1,029	0.24 %	106	0.12 %	-0.79	0.427
80	7,879	1.90 %	2,211	2.61 %	4.61	0.000
81	2,733	0.65 %	406	0.47 %	-1.16	0.245
82	4,417	1.06 %	825	0.97 %	-0.59	0.554
83	7,811	1.88 %	2,028	2.39 %	3.32	0.001
84	3,521	0.84 %	433	0.51 %	-2.19	0.029
85	3,755	0.90 %	709	0.83 %	-0.44	0.657
86	7,398	1.78 %	1,589	1.87 %	0.59	0.553
87	3,001	0.72 %	475	0.56 %	-1.05	0.292
88	4,363	1.05 %	873	1.03 %	-0.14	0.888
89	3,553	0.85 %	867	1.02 %	1.08	0.282
90	10,368	2.50 %	1,830	2.16 %	-2.22	0.026
91	9,198	2.21 %	2,792	3.29 %	7.02	0.000
92	1,019	0.24 %	180	0.21 %	-0.21	0.830
93	3,665	0.88 %	786	0.92 %	0.28	0.777
94	4,052	0.97 %	896	1.05 %	0.52	0.603
95	1,325	0.31 %	226	0.26 %	-0.34	0.733
96	2,926	0.70 %	563	0.66 %	-0.27	0.790
97	8,331	2.01 %	1,471	1.73 %	-1.78	0.076
98	8,171	1.97 %	1,676	1.97 %	0.05	0.960
99	848	0.20 %	251	0.29 %	0.59	0.554
Missing	1,247	0.30 %	366	0.43 %	0.85	0.397
0	326	0.07 %	60	0.07 %	-0.05	0.960
Total	414,460	100 %	84,683	100 %		

Additional Analyses

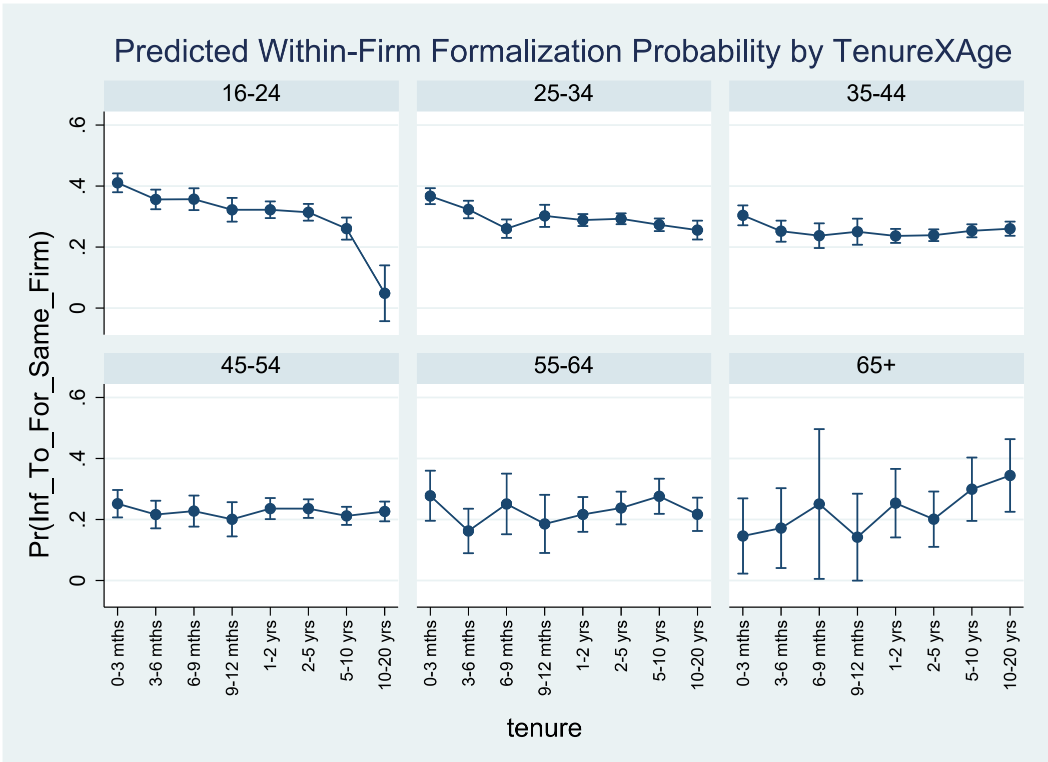
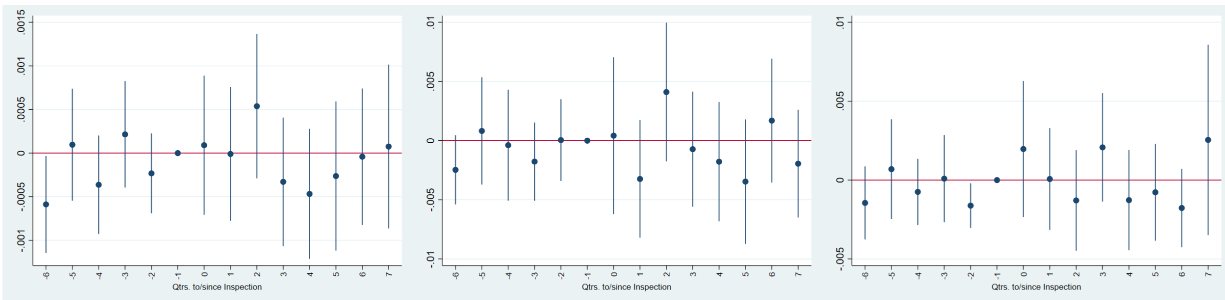
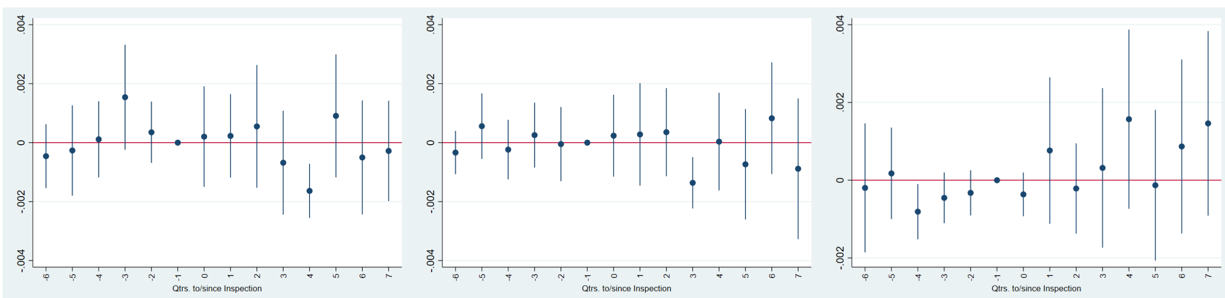


Figure D1. : Within-Firm Formalization Probability

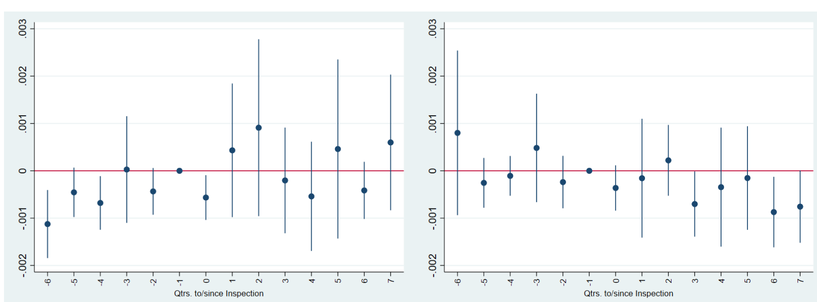
Notes to Figure D1: We calculate the predicted probability of "organic" within-firm informal to formal job transitions using a log-linear probability model with age and tenure groups, an interaction of these two variables, as regressors, We also control for workers' years of education, firm size as reported by the worker, firm and year fixed effects. The figure shows the marginal transition probability for each worker age and tenure groups.
Source: Own calculations using data from ENOE. The sample includes all individuals, between the ages of 18 and 60, who are informally employed at a formal firm for at least one of the waves in ENOE 2005 to June 2016. We exclude domestic workers and agriculture.



(a) All Firms Matched to the DNE (b) Less than 5 Employees (c) 6-10 Employees



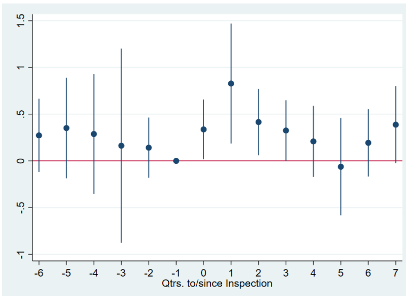
(d) 11-25 Employees (e) 26-50 Employees (f) 51-100 Employees



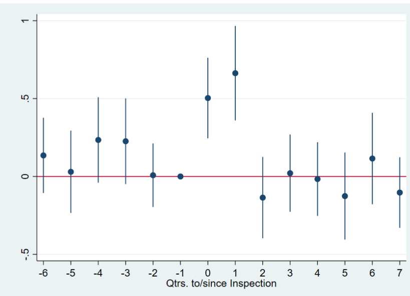
(g) 101-250 Employees (h) 251+ Employees

Figure D2. : Inspection's Effect on the Probability of Temporary Exit from Formal Sector

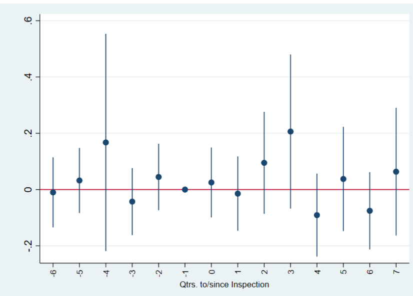
Note: Notes to Figure D2: These figures display the effect of a firm's first inspection on the probability of exit from the formal sector. The dependent variable is an indicator equal to 1 if the employer reports no formal workers in the current period and at least one in the period prior. $q = 0$ indicates the quarter of inspection. We create these charts using linear probability model specified in equation 1, estimated separately for each size group. Errors are clustered at the firm level.
Source: Own calculations using employer-level panel from IMSS records. The sample includes formal establishments' matched in the DNE and IMSS administrative records between 2005-2016.



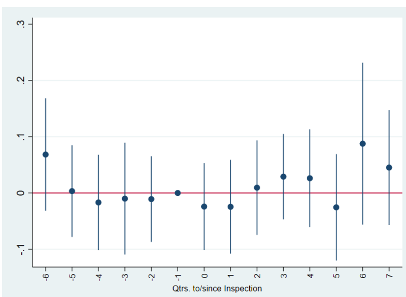
(a) Less than 5 Employees



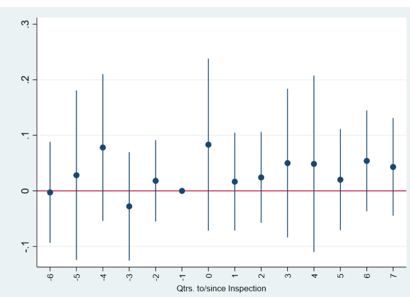
(b) 6-10 Employees



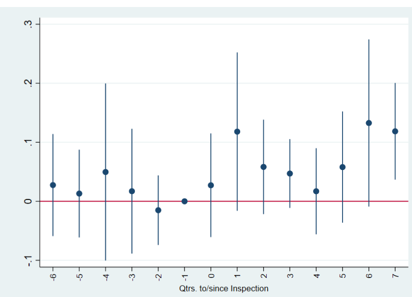
(c) 11-25 Employees



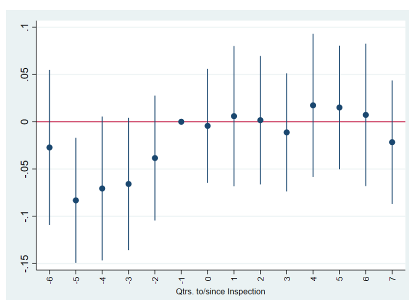
(d) 26-50 Employees



(e) 51-100 Employees



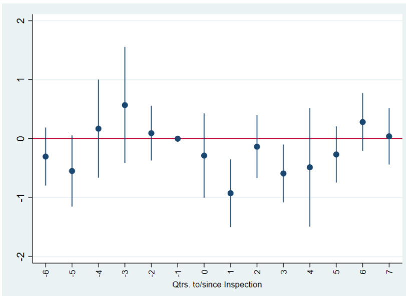
(f) 101-250 Employees



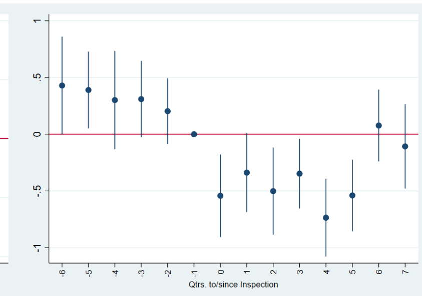
(g) 251+ Employees

Figure D3. : Effect of Firm's First Inspection on Formal Job Hires from Outside the Formal Sector by Firm Size

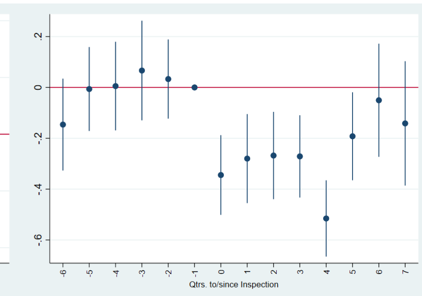
Notes to Figure D3: These figures display the effect of a firm's first inspection on establishments' inflows of workers who were not employed at a different formal firm in the previous 6 months. Firm size is measured as the average number of formal workers during the year of activity in the formal sector prior to the first inspection.
Source: Own calculations using employer-level panel from IMSS records. The sample includes formal establishments' matched in the DNE and IMSS administrative records between 2005-2016.



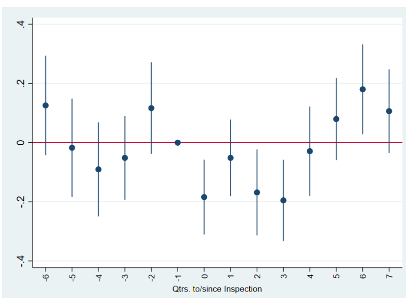
(a) Less than 5 Employees



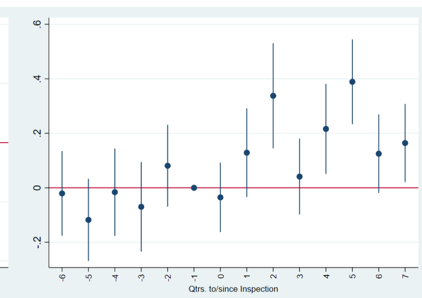
(b) 6-10 Employees



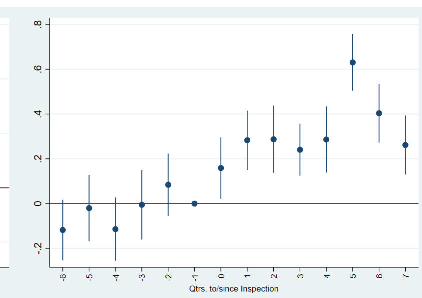
(c) 11-25 Employees



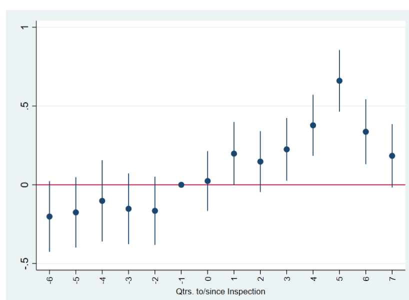
(d) 26-50 Employees



(e) 51-100 Employees



(f) 101-250 Employees



(g) 251+ Employees

Figure D4. : Effect of Firm's First Inspection on Formal Job Hires from Within the Formal Sector by Firm Size

Notes to Figure D4: These figures display the effect of a firm's first inspection on establishments' inflows of workers who were employed at a different formal firm in the previous 6 months. Firm size is measured as the average number of formal workers during the year of activity in the formal sector prior to the first inspection.

Source: Own calculations using employer-level panel from IMSS records. The sample includes formal establishments' matched in the DNE and IMSS administrative records between 2005-2016.