

Impact of Education Level Changes on Inequality

Andrea Assenza, Anshruta Thakur, Marlon Morales, Mauricio Vargas, Nikolaos Papadatos

Fall 2024



Overview

Introduction

Data Description

Methodology

Main Results

Conclusions

Research Question

What is the impact of a substantial change in the share of the population with no more than 12 years of schooling on between-education inequality?

Methodological Approach

The exercise employs a difference-in-differences (diff-in-diff) framework, using the change in the share of the population with less than 12 years of education as the endogenous variable, instrumented using the Mexican population growth between 1990 and 2000 in different regions. Key controls are derived from 1990 data.

Key Findings

- No significant effects were observed on differences in wage growth rates between highly educated and less-educated workers.
- Unemployment rates for highly educated workers increased significantly more than for less-educated workers.
- Rates of labor force inactivity (NILF) grew significantly more for highly educated workers compared to less-educated workers.

The two main data sources used in this study are:

1. **Data I:** The 1990 and 2000 censuses from the IPUMS-USA (Integrated Public Use Microdata Series from United States).
 - ▶ <https://usa.ipums.org/usa/>
2. **Data II:** The “ConsPUMA” (Consistent Public Use Microdata Areas), which is a geographic aggregation of PUMAs (Public Use Microdata Areas).

- **Step I:** Construct native labor-market outcomes by ConsPUMA “c”, year “y” (1990 and 2000), and education “e” (e = H for ≥ 12 years schooling vs. e = L for < 12 years)
- **Step II:** Construct working-age population with (e = H) and without (e = L) more than 12 years of schooling by cy, N_{cy}^e .
- **Step III:** Measure change in share with ≤ 12 years of schooling

$$x_c = \frac{N_{c2000}^L}{N_{c2000}^L + N_{c2000}^H} - \frac{N_{c1990}^L}{N_{c1990}^L + N_{c1990}^H}$$

- **Step IV:** Construct instrument:

$$z_c = \left(\frac{I_{c,1990}^{\text{Mex}}}{I_{1990}^{\text{Mex}}} \right) \left(I_{2000}^{\text{Mex}} - I_{1990}^{\text{Mex}} \right) = \frac{N_{c1990}^L}{N_{c1990}^L + N_{c1990}^H} \frac{I_{2000}^{\text{Mex}} - I_{1990}^{\text{Mex}}}{I_{1990}^{\text{Mex}}}$$

- ▶ $I_{c,y}^{\text{Mex}}$ = Mexican population of c in year y
- ▶ I_y^{Mex} = total Mexican population of in y
- **Step V:** Using 2SLS, project changes in conspuma relative outcomes for higher vs. lower education on x_c , instrumenting with z_c (with controls defined in 1990).

Given a native outcome w_c , we define:

- $w_{c,1990}^H$ as the native outcome in 1990 for highly educated workers.
- $w_{c,1990}^L$ as the native outcome in 1990 for less-educated workers.
- $w_{c,2000}^H$ as the native outcome in 2000 for highly educated workers.
- $w_{c,2000}^L$ as the native outcome in 2000 for less-educated workers.

The dependent variable y_c is defined as:

$$y_c = (w_{c,2000}^H - w_{c,1990}^H) - (w_{c,2000}^L - w_{c,1990}^L) \quad (1)$$

Given the dependent variable y_c , we propose the following diff-in-diff model:

$$y_c = \alpha + \beta X_c + W_c' \gamma + \epsilon_c \quad (2)$$

where W_c represents the control vector defined in 1990, and X_c is the endogenous variable with an instrument Z_c , as detailed in the previous section.

Table 1: First-Stage 2SLS Results

Endogenous Variable: Change in the share of the population with less than 12 years of education			
		(1)	(2)
	Partial F Statistic	22.376	44.856
	P-value	0.0000	0.0000
Constant		-0.0362*** (0.0027)	-0.2737*** (0.0311)
Share of the college-educated population in 1990			0.0130 (0.0229)
Share of employment among the foreign-born population 1990			0.0607*** (0.0119)
Share of employment among women in 1990			0.0353 (0.0775)
Share of population with high school education in 1990			0.2617*** (0.0170)
Logarithm of total population in 1990			0.0000 (0.0013)
Share of employment in manufacturing in 1990			-0.0431*** (0.0139)
Instrument		0.2015*** (0.0426)	0.2276*** (0.0340)
R^2		0.1834	0.7092
Partial R^2		0.1834	0.2424

Notes: $N = 741$. Robust standard errors (in parentheses) are clustered at the state level. A Partial F-statistic below 10 is typically considered weak evidence of instrument relevance. The p-value is calculated using a $\chi^2(1)$ distribution. Weighted regression based on the population size in each "CONSPUMA."

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Main Results

Instrumental Relevance

The results in Table 1 present the partial F -statistic corresponding to the first stage of the 2SLS regression. According to the commonly accepted rule by Staiger and Stock (1997), and considering that the value of the partial F -statistic exceeds 10 in both exercises, it is concluded that the instrument is relevant for the measurement of the change in the proportion of individuals with less than 12 years of schooling.

Main Results

Causal Effect on Native Population Outcomes

Table 2: Results of the 2SLS on the Difference Between High and Low Education Levels for Various Outcomes of the Native Population

Native Population Outcome	Growth Rate of Wages		Growth Rate of Unemployment		Growth Rate of NILF	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant Term	-0.010*** (0.002)	0.120* (0.068)	0.163*** (0.017)	-0.047 (0.198)	0.0122*** (0.0020)	0.0749 (0.0516)
Share of the college-educated population in 1990		-0.012 (0.043)		0.254** (0.129)		-0.2459*** (0.0522)
Share of employment among the foreign-born population in 1990		-0.076* (0.039)		0.200** (0.079)		-0.0692** (0.0275)
Share of employment among women in 1990		-0.353*** (0.080)		-0.225 (0.354)		0.1914* (0.1096)
Share of population with high school education in 1990		0.011 (0.063)		0.230* (0.125)		-0.0434 (0.0511)
Logarithm of total population in 1990		0.002 (0.001)		-0.002 (0.005)		-0.0033** (0.0017)
Share of employment in manufacturing in 1990		0.044* (0.025)		0.328*** (0.093)		0.0225 (0.0273)
Change in the share of the population with less than 12 years of education	-0.074 (0.083)	0.154 (0.207)	1.902*** (0.503)	1.221** (0.518)	0.4093*** (0.1279)	0.9830*** (0.1379)

Notes: $N = 741$. Robust standard errors (in parentheses) are clustered at the state level. Weighted regression based on the population size in each CONSPUMA.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Main Results

Causal Effect on Native Population Outcomes

The results in Table 2 show that, given an increase in the change in the share of the population with less than 12 years of education (x):

- There is no significant effect on the relative difference between the growth rates of wages for highly-educated and low-educated native workers (Columns 1 and 2).
- The growth in the unemployment rate of highly-educated native workers is significantly greater than that of low-educated workers (Column 4).
- The growth in the labor inactivity rate of highly-educated native workers is significantly greater than that of low-educated native workers (Column 6).

Main Results

Causal Effect on Native Population Outcomes

Defining a region as highly exposed if it ranks in the 75th percentile ¹ of the change in population share with less than 12 years of education, while a low-exposure region corresponds to the 25th percentile:

- The unemployment growth rate for highly educated native workers relative to less-educated native workers in highly exposed regions is projected to be 4.58 percentage points higher than in low-exposure regions.
- The NILF growth rate for highly educated native workers relative to less-educated native workers in highly exposed regions is projected to be 3.68 percentage points higher than in low-exposure regions.

¹Percentiles weighted by population size.

- **Instrument Relevance:** The instrument for the change in the population with less than 12 years of education is relevant, as the Partial F-statistics far exceed the threshold of 10.
- **Main Findings:**
 - ▶ *Wage Growth:* No significant effects on the differences in wage growth rates between highly educated and less-educated workers were observed given an increase in the low-education population.
 - ▶ *Unemployment Growth:* The unemployment rate for highly educated workers increased significantly more than for less-educated workers following an increase in the low-education population.
 - ▶ *Inactivity Growth:* Rates of labor force inactivity (NILF) grew significantly more for highly educated workers than for less-educated workers following an increase in the low-education population.
- **Regional Disparities:**
 - ▶ Highly exposed regions (top 75th percentile of changes in low-education population) show unemployment growth for highly educated workers exceeding that of less-educated workers by 4.58 percentage points.
 - ▶ Inactivity rates for highly educated workers in these regions exceeded those of less-educated workers by 3.68 percentage points.

