1 Introduction

A growing body of research suggests that lack of state capacity | in particular, di culty in raising taxes to fund public goods | is an important constraint on the growth of developing countries (Burgess and Stern, 1993; Besley and Persson, 2013). Developing countries generally have low ratios of tax revenues to GDP and large informal sectors. Mexico, the focus of our study, is no exception: it has the lowest tax revenue share of GDP in the OECD, between 15 and 20 percent during the period we study, and the informal sector has been estimated to make up 40 percent or more of total output (OECD, 2011b; IMF, 2010; Schneider and Enste, 2000). Given

raising

and in rm output, for reasons that may include the greater di culty of maintaining collusion in larger rms, as argued in a recent paper by Kleven, Kreiner, and Saez (2009), or simply the greater visibility of larger rms to auditors. The nding that compliance is increasing in rm size is consistent with the suggestion of Hsieh and Klenow (2014) and Hsieh and Olken (2014) that the burden of taxation in developing countries falls more heavily on larger rms, and that this may part of the explanation for the disproportionately large number of small rms in Mexico and other developing countries.

We also show that evasion responded to an important change in the Mexican social security system in the way that economic theory would predict. We focus on a pension reform that introduced a system of personal retirement accounts, passed by the Mexican Congress on December 21, 1995 and implemented on July 1, 1997. As discussed in more detail below, prior to the reform the social security bene ts of most workers were largely insensitive to the wages reported by rms on their behalf (as long as they reported at least the minimum allowable wage). The reform tied individual pensions more closely to rms' wage reports and made it easier for employees to observe those reports. Workers already in the traditional system prior to July 1, 1997 retained the right to choose, at the time of retirement, the pension that they would have received under the pre-reform regime. Because older workers had little time to accumulate su cient balances in their personal accounts, their expected pension was higher under the old regime. Younger workers had a greater expectation of being better o under the new regime and hence had stronger incentives to ensure accurate reporting. We use this di erential impact by age as the basis for a di erence-in-di erences estimation strategy. Consistent with our theoretical model, evasion declines relatively more for younger age groups. The fact that the discrepancies between the two data sources move in the way predicted by theory reinforces the view that they are a meaningful measure of evasion, and the result suggests that giving employees incentives and information to improve the accuracy of employer reports can be an e ective way to improve payroll-tax compliance.

A key limitation of our study is that, although evasion decisions are taken at the level of individual rms, the household labor-force survey does not contain rm identi ers and we are not able to construct measures of evasion at the rm level. Instead, we construct measures of evasion at the level of cells de ned by di erent combinations of metropolitan areas, sectors, rm-size categories and age groups, depending on the speci cation. A second limitation is that it is di cult to separate the e ects of the change of incentives and the change of information (discussed in more detail in Section 2) in the pension reform. It seems likely that the change in incentives was more important than the change in information: if incentives had not changed, and bene ts had remained largely insensitive to wage reports, it is not clear why information alone would have led to changes in compliance. But the \experiment" we consider combined both elements, and the e ects we estimate should be interpreted as the combined e ects of both.[1]

This paper is related to a number of di erent literatures. Research in development economics on the non-compliance of rms with tax regulations has tended to focus on the the failure of rms to register with tax authorities, which we might term the extensive margin of compliance (Gordon and Li, 2009; McKenzie and Sakho, 2010; de Mel, McKenzie, and Woodru , 2012). In this paper, by contrast, we focus on an intensive margin of compliance: the extent of compliance by formally registered rms, reporting wages for formally registered workers.

There is a small literature on salary misreporting, including Nyland, Smyth, and Zhu (2006), Tonin (2011), and Bergolo and Cruces (2012), and Mao, Zhang, and Zhao (2013). This paper appears to be the rst to analyze how tying bene ts more closely to reported wages can contribute to improved compliance.[2]

Andreoni, Erard, and Feinstein (1998), Slemrod and Yitzhaki (2002), and Saez, Slemrod, and Giertz (2012).

2 Institutions: The Mexican Social Security System

The schedule re ects a complicated set of formulas determining contributions to the various components of the IMSS system, principally health care, pensions, and child care.[8] The gure illustrates that the most signi cant changes in the schedule are for the highest-wage workers, earning above 500 pesos per day, due to changes in the maximum taxable income over the period, from 10 times to 25 times the minimum wage in Mexico City.[9] The topcodes apply to no more than 5 percent of wage-earners in any year and will play little role in our analysis. The total employer contribution varied between 18 percent and 22 percent of the wage over the range in which almost all workers fall. There was an increase in the employer contribution from 1990 to 1993, and then the reform in 1997 introduced a kink in the schedule, which raised contributions disproportionately on the lowest-wage workers. Figures 2 displays worker contributions, which vary between 2 percent and 5 percent over the relevant range and declined with the 1997 reform. Overall, while there were changes in the contribution schedules, these were relatively modest over the relevant wage range.[10] Looking ahead to the empirical strategs

2.3 Pension Bene ts

2.3.1 Pre-reform (pay-as-you-go) system

Under the pre-reform regime, workers became vested in the system after 10 years of contributions, and were then entitled to receive at least the minimum pension. Pensions were calculated on the basis of the nal average wage, de ned as the average nominal wage in the ve years preceding retirement. Panel A of Figure 3 illustrates the expected daily pension as a function of the nal average wage for workers with 10, 20 and 30 years of contributions in selected years. The schedules combine a minimum pension guarantee with a bene t proportional to an individual's wage. At rst glance, the pension values illustrated in Panel A appear to be sensitive to the reported nal average wage, but it is important to note that in the years leading up to the reform in ation had severely eroded the real value of wages and pensions, such that a large majority of workers had nal average wages in the region in which the minimum was binding. In ation exceeded 50 percent in every year in the volatile 1982-1988 period, and exceeded 100 percent in 1987 and 1988; it was above 25 percent in a number of subsequent years (1990-1991 and 1995-1996). (See Appendix Table A4 (online).) In response to public pressure, the Mexican Congress in 1989 increased the minimum pension to 70 percent of the minimum wage and indexed it to the minimum wage going forward, without raising the value of pensions greater than the minimum.[13] The congress subsequently raised the value of the minimum pension relative to the minimum wage, until it reached 100 percent of the minimum wage in Mexico City in 1995.

As a consequence of the erosion of the real value of pensions above the minimum and the legislative interventions to raise the minimum, the fraction of workers who expected to receive the minimum pension remained high throughout the pre-reform period. Panel B of Figure 3 plots the real value of the pension for male workers with 10, 20 or 30 years of contributions against the nal average wage percentile of 60-65 year old men in the IMSS data, for selected years.[14] In 1990, approximately 80 percent of male retirees with 10 years of contributions received the minimum pension. The corresponding numbers for male workers with 20 or 30 years of contributions were 70 percent and 60 percent respectively. In 1997, just prior to the implementation of the pension reform, nearly all workers with 10 years of contributions, roughly 50 percent of those with 20 years, and 40 percent of those with 30 years could expect to receive the minimum pension.[15]

[13] In 1991, bene ts were indexed to the minimum wage, which slowed the erosion of the values of pensions above the minimum. That is, if a worker's nal average wage was twice the minimum wage in 1991, the pension payment in 1992 was calculated on the basis of twice the minimum wage. The real minimum wage declined steadily over the period (see Appendix Table A4 (online)) so the slowing of the erosion of pensions as a result of this change was modest.

[14] To calculate the nal average wage percentile, we calculate the nominal wage at each percentile of the IMSS wage distribution for 60-65 year old men in each of preceding ve years, then take the average for each percentile.

[15] In addition, there was a penalty for retirement before age 65 of 5 percent per year (i.e. a worker who retired

bene ts under the PAYGO scheme or the PRA scheme. The PAYGO pension is calculated as if workers' post-reform contributions were under the old regime. If a transition worker opts for the PAYGO pension, IMSS appropriates the balance of his or her personal retirement account. The only option for new workers is the PRA.[19]

To illustrate the impact of the reform on pension wealth, we conduct a simulation of pension wealth under the two regimes, based on a similar simulation by Aguila (2011). In carrying out the simulation, we choose a relatively optimistic annual return on the personal accounts: 8.59 percent, the average return from 1998-2002, as in the more optimistic of the two scenarios considered by Aguila (2011). We also assume that participants expected the real value of the minimum wage to decline, as it had done for more than a decade (see Appendix Table A4). Assumptions of lower interest rates and less rapid declines in the real minimum wage would be less favorable to the PRAs. Details of the simulation are in Appendix A.3 (online).

One way to see the di erences in incentives by age in the system is to compare pension wealth for workers of di erent ages in 1997. Table 1 displays the real present value of pension wealth by wage level for male workers of di erent ages in 1997, all of whom began working at age 25 and expect to continue working until age 60, assuming real wages are constant over their lifetimes. Numbers in italics (and in blue where color is available) indicate that the PRA pension is more valuable than the PAYGO pension. The message of the simulation is clear: the PRA pension is expected to be more valuable only for younger workers who expect to contribute to the personal account for 25 or more years, and among these workers the PRA pension is relatively more attractive for higher-wage workers.[20]

We do not attempt to infer from the simulation exact crossing points at which the PRA becomes preferable to the PAYGO pension; any such calculation would be sensitive to assumptions about the path of interest and in ation rates, and it is not clear that workers are sophisticated in calculating the precise values of pensions under the di erent systems. The basic message of the simulation, which we believe was understood by participants at the time of the reform, is that for most workers, conditional on qualifying for the minimum pension under the old regime,

[19] Under the personal-account system, individuals have three options upon retirement. One is to receive programmed withdrawals from the individual's AFORE, where the withdrawal amount is calculated based on the account balance as well as the age and life expectancy of the individual and dependents. (A worker who receives the minimum pension must choose this option.) A second option is to purchase an annuity from a private insurance company that guarantees a xed monthly pension. A third option, available to workers with a personal-account balance exceeding 130 percent of the cost of an annuity providing a monthly payment equal to the minimum pension, is to take a lump-sum payment upon retirement.

[20] Another way to see the e ect of the reform is to consider the values of the pensions for di erent numbers of years of expected contributions, for a worker who entered the system on June 30, 1997, as presented in Appendix Table A5. Note that workers with fewer than 10 years of contributions are better o under the new regime, since they receive no pension under the old regime but a small pension under the new regime. But conditional on a worker having at least 10 years of contributions, we again see that the attractiveness of the PRA pension is increasing in the number of years of contributions and the wage. The median wage for male workers is just above 100 pesos/day, and for a worker at this level the PRA only becomes more attractive if he expects to contribute for more than 25 years.

the OECD in 1992 found that, in part due to various loopholes, 70 percent of corporate tax declarations reported no taxable income (OECD, 1992). By all accounts, tax evasion remains high (OECD, 2011a). In addition, the social security agency and the Mexican tax authority first signed an agreement to share data in June 2002; thus for almost all of the period under study, there was no chance that information reported to the social security agency would affect the corporate tax burden. It appears, in other words, that evaded payroll taxes were not offset by increases in other taxes.

Also, it does not appear that individual income taxes provided a strong disincentive to most workers to have their wages reported accurately. Mexico provides extensive tax credits for low-wage workers, originally instituted to offset the regressive effects of VATs, with the consequence that many workers legally pay no income tax, or even receive funds from the tax authority (i.e. face a negative income tax.) In 1997, for instance, individuals making less than 3.2 times the minimum wage in Mexico City faced a zero or negative tax rate (OECD, 1999, p. 80).

3 Conceptual Framework

To organize our empirical analysis, we have developed a simple partial-equilibrium model of the compliance decisions of heterogeneous firms, in which employees and firms collude in under-reporting (as in Yaniv (1992)) and firms are monopolistically competitive and differ in productivity (as in Melitz (2003)). The model shares with a number of existing models that less-able entrepreneurs, whose firms are smaller, comply less than more-able entrepreneurs (Rauch, 1991; Dabla-Norris, Gradstein, and Inchauste, 2008; De Paula and Scheinkman, 2011; Galiani and Weinschelbaum, forthcoming) but differs in that we consider partial compliance: wage under-reporting by formally registered firms, as opposed to a binary decision about whether to register.[24] To save space in the main text, we have put the full model in Appendix B; here we briefly summarize the main ideas.

Let w_r be the pre-tax wage reported by a firm to the government, w_u the unreported wage (paid \under the table"), and the tax rate (the sum of firm and worker contributions). Then the the net take-home wage received by workers is $w_{net} = w_u + (1 \quad)w_r$. Rearranging,

$$w_u = w_{net} \quad (1 \quad)w_r \tag{1}$$

[24] Three other recent papers discuss heterogeneity of firms' tax-compliance decisions. Kleven, Kreiner, and Saez (2009) consider a particular mechanism that generates greater compliance among larger firms | the increasing difficulty of maintaining collusion as the number of employees increases | but do not focus on differential responses to tax or benefit changes. Besley and Persson (2013, pp. 103-105) note that if compliance costs depend on firm size, then firm heterogeneity will matter for compliance, without taking a position on the source of the firm heterogeneity or on the implication for responses to tax changes. Dharmapala, Slemrod, and Wilson (2011) consider the optimal taxation of firms in a setting with firm heterogeneity and the implications for firm size distributions, but do not focus on wage under-reporting.

In the empirics, w_r will correspond to the wage reported by the rm in the administrative records of the social security agency and w_{net} to the take-home pay reported by workers in the ENEU household survey. As mentioned above and discussed in more detail below, we do not observe w_{net} at the rm level, but we do observe it | and will be able to construct measures of the unreported wage | at a more aggregated level.

We assume that the cost of evasion is given by $xc(w_u)$, where $c'(w_u) > 0$, $c''(w_u) > 0$ and x is the output of the rm. One possible justi cation for this assumption is simply that auditors are more likely to audit larger rms because their operations are more visible, as suggested by Besley and Persson (2013, p. 66) | a conjecture that appears anecdotally to be relevant in Mexico. Another is the argument of Kleven, Kreiner, and Saez (2009) that collusion in under-reporting is more di cult to sustain in larger rms. Whatever the underlying mechanism, the assumptions on the cost-of-evasion function give us our rst key theoretical implication: in equilibrium, more productive rms, which are larger, choose to evade less.

In our static setting, we model the per-period value of the future pension bene t as bw_r, where we call b the \bene t rate." We assume that $b <$, which corresponds to the Mexican institutional setting, where the tax payment includes contributions for health care (which are not sensitive to reported wages) as well as pension bene ts (which may be). This assumption means that there is a rent to not reporting wages at the margin (some of which may be shared with employees); rms will weigh their share of this rent against the costs of evasion. The total e ective wage, inclusive of pension bene ts, which we denote by w_e, is then (using (1)):

$$w_e = w_{net} + bw_r = w_u + (1 \quad (\quad b))w_r \tag{2}$$

We assume that the labor market is competitive and that workers' labor supply responds to the e ective wage, w_e.[25] It can be shown that an increase in the bene t rate, b, will lead rms to rely more heavily on the reported wage, w_r, in the compensation package to achieve a given market-clearing e ective wage. This is our second key theoretical implication: an increase in the pension bene t rate will lead to a decrease in the unreported wage, w_u, within each rm. The model considers homogeneous workers, but could be easily extended to consider more than one type of worker, who di er in the bene t rate they face. We would then expect the unreported wage, w_u, to decline more for workers who face a greater increase in the bene t rate, b.[26]

An important issue in this context is the incidence of the change in the pension bene t rate on wages. Theoretically, it is possible to show that, for a nite labor-supply elasticity, the e ective wage, w_e, is increasing in the bene t rate, b. If b rises, the government ends up paying a larger

[25] We assume that workers observe both w_{net} and w_r, and hence w_u and w_e. In this sense, workers collude in under-reporting in that they are aware of it and do not report it.

[26] An additional implication of the model is that a decrease in the tax rate, , has an analogous e ect to an increase in the bene t rate on compliance; we return to this below.

share of the effective wage and some of this increased contribution redounds to workers. But in general it is not possible to sign the effects of the reform on the observable wage measures, the firm-specific reported wage, w_r, or the firm-specific take-home wage, w_{net}, for reasons discussed in the appendix. It is worth emphasizing, however, that in the model the response of w_u to the policy change does not depend on the incidence of the policy change on w_e, w_r or w_{net}. In this sense, the model suggests that it is reasonable to examine the effect of the policy change on evasion separately from the question of incidence, which is how we proceed in the empirical analysis.

4 Data

The establishments' wage reports are drawn from IMSS administrative records. All private Mexican employers are in principle legally obligated to report wages for their employees, and pay social-security taxes on the basis of the reports. The IMSS dataset contains the full set of wage reports for employees in registered, private-sector establishments over the period 1985-2005.[27] The dataset contains a limited set of variables: age, sex, daily wage (including benefits), state and year of the individual's first registration with IMSS, an employer-specific identifier, and industry and location of the employer. Wages are reported in spells (with a begin and end date for each wage level) and in theory we could construct a day-by-day wage history for each individual. To keep the dataset manageable, we extract wages for a single day, June 30, in each year. Prior to 1997, records for temporary workers were not collected in digital form. To ensure comparability before and after 1997, we focus on workers identified in the IMSS data as permanent, defined as having a written contract of indefinite duration.

We select ages 16-65. To maintain consistency across years, we impose the lowest real value of the IMSS topcode for wage reporting (which occurred in 1991) in all years. We drop establishments with a single insured worker, since these are likely to be self-employed workers. In the interests of comparability with the ENEU household data, we include only the metropolitan areas included in the ENEU samples (described below). We also focus on sectors for which we are confident that IMSS is the only available formal-sector social insurance program: manufacturing, construction, and retail/hotel/restaurants. Other broad sectors contain a substantial share of public employees, who are typically covered by a separate system.[28] We focus primarily on men, for the reasons discussed in Section 2 above. (Results for women are reported in Appendix D (online).) When individuals have more than one job, we select the highest wage job. We refer to the sample selected following these criteria as our IMSS baseline sample. Further details on sample selection and data processing in Appendix C (online).

[27] The data have been used in several previous papers, including Castellanos, Garcia-Verdu, and Kaplan (2004), and Fr as, Kaplan, and Verhoogen (2009).

[28] We focus on manufacturing, construction, and retail/hotel/restaurants in part so that we can be confident that respondents to the household survey are not mistaking coverage under the public-sector system for IMSS coverage.

The household data we use are from the *Encuesta Nacional de Empleo Urbano* (ENEU) [National Urban Employment Survey], a household survey modeled on the Current Population Survey (CPS) in the United States, collected by the *Instituto Nacional de Estad sticas y Geograf a* (INEGI), the Mexican statistical agency. The original ENEU sample, beginning in 1987, focused on the 16 largest Mexican metropolitan areas; although the coverage expanded over time, to maximize the number of pre-reform years we focus on the original 16 areas. As in the IMSS data, we include male workers ages 16-65, focus on the second quarter of each year, exclude self-employed workers, impose the 1991 IMSS topcode in all years, and include only manufacturing, construction, and retail/hotels/restaurants. When individuals report having more than one job, we use the information only from their main job. All calculations below use the sampling weights provided by INEGI.

A very useful feature of the ENEU for our purposes is that it asks respondents whether they receive IMSS coverage as an employment bene t. Beginning in the third quarter of 1994, the ENEU also asked respondents whether they had a written contract of inde nite duration, the

sample, containing all non-self-employed men satisfying the age and sector criteria. Comparing columns 3 and 4, we see that ENEU workers with IMSS coverage tend to be higher-wage and more likely to work in large establishments than workers without IMSS coverage. Column 5 contains the sample that in principle should be the best match for the IMSS baseline sample: ENEU workers who report receiving IMSS coverage and having a written contract of inde nite duration | that is, who satisfy the de nition of \permanent" used by IMSS. The average wage for this ENEU sample is greater than for the IMSS baseline sample, consistent with our argument below that there is under-reporting of wages in the IMSS data. Because the contract-type variable is available only beginning in 1994, however, we have prohibitively few years of pre-reform data for this sample. Instead, we will focus hereafter on the Column 6 sample, ENEU workers who report receiving IMSS coverage and working full-time (i.e. at least 35 hours in the previous week), which can be de ned consistently over the entire period. We refer to the Column 6 sample as our ENEU baseline sample.

It is important to recognize that there are a number of reasons why the IMSS and ENEU baseline samples may di er. Some temporary workers may work full-time, and some permanent workers may work part-time. Comparing Columns 5 and 6 in Table 2 for the year 2000, we see that average wages are signi cantly lower in the Column 6 sample; this is attributable to the facts that temporary full-time workers earn relatively low wages and that permanent part-time workers earn relatively high wages on average. It may also be that rms interpret \permanent" to mean something di erent from the legal de nition (i.e written contract of inde nite duration) when reporting wages. In addition, patterns of non-response may di er between the IMSS and ENEU samples. It is well known, for instance, that richer households tend to be less likely to respond to income questions in household surveys (Groves and Couper, 1998). The weighted employment totals from the ENEU data in Columns 5 and 6 are below the IMSS totals in Column 1; this may in part re ect such non-response.[31]

To further explore the employment discrepancy, Figure 5 plots employment totals over the 1988-2003 period for the same samples as in Table 2. We see that over most of the period the number of workers in the IMSS sample is slightly *greater* than the numbers in any of the ENEU samples. In addition to non-response in the ENEU, this di erence likely re ects that fact that the IMSS sample is based on place of work while the ENEU sample is based on place of residence; hence people who commute in to metropolitan areas are included in the IMSS data but not in the ENEU. Another possibility is that some respondents are unaware that they receive IMSS coverage from their employer, or believe that they are covered by the public-sector social security agency (known by the acronym ISSSTE) when in fact they are covered by IMSS. For our purposes, however, the most important lesson of the gure is that there does not appear to have been a

[31] Note, however, that non-response by richer households will tend to lead us to understate evasion, making it more di cult for us to pick up statistically signi cant di erences in cross-section.

large change over time in the extent of the employment discrepancy between the IMSS and ENEU samples in response to the pension reform. Nor does it appear that there was a signi cant large in ow to (or out ow from) formal employment in response to the pension reform.

To further explore the comparability of the IMSS and ENEU samples, we conduct two additional checks. First, Table 3 compares the distributions in each sample across two dimensions that will be important in our analysis, age and rm size. In order to ensure that we have su cient sample size in the ENEU to calculate the evasion measures below, we group individuals into ve age categories (ages 16-25, 26-35, 36-45, 46-55, 56-65). Comparing the rightmost columns for the two panels, which indicate the share of employment in each rm size category as a share of total employment, it appears that rm sizes in the ENEU are skewed slightly away from the smallest and toward the largest size category (although there is non-monotonicity at intermediate sizes.) This may be because respondents in the household survey do not distinguish between employees directly hired by their employer and sub-contracted employees, or simply that respondents systematically overestimate employment. It may also be that rms under-report employment to IMSS, although the patterns of employment di erences in Table 2 and Figure 5 tend to cast doubt on this interpretation. The distributions of employment across age groups conditional on a particular rm-size category also reveal some di erences. In general, in the ENEU it appears that employment in smaller rms is shifted a bit toward younger workers relative to the IMSS (with the opposite shift among larger rms). But the overall distributions across age categories (in the \all rm sizes" rows) appear to be fairly similar.

ature is that the larger the number of reports the tax authorities receive on rms' tax liabilities, the more di cult it is for rms to evade (Kopczuk and Slemrod, 2006).[36] The plant-level datasets from INEGI do not contain individual-level wage information on the full distribution of wages, but it is nevertheless instructive to consider the reports of covered plants to the social security agency. Figure 11 plots the IMSS wage distribution for workers in manufacturing establishments that also

corresponding to ages 16-25, and the last column, corresponding to ages 56-65. There is a clear decline in bunching and shift to the right of the IMSS distribution for the youngest age group. For the oldest age group, there is little evident decline in bunching or shift to the right in the IMSS distribution.

It is important to note that there were a number of macroeconomic events that a ected wages over our period and these may have a ected the IMSS and ENEU distributions di erently. In particular, the peso crisis of late 1994 and 1995 lead to a decline in average real wages of approximately 25 percent. It took several years for wages to regain their 1994 levels in real terms.

Dharmapala, D., J. Slemrod, and J. D. Wilson (2011): \Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-Level Administrative Costs," *Journal of Public Economics*, 95(9-10), 1036 { 1047.

Duarte, F., and J. Hastings (2010): \Fettered Consumers and Sophisticated Firms: Evidence from Mexico's Privatized Social Security Market," Unpub. paper, March.

Fisman, R., and S.-J. Wei (2004): \Tax Rates and Tax Evasion: Evidence from Missing Imports in China," *Journal of Political Economy*, 112(2), 471{500.

Fr as, J. A., D. S. Kaplan, and E. A. Verhoogen (2009): \Exports and Wage Premia: Evidence from Mexican Employer-Employee Data," Unpub. paper, Columbia University.

Galiani, S., and F. Weinschelbaum (forthcoming): \Modeling Informality Formally: Households and Firms," Forthcoming, *Economic Inquiry*.

Gordon, R., and W. Li (2009): \Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation," *Journal of Public Economics*, 93(7-8), 855 { 866.

Gorodnichenko, Y., J. Martinez-Vazquez, and K. S. Peter (2009): \Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion Response and Welfare E ects in Russia.," *Journal of Political Economy*, 117(3), 504 { 554.

Grandolini, G., and L. Cerda (1998): \The 1997 Pension Reform in Mexico," World Bank Policy Research Working Paper no. 1933.

Groves, R. M., and M. P. Couper (1998): *Nonresponse in Household Interview Surveys*. John Wiley & Sons, New York.

Hsieh, C.-T., and P. J. Klenow (2014): \The Life Cycle of Plants in India and Mexico," *Quarterly Journal of Economics*.

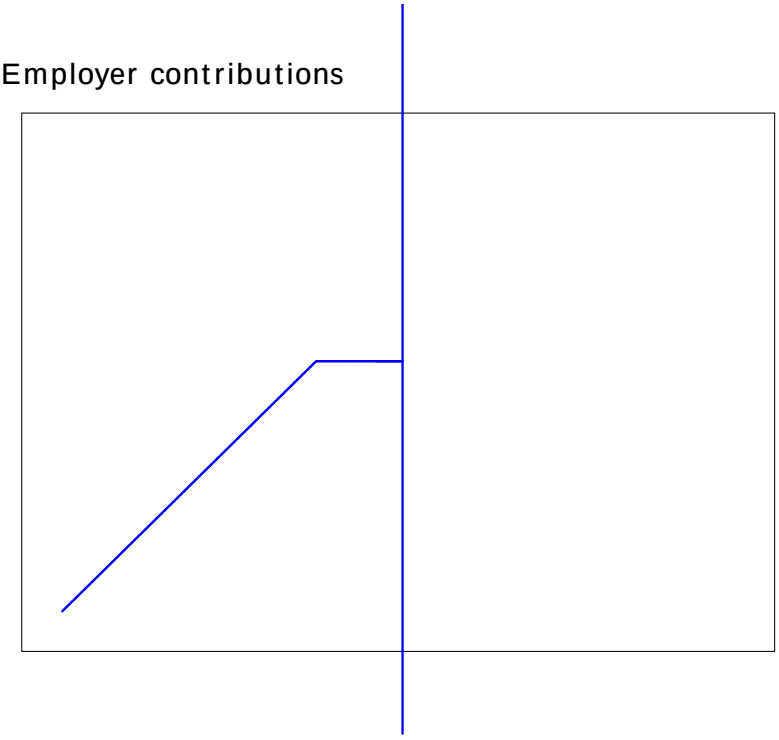
Hsieh, C.-T., and B. A. Olken (2014): \The Missing\Missing Middle"," *Journal of Economic Perspectives*, 28(3), 89{108.

Hurst, E., G. Li, and B. Pugsley (2011): \Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed," Unpub. paper, University of Chicago.

IMF (2010): *Government Finance Statistics Yearbook*. International Monetary Fund Statistics Department.

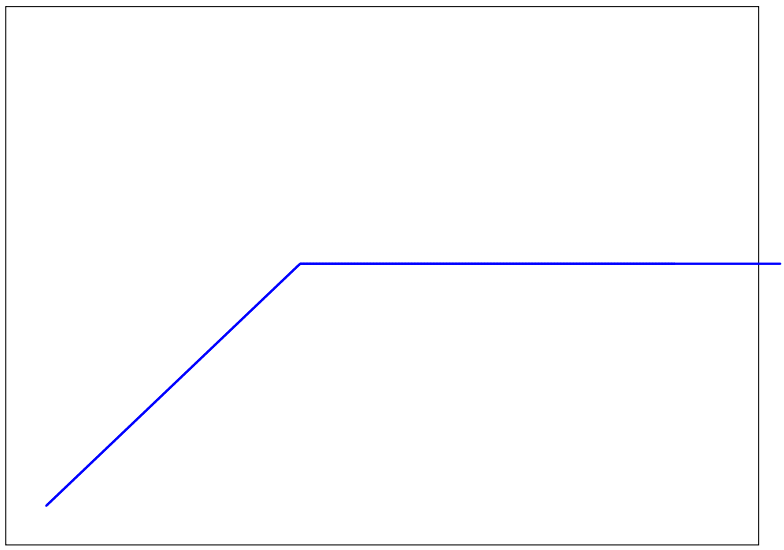
IMSS (2011): \Informe al Ejecutivo Federal y al Congreso de la Union sobre la Situation Financiera y los Riesgos del Instituto Mexicano del Seguro Social, 2010-2011," Mexico City, Mexico.

Figure 1. Employer contributions



Notes: Variation in IMSS employer contribution rates at levels above 500 pesos/day are primarily due to changes in topcodes, which varied from 10 to 25 times the minimum wage in Mexico City over the period. Average 2002 exchange rate: 9.66 pesos/dollar.

Figure 2. Employee contributions

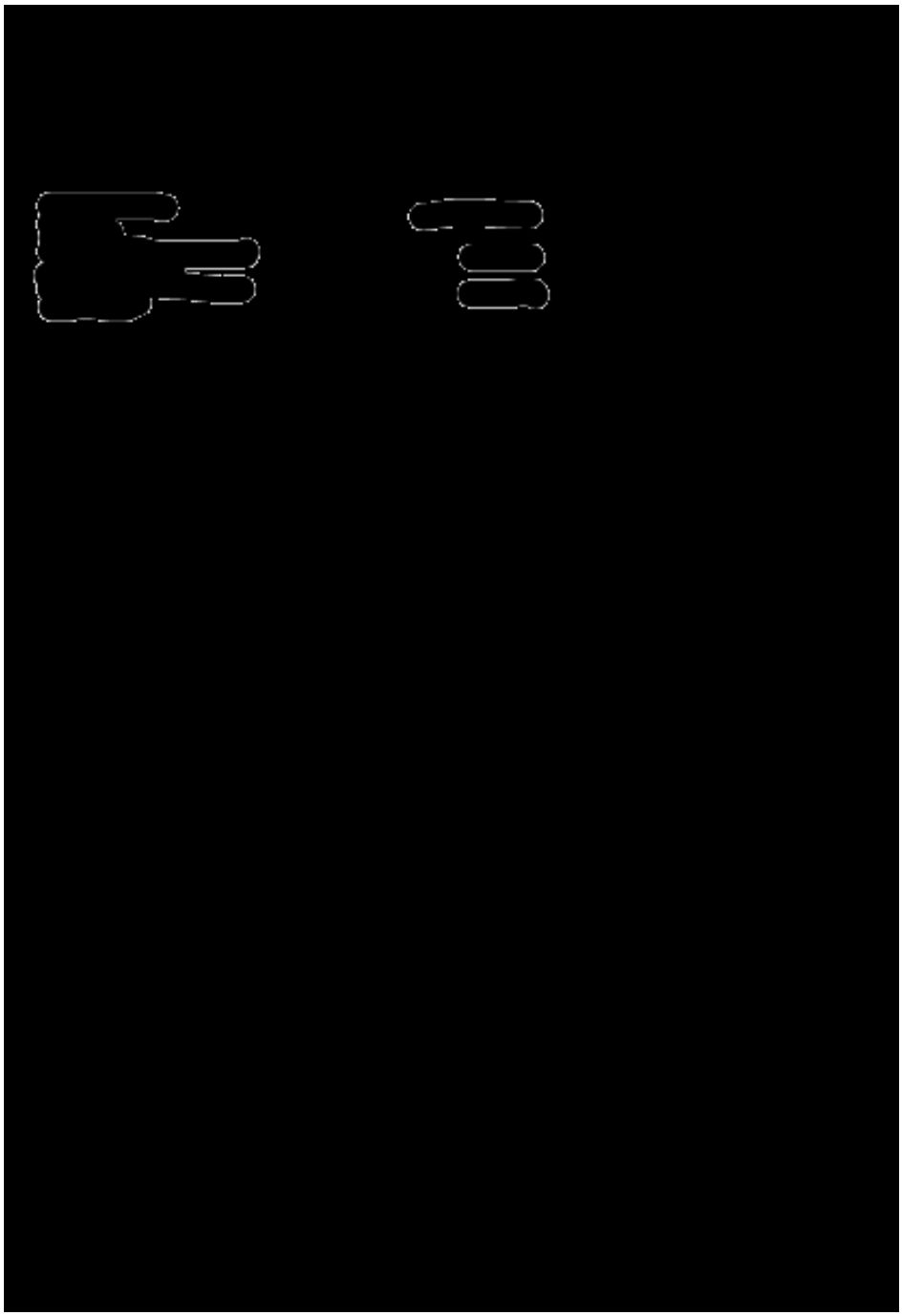


Notes: Variation in IMSS worker contribution rates at levels above 500 pesos/day are primarily due to changes in topcodes, which varied from 10 to 25 times the minimum wage in Mexico City over the period. Average 2002 exchange rate: 9.66 pesos/dollar.

Figure 3. Pension values, selected years, men

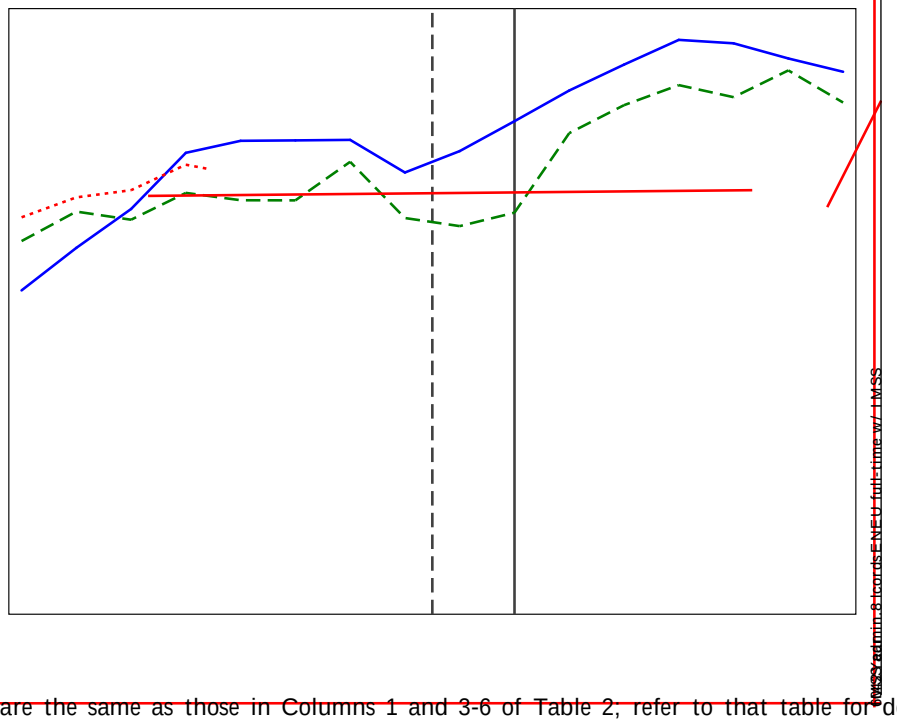
Notes: Final average wage (2002 pesos/day) is average nominal daily wage over ve years prior to retirement, de ated to constant 2002 pesos. Figure indicates pension values for individuals with 10, 20 and 30 years of contributions to IMSS. In Panel B, we calculate the nominal wage at each quantile of the IMSS wage distribution for 60-65 year old men in each year and take the average for that quantile over the preceding ve years. Panel C is constructed similarly using wage distributions from the ENEU baseline samples. See Section 4 for details of samples and Section 2.3 for details on pension bene ts. Average 2002 exchange rate: 9.66 pesos/dollar.

Figure 4. Account statement



Notes: The box at top right (\Cuanto tengo en mi cuenta individual") reports total balance. The rst row of boxes in the middle section (\Mi ahorro para el retiro") pertains to the retirement pension and reports previous balance (\Saldo anterior"), new contributions (\Aportaciones"), withdrawals (\Retiros"), interest earned (\Rendimientos"), AFORE commission charged (\Comisiones"), and nal balance (\Saldo nal"). The second and third rows in the middle section report balances in the individual's voluntary savings account and housing account. The bottom section reports 3-year returns and commissions for each AFORE, as well as the average 5-year net return (at left).

Figure 5. Employment, IMSS admin. records vs. ENEU household data, men

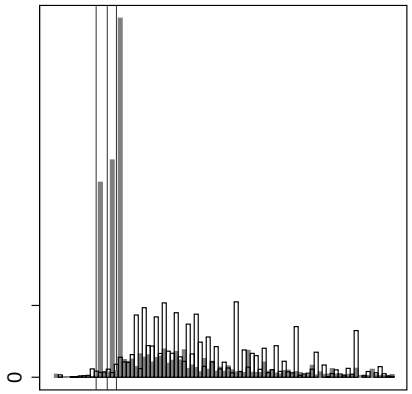


Notes: Samples are the same as those in Columns 1 and 3-6 of Table 2, refer to that table for details. ENEU totals are calculated using sampling weights. The dashed vertical line indicates the date the pension reform was passed by Congress (Dec. 21, 1995); the solid vertical line indicates the date the reform took e ect (July 1, 1997). Observations correspond to the second quarter of each year. See Section 4 and Appendix C (online) for details of sample selection.

Figure 7. Wage histograms, men, 1990

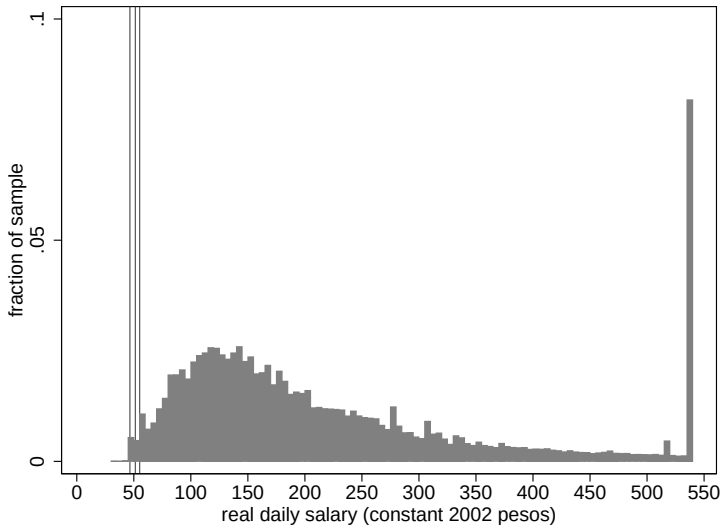
Notes: Samples are IMSS and ENEU \baseline" samples of men. Data in both samples are from second quarter of 1990. As in Figure 6A, the IMSS wage is the real daily

Figure 9. Wage histograms by rm size, men, 1990, low wage levels



Notes: Histograms are similar to those in Figure 8. Vertical lines indicate minimum wages in the three minimum-wage zones in Mexico (A, B, C). Bins are 2 pesos wide. Average 2002 exchange rate: 9.66 pesos/dollar. See Section 4 and Appendix C (online) for further details.

Figure 11. Wage histograms, men, 1993, IMSS data, establishments linked to EIA

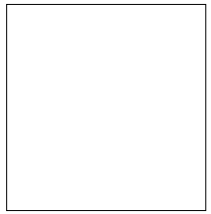


Notes: Sample is permanent male workers ages 16-65 in IMSS data in 2389 establishments that can be linked to a balanced 1993-2003 panel from the *Encuesta Industrial Anual (EIA)* [Annual Industrial Survey], which excludes assembly-for-export *maquiladora* plants. Data are from second quarter. Vertical lines indicate minimum wages in the three minimum-wage zones in Mexico (A, B, C). Bins are 5 pesos wide. See Section 4 and Appendix C (online) for further details of data processing.

, establishments linked to

Notes: Sample is permanent male workers ages 16-65 in IMSS data in 520 establishments that can be linked to a balanced 1993-2003 panel from the *Estadsticas Mensuales de la Industria Maquiladora de Exportacion (EMIME)* [Monthly Statistics on Maquiladora Export Industry], a dataset made up exclusively of assembly-for-export *maquiladora* plants. Data are from second quarter. Vertical lines indicate minimum wages in the three minimum-wage zones in Mexico (A, B, C). Bins are 5 pesos wide. See Section 4 and Appendix C (online) for further details of data processing.

Figure 13. Wage densities by age group, 1990, 1997, 2003, men



Notes: The wage variables are the real daily take-home wage from ENEU and real daily *post-tax* reported wage from IMSS. Densities are estimated and an Epanechnikov kernel with bandwidth 3 pesos for IMSS data and 6 pesos for ENEU data (using Stata kdensity command). Wages are in 2002 pesos. Average 2002 exchange rate: 9.66 pesos/dollar. Rows correspond to years 1990, 1997, 2003; columns to age groups 16-25, 26-35, 36-45, 46-55, 56-65.

Figure 14. Wage gaps (medians) by age group, men, deviated from metro-year means

Table 1. Pension wealth simulation, by age in 1997, male worker with 35 years of expected contributions

Age in 1997	Years of Expected PRA Contributions	Plan	Real Daily Wage					
			43	100	200	300	500	1079
25	35	PRA	398.6	*815.0*	*1626.2*	*2437.3*	*4059.7*	*8751.9*
		PAYGO	398.6	*398.6*	*603.8*	*890.2*	*1483.6*	*3200.1*
30	30	PRA	398.6	*523.4*	*1044.3*	*1565.3*	*2607.1*	*5620.5*
		PAYGO	398.6	*398.6*	*603.8*	*890.2*	*1483.6*	*3200.1*
35	25	PRA	398.6	398.6	*659.1*	*987.8*	*1645.3*	*3546.9*
		PAYGO	398.6	398.6	*603.8*	*890.2*	*1483.6*	*3200.1*
40	20	PRA	398.6	398.6	403.9	605.4	1008.4	2173.9
		PAYGO	398.6	398.6	603.8	890.2	1483.6	3200.1
45	15	PRA	398.6	398.6	398.6	398.6	586.6	1264.7
		PAYGO	398.6	398.6	603.8	890.2	1483.6	3200.1
50	10	PRA	398.6	398.6	398.6	398.6	398.6	662.6
		PAYGO	398.6	398.6	603.8	890.2	1483.6	3200.1
55	5	PRA	398.6	398.6	398.6	398.6	398.6	398.6
		PAYGO	398.6	398.6	603.8	890.2	1483.6	3200.1

Notes: Values are real present discounted value of the future stream of pension bene ts in thousands of 2002 pesos, for a male worker who began contributing at age 25 and expects to continue until age 60. Numbers in italics (and blue where color is available) indicate that personal retirement account (PRA) has a higher expected payo than the pre-reform pension (PAYGO). Average 2002 exchange rate: 9.66 pesos/dollar. 43 pesos is real daily minimum wage (in Mexico City)

Table 2. Comparison of IMSS baseline sample and various ENEU samples, men

	IMSS baseline sample (1)	full ENEU sample (2)	ENEU w/ IMSS (3)	ENEU w/o IMSS (4)	ENEU permanent w/ IMSS (5)	ENEU full-time w/ IMSS (6)
A. 1990						
real avg. daily post-tax wage	121.02	163.88	172.98	143.88		166.73
	(0.07)	(1.58)	(1.94)	(2.62)		(1.85)
age	31.75	31.46	32.13	29.98		32.22
	(0.01)	(0.15)	(0.17)	(0.29)		(0.17)
fraction employed in ests >100 employees	0.52	0.43	0.55	0.18		0.55
	(0.00)	(0.01)	(0.01)	(0.01)		(0.01)
N (raw observations)	1691417	16169	11592	4577		10978

Table 3. Age composition by rm size category, 1990, men

	Age category (employment as % of row total)					
	16-25	26-35	36-45	46-55	56-65	employment as % of column total
A. IMSS						
1-10 employees	29.9	32.6	19.8	11.9	5.8	14.5
11-50 employees	33.6	32.2	18.7	10.6	4.9	22.6
51-100 employees	35.0	32.5	18.5	9.8	4.2	10.8
101-250 employees	36.3	33.3	17.8	9.0	3.5	14.7
> 250 employees	37.7	34.8	17.5	7.6	2.5	37.5
all rm sizes	35.1	33.4	18.3	9.3	3.8	
B. ENEU						
1-10 employees	35.9	28.3	18.0	12.5	5.3	12.4
11-50 employees	33.5	33.3	18.4	10.3	4.5	21.0
51-100 employees	35.6	33.4	15.2	10.7	5.1	11.6
101-250 employees	30.2	31.2	21.5	12.4	4.7	10.5
> 250 employees	34.0	33.4	21.5	8.5	2.7	44.5
all rm sizes	33.9	32.5	19.7	10.1	3.9	

Table 4. Cross-sectional patterns of evasion, 1990, men

	wage gap (medians)			wage gap (means)			exc. mass (15th percentile)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
age 26-35	-0.054*		-0.053**	-0.081***		-0.081***	-0.154***		-0.154***
	(0.029)		(0.021)	(0.024)		(0.019)	(0.017)		(0.013)
age 36-45	-0.072**		-0.072***	-0.150***		-0.150***	-0.170***		-0.170***
	(0.034)		(0.027)	(0.028)		(0.024)	(0.017)		(0.014)
age 46-55	-0.029		-0.025	-0.154***		-0.151***	-0.154***		-0.152***
	(0.035)		(0.031)	(0.031)		(0.027)	(0.018)		(0.015)
age 56-65	-0.020		-0.026	-0.167***		-0.174***	-0.117***		-0.119***
	(0.044)		(0.040)	(0.038)		(0.035)	(0.020)		(0.017)
11-50 employees		-0.333***	-0.333***		-0.177***	-0.176***		-0.155***	-0.154***
		(0.026)	(0.024)		(0.025)	(0.023)		(0.011)	(0.010)
51-100 employees		-0.475***	-0.469***		-0.283***	-0.280***		-0.247***	-0.242***
		(0.033)	(0.031)		(0.030)	(0.028)		(0.015)	(0.014)
101-250 employees		-0.395***	-0.374***		-0.245***	-0.233***		-0.235***	-0.224***
		(0.039)	(0.037)		(0.035)	(0.032)		(0.018)	(0.016)
> 250 employees		-0.500***	-0.464***		-0.233***	-0.200***		-0.288***	-0.268***
		(0.035)	(0.034)		(0.030)	(0.030)		(0.018)	(0.017)
construction			0.134***			0.122***			0.064***
			(0.029)			(0.025)			(0.014)
retail/services			-0.074***			-0.110***			-0.043***
			(0.024)			(0.021)			(0.011)
constant	0.559***	0.855***	0.633***	0.501***	0.577***	0.506***	0.519***	0.578***	0.566***
	(0.017)	(0.018)	(0.047)	(0.016)	(0.018)	(0.039)	(0.010)	(0.007)	(0.019)
metro area e ects	N	N	Y	N	N	Y	N	N	Y
R-squared	0.00	0.20	0.31	0.03	0.09	0.27	0.09	0.24	0.44
N	1062	1062	1062	1062	1062	1062	1062	1062	1062

Notes: Data are from IMSS and ENEU baseline samples, collapsed to metro area/age group/ rm-size category/sector level for 1990. The omitted category for age is 16-25, for rm size is 1-10 employees, and for sector is manufacturing. The wage gap (medians) is log median real daily take-home wage from the ENEU minus log median real daily post-tax reported wage from IMSS, calculated. Wage gap (means) is analogous, using mean in place of median. Excess mass is calculated as described in Section 5 and Figure 10. In calculating evasion measures, we pool ENEU data across quarters within year. *** 1%, ** 5%, * 10% level. See Section 4 and Appendix C (online) for further details of data processing.

Table 5. Di erential e ects of pension reform on evasion, men

	wage gap (medians) (1)	wage gap (means) (2)	excess mass (15^{th} perc.) (3)
1(age <= 55)*1988	0.050	-0.069**	0.019
	(0.037)	(0.031)	(0.020)
1(age <= 55)*1989	0.030	-0.077**	0.004
	(0.042)	(0.036)	(0.016)
1(age <= 55)*1990	0.038	-0.089**	-0.002
	(0.040)	(0.038)	(0.015)
1(age <= 55)*1991	0.047	-0.069*	-0.020
	(0.039)	(0.040)	(0.014)
1(age <= 55)*1992	0.069	-0.016	0.007
	(0.043)	(0.042)	(0.016)
1(age <= 55)*1993	0.067*	-0.032	0.014
	(0.040)	(0.038)	(0.017)
1(age <= 55)*1994	0.011	-0.062*	0.019
	(0.045)	(0.035)	(0.016)
1(age <= 55)*1995	0.106**	-0.029	0.017
	(0.045)	(0.031)	(0.017)
1(age <= 55)*1996	-0.019	-0.087*	-0.023
	(0.040)	(0.046)	(0.017)
1(age <= 55)*1998	-0.042	-0.093***	-0.023*
	(0.037)	(0.035)	(0.014)
1(age <= 55)*1999	-0.048	-0.129***	-0.035**
	(0.041)	(0.036)	(0.015)
1(age <= 55)*2000	-0.041	-0.133***	-0.034**
	(0.039)	(0.029)	(0.013)
1(age <= 55)*2001	-0.095**	-0.181***	-0.045***
	(0.047)	(0.039)	(0.015)
1(age <= 55)*2002	-0.137***	-0.218***	-0.032**
	(0.039)	(0.034)	(0.015)
1(age <= 55)*2003	-0.087**	-0.204***	-0.029*
	(0.040)	(0.035)	(0.015)
age group-metro area e ects	Y	Y	Y
metro-year e ects	Y	Y	Y
R-squared	0.92	0.89	0.96
N	1280	1280	1280

Notes: Data are from IMSS and ENEU baseline samples, collapsed to metro area/age group/year level. Wage gap (medians) is log median real daily net wage from ENEU minus log median post-tax daily wage from IMSS. Wage gap (means) is de ned analogously, using means in place of medians. Excess mass is calculated as described in Section 5 and Figure 1025ti calculated as described int 71t 71withi]0 d 0 red1 re [-301(10)]TJ0 g 0 G 0 -4(10)]TJ0 g 0 G 0ro a3(A)-2pp

Table 6. Di erential e ects of pension reform on employment gap, men

	dep. var.: log(empl., ENEU) - log(empl., IMSS)	
	(1)	(2)
1(age <= 55)*1988	0.035	0.035
	(0.100)	(0.090)
1(age <= 55)*1989	-0.040	-0.040
	(0.104)	(0.087)
1(age <= 55)*1990	-0.065	-0.065
	(0.097)	(0.091)
1(age <= 55)*1991	-0.100	-0.100
	(0.109)	(0.098)
1(age <= 55)*1992	-0.044	-0.044
	(0.100)	(0.083)
1(age <= 55)*1993	-0.090	-0.090
	(0.092)	(0.076)
1(age <= 55)*1994	0.231**	0.231***
	(0.101)	(0.082)
1(age <= 55)*1995	-0.017	-0.017
	(0.108)	(0.093)
1(age <= 55)*1996	-0.003	-0.003
	(0.102)	(0.092)
1(age <= 55)*1998	-0.042	-0.042
	(0.104)	(0.092)
1(age <= 55)*1999	-0.027	-0.027
	(0.106)	(0.096)
1(age <= 55)*2000	0.011	0.011
	(0.094)	(0.084)
1(age <= 55)*2001	-0.009	-0.009
	(0.105)	(0.098)
1(age <= 55)*2002	-0.087	-0.087
	(0.103)	(0.089)
1(age <= 55)*2003	-0.033	-0.033
	(0.091)	(0.080)
age group e ects	Y	
age group-metro area e ects	N	Y
metro-year e ects	Y	Y
R-squared	0.55	0.68
N	1280	1280

Notes: Samples are IMSS and ENEU baseline samples, collapsed to metro area/age group/year level. *** 1%, ** 5%, * 10% level. See Section 4