School-Linked Social Insurance and Labor Market Effects: Evidence from Mexico

Alain Pineda*

February 20, 2024

Click here for the latest version

Abstract

In 2016, Mexico initiated a school-linked social insurance policy, providing more than 6 million public high school and college students with basic medical benefits through the national social insurance institution (IMSS). This policy is especially pertinent in a context where 60% of Mexico's workforce is employed informally, and policymakers and international institutions commonly worry that social protections not linked to formal employment might incentivize informality. With significant state-level heterogeneity in informality rates ranging from 40% to 80%, this paper combines differences across cohorts in exposure to the policy with differences across states in initial informality conditions to scrutinize the policy's impact on early labor market outcomes. The results indicate that in states with higher initial informality rates, exposed cohorts exhibit increased labor force participation and formality rates. Specifically, a 1 percentage point rise in initial state informality is associated with a 0.8 to 1.3 percentage point increase in formality among the impacted groups. The policy does not significantly influence school enrollment or monthly income, and its effects are not gender-specific. While the formalization effect recedes post COVID-19, I find no evidence that the policy skews young workers toward informality in the early stages of their labor trajectories.

*Stanford University, Department of Economics. alainpin@stanford.edu. I thank Nick Bloom, Caroline Hoxby, and Ran Abramitzky for invaluable guidance and support. I gratefully acknowledge financial support from the Knight-Hennessy Scholars program and the E.S. Shaw and B.F. Haley Fellowship for Economics through a grant to the Stanford Institute for Economic Policy Research.

1 Introduction

Informal work is the set of productive activities performed by individuals that is not properly registered by the government. According to the International Labour Organization, more than 2 billion workers worldwide are in informal employment, representing over 60% of the workforce (ILO, 2018). While informal employment is present across all countries, it is a key characteristic of developing labor markets. High rates of informality are associated with higher incidence of poverty and job insecurity. On a macro level, informality represents an important obstacle for tax collection and social insurance provision. Informal workers usually lack access to social protection programs that provide a safeguard against negative shocks such as disease or unemployment. A common concern among policymakers is the effect on informality of policies that provide social insurance not linked to employment, as they could introduce distortions in the incentives to work formally vs. informally. This has gained more relevance as working towards universal social protection schemes has become a policy objective, showcased by the Global Partnership for Universal Social Protection to Achieve the Sustainable Development Goals (USP2030), led by the World Bank Group and the International Labour Organization. It is important to understand how informality interacts with public policy in order to identify strategies that can increase social protection of vulnerable populations and support the transition to the formal economy.

In this paper, I study a school-linked social insurance policy enacted in Mexico. In 2016, students enrolled in public high schools began receiving medical benefits associated with social insurance and formal jobs while they remained in school. Using information from the quarterly National Survey of Occupation and Employment (ENOE), I evaluate the impact of gaining access to these benefits on early outcomes of the labor market of the exposed population. I am particularly interested in assessing whether there is evidence of a distortion towards informality from this policy. Early experiences in the labor market are important as there is evidence of high persistence of informality (Akay and Khamis, 2012) and of *scarring* effects of youth informality (Cruces et al., 2012) on future employment and earnings. The Mexican labor market provides a good scenario for this study, as it has a large share of informal employment (close to 60%), with significant heterogeneity across states.

I combine differences across states in the prevalence of informality prior to the social insurance expansion with differences in exposure to the policy across birth cohorts. I

find that following the expansion of medical benefits associated with social insurance, exposed cohorts in states with higher initial informality show higher participation and formality rates. Specifically, estimates suggest that an increase of 1 percentage point in the initial informality rate at the state level is associated with an increase of 0.4 to 0.6 percentage points in the participation rate and an increase of 0.8 to 1.3 percentage points in the formality rate among exposed cohorts. However, these effects dissipate after the COVID-19 pandemic. I find no significant impact on school enrollment and employment rates or monthly income. Furthermore, the positive impacts on participation and formality are similar for workers regardless of school enrollment or the formality of the head of the household. This suggests that students who received medical benefits through their school were not more likely to have an informal job despite the reduction in the relative cost of an informal job versus a formal job. The results are similar between men and women and robust to an alternative definition of informality that excludes the self-employed.

The remainder of the paper is organized as follows. In Section 2, I describe the policy and the potential mechanisms. In Section 3, I briefly discuss the relevant literature. Section 4 is devoted to the description of the data, research design, and presentation of the main results. Section 5 presents results by school enrollment, gender, and an alternative definition of informality. Section 6 concludes.

2 Expansion of Social Insurance Benefits for Students

In late 2015 the federal government of Mexico established a health insurance plan for all students enrolled in public high schools and universities. Students would be able to enroll in the national social insurance institution (IMSS) through a unique social security number that would remain the same for the rest of their lives.[1] The students had the right to start receiving basic medical benefits while remaining enrolled in a public school. Concretely, by being enrolled in IMSS, students had free access to medical consultations, lab tests, medication provision, certain surgical interventions, pregnancy support, and guidance on nutrition, addictions, and sexual education.

The following year there was a widespread campaign titled "Tienes IMSS" (You Have IMSS) where the federal government encouraged state governments to enroll their stu-

[1]The social security number in Mexico is not used as the main identification number. However, it is required to be able to be employed formally.

dents in the social insurance institution as well as promoting the use of the benefits among students. There was a widespread media campaign and big political events with the presence of the president, cabinet members, and governors of the states announcing the enrollment progress. Between March and August 2016, over 6 million students were enrolled in IMSS, covering the universe of students in public high schools and colleges. Figure A.1 in the Appendix shows the enrollment evolution. The coverage rate of the population in high school age is 74% and out of all high school students, about 80% of them are enrolled in a public institution.[2] Thus, this policy affected a significant portion of the young population. In the Mexican context, the main difference between formal and informal employment is the lack of access of the latter to social insurance benefits (mainly health insurance and retirement funds). Having access to healthcare is considered to be the top benefit of having a formal job.[3] Given the close link between employment status and healthcare access in the Mexican context, this policy gives rise to the question of its impact on labor market outcomes. Of particular interest is the informality rate among the affected population. Namely, the research question is what is the impact of providing access to social insurance benefits attached to school enrollment on early labor market outcomes. Is there evidence of distortions towards informality?

Ex-ante, this policy has an uncertain impact on the early labor market outcomes of the exposed population. I will lay out three potential mechanisms arising from this policy. First, the new medical benefits attached to school enrollment increased the value of education. Therefore, we would expect students to remain longer at school and could delay their entry to the labor market, thus lowering their participation.

Second, the new medical benefits could be seen as a complement of employment compensations for students. Given that one of the main disadvantages of an informal job is the lack of social insurance benefits, this policy reduced the relative cost of working at an informal job as a student, given that now they have access to some of these benefits by being enrolled in school. It could even be the case that students use the medical benefits as leverage to request a higher wage in substitution of benefits facilitated by the employer, increasing the value of working while being a student. Employers could also exploit the complementarities of the new medical benefits and decrease the formal job offers relative to informal ones. Following this potential mechanism, we could expect an increase in the participation and a decrease in the formality of the exposed population. An unintended

[2] Secretaría de Educación Pública. *Principales Cifras del Sistema Educativo Nacional 2015-2016.*
[3] *Módulo de Trayectorias Laborales (MOTRAL) 2015.*

consequence of this policy could be an implicit subsidy to informality.

Third, there is an important information and salience component of the policy regarding social insurance benefits attached to formality. If prior to the policy students did not know what were the concrete benefits associated with social insurance and formal jobs, then an increased salience of them might make the students more likely to look for a formal job in order to retain the benefits after they stop attending school. Increased information and salience of their legal rights could also have a potential positive impact on the bargaining power of the affected population when negotiating employment compensation with potential employers. Through this mechanism, we would expect an increase in the formality of the exposed population. Given the ex-ante uncertainty, it remains as an empirical exercise to assess what is the impact of the policy on the labor market outcomes of the affected population.

3 Literature Review

This paper is most closely related to the literature studying the relationship between social protection programs that are not attached to employment and labor market outcomes, where there is a debate of their impact. In developed settings, the effect of welfare programs on labor supply has been extensively studied. The majority of the studies consider conditional cash transfer (CCT) programs and find evidence of a disincentive to work among beneficiaries (Moffitt (2002), Meghir and Phillips (2008)). In the case of Latin America, Alzúa et al. (2013) find evidence for three CCT programs of negative but small and insignificant effects on the labor supply of the targeted population. However, they point to differentiated effects across genders and within the household. Alternatively, Gerard et al. (2021) find that an expansion of *Bolsa Familia* in Brazil increased local formal employment.

In the case of Mexico, there have been several studies that focused on the impact of the *Seguro Popular* program, which provided medical benefits to poor households. On one side, one of the main exponents of the argument that certain forms of social insurance can cause substantial increases in informality is Santiago Levy (2007, 2008). He argues that *Seguro Popular* acted as an implicit subsidy to informal employment and resulted in a significant reduction of formal jobs. Consistent with this, Bosch and Campos-Vazquez (2014) find evidence of a reduction in employers and employees formally registered. On the other hand, other authors argue against this idea, and most of the empirical literature

does not find a significant impact on informality (Esquivel and Ordaz-Díaz (2009), Arias et al. (2010), Campos-Vázquez and Knox (2013), Azuara and Marinescu (2013) Alonso-Ortiz and Leal (2018)). More recently, Seira et al. (2023) use administrative data and improved econometric methods, and find no evidence of a decrease in formal employment. This paper contributes to the literature by analyzing a social program that is not linked to employment but that could affect the incentives toward informality at a critical period, the start of labor market trajectories.

Additionally, this paper is related to a strand of literature that emphasizes the importance of early labor market experiences. There is widespread evidence regarding *scarring* effects of youth unemployment spells, otherwise known as "unlucky cohorts" that face adverse conditions at labor market entry (Arulampalam (2001), Bell and Blanchflower (2011), Schwandt and von Wachter (2019)). In the context of Latin America, Cruces et al. (2012) find that youth informal employment spells have an impact on persistent informality as well as wage penalty effects. Furthermore, Berniell et al. (2023) provide evidence of important gender differences in the region, with women from "unlucky cohorts" showing a higher employment rate and earnings in the future, consistent with the theory that women act as secondary workers in hard times. Particularly important for my study is the consideration of working while in school. Le Barbanchon et al. (2023) provide some experimental results of employment lotteries in Uruguay. They find a positive income effect two years after the program, while previous non-experimental evidence provided mixed results (Ruhm (1997), Hotz et al. (2002), Ashworth et al. (2021)).

It is also important to note that there are two different traditional schools of thought with respect to the barriers to entry to formal employment. The traditional, dualist school of thought argues in favor of segmented labor markets, based on the Harris-Todaro model (1970) and considers informal workers as being informal in an involuntary way. Alternatively, a revisionist school of thought argues in favor of integrated labor markets, where informality is caused by attempts of formal firms to reduce labor costs and increase competitiveness, and by entrepreneurs who choose to operate informally (Portes et al. (1989), Maloney (2004), Bosch and Maloney (2010)). Therefore, in this school of thought, informal workers are informal voluntarily. Gabriel Ulyssea's work (2018, 2020) has considerably expanded our conceptual and empirical understanding of informality. He distinguishes two margins of informality: the *extensive margin*, where firms decide whether to register formally or not, and the *intensive margin*, where a formal firm can hire informal workers by keeping them "off the books".

For my conceptual framework, it is important to consider the *intensive margin* of informality, since I give an important weight to the decision process of a worker participating in the labor market formally or informally. Given that by law any employee must have access to social insurance benefits through their employer, I argue that the school-linked social insurance policy increased the bargaining power of employees negotiating contracts with employers, but it also decreased the relative cost of working informally while being a student. This is in line with the results of Samaniego de la Parra and Fernández Bujanda (2023), which document that formal firms in Mexico offer both informal and formal jobs and are more likely to increase formality after random workplace inspections.

4 Impact on Labor Market Outcomes

4.1 Data

The main data source for this paper is the Mexican quarterly survey on employment called ENOE (*Encuesta Nacional de Ocupación y Empleo*).[4] This survey has information on approximately 300 thousand workers each quarter, is representative at the national and state level, and allows me to have information for both formal and informal employees. The design of the survey is a rotating panel, so each individual is followed for 5 quarters, and each quarter, one fifth of the sample is replaced. Ideally I would have longitudinal data for individuals, so I could compare the population affected, exploiting the variation on the timing of implementation and using a more conventional difference-in-differences strategy at the individual level. However, given that I only have the limited longitudinal data of the traditional employment survey and the fact that I cannot identify exposure to the policy at the individual level (I cannot observe whether a person attended a public or private school), a way to overcome this is to exploit differences across exposure to the policy by cohorts and differences across states regarding the intensity of informality prior to the implementation of the policy. Therefore, I can compare the cohorts affected by this policy with the previous similar cohorts that did not have access to the social insurance benefits through their education and observe whether there are significant differences in occupational outcomes, varying by initial informality. The main period of analysis corresponds to the year 2019.

[4]Due to the COVID-19 pandemic there is no information for the second quarter of 2020. For the period between the third quarter of 2020 and the fourth quarter of 2022 the information comes from the ENOE New Edition (ENOE^{N}). This temporary survey maintains the conceptual, statistical, and methodological design of ENOE, but combines in-person interviews with phone interviews. Throughout this period, in-person interviews represent between 79% and 100% of all interviews in ENOE^{N}.

4.2 Research Design and Main Results

I will focus on the policy affecting high school which corresponds to years 10 through 12 of the American system. I focus on high school cohorts mainly because there is a more straightforward mapping between birth cohorts and high school years. Usually, high school students are between 15 and 18 years old. The number of years of college in Mexico varies by major and institution, and the age of the students is subject to higher variation. Furthermore, the coverage rate of the population in college years drops significantly to 30%, less than half of that for the population in high school age.[5] The birth cohort will determine an individual's exposure to the policy, so focusing on high school leads to a cleaner specification. Incorporating the college component of the policy and studying longer term impacts of the policy is left for future work.

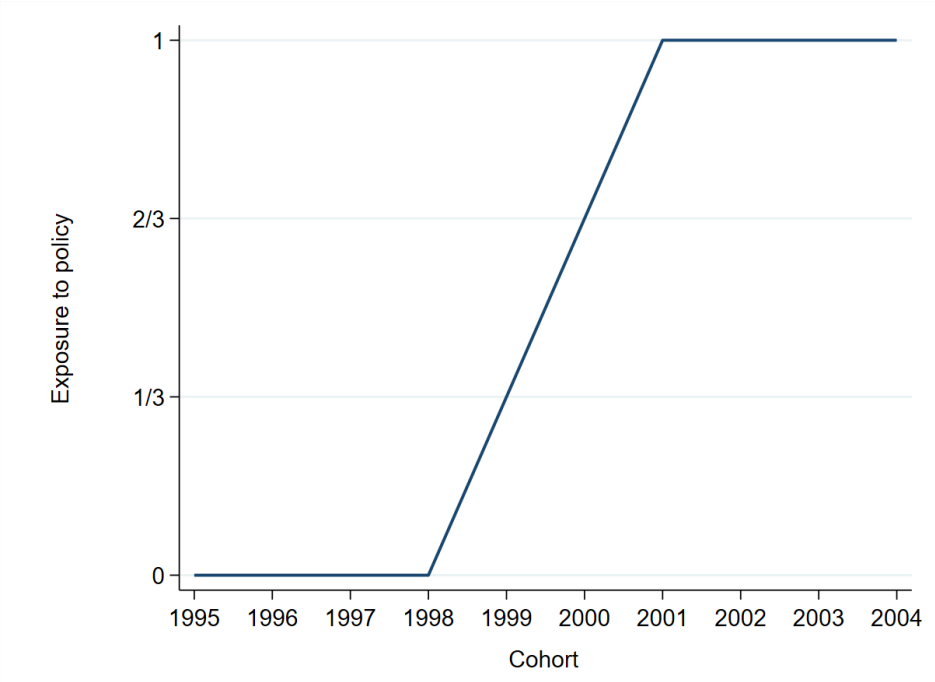
The main cohorts studied will be those born between 1995 and 2004.[6] This is because the cohorts of interest will be the ones that were affected by the policy and those that barely missed it due to graduating from high school before 2016. The main analysis is going to be done with data from 2019 such that the youngest cohorts are 15 years old (the minimum legal working age). Given that the policy is fairly recent, and fully affected cohorts are still in school age, I focus on these young cohorts to maintain comparability. Aditionally, while the policy is still in place at the time of the writing of this paper, students now need to voluntarily request their social security number in order to get access to the medical benefits from IMSS. In 2016, all students enrolled in public high school were issued their social security number. The exogeneity of this shock affecting all students enrolled in any year of public high school during that year is relevant to identify the impact of the policy on early labor market outcomes

Figure 1 plots the exposure to the policy during high school by cohort, defined by the percentage of years where students had access to social insurance benefits. Cohorts that graduated high school before 2016 (i.e. were older than 18 in 2016) were not exposed at all to the policy, while cohorts that were 15 or younger in 2016 entered high school after the full implementation of the program and were fully exposed. There were two cohorts that were partially exposed, corresponding to those students who only had the social insurance benefits for one or two years of their high school.

[5] Secretaría de Educación Pública. *Principales Cifras del Sistema Educativo Nacional 2015-2016*.

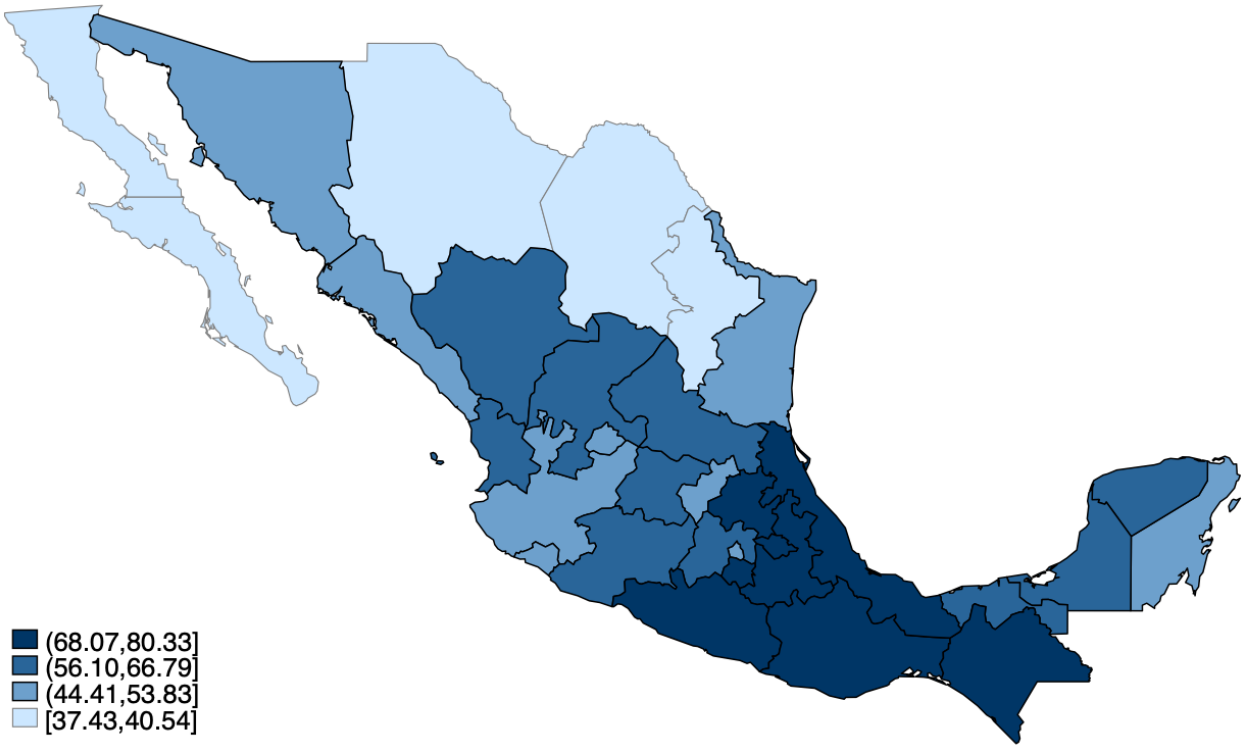
[6] Each quarter there is information for around 7,000 individuals per cohort, representing over 2 million individuals per cohort at a national level.

Figure 1: Exposure to policy (high school) by cohort



To better assess the impact of the expansion of social insurance benefits for students on labor market outcomes, I will consider the extent of informality prior to the implementation of the policy. The national average of almost 60% of the workforce being informal masks significant heterogeneity in the informality rate across states in Mexico. The richer states bordering the United States have informality rates of around 40% whereas the poorest, southern states have informality rates as high as 80%. I am going to follow a research design close to that implemented by Bleakley (2010). He studies the impact of malaria eradication campaigns in the Americas by comparing outcomes across cohorts separating along the timing of birth relative to the campaigns and the degree of pre-campaign malaria intensity. This is also similar to Duflo (2001), where she studies the schooling and labor market consequences of school construction in Indonesia. In my setting, the campaign is defined as the access to social insurance benefits through school enrollment, and the separation of intensity will be regarding pre-policy informality, defined as the average of the informality rate at the state level between 2013 and 2015. Figure 2 plots the pre-policy informality rate by state. The main labor market outcomes I will be studying are the participation rate (share of population working or actively looking for work), the employment rate (share of labor force that is employed), the formality rate (share of employed population in a formal job), and the monthly labor income.

Figure 2: Pre-policy informality rate by state (%)
2013-2015 average



The identifying assumption underlying my analysis is that in the absence of the policy, the average change in the unexposed cohorts among low informal states represents the counterfactual change in the exposed cohorts in high informal states across the outcomes studied. Given that high informal states are also poorer, a main threat to identification will be mean reversion. I will perform tests to the identifying assumption after presenting the main results.

4.2.1 Long Differences

First, I am going to simplify the analysis and aggregate the different cohorts into two groups: those who were fully exposed to the policy and those who were not exposed at all. I will compare the main labor market outcomes across these groups of cohorts for each state to obtain long difference estimates. Therefore, the basic equations to be estimated will be

$$Y_{s,post} - Y_{s,pre} = \beta I_{s,pre} + X_{s,pre}\Gamma + \alpha + \varepsilon_{s,post},$$

where Y is a labor market outcome of state s, and the time subscripts $post/pre$ indicate

whether the cohorts were fully exposed or not to the policy. The preceding informality rate is $I_{s,pre}$, defined as the average state informality rate between 2013 and 2015. The X variables are a set of controls, and α is a constant term. The main outcomes of interest are the participation, employment, and formality rates, and monthly labor income. For the set of controls I include state GDP per capita in 2013, mean educational attainment in 2015 and regional controls, grouping the 32 states into 8 different regions. The equation will be estimated using weighted least squares with data from the four quarters of the employment survey of 2019, where the weights are defined by the square root of cell sizes to take into account the different precision of the estimations of the cohort means.

The long differencing estimates are presented in Table 1. Panel A shows the estimates only with the preceding informality rate as an independent variable, and Panel B includes the set of controls. From the mean differences between cohorts fully exposed and not exposed at all, we can see that the exposed cohorts have significant lower participation, and formality rates, as well as average income. These patterns are expected as fully exposed cohorts are younger.

From the estimated coefficients, we can interpret that one additional percentage point in the preceding informality rate is associated with an increase of 0.59 percentage points in the participation rate, an increase of 0.15 percentage points in the employment rate, and an increase of 0.87 percentage points in the formality rate among the exposed cohorts. The impact on monthly income is small and not statistically significant. The sample of income results is smaller due to higher non-response rates in this question of the survey. These reduced form results suggest that in states with high informality rates, cohorts that were exposed to the policy experienced an increased participation in the labor market, with higher employment and increased formality, relative to cohorts in less informal states.

The main analysis is done with information from 2019 to evaluate the medium term effects of the policy for the fully exposed cohorts in 2016, right before the onset of the COVID-19 pandemic in 2020. To consider the full picture, figure 3 plots the estimated coefficients ($\hat{\beta}_t$) of the long differences equation for each quarter t from 2016 to 2022. We can observe that the positive effects on the participation and formality rate are relatively consistent from 2016 to 2019. After the COVID-19 pandemic, the estimated coefficients for both of these indicators are close to zero and no longer statistically significant.

Table 1: Cross Cohort Long Differences in 2019

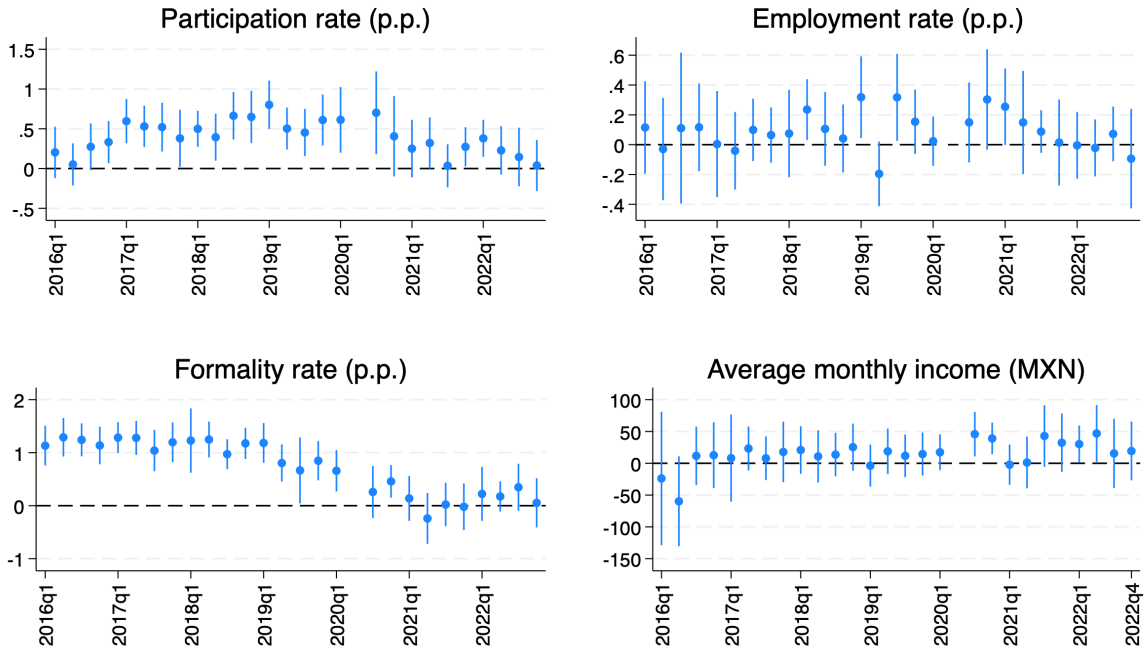
	(1) Participation rate	(2) Employment rate	(3) Formality rate	(4) Monthly income
Panel A				
Pre-policy informality rate	0.50***	0.16***	0.79***	30.07***
	(0.03)	(0.02)	(0.04)	(2.71)
Panel B				
Pre-policy informality rate	0.59***	0.15*	0.87***	10.32
	(0.07)	(0.07)	(0.10)	(7.63)
Log GDP per capita in 2013	6.04***	1.58	1.91	-122.39
	(1.05)	(1.27)	(1.22)	(97.52)
Educational attainment in 2015	-3.29***	-0.37	2.25	-350.76**
	(0.86)	(0.96)	(1.27)	(104.34)
Regional controls	✓	✓	✓	✓
Quarter FE	✓	✓	✓	✓
Mean dependent variable	-37.89	0.08	-32.75	-1,930.18
Observations	128	128	128	128

Robust standard errors in parentheses. The square root of cell sizes is used to construct weights for the observations. Cell sizes are identical across specifications, except for the income one as we restrict to individuals reporting strictly positive income.
***p<0.01, **p<0.05, *p<0.1

It is not clear how to interpret these estimates as the pandemic severely disrupted labor markets. There is evidence that in Latin America there was an initial drop in participation, employment, and informality followed by a rebound in informality, with the informal sector acting as an important margin of adjustment (Alvarez and Pizzinelli (2021), Leyva and Urrutia (2023)). Furthermore, Osuna-Gómez (2023) finds that people that entered the labor market after the Great Recession lost formal employment during the pandemic at higher rates than previous cohorts.

It could be the case that the pandemic shock reverted any effects that were driven by the school-linked social insurance policy on labor market outcomes. However, there are additional concerns for more recent estimates as we are tracking cohorts at the state level and not individuals, and we could worry about potential migration patterns and readjustments between different local labor markets. The employment and income estimates are mostly small and not significant.

Figure 3: Estimated coefficients of pre-policy informality rate on cross cohort long differences, $\hat{\beta}_t$



Note: No available data for 2020q2.

4.2.2 Panel Estimates

The long differencing estimates have the advantage of assessing the medium term impact of the school-linked social insurance policy between cohorts that were fully exposed and not exposed at all. However, they do not account for pre-existing trends, which are going to be addressed with panel estimates constructed with information for all the quarters for the period spanning from 2016 when the policy started to 2019. First, to incorporate cohorts that were partially exposed to the policy I will estimate pooled regressions of the form

$$Y_{kst} = \tilde{\beta} I_s \times \text{Exp}_k + \delta_k + \delta_s + \delta_t + \sum_i \left(x_s^i \times \text{Exp}_k\right) \gamma_i + \nu_{skt},$$

where Exp_k is exposure to the policy for cohort k as defined by figure 1, I_s is the pre-policy informality rate in state s, the x_s^i are state-specific controls, and $\delta_k, \delta_s, \delta_t$ are fixed effects for cohort, state, and quarter, respectively.

Second, I will compare changes in outcomes by cohort across states with distinct pre-

policy informality rates in order to assess the contribution of the policy to the observed changes. The regressions to be estimated are

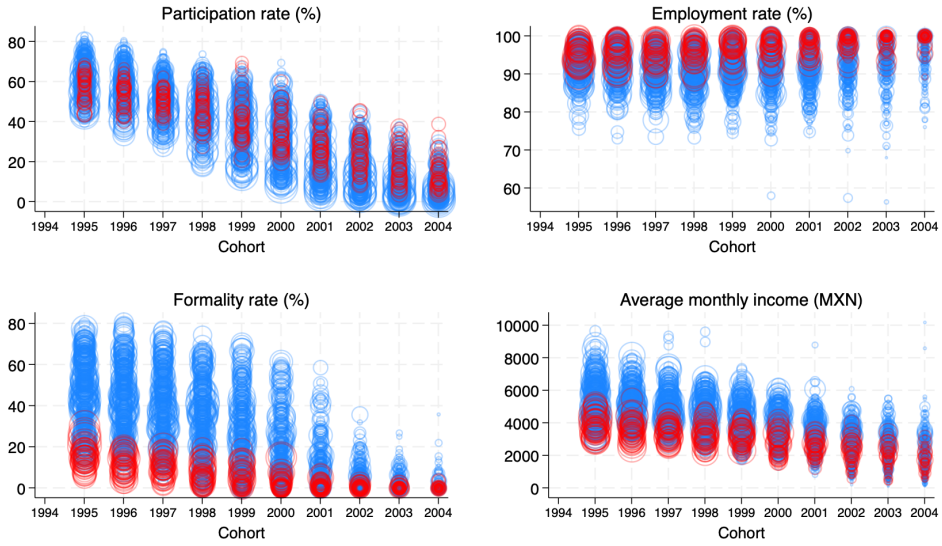
$$Y_{kst} = \beta_k I_{s,pre} + \delta_k + X_s \Gamma_k + \nu_{skt},$$

where β_k is a cohort specific coefficient on informality, X_s is a vector of region controls, and δ_k and Γ_k are cohort-specific intercept and slope coefficients. By estimating the equation through weighted least squares for each cohort, I will generate a series of estimates across cohorts. The series of cohort estimates will help me produce graphical evidence of the impact of the policy.

Before presenting the estimates of the pooled and cohort-specific regressions, Figure 4 plots the summary statistics for the panel constructed by cohort in states and quarters between 2016 and 2019. Each observation represents the value of the dependent variable for each cohort-state-quarter tuple. The size of the observations is determined by the size of the population represented. I single out in red the three poorest states (Guerrero, Oaxaca, and Chiapas), which also happen to be the southernmost states. We can see expected patterns in the data, such as lower participation rates and higher informality rates for younger cohorts. Among the poorest states it is interesting to observe that for the oldest cohorts analyzed, the participation rate is lower relative to richer states but for the youngest cohorts, relative to richer states, the participation rate is higher. It is also worth noting that the employment rate is consistently higher for the poorest states across all cohorts, but the quality of jobs within these states is worse, as seen by the lower formality rate and lower monthly income.

Table 2 presents the estimates for the pooled regressions. The regressions are estimated through weighted least squares with the weights constructed from the square root of cell size. Standard errors are clustered at the state level. We can observe that these estimates are consistent with the results of the long differencing. The magnitude of the estimate for the participation rate decreased but remains significant. The estimate for the employment rate is very similar, whereas the estimate for the formality rate increased to almost 1.3 percentage points associated with an increase of 1 percentage point in the initial informality rate. This means that taking into account the panel structure and the partial exposed cohorts increases the differential positive impact of the policy on the formalization of exposed cohorts in high informal states relative to low informal states. The impact on income is statistically significant, although it is still very small (an increase of around 1.4 USD in monthly income).

Figure 4: Summary statistics of panel by cohort across states and quarters, 2016 - 2019



Note: Each observation represents the value of the dependent variable for each cohort by state-quarter. The red observations denote the three poorest states (Guerrero, Oaxaca, and Chiapas). The size of the observations is determined by the size of the population for each cohort-state-quarter tuple.

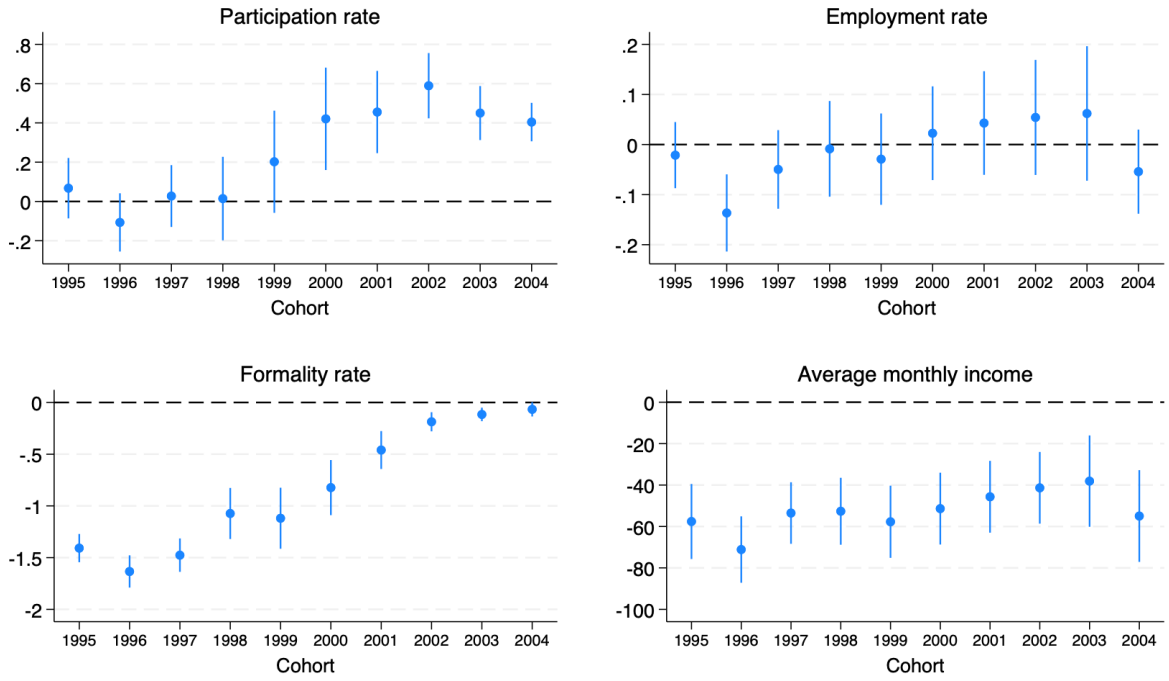
Table 2: Panel Estimates of the Effect of Exposure to Policy, 2016 - 2019

	(1) Participation rate	(2) Employment rate	(3) Formality rate	(4) Monthly income
Pre-policy informality	0.44***	0.15***	1.28***	21.33***
rate × Exposure	(0.05)	(0.03)	(0.06)	(7.05)
Log GDP per	2.01	0.43	3.67**	-61.73
capita in 2013 × Exposure	(1.01)	(1.02)	(1.01)	(114.12)
Educational attainment	-0.33	1.86**	1.86	-72.11
in 2015 × Exposure	(0.67)	(0.55)	(1.22)	(86.32)
Cohort FE	✓	✓	✓	✓
State FE	✓	✓	✓	✓
Quarter FE	✓	✓	✓	✓
Mean dependent variable	33.31	93.84	17.99	4,299.21
Observations	5,094	5,091	5,091	5,046

Standard errors clustered by state in parentheses. The square root of cell sizes are used as weights for the observations. Cell sizes are identical across specifications, except for the income one as we restrict to individuals reporting strictly positive income.
***p<0.01, **p<0.05, *p<0.1

Figure 5 plots the cohort-specific estimates on the pre-policy informality rate for the different outcomes of interest. The pattern of estimates is broadly consistent with the exposure to the policy, as defined in figure 1. The patterns are more clear for the participation and formality rates. This graphical evidence points to the relevance of the policy explaining the results found in the long-differencing and pooled regression analyses.

Figure 5: Cohort-specific estimates on pre-policy informality rate, $\hat{\beta}_k$



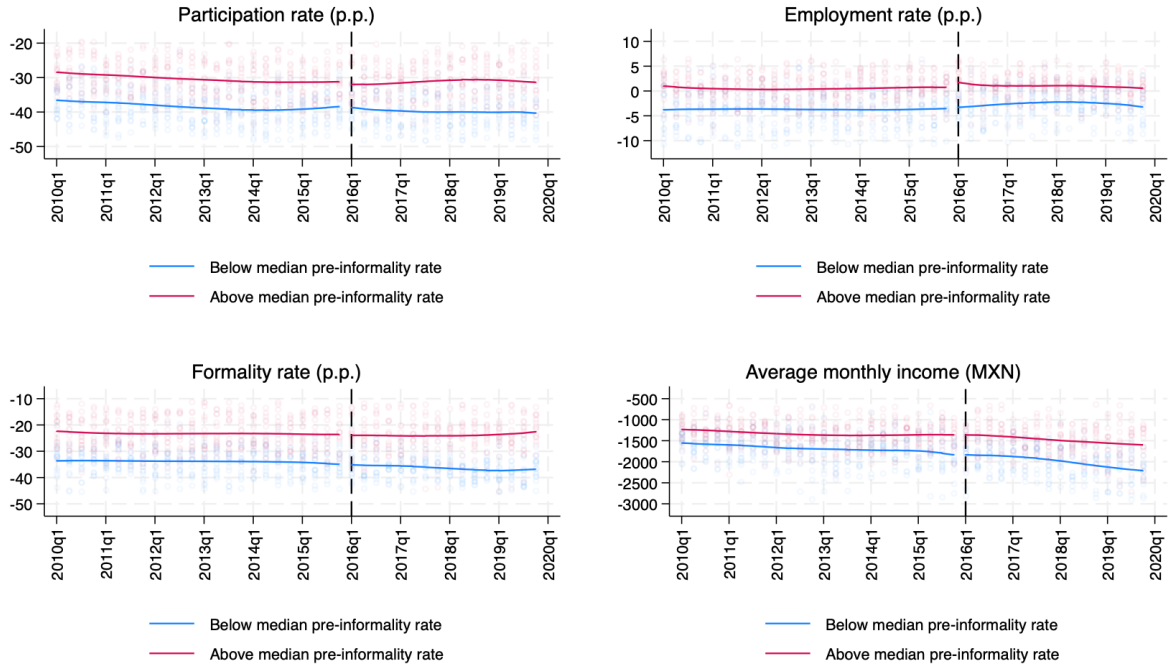
4.3 Tests to Identifying Assumption

The identifying assumption of my empirical strategy should not be taken for granted. An alternative explanation to the main results would be that of mean reversion. It could be the case that even in the absence of the policy, we would have observed an increase in the participation, employment, and formality of younger cohorts in high informal states as these very informal, poor areas catch up with the more formal, richer states.

As a first approach to test the identifying assumption, figure 6 plots the trends in the long

differences between the correspondent treatment (15-18) and control (21-24) age groups of the main analysis since 2010. For each quarter, I calculate the long differences across the main outcomes for each state. I then divide states between those below and above median pre-policy informality rate, and plot the trends of the long differences before and after the implementation of the policy as estimated through a locally weighted scatterplot smoothing (LOWESS). A visual analysis of the pre-trends does not identify significant differences, suggesting that the differences across cohorts in low informal states could be an appropriate counterfactual for the differences across cohorts in high informal states. Additionally, Table A.1 in the Appendix shows that the long differences results are robust to controlling for the long differences three years prior.

Figure 6: Long Differences Trends (difference between 15-18 and 21–24 cohorts by state)



As a further test for the identifying assumption, I run a set of placebo regressions to address the concerns on mean reversion. I artificially assign full exposure to my main control group (those born between 1995 and 1998) and compare them with older cohorts (those born between 1991 and 1994). Table 3 shows the results. If we are concerned with mean reversion, we would worry to see a similar pattern as the main results, consistent

with younger cohorts in high informal states catching up with older cohorts in less informal areas. However, if anything, these placebo regressions give evidence of the contrary. Among slightly older cohorts that were not exposed to the policy, there had been a divergence in the participation and formality of younger cohorts in high informal states, relative to older cohorts in less informal states. These results alleviate concerns of mean reversion and suggest that the main results are capturing the causal impact of the policy.

Table 3: Placebo Regressions

	(1) Participation rate	(2) Employment rate	(3) Formality rate	(4) Monthly income
Panel A: Long Differences, 2019				
Pre-policy informality rate	-0.11	-0.01	-0.24**	-7.16
	(0.06)	(0.04)	(0.08)	(7.73)
Full set of controls	✓	✓	✓	✓
Mean dependent variable	-12.47	-2.17	-8.07	-1,050.38
Observations	128	128	128	128
Panel B: Pooled Regressions, 2016 - 2019				
Pre-policy informality rate × Exposure	-0.09**	0.04*	-0.10	10.83**
	(0.03)	(0.02)	(0.05)	(3.23)
Full set of controls	✓	✓	✓	✓
Mean dependent variable	60.42	93.00	40.45	5,288.04
Observations	4,096	4,096	4,096	4,096
Panel C: Pooled Regressions, 2016				
Pre-policy informality rate × Exposure	-0.01	0.08**	0.14	10.11**
	(0.04)	(0.03)	(0.09)	(2.94)
Full set of controls	✓	✓	✓	✓
Mean dependent variable	54.39	91.88	34.81	4,490.38
Observations	1,024	1,024	1,024	1,024

Standard errors in parentheses (robust standard errors for long differences and clustered standard errors by state for pooled regressions). The square root of cell sizes is used to construct weights for the observations. Cell sizes are identical across specifications, except for the income one as we restrict to individuals reporting strictly positive income.
***p<0.01, **p<0.05, *p<0.1

5 School Enrollment and Robustness

An important consideration is whether this policy had an impact on the schooling decisions of individuals. The new medical benefits associated with being enrolled in a public high school or college can increase school enrollment and attainment to the detriment of participation in the labor market at younger ages, which can be actually a desirable outcome from both the perspective of the individual and society as a whole. The results of the previous section of an increased participation rate among exposed cohorts in high informal states suggest that this is not the case. In this section, I will explore more thoroughly the impact on school enrollment as well as differentiate the main results by school enrollment to better understand the mechanisms at play. I will then look at the heterogeneous results and robustness to an alternative definition of informality.

5.1 School enrollment

First, I will explore the impact of the social insurance expansion policy on school enrollment rates of the cohorts analyzed. Table 4 shows estimates of the long differences and panel equations presented in sections 4.2.1 and 4.2.2, using school enrollment rates as the dependent variable. There is no significant evidence of an impact of the policy on schooling rates when comparing exposed and unexposed cohorts in high informal states relative to low informal states.

Table 4: Long Differences and Panel Estimates of School enrollment Rates

	(1)	(2)	(3)
	Long Differences	Long Differences	Panel Estimates
Pre-policy informality rate	-0.05	-0.21*	-0.12
	(0.04)	(0.08)	(0.09)
Log GDP per capita in 2013		-9.36***	-1.76
		(1.72)	(1.60)
Educational attainment in 2015		4.41***	-2.55
		(1.26)	(1.38)
Regional controls		✓	
Cohort FE			✓
State FE			✓
Quarter FE		✓	✓
Mean dependent variable	46.42	46.42	63.97
Observations	128	128	5,120

Standard errors in parentheses. The square root of cell sizes are used as weights for the observations. For column (3) the independent variables are interacted with the policy exposure function as in Table 2.
***p<0.01, **p<0.05, *p<0.1

Now, I will distinguish the main results of the previous section between workers attending school at the time of the survey and those not attending school. Table 5 presents the estimates of long differences on the main outcomes of the labor market distinguishing by school enrollment. Panel A shows the main results, considering the full sample, which are exactly those presented previously in Table 1. Panel B shows the long differences estimates using the sample of people not attending school and Panel C shows the results using the sample of students at the time of the survey. Approximately 53% of the full sample studied is enrolled in school. We can see that the results for participation rate, formality rate, and monthly income are higher among the sample that is attending school. In the case of the formality rate, one percentage point increase in the pre-policy informality rate is associated with an increase of 0.95 percentage points in the formality rate of the cohorts affected by the policy.

To see this graphically, figure 7 shows the estimated coefficients of the pre-policy informality rate on the main outcome variables distinguishing by school enrollment. This figure presents the results from both the long differences and panel specifications. For the case of the panel specification the coefficient corresponds to the interaction of the pre-policy informality rate and the exposure function. We confirm that the estimated coefficients are higher for the subsample of people that are attending school, across the specifications and main outcomes (with the exception of the employment rate).

We can conclude from this that the main results are consistent across the subsamples of people that are enrolled and not enrolled in school at the time of the survey. In particular, these results provide suggestive evidence that students affected by the social insurance expansion policy are not more likely to be employed informally, despite the reduction in the relative cost of an informal job while also being enrolled in school. This is more relevant as there is evidence that there was an increase in the participation rate among exposed cohorts in states with higher informality. It is uncertain whether it is desirable to have more individuals start working while they study. Employment spells while studying could provide valuable training experiences. On the other hand, it could deter from further investing in schooling (Atkin, 2016). However, if anything, it is desirable to have students work formally vs. informally at the beginning of their labor market trajectories as there are negative employment and earning effects associated with youth informality spells.

Table 5: Cross Cohort Long Differences by School enrollment, 2019

	(1) Participation rate	(2) Employment rate	(3) Formality rate	(4) Monthly income
Panel A: Full Sample				
Pre-policy informality rate	0.59***	0.15*	0.87***	10.32
	(0.03)	(0.07)	(0.10)	(7.63)
Log GDP per capita in 2013	6.04***	1.58	1.91	-122.39
	(1.05)	(1.27)	(1.22)	(97.52)
Educational attainment in 2015	-3.29***	-0.37	2.25	-350.76**
	(0.86)	(0.96)	(1.27)	(104.34)
Regional controls	✓	✓	✓	✓
Mean dependent variable	-37.89	0.08	-32.75	-1,930.18
Panel B: Sample Not Attending School				
Pre-policy informality rate	0.37***	0.10	0.83***	19.31
	(0.10)	(0.09)	(0.12)	(10.45)
Log GDP per capita in 2013	-1.02	2.43	2.32	-242.00
	(2.11)	(1.46)	(1.50)	(155.41)
Educational attainment in 2015	-2.65	-1.80	2.59	-119.53
	(1.44)	(1.19)	(1.50)	(153.14)
Regional controls	✓	✓	✓	✓
Mean dependent variable	-17.25	-1.13	-31.43	-1,493.67
Panel C: Sample Attending School				
Pre-policy informality rate	0.56***	0.20	0.95***	29.13
	(0.09)	(0.10)	(0.14)	(16.44)
Log GDP per capita in 2013	4.06**	1.20	-2.91	-359.45
	(1.41)	(1.74)	(2.53)	(342.45)
Educational attainment in 2015	-0.79	1.52	5.07*	24.63
	(1.30)	(1.57)	(2.31)	(274.43)
Regional controls	✓	✓	✓	✓
Mean dependent variable	-21.95	0.68	-30.78	-2,120.30
Observations	128	128	128	128

Robust standard errors in parentheses. The square root of cell sizes are used as weights for the observations.
***p<0.01, **p<0.05, *p<0.1

Figure 7: Estimates on pre-policy informality rate by school enrollment

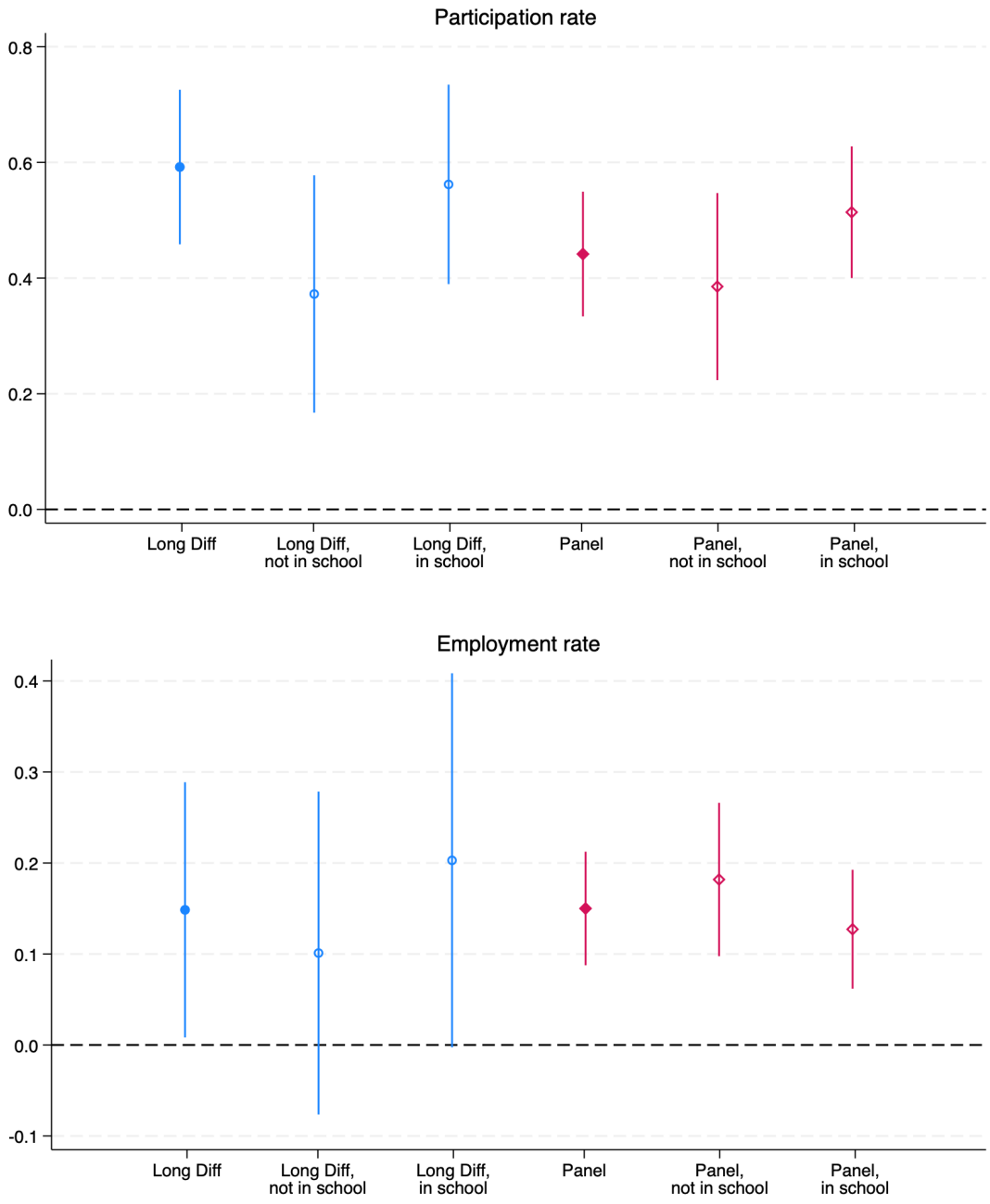
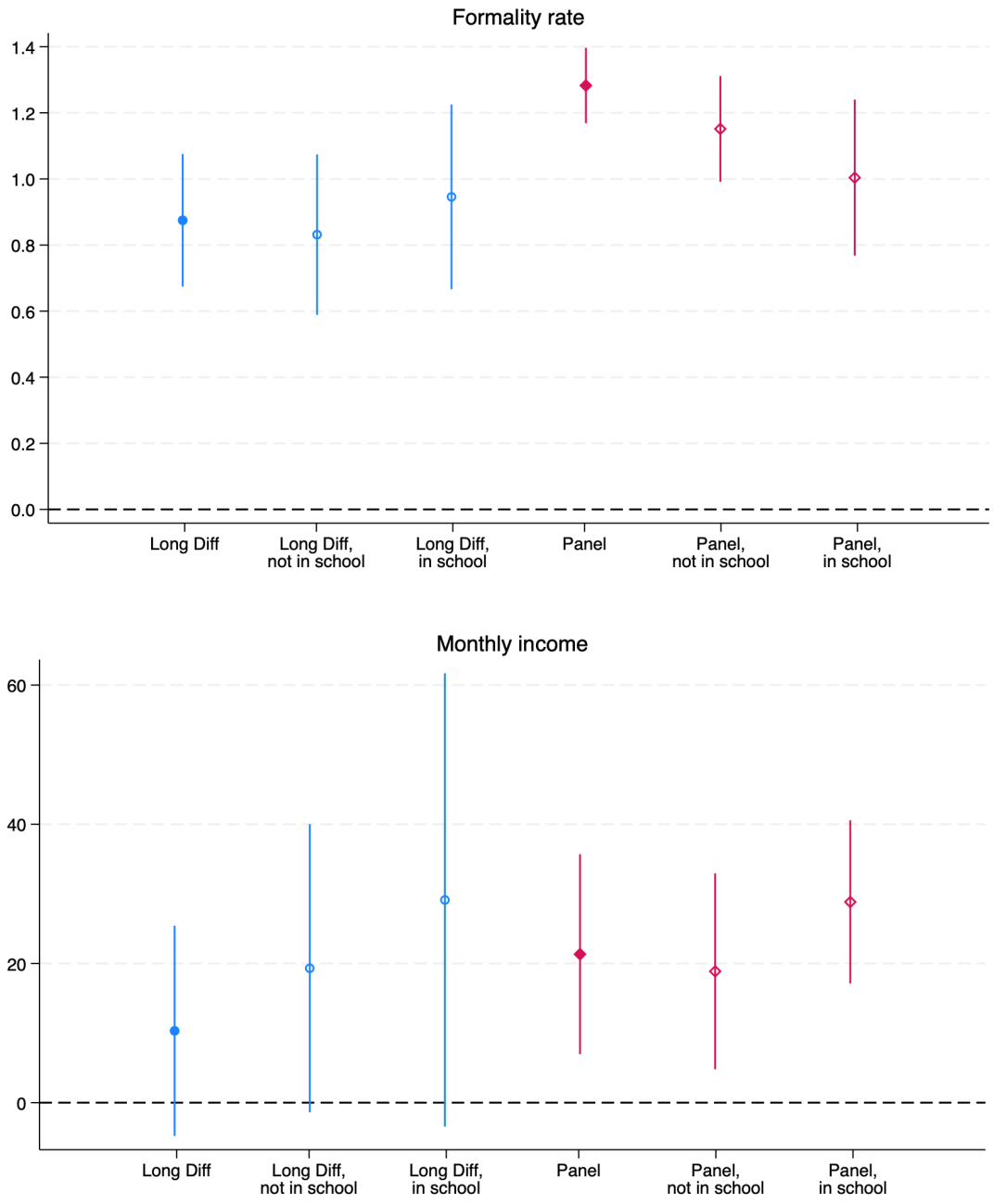


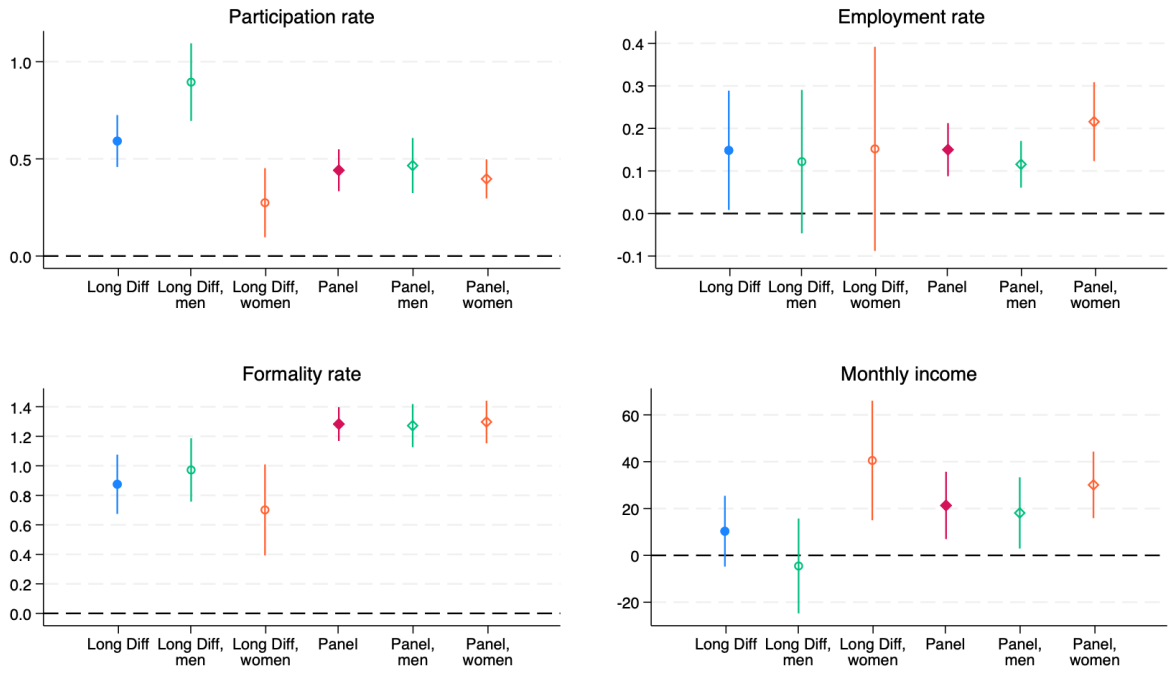
Figure 7: Estimates on pre-policy informality rate by school enrollment (cont.)



5.2 Heterogeneous Results

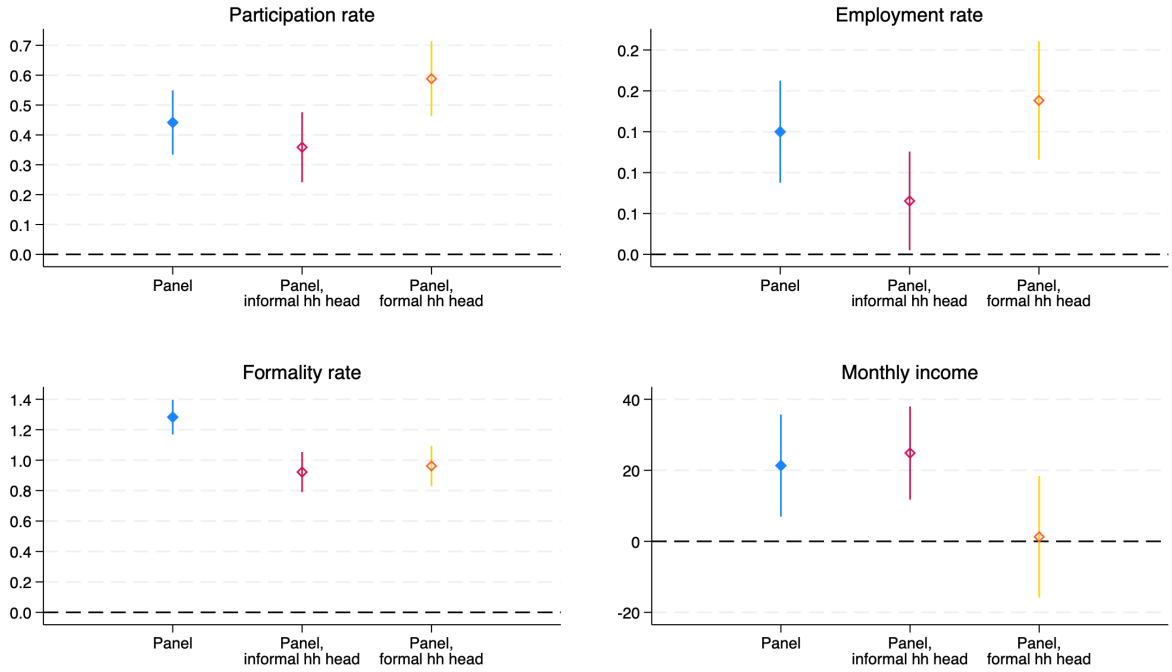
I distinguish the main results among the sample of men and women to explore any differential impact of the policy by gender. The sample is roughly even split between men and women. Figure 8 presents the graphical results of the coefficient on the pre-policy informality rate between specifications and by gender. From these results, we cannot observe a stark difference in the impact of the policy between men and women. However, it is the case that the point estimates for the participation and formality rate are higher for men than for women, whereas the point estimates for the employment rate and monthly income are higher for women. Overall, I do not find significant differences in the impact of the policy by gender.

Figure 8: Estimates on pre-policy informality rate by gender



I further make a distinction by the formality of the head of household. Figure 9 shows that while the impact on the participation rate is larger for individuals with a formal head of household, the formalization effect is similar regardless. The fact that the results are not local to fully informal households highlights the importance of having access to social insurance benefits directly, not only through a parent.

Figure 9: Estimates on pre-policy informality rate by formality of head of household (hh)

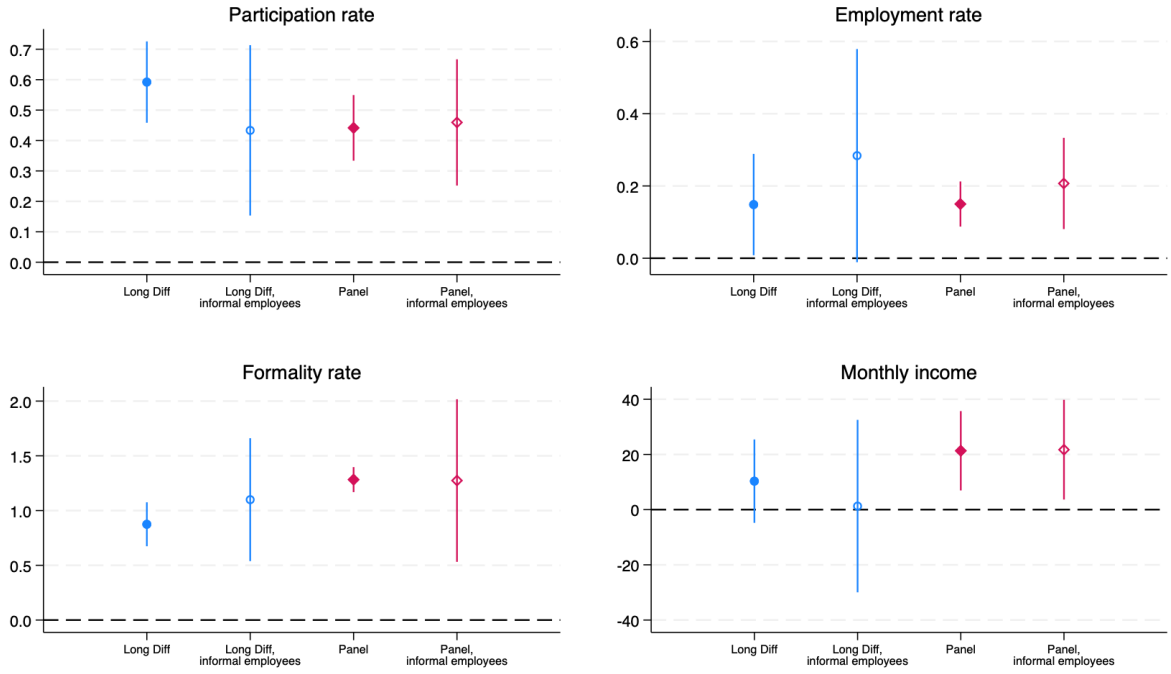


5.3 Alternative Definition of Informality

Up to this point, I have been using the official definition of informal employment established by the Mexican National Institute of Statistics and Geography (INEGI). This definition includes both employees and self-employed individuals. I will explore how do the main results change if we focus only on subordinate employment, i.e., employees without access to social insurance. I use this alternative definition of informality that only considers employees for the pre-policy period as well as for the main period of analysis.

Figure 10 presents the main results of the long differences and the panel specifications with full controls along with the results using the alternative definition of informality, considering only employees. As expected, the alternative estimates on the pre-policy informality rate become noisier but are similar in magnitude to the main estimates. The panel estimates suggest a positive and significant impact of the policy on the participation and formality rates among exposed cohorts in highly informal states.

Figure 10: Estimates on pre-policy informality rate, alternative definitions of informality



6 Conclusion

Informal employment remains a predominant characteristic of the labor markets in developing countries. Informality rates are higher among younger populations and throughout earlier jobs in the career trajectories. Given the potential scarring effects of young informal employment on future formality and income it is important to understand how public policy affects the formalization of young workers. Policies increasing the access to social insurance without being tied to employment have been implemented with differing effects on formalization.

This paper studies the impact of an expansion of medical benefits associated to social insurance in Mexico in 2016 among public high school students. Exploiting variation in the exposure to the policy by year of birth and in the initial informality rate across states, I find that the policy had a positive impact on the participation and formality rates of ex-

posed cohorts in high informal states relative to less informal ones. These effects are similar across the population by school enrollment, suggesting that students do not consider the policy as a relevant subsidy to informal employment through the complementarity of the medical benefits attached to their education. I find no significant differences between men and women, and the results are robust to excluding self-employed informal workers.

There is no significant evidence of an impact on school enrollment, suggesting that young people affected by the policy did not consider the new medical benefits attached to public education as an important determinant of their schooling decisions. While the results of increased participation may be ambiguous given the young age of the affected population, it doesn't seem to be in the detriment of their education. Furthermore, the increase in participation was accompanied by higher formal employment. The increase in formality is sizable among the exposed cohorts in high informal states, with an additional percentage point in the informality rate of the state prior to the policy being associated with an increase of between 0.8 and 1.3 percentage points in the formality rate. A back of the envelope estimate suggests a formalization of up to 100 thousand young people in the 3 poorest and most informal states.

Further research is needed to fully assess the impact of providing school-linked social insurance. It is important to incorporate the impacts on labor market outcomes with the health benefits among the exposed population. The findings of this paper suggest that it is possible to work towards advancing universal social protection systems with supplementary programs that are well-designed, without having distortionary effects that shift workers towards informality.

It is crucial to advance the study of the role of early labor market experiences, particularly that of informal jobs in the context of developing countries. Moreover, it is important to improve the understanding of the individual's decision process on formalization and the different barriers they face, as well as the interactions with the firm's hiring process. Finally, given the relative stability of informality rates in countries such as Mexico it is relevant to assess which policies foster formalization and which hinder it. Improving the working conditions of millions of people in the developing world remains one of the main challenges from a policy perspective.

References

Akay, A. and Khamis, M. (2012). The Persistence of Informality: Evidence from Panel Data. *Informal Employment in Emerging and Transition Economies (Research in Labor Economics, Volume 34). Emerald Group Publishing Limited*, pages 229–255.

Alonso-Ortiz, J. and Leal, J. (2018). Cross-Subsidies, and the Elasticity of Informality to Social Expenditures: The Case of Mexico's Seguro Popular. *Review of Income and Wealth*, 64(2):482–512.

Alvarez, J. and Pizzinelli, C. (2021). COVID-19 and the Informality-driven Recovery: The Case of Colombia's Labor Market.

Alzúa, M. L., Cruces, G., and Ripani, L. (2013). Welfare Programs and Labor Supply in Developing Countries: Experimental Evidence from Latin America. *Journal of Population Economics*, 26(4):1255–1284.

Arias, J., Azuara, O., Bernal, P., Heckman, J. J., and Villarreal, C. (2010). Policies to Promote Growth and Economic Efficiency in Mexico. Technical report, National Bureau of Economic Research.

Arulampalam, W. (2001). Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages. *The Economic Journal*, 111(475):F585–F606.

Ashworth, J., Hotz, V. J., Maurel, A., and Ransom, T. (2021). Changes across Cohorts in Wage Returns to Schooling and Early Work Experiences. *Journal of Labor Economics*, 39(4):931–964.

Atkin, D. (2016). Endogenous Skill Acquisition and Export Manufacturing in Mexico. *American Economic Review*, 106(8):2046–85.

Azuara, O. and Marinescu, I. (2013). Informality and the Expansion of Social Protection Programs: Evidence from Mexico. *Journal of Health Economics*, 32(5):938–950.

Bell, D. N. and Blanchflower, D. G. (2011). Young People and the Great Recession. *Oxford Review of Economic Policy*, 27(2):241–267.

Berniell, I., Gasparini, L., Marchionni, M., and Viollaz, M. (2023). Lucky Women in Unlucky Cohorts: Gender Differences in the Effects of Initial Labor Market Conditions in Latin America. *Journal of Development Economics*, 161:103042.

Bleakley, H. (2010). Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure. *American Economic Journal: Applied Economics*, 2(2):1–45.

Bosch, M. and Campos-Vazquez, R. M. (2014). The Trade-Offs of Welfare Policies in Labor Markets with Informal Jobs: The Case of the "Seguro Popular" Program in Mexico. *American Economic Journal: Economic Policy*, 6(4):71–99.

Bosch, M. and Maloney, W. F. (2010). Comparative Analysis of Labor Market Dynamics Using Markov Processes: An Application to Informality. *Labour Economics*, 17(4):621–631.

Campos-Vázquez, R. M. and Knox, M. A. (2013). Social Protection Programs and Employment: The Case of Mexico's Seguro Popular Program. *Economía Mexicana. Nueva Época*, 22(2):403–448.

Cruces, G., Ham, A., and Viollaz, M. (2012). Scarring Effects of Youth Unemployment and Informality: Evidence from Argentina and Brazil.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91(4):795–813.

Esquivel, G. and Ordaz-Díaz, J. L. (2009). *¿Es Correcto Vincular la Política Social a la Informalidad en México?: Una Prueba Simple de las Premisas de esta Hipótesis*. CEPAL.

Gerard, F., Naritomi, J., and Silva, J. (2021). Cash Transfers and Formal Labor Markets: Evidence from Brazil.

Harris, J. R. and Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, 60(1):126–142.

Hotz, V. J., Xu, L. C., Tienda, M., and Ahituv, A. (2002). Are There Returns to the Wages of Young Men from Working While in School? *The Review of Economics and Statistics*, 84(2):221–236.

ILO (2018). Women and Men in the Informal Economy: A Statistical Picture. Technical Report 46, International Labour Organization.

Le Barbanchon, T., Ubfal, D., and Araya, F. (2023). The Effects of Working While in School: Evidence from Employment Lotteries. *American Economic Journal: Applied Economics*, 15(1):383–410.

Levy, S. (2007). ¿Pueden los Programas Sociales Disminuir la Productividad y el Crecimiento Económico? Una Hipótesis para México. *El Trimestre Económico*, 74(295):491–540.

Levy, S. (2008). *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*. Rowman & Littlefield.

Leyva, G. and Urrutia, C. (2023). Informal Labor Markets in Times of Pandemic. *Review of Economic Dynamics*, 47:158–185.

Maloney, W. F. (2004). Informality Revisited. *World Development*, 32(7):1159–1178.

Meghir, C. and Phillips, D. (2008). Labour Supply and Taxation. Technical report, IFS Working Papers.

Moffitt, R. A. (2002). Welfare Programs and Labor Supply. *Handbook of Public Economics*, 4:2393–2430.

Osuna-Gómez, D. (2023). The Impact of the COVID-19 Pandemic on Post-Great Recession Entrants: Evidence from Mexico. *Labour Economics*, 81:102337.

Portes, A., Castells, M., and Benton, L. (1989). *The Informal Economy: Studies in Advanced and Less Developed Countries*. A Johns Hopkins paperback. Johns Hopkins University Press.

Ruhm, C. J. (1997). Is High School Employment Consumption or Investment? *Journal of Labor Economics*, 15(4):735–776.

Samaniego de la Parra, B. and Fernández Bujanda, L. (2023). Increasing the Cost of Informal Workers: Evidence from Mexico. *American Economic Journal: Applied Economics (Forthcoming)*.

Schwandt, H. and von Wachter, T. (2019). Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets. *Journal of Labor Economics*, 37(S1):S161–S198.

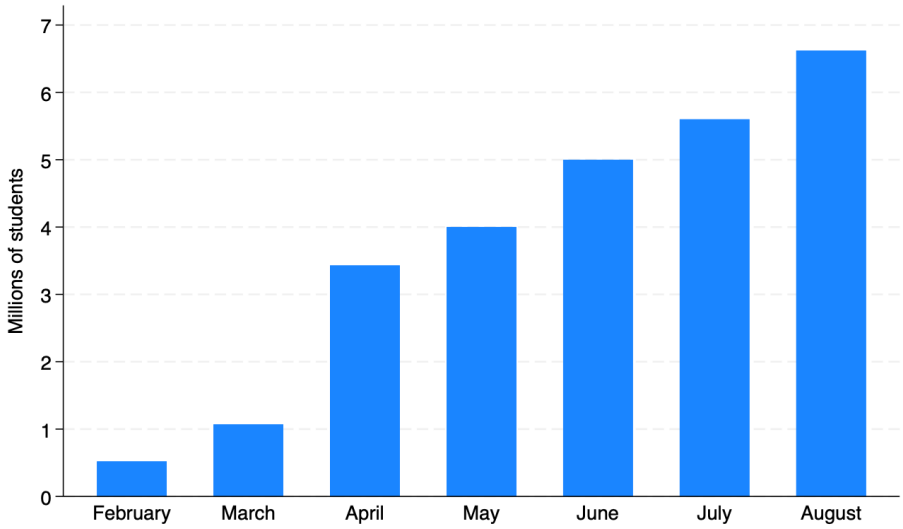
Seira, E., Meza, I., González-Pier, E., and Alcaraz Prous, E. (2023). Did Mexico's Seguro Popular Universal Health Coverage Programme Really Reduce Formal Jobs? Technical Report 46, WIEGO.

Ulyssea, G. (2018). Firms, Informality, and Development: Theory and Evidence from Brazil. *American Economic Review*, 108(8):2015–47.

Ulyssea, G. (2020). Informality: Causes and Consequences for Development. *Annual Review of Economics*, 12(1):525–546.

A Additional Material

Figure A.1: Evolution of social security numbers issued to students, 2016



Source: IMSS. *Informe de Labores y Programa de Actividades 2016 - 2017*.

Figure A.2: "Tienes IMSS" campaign flyer

Source: IMSS. Official Twitter account (April 18, 2016).

Figure A.3: Political event announcing the affiliation of students to IMSS in Mexico City

Source: IMSS. *El Seguro Médico para estudiantes es un escudo para prevenir enfermedades crónicas: Mikel Arriola*

Figure A.4: Alternative definition of pre-policy informality rate by state (%), only employees
2013-2015 average

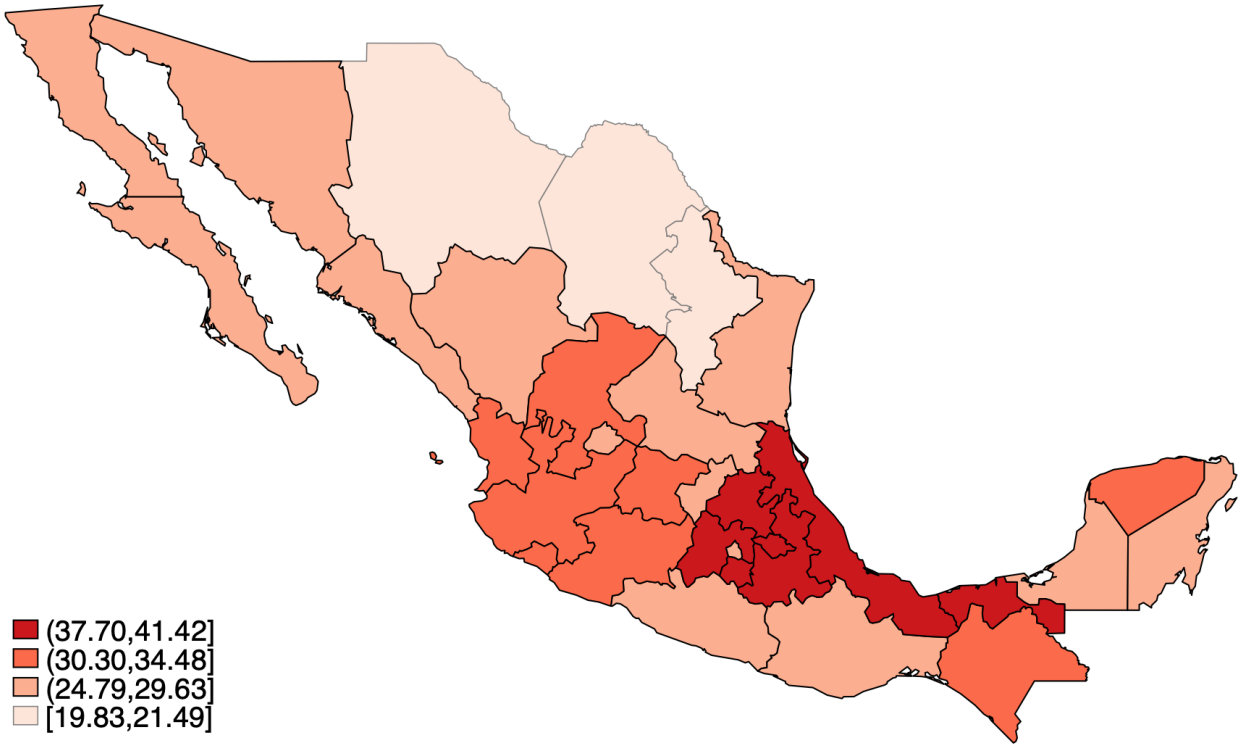


Table A.1: Long Differences Robustness, 2019

	(1) Participation rate	(2) Employment rate	(3) Formality rate	(4) Monthly income
Pre-policy informality rate	0.58***	0.15*	0.37**	17.94*
	(0.09)	(0.07)	(0.12)	(7.28)
Log GDP per capita in 2013	5.95***	1.72	-2.60	-36.64
	(1.32)	(1.24)	(1.45)	(90.99)
Educational attainment in 2015	-3.27***	-0.58	4.12**	42.22
	(0.92)	(1.07)	(1.26)	(115.44)
Regional controls	✓	✓	✓	✓
Quarter FE	✓	✓	✓	✓
Long Diff in 2015	0.02	-0.12	0.96***	0.91***
	(0.16)	(0.19)	(0.16)	(0.20)
Mean dependent variable	-37.89	0.08	-32.75	-1,930.18
Observations	128	128	128	128

Robust standard errors in parentheses. The square root of cell sizes is used to construct weights for the observations. Cell sizes are identical across specifications, except for the income one as we restrict to individuals reporting strictly positive income.
***p<0.01, **p<0.05, *p<0.1