

Cortez, Willy W.; Nowak-Lehmann D., Felicitas

Working Paper

Can informality explain inequality? The case of Mexico, 1987-1999

IAI Discussion Papers, No. 222

Provided in Cooperation with:

Ibero-America Institute for Economic Research, University of Goettingen

Suggested Citation: Cortez, Willy W.; Nowak-Lehmann D., Felicitas (2012) : Can informality explain inequality? The case of Mexico, 1987-1999, IAI Discussion Papers, No. 222, Georg-August-Universität Göttingen, Ibero-America Institute for Economic Research (IAI), Göttingen

This Version is available at:

<http://hdl.handle.net/10419/92999>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

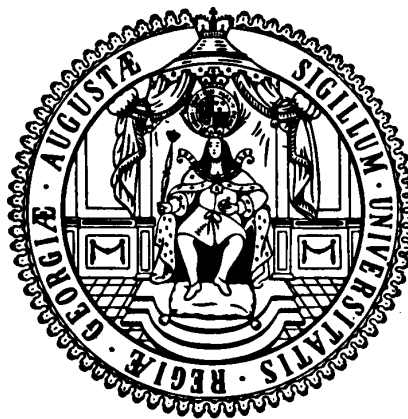
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**Ibero-Amerika Institut für Wirtschaftsforschung
Instituto Ibero-Americano de Investigaciones Económicas
Ibero-America Institute for Economic Research
(IAI)**

**Georg-August-Universität Göttingen
(founded in 1737)**



Diskussionsbeiträge · Documentos de Trabajo · Discussion Papers

Nr. 222

**Can Informality Explain Inequality?
The Case of Mexico, 1987-1999**

Willy W. Cortez, Felicitas Nowak-Lehmann D.

November 2012

Can Informality Explain Inequality? The Case of Mexico, 1987-1999

Willy W. Cortez¹

Felicitas Nowak-Lehmann²

Abstract

In this paper we contrast the impact of the income gap between formal and informal sector on Mexico's income inequality to alternative explanations. Unlike previous studies we take a time series approach to analyze the determinants of income inequality for the period 1987-1999. We find that an increase in the wage gap between the formal and the informal sector and in the wage gap between the higher educated and the lower educated people leads to an increase in inequality. Inequality seems to be counter-cyclical, that is, an increase in output leads to a reduction in inequality. An unexpected result is that inequality declines when output is below its long term trend. We obtained mixed results about the role played by prices and inflation on inequality.

Draft*
(comments are welcome)

University of Goettingen, Germany

November, 2012

¹ Visiting Fellow at the Ibero-America Institute for Economic Research, University of Goettingen. Department of Quantitative Methods, University of Guadalajara, Mexico. I would like to thank Mexico's National Council of Science and Technology (CONACYT) for financial support (wcortez@cucea.udg.mx).

² Researcher Ibero-America Institute for Economic Research, University of Goettingen, Germany (fnowak@uni-goettingen.de).

I. Introduction

For the past several decades, Mexico's economic development has been accompanied by two striking features: high income inequality and a large informal sector. Both of these features have evolved with no clear trend along Mexico's economic cycles. Yet there are no systematic studies that have analyzed their likely interrelations. The Mexican literature on these two areas, even though very prolific, has not been able to assess the possible interconnections that may exist between them. In general, they seem to have gone in different directions: while the majority of studies on informality have been mainly descriptive, studies on income inequality have provided some explanations about the factors behind the latter's observed changes.³

The predominant view about the changes in income inequality attributes them mainly to two factors. On the one hand, changes in the relative demand and supply of skilled labor caused by trade liberalization. On the other hand, institutional changes in the labor market, caused by the increasing flexibilization of Mexico's labor market and the reduction of the labor unions' bargaining power.

Initial studies on informality were more concerned with testing the dual labor market hypothesis, that is, they concentrated on testing the extent by which the informal sector was a secondary market characterized by low productivity and hence low wages. An implicit assumption of the early studies was that the size of the informal sector was inversely related to the performance of the economy, i.e., during expansions the informal sector

³ Two additional characteristics of these studies are: their cross section nature and they have usually been comparative analysis that include few points in time.

would decline, while during recessions it would expand. In other words, the size of the informal sector was counter-cyclical. More recent studies have provided further evidence about the relationship between the formal and informal sectors; in particular, they have estimated the degree of labor mobility between formal and informal sectors. An important conclusion drawn from these recent studies is that under some circumstances an individual may choose to be in the informal sector because it maximizes her utility.

A possible explanation for the lack of empirical studies about the relationship that may exist between informality and income distribution is in part due to the lack of theoretical models. Winkelried (2005), however, provides a link between income distribution and informality. He develops a theoretical model where the size of informality depends on income inequality. In his model, informality is the dependent variable whereas income inequality is the explanatory variable. The idea is that the firm's decision to be in the formal or informal sector depends on the size of the demand it faces and given that demand depends on income distribution, the latter determines the firm's decision. Using data from Mexico's labor market he finds evidence in favor of his hypothesis i.e., higher inequality tends to increase the size of the informal sector.

However, the causality between income distribution and informality is not necessarily from the former to the latter. Bhattacharya (2007), for example, using a general equilibrium model, carries out a simulation exercise for a developing economy characterized by the existence of an informal sector, income inequality and rural-urban migration. He is interested in determining the conditions under which the economy would obtain an inverted U-shaped curve of income distribution. He finds that the behavior of income

inequality (measured by the Gini coefficient) depends crucially on the wage gap between the formal and informal sectors, which, in turn, depends on a number of other variables.⁴

Unlike previous studies on income inequality, we perform a time series analysis of the determinants of income inequality. In particular, we analyze the relationship between inequality and informality and how this relationship may change over the business cycle. The analysis includes an evaluation of the role played by the rates of return to education, inflation and unemployment on the behavior of inequality over time.

Our paper contains four additional sections. In section II we present a brief literature review about studies done on income inequality and informality for the Mexican economy. In section III we present some stylized facts of the informal sector and income inequality for the period 1987:01-1999:04. In section IV we describe the data used and provide descriptive statistics.. Section V presents the econometric model to carry out the empirical analysis and the main results, while section VI concludes.

II. On Income Inequality and Informality in Mexico

There is evidence that at least for the last 70 years income inequality in Mexico has been very high by international standards. In effect, one of the earliest studies done on Mexico's income inequality identified that in 1950, Mexico had higher inequality than the USA or United Kingdom 20 years earlier (Martinez de Navarrete, 1960). Since then income inequality in Mexico has shown fluctuations with no clear trend.

During the fast industrialization period, 1950-1970-, income inequality also increased.⁵ The increase in inequality seems to have responded to the higher growth rate of wages of

⁴ like skilled-based technological change, trade union pressure, government policies, among other variables.

skilled labor.⁶ Afterward, came a period when inequality declined which made some people think that Mexico was a case where Kuznets' hypothesis was accomplished (Székely, 2005). However, a new upward trend began from the mid-eighties that lasted until the second half of the nineties. Subsequently, income inequality began a new declining trend that seems to have ended by 2005. In short, during the past several decades, the overall behavior of income inequality is one characterized by cyclical fluctuations. As a matter of fact, the current rate of inequality is not much lower than it was 70 years ago.

A key characteristic of the income inequality studies is that there is not a consensus about the actual magnitude of inequality because it varies according to the index used⁷, the definition of income used⁸, the period analyzed and the region studied, among other things. Two of the leading explanations about the observed changes in inequality since the mid-eighties is that they have been caused by changes in relative demand and supply of skilled labor caused, in turn, by trade liberalization and technical change (Campos et al 2012) and changes in labor markets institutions, particularly the flexibilization of labor contracts and the reduction of labor unions' bargaining power (Cortez, 2001).

Unlike inequality studies, research on informality has drawn less attention from economics analysts. Most studies about the informal sector have been focused on determining its size and to measure the wage gap between formal and informal sectors after controlling for workers' personal characteristics; that is, they have been interested in testing the dual labor markets hypothesis. Recent papers estimate the transition rate

⁵ Nugent y Tarawneh (1982)

⁶ Gollas (1983) presents evidence that the overall behavior of inequality during those years responded to the behavior of labor income.

⁷ The Gini coefficient, Theil Index, Atkinson Index, SD of Log Income, etc

⁸ Labor income, total income, monthly income, hourly wage rate, per-capita, household income, adjusted household, and so on.

between formal and informal employment on the assumption that workers move from formal to informal employment (and vice-versa) according to labor market conditions (Calderon, 2000; Alcaraz 2009; Rodriguez-Oreggia, 2007). One of the conclusions we can draw from these studies is that formal and informal sectors are intertwined. Furthermore, the view that people have informal jobs because they did not have another option is increasingly challenged. This implies that for some workers to be in the informal sector is a rational decision because it maximizes their expected utilities.

Leal Ordoñez (2010), in turn, estimates the impact of eliminating the informal sector of the economy. Using a Dynamic System General Equilibrium (DSGE) model, he calculates that if the government place a complete enforcement policy so that informality is totally wiped up, labor productivity would increase by 17%, taxes would be reduced by 36% of what it is now; whereas total factor productivity would increase by 4% and wages by 9%.

Informality is usually measured by two imperfect proxies. The first one considers firms with less than 5 workers to be informal, while the other definition believes workers without social security benefits, -like health insurance and retirement funds-, is a better indicator of informality.⁹ In any event, Marcouiller et al (1997) found that when using the size of the firm definition the proportion of workers in the informal sector were about 30.8%, while when using the social security benefits definition about 43.2% of the labor force was in the informal sector.¹⁰ They also found that after controlling for personal characteristics, there was a wage premium associated with the informal sector. Rodriguez-Oreggia (2007), in turn, found that about 60% of Mexican working population did not have

⁹ It should be pointed out that the latter responds to a definition of informal worker more than an informal firm. Under this definition then an informal worker could be employed by a formal firm.

¹⁰ These percentages correspond to the third quarter of 1990 using the ENEU.

social security benefits. These estimates indicate that the social benefits proxy might be a more biased indicator of informality than the firm's size.¹¹

Cortez and Islas-Camargo (2009), using a definition based on whether the firm has name or registry, estimated that during 1994-2004, workers laboring in these informal firms represented about 19% of economically active population.

III. Some Stylized Facts about Informal Sector and Income Inequality in Mexico¹²

We start by presenting some descriptive statistics about the informal sector such as women's participation in the labor market, the employment distribution by firms' size and economic sectors, educational distribution and relative monthly income. Then we present some statistics of income inequality.

Table 1 shows the distribution of workers by gender in both sectors of the economy. First of all, about 20 percent of total employed workers are in the informal sector.¹³ Moreover, within the formal sector about a third of workers are women, whereas in the informal sector the proportion is a little higher, about 39%. In both sectors, -formal and informal ones-, women participation in the labor market shows steady increases.

¹¹ Several studies indicate that since the late eighties, Mexican firms have used a number of schemes to lower their labor costs. These have included subcontracting, short term contracts which did not include social benefits to workers.

¹² The statistics we present are for workers between 14 and 65 years old and are based on the quarterly National Survey of Urban Employment (ENEU), 1987Q1-1999Q4. The definition of informality is based on whether or not the enterprise has name or registry. Unlike other studies that define informality according to whether the worker has Social Benefits or not. These two ways to measure informality do not necessarily coincide (See for instance, Rodriguez-Oreggia, 2007).

¹³ In the Appendix, Figure A1, we illustrate the behavior of the employment in the informal sector as a proportion of total employment for the period 1987:01-1999:04.

Table 1: Workers' Distribution by Gender and (in)Formality

Period	Sector	Female	Male	N
1987 Q3	Formal	32.2	67.8	6243092
	Informal	37.0	63.0	1894552
1990 Q3	Formal	32.6	67.4	7305706
	Informal	37.7	62.3	1751792
1993 Q3	Formal	33.2	66.8	9500143
	Informal	39.8	60.2	2343077
1996 Q3	Formal	35.9	64.1	10930928
	Informal	42.5	57.5	2996769

Source: ENEU (DataBase 1987:03, 1990:03, 1993:03, 1996:03)

In the Appendix A we present the employment distribution by firms' size and economic sectors for both formal and informal enterprises. With regards to firms' size, Tables A1 and Table A2 confirm what is conventional wisdom; that is, informal employment is heavily concentrated in micro-enterprises (firms with less than 5 workers): more than 90 percent of informal employment is absorbed by this type of firms. Formal employment, on the other hand, shows a less skewed distribution than the informal one. In fact, it shows a bi-modal distribution in the sense that the larger percentage of employment takes place in small and medium size firms (38%) and in firms that employ more than 250 workers (42%).

When looking at the employment distribution by economic sectors, Tables A3 and A4 also corroborate what we already know from existing studies. Informal employment is concentrated in two economic sectors: Social and Personal Service and Commerce and Restaurants. On average, they occupy at about 69% of informal workers. We also see informal enterprises in two other sectors: Manufacturing and Construction. On average,

about 25% of informal employment occurs in these sectors. Formal employment, in turn, is less concentrated across economic sectors. The three leading sectors are Manufactures, Social and Personal Services and Commerce and Restaurants which employ about 27%, 28% and 22% of formal employment respectively. Transportation and Financial Services come in a second tier group as they employ a lower percentage. On average they both employ about 14% of formal workers.

As mentioned earlier there are two competing views about the nature of the informal sector. On the one hand, there is the notion that informality exists because the formal sector is unable to accommodate a growing labor force. As such, it represents a second best option for people that face unemployment. On the other hand, there is the view that under some conditions, informality provides a higher utility to some individuals than formal employment. In this case, employment in the informal sector is a rational choice.

From either perspective's point of view, one should not expect that the educational distribution in both sectors be much different one from the other. The latter, however, is contrary to what early supporters of the dual labor market would argue, i.e., that in the informal sector are those who cannot find employment in the formal sector because of their low levels of education, their age or gender.¹⁴

Figures 1a and 1b show the educational distribution in the formal and informal sectors for male and female workers, respectively.¹⁵ The vertical axis measures the percentage of workers for each educational level within each sector (formal and informal). Overall, the educational distribution in the case of male and female workers presents the same

¹⁴ It has been argued that young people and old people as well as women are more likely to be employed in the informal sector.

¹⁵ These are the average distribution of four points in time (1987q3, 1990q3, 1993q3 and 1996q3). Although the sample size increased during those years, the percentage distribution remained more or less constant.

characteristics. Workers in the formal sector have, on average, a higher level of education than workers in the informal sector. It should be noted however, that the percentage of workers with High School or even College in the informal sector is not insignificant.¹⁶

Figure 1a

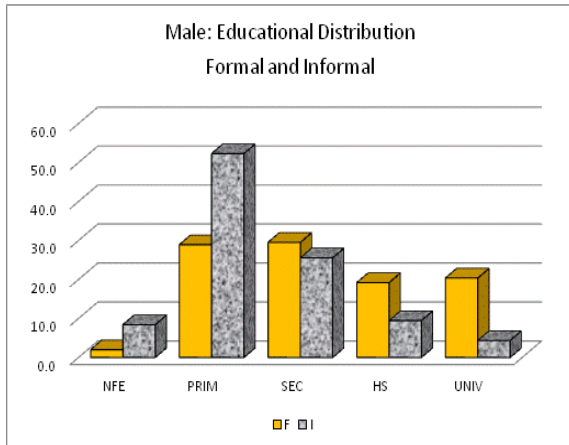
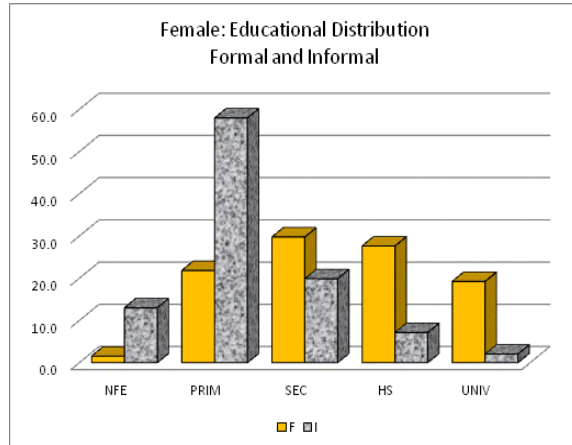


Figure 1b



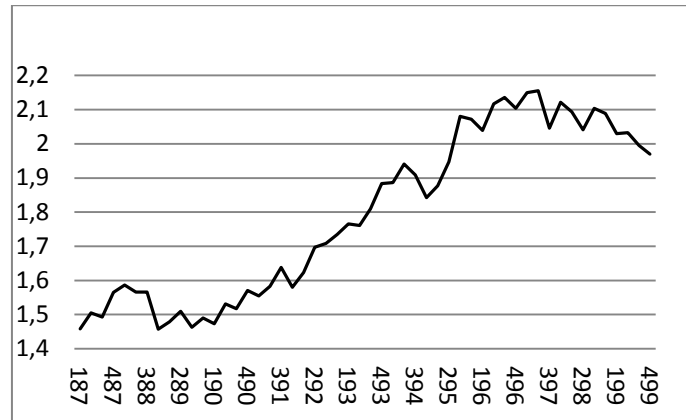
Source: Own estimates using ENEU (several years).

We now turn to a brief discussion about the income gap between formal and informal employment. Figure 2 shows the ratio of monthly income between the formal and informal employment during 1987:01-1999:04. In general, we observe a deteriorating trend of the monthly income of the informal sector compared to the formal one. As we can observe the figure below, its increasing trend started in the early nineties, a period characterized by a growing economy. The increasing trend had a sudden break during 1994-5, caused mainly by the Mexican crisis of that particular year. The ratio resumed its upward trend although at a slower rate during 1996. Between 1997 and 1999, we observe a reversal of the initial upward trend of the ratio. Overall, between 1989:01-1999:04, monthly income in the

¹⁶ In the case of male workers the percentage is as follows: Secondary (15.5%), High School (9.4%), College (4.3%), whereas in the case of female workers the percentage are: Secondary (19.9%), High School (7.1%) and College (2%).

formal sector grew faster than income in the informal sector, so that the relative income of the formal sector grew at an average quarterly rate of 0.67%.

Figure 2: Relative Income: Formal and Informal Sector



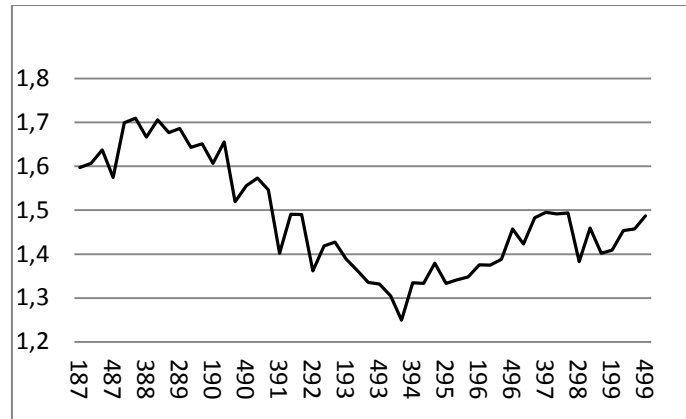
Source: own estimates using ENEU (several years)

Within the informal sector we identify two types of workers: those who said they were self-employed or hired people (let's call them "informal bosses") and those who responded they are employees and work for someone else ("Informal employees"). Figure 3 exhibits the ratio of the monthly income of these two types of informal workers. In general, informal bosses earn more than informal employees. The income differential however declined during the period that the economy grew i. e., early nineties. It reaches its lowest point by the end of 1994, which is the period where the Mexican crisis began and starts a new upward trend in 1996.

When we compare the behavior of the income gap between the formal and informal sector with the income gap between the two types of informality, we observe that the increasing income gap of the former coincided with the declining trend of the latter. This

seem to indicate that the formal-informal income gap is mainly determined by the relative income of the informal bosses (with respect to informal employees).

Figure 3: Relative Income: Informal Sector



Source: own estimates using ENEU (several years)

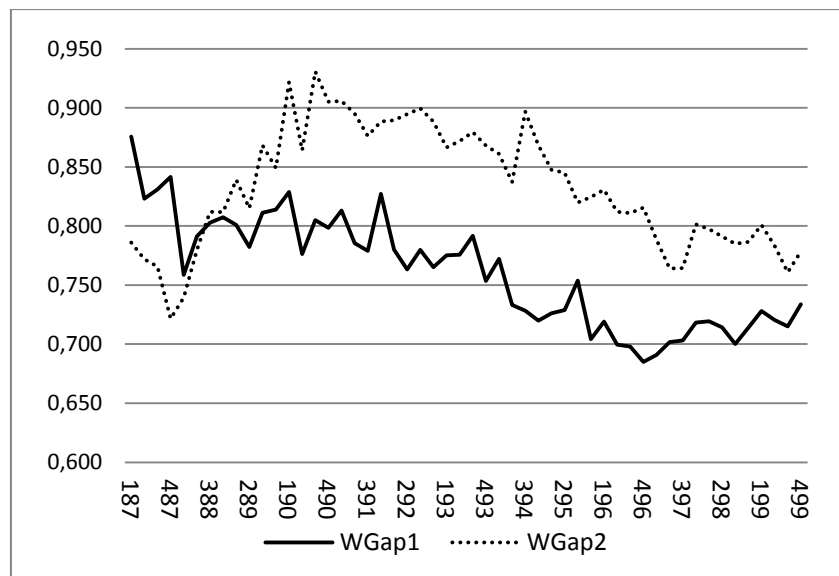
Thus the question about the wage premium of informal employment arises. To calculate it, we estimate a quarterly Mincerian wage equation. We control for workers' individual characteristics as well as for other factors like city, firms' size, economic sector and type of payment. Within the informal sector we identify two types of informal workers. On the one hand, there is the self-employed or subcontractor who hires other informal workers (I1), and, on the other hand, there is the informal employee, who works for somebody else (I2). Thus,

$$(1) \quad \ln W_i = \alpha_1 G_i + \alpha_2 A_i + \alpha_3 A_i^2 + \sum_j^J \beta_j e_{j i} + \sum_k^K \delta_k c_{k i} + \sum_m^M \gamma_m S_{m i} + \sum_n^N \theta_n F_{n i} + \sum_p^P \vartheta_p P_{p i} + \sum_q^Q \pi_q I_{q i} + \mu_i$$

Where W_i is real monthly income of individual i th, G_i is a dummy variable that defines sex; A_i is age of the i th individual; $e_{j i}$ is the j th educational level of individual i th; $c_{k i}$ is a set of

dummy variables that controls for the city where the i th individual lives; S_{mi} is a set of dummy variables that controls for economic sectors where the individual works; F_n is a set of dummy variables that controls for firms' size; P_{pi} is a set of dummy variables that control for the type of labor contract the worker has; and I_{qi} is a dummy variable that controls for informality ($q = I1$ and $I2$). In the Appendix B we further describe each of the variables used to estimate the wage differential of the informal sector. One characteristic of the ENEU is that the number of cities included in the sample increased over the years. It began with only 16 cities, which by 1999 had increased to 45. To have a homogeneous time series data set we restrict the analysis to the original 16 cities.¹⁷

Figure 4: Estimated Income Gap¹⁸



Source: own estimates using ENEU (several years)

We can observe from figure 4, the evolution of the (estimated) income gap between formal workers and informal self-employee (or informal bosses), WGap1, and formal

¹⁷ In the Appendix B we list the 16 cities included in the analysis.

¹⁸ We obtained the wage differential by applying antilog to the estimated income contribution of the informal sector.

workers and informal employees, WGap2. Three results are worth mentioning. First, the income gap is higher between formal workers and informal employees than between formal workers and informal bosses. Second, throughout the nineties both income gaps have steadily declined. Third, the last two years of the nineties there seems to be a tendency towards convergence of both income gaps.

Income Distribution

We now describe the behavior of income inequality during the period 1987:01-1999:04. We estimate two indicators of inequality: the log standard deviation and the Gini coefficient. We use monthly real labor income to estimate inequality.¹⁹ Figures 4 and 5 corroborate what is also conventional wisdom: inequality started its increasing trend in the late 1980s and continued until the mid nineties. By the late nineties a declining trend began.

Figure 4: SD of Ln Income

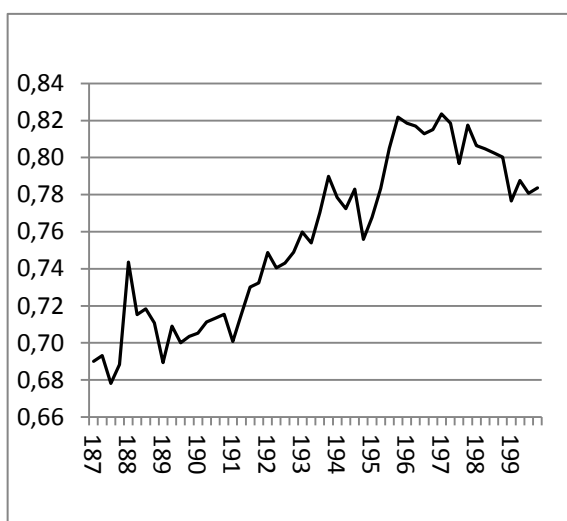
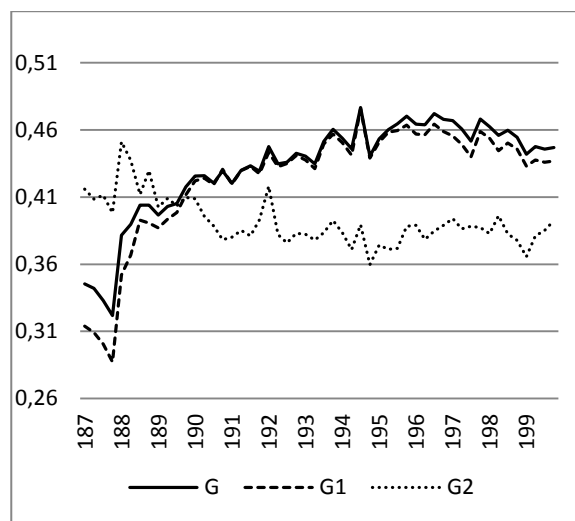


Figure 5: Gini



¹⁹ Income were deflated by consumer price index (2QJune 2002=100).

Source: own estimates using ENEU (several years)

Figure 5 in addition to the economy-wide Gini coefficient also shows the Gini coefficient for the formal (G1) and informal sectors (G2) as well. Three results are worth noting. First, income inequality in the formal is much higher than what it is in the informal sector. Second, economy-wide inequality seems to be explained mainly by what is going on in the formal sector. Third, the Gini coefficient of the informal sector shows a slight declining trend during the period when overall inequality was increasing.

In the next section we further explore the role of the informal sector on overall inequality. Unlike previous studies about income inequality we carry out a time series analysis and test alternative explanations about the factors behind income inequality in Mexico. In particular, we contrast the role played by the rate of return of education, as well as by the wage differential between formal and informal employment and how these impacts may change over the business cycle.

IV. An empirical model of income inequality

We begin by assuming that income distribution is somewhat affected by the wage differential between workers in the formal and informal sectors. Mexico does not have unemployment benefits, thus a worker cannot be out of work for very long time. If she/he cannot find a job in the formal sector, she/he would have to enter the informal sector which would induce further changes in inequality.

To the extent that in general labor market conditions depend on the performance of the economy, we include the output as one variable that explains inequality over time (Y). It is unclear whether an expansion would reduce or increase inequality. If the expansion induces an increase in the relative demand for skilled labor, then we should expect a positive relation, i.e. an increase in inequality. On the other hand, if the expansion pulls people out of the informal market or generates an increase in the relative demand for unskilled labor, then one would expect a negative effect on inequality, i.e. a decrease of inequality.

There is a consensus among researchers that one of the leading factors behind the changes in income inequality in Mexico is related to the changes in the return of skilled labor in contrast to unskilled labor. Assuming that the level of skill is proportional to the level of education, we can argue that income inequality is affected by the wage differential between different levels of education (WGE).

Besides, two additional variables are considered in the model : the unemployment rate and the price level. The rate of inflation has distributional impacts which affects income distribution through several mechanisms. For instance, income distribution is somewhat affected by the wage bargaining that occurs as a result of inflation. However, wage increases are not homogeneous across economic regions, economic sectors, much less across formal and informal workers. It depends on the bargaining power of trade unions. Another possible channel is through the changes caused by the change in the minimum wage which is indexed to the inflation rate. The net effect however is uncertain for it depends on the bargaining position of the different trade unions. The model is thus

$$(2) \quad G_t = a \cdot WG_t^\varphi \cdot Y_t^\beta \cdot WGE_t^\gamma \cdot U_t^\delta \cdot \Delta P_t^\eta \cdot e^{u_t}$$

Where a is a constant term, WG_t is the wage gap between the formal and informal sectors, Y_t is the seasonally adjusted GDP in real terms, WGE_t is the wage differential between skilled and unskilled workers, U_t is the rate of unemployment, ΔP_t is the inflation rate measured by the consumer price index. We assume these factors to influence inequality in a non-separable and non-additive way and formulate the following multiplicative model. We found all variables to be in a long-run equilibrium, i.e. cointegrated.

To carry out the empirical analysis we use the National Survey of Urban Employment (ENEU) from 1987Q1 to 1999Q4 published quarterly by Mexico's National Institute of Statistics, Geography and Informatics (INEGI). We limit the sample to include individuals between 14-65 years old and include only those who work for a monetary income.

V. Estimation Technique and Main Results

To facilitate estimation, the model is linearized by taking logarithms on both sides of the equation so that the equation reads as follows

$$(3a) \quad \ln(G_t) = \ln a + \varphi \cdot \ln(WG_t) + \beta \cdot \ln(Y_t) + \gamma \cdot \ln(WGE_t) + \delta \cdot \ln(U_t) + \eta \cdot \ln(\Delta P_t) + \mu_t$$

$$(3b) \quad g_t = \alpha + \varepsilon_1 wg_t + \beta_1 y_t + \phi_1 wge_t + \chi_1 u_t + \delta_1 \Delta p_t + \mu_t$$

with variables in small letters being the ln-values of the original variables, μ_t is supposed to be iid and uncorrelated over time. It is further postulated that $\text{cov}(X_{it}, \mu_t) = 0$. However, there are basically two econometric issues that deserve discussion at this point. The first and most important issue is the endogeneity of the explanatory variables $\text{cov}(X_{it}, u_t) \neq 0$, which causes feedback effects between inequality and the right-hand side variables. The second issue is whether and how to deal with omitted variables and whether autocorrelation should be controlled for.

The issue of endogeneity of the explanatory variables

Neglecting endogeneity leads to biased and even inconsistent estimates so that a strategy to tackle endogeneity is called for. In our case, some right-hand side variables (growth, unemployment and the wage-gap between the formal and the informal sector, wage differential between different educational levels and unemployment) might be endogenous (affected by a common event or omitted variables, such as the institutional framework) or stand in a bi-directional relationship with inequality (a high level of inequality might speed up or reduce growth; a high level of inequality might shape institutions that only care about the people who are “in” and have formal jobs thus increasing the wage gap between the formal and informal sector etc.). To control for endogeneity, we estimate the model by means of dynamic ordinary least squares, DOLS, (Stock and Watson, 1993; Wooldridge, 2009). DOLS is also known as the leads and lags approach proposed by. It can be shown that by decomposing the error term and inserting the leads and lags of the right-hand side variables in first differences, the explanatory variables become (super-) exogenous and the regression results thus become unbiased.

Controlling for endogeneity requires the decomposition of the error term of equation (3b) into the endogenous changes of the right-hand side variables, which are correlated with μ_t 's.²⁰ This leads to the following equation in which all explanatory variables from the baseline model can be considered exogenous:

$$(4) \quad g_t = \alpha + \varepsilon_1 w g_t + \beta_1 y_t + \gamma_1 w g e_t + \chi_1 u_t + \delta_1 \Delta p_t + \sum_{p=-1}^{p=+1} \varepsilon_{2p} \Delta w g_{t-p} + \sum_{p=-1}^{p=+1} \beta_{2p} \Delta y_{t-p} + \sum_{p=-1}^{p=+1} \gamma_{2p} \Delta w g e_{t-p} + \sum_{p=-1}^{p=+1} \chi_{2p} \Delta u_{t-p} + \sum_{p=-1}^{p=+1} \delta_{2p} \Delta^2 p_{t-p} + v_t$$

with α representing the “natural rate of inequality” and Δ indicating that the variables are in first differences; Δ^2 indicates that the variables are in second difference, the error term v_{it} should fulfil the requirements of the classical linear regression model.

The wage gap between formal and informal sector, gdp, the wage differential by educational level, inflation etc. all become exogenous and the coefficients $\varepsilon_1, \beta_1, \gamma_1, \chi_1$, and δ_1 follow a t-distribution. This property allows us to draw statistical inferences on the impact of the explanatory variables on overall inequality. On the other hand, $\varepsilon_2, \beta_2, \gamma_2, \chi_2$ and δ_2 are coefficients that belong to the endogenous part of the explanatory variables and do not follow a t-distribution. Nevertheless, this does not affect our inferences since we are not interested in the influence of these “differenced variables” on inequality.

Yet the application of DOLS is not so standard and a prerequisite for using the DOLS approach is that the variables entering the model are non-stationary and that the series are

²⁰ Usually the leads and lags of the variables in first differences are inserted (a classical DOLS procedure). We apply a simple, reduced DOLS, which contains only the first lags and leads of the variables in first differences. This reduction was necessary due to the limited number of observations.

cointegrated. In the Appendix D we present the results of the unit root and the cointegration test.

Dealing with omitted variables and unexpected events/shocks

Having found cointegration, we can be sure that omitted variables do *not* systematically influence our long-run relationship between inequality and macroeconomic and structural variables. The error term is stationary [I(0)], a characteristic of cointegration. The finding of cointegration is not compatible with having important omitted variables explaining and impacting on the long-run relationship that we have established in equation (4).

Nonetheless, the error term might still contain some unexpected events/shocks.²¹ Indeed we observe positive temporary swings (positive autocorrelation) in the error terms that are controlled via a Feasible Generalized Least Squares procedure (FGLS). It is realistic to assume that the disturbances at time t are somehow related to past values of the disturbance term. Correcting for swings in the error term leads to the following equation:

$$\begin{aligned}
 g_t^* &= \alpha^* + \varepsilon_1 w g_t^* + \beta_1 y_t^* + \gamma_1 w g e^* + \chi_1 u_t^* + \delta_1 \Delta p_t^* + \sum_{p=-1}^{p=+1} \varepsilon_{2p} \Delta w g_{t-p}^* \\
 &+ \sum_{p=-1}^{p=+1} \beta_{2p} \Delta y_{t-p}^* + \sum_{p=-1}^{p=+1} \gamma_{2p} \Delta w g e_{t-p}^* + \sum_{p=-1}^{p=+1} \chi_{2p} \Delta u_{t-p}^* + \sum_{p=-1}^{p=+1} \delta_{2p} \Delta^2 p_{t-p}^* + v_t^*
 \end{aligned}
 \tag{5}$$

with (*) indicating that the variables have been transformed (purged from autoregressive processes) and that the error term v_t^* fulfils the requirements of the classical linear regression model (it has been made free from autocorrelation), i.e., $x_{i t}^* = x_{i t} - \rho x_{i t}$ and

²¹ Our finding from the cointegration test tells us that these shocks are only of a temporary nature.

$v_t^* = v_t - \rho_1 v_{t-1}$. Since the coefficient ρ_1 is usually unknown (as in our case), it has been estimated by means of, e.g. the Cochrane-Orcutt method, an FGLS procedure.

To estimate equation (5) we operationalize some variables. First, the income gap between formal and informal sectors is measured by the ratio of monthly income between both sectors (*wg*). Second, output is measured by gross domestic production (in millions of pesos Of 2002). To measure if the impact of output on inequality changes over the business cycle we added a dummy variable called "*crisis*" which takes the value of 1 when the difference between potential output and observed output is positive and 0 otherwise.²² To measure the wage differential across educational level, we estimated four indexes: two for the formal sector (*wge_f40*, *wge_f31*) and two for the informal sector (*wge_i40*, *wge_i31*). Within the formal and informal sector, we measure the ratio of monthly income of worker with college education and no formal education, *-wge_f40* and *wge_i40-* and the ratio of monthly income between high school and primary education, *-wge_f31* and *wge_i31* respectively. We also include open unemployment, consumer price index and inflation as explanatory variables.

With the exception of the dummy "*crisis*", we tested all other variables for unit roots. We conclude that the majority of variables are non-stationary, integrated of order 1, except the index of the consumer price level which is integrated of order 2. Inflation rate is integrated of order 1. Also, after analyzing the correlation among the variables, we decided to exclude unemployment because it was highly correlated with output level.²³

²² Potential output was estimated using the HP filter.

²³ The results of the Augmented Dickey Fuller (ADF) test are in the Appendix D.

Table 2 presents the empirical results for the 1987-1999 period²⁴. To see the stability of the estimated parameters we added one variable at a time. The results indicate that in effect the estimated parameters are stable and do not change dramatically as we included more variables to the model.

Table 2: Empirical Results (DOLS)

Variable	(1)	(2)	(3)	(4)	(5)	(6)
<i>c</i>	-0.9512 (-11.716)	2.5241 (1.5171)	2.1623 (2.8447)	3.9300 (3.6506)	4.3491 (2.3145)	5.64 (9.68)
<i>wg</i>	0.2409 (2.003)	0.3868 (3.3235)	0.3118 (5.5995)	0.1098 (2.2856)	0.3769 (5.9328)	0.3706 (5.93)
<i>y</i>		-0.2537 (-2.0913)	-0.2359 (-4.3125)	-1.3743 (-4.6634)	-0.3915 (-2.8302)	-0.4870 (-11.594)
<i>crisis</i>		-0.0139 (-1.3663)	-0.0075 (-1.2414)	-0.0064 (-2.1536)	-0.0283 (-5.7633)	-0.0299 (-8.2260)
<i>wge_f40</i>			0.1313 (3.0918)	0.2120 (7.8691)		
<i>wge_f31</i>					0.4293 (3.2468)	0.4260 (3.3721)
<i>wge_i31</i>					0.1443 (2.8802)	0.1591 (3.6961)
<i>p</i>				0.0556 (3.2512)	-0.015 (-0.7354)	
Δp						-0.1367 (-0.8892)
<i>leads & lags</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>AR</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Cointegration</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Adj_r²</i>	0.8302	0.8309	0.8554	0.9041	0.9459	0.9499
<i>F Stat</i>	26.675	16.8756	13.5272	17.1545	26.1925	28.3162
<i>DW</i>	1.9299	2.0672	2.1963	2.1131	2.3943	2.4286

Source: Own estimates.

Due to space limitation we do not include the estimated parameters, but all models include leads and lags as well as autorregressive components. We also tested for co-

²⁴ To eliminate an outlier problem, forming first differences and lagging of variables, the regressions actually start in 1989q1.

integration which they passed i.e, we accepted the Null that at least there was one co-integrating vector in the different models presented in Table 2.

The results indicate that the wage gap between formal and informal employment is positively associated with income inequality; that is, increases in the income gap between these two sectors results in increases in the gini coefficient. Output, on the other hand, is inversely related to inequality, i. e., increases in output leads to declines in inequality. However, our results also indicates that inequality declines in periods of crisis. This is an unexpected result given that inequality seems to be counter-cyclical.

About the role that income gap across educational level plays on inequality, we tested two indicators: the gap between college educated workers with respect to those with no formal education (wge_{40}) and the gap between worker with high school education and those with primary education (wge_{31}). We estimated these ratios for the formal and informal sector respectively. The income gap between college educated and workers with no formal education in the informal sector (wge_{i40}) was not significant in any specification; thus was excluded from the analysis. Model (5) measures the impact of the income differential between high school education and primary education on inequality. In this case, we did find evidence that this income gap in both the formal and informal sectors has a positive effect on inequality.

Models (4) and (5) include price index as additional explanatory variable, while model (6) includes inflation rate instead.²⁵ We did not find conclusive evidence about the impact of price index or inflation on inequality.

²⁵ We decide to include inflation rate instead of the price index because the former is I(1) while the latter is I(2).

The role played by the wage differential across different educational levels on income inequality is well documented in the Mexican literature and our results agree with them. However, when analyzing the behaviour of inequality over time we find that the overall performance of the economy is also important. Furthermore, our results suggest that changes in wage differential across educational level did not necessarily respond to changes in aggregate output. One possible explanation is that changes in wage differential are more related to the process of trade liberalization that the Mexican economy experienced during the late 80s and early 90s than to the overall performance of the economy.

VII. Some concluding remarks

In sum, we find that an increase in the wage gap between the formal and the informal sector and in the wage gap between the higher educated and the lower educated people leads to an increase in inequality. Inequality seems to be counter-cyclical, that is, an increase in output leads to a reduction in inequality under normal economic conditions. An unexpected result is that inequality declines when output is below its long term trend and when we observe a severe recession. It is unexpected because given that inequality declines with output growth, it also declines when the economy is below its long term trajectory. We obtained mixed results about the role played by prices and inflation on inequality.

References

- Bhattacharya, Prabir C. (2007) "Informal Sector, Income Inequality and Economic Development", Discussion Paper 2007/09, September, Centre for Economic Reform and Transformation, School of Management and Languages, Heriot-Watt University, Edinburgh, U.K.
- Campos, Raymundo; Esquivel, Gerardo; Lustig, Nora (2012) "The Rise and Fall of Income Inequality in México, 1989-2010", Working Paper No. 2012/10, January, UNU WIDER.
- Cortez, Willy W. (2001) "What is behind Increasing Wage Inequality in Mexico?", World Development, Vol. 29, Num. 11, November, pp. 1905-1922.
- Cortez, Willy W. and Alejandro Islas-Camargo (2009) NAIRU y Política Monetaria en México: 1987-2004, Universidad de Guadalajara, Guadalajara, México.
- Gollás, Manuel (1983) "La desigualdad del ingreso familiar en México: origen y causas", Vol. L (1), Num. 197, Enero-Marzo, pp. 329-347.
- Krstic, Gorana and Peter Sanfey (2010) "Earnings inequality and the informal economy: evidence from Serbia", European Bank for Reconstruction and Development, Working Paper 114, February.
- Leal Ordoñez, Julio C. (2010) "Informal Sector, Productivity and Tax Collection", Unpublished paper, Centro de Investigación y Docencia Económicas (CIDE), October.
- Maloney, William F. (1998) "The Structure of Labor Markets in Developing Countries", Policy Research Working Paper No 1940, The World Bank, June.
- Mann, Arthur (1979) "Un modelo de la distribución del ingreso en México", El Trimestre Económico, Vol. XLVI (1), Num. 181, Enero-Marzo, pp. 169-182.

- Martínez de Navarrete, Ifigenia (1960) "La Distribución del Ingreso y el Desarrollo Económico de México", Instituto de Investigaciones Económicas, Escuela Nacional de Economía, Universidad Nacional de Autónoma de México, México D. F.
- Marcouiller, Douglas; Veronica Ruiz de Castilla; Christopher Woodruff (1997) "Formal Measures of the Informal Sector Wage Gap in Mexico, El Salvador and Peru", *Economic Development and Cultural Change*, Vol. 45 Num. 2, January, pp. 367-392.
- Nugent, Jeffrey; Tarawneh, Fayez A. (1982) "Anatomía de los cambios ocurridos en la distribución del ingreso y de la pobreza entre la población económicamente activa de México en el periodo 1950-1970", *El Trimestre Económico*, Vol. XLIX (3), Num. 195, Julio-Septiembre, pp. 731-769.
- Rodríguez-Oreggia, Eduardo (2007) "The Informal Sector in Mexico: Characteristics and Dynamics", *Social Perspectives*, Vol. 9, Num. 1, Spring, pp. 89-175.
- Stock, J. H., Watson, M.W., 1993. A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica* 61, pp. 783-820.
- Székely, Miguel (2005). "Pobreza y desigualdad en México entre 1950 y 2004". *El Trimestre Económico*, LXXII(288), Octubre-Diciembre pp. 913-929.
- Winkelried, Diego (2005) "Income Distribution and the Size of the Informal Sector", mimeo, St. John's College, University of Cambridge, October.
- Wooldridge, J.M., 2009. *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, Mason (OH).

Appendix A

Table A1: Formal Employment by Firm's Size

(Percentage)

	1987q3	1990q3	1993q3	1996q3	1999q3
Self-empl	3.8	6.2	5.6	5.6	5.8
2-5	14.3	16.2	19.2	21.2	19.6
6-50	20.8	21.1	20.8	21.6	21.8
51-100	6	6.3	6	5.8	5.8
100-250	5.5	5	4.4	4.2	3.8
251 +	49.6	45.3	43.9	41.6	43.3

Source: Encuesta Nacional de Empleo Urbano (several years)

Table A2: Informal Employment by Firm's Size

(Percentage)

	1987q3	1990q3	1993q3	1996q3	1999q3
Self-empl	43.0	48.1	44.6	58.9	58.9
2-5	48.2	45.9	48.9	36.1	36.4
6-50	8.2	5.2	5.7	3.6	3.5
51-100	0.3	0.3	0.2	0.1	0.2
100-250	0.1	0.1	0	0.2	0
251 +	0.2	0.4	0.6	1.1	0.9

Source: Encuesta Nacional de Empleo Urbano (several years)

Table A3: Formal Employment by Economic Sector

(Percentage)

	1987q3	1990q3	1993q3	1996q3	1999q3
Agr	0.3	0.4	0.5	0.6	0.6
Min	0.9	0.4	0.3	0.3	0.3
Man	30.1	29.1	25.8	24.5	27.0
Cons	3.1	3.0	3.8	2.7	2.9
Util	1.0	0.8	0.6	0.9	0.8
Comm	20.2	22.2	23.7	25.7	25.3
Trans	6.0	6.9	7.7	7.4	7.8
F. Serv	6.8	6.8	7.4	8.2	7.7
S. Serv	31.3	29.8	29.6	29.0	26.9
Other	0.3	0.6	0.5	0.6	0.8

Source: Encuesta Nacional de Empleo Urbano (several years)

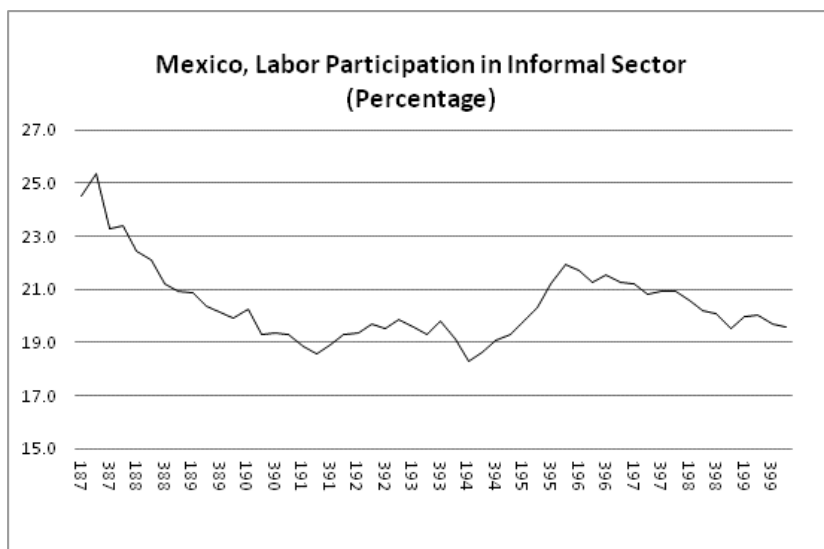
Table A4: Informal Employment by Economic Sector

(Percentage)

	1987q3	1990q3	1993q3	1996q3	1999q3
Agr	4.0	2.8	2.4	2.7	2.1
Min	0.0	0.0	0.0	0.0	0.0
Man	11.1	8.6	10.6	10.7	10.6
Cons	12.8	14.3	16.3	14.2	18.3
Util	--	--	--	--	--
Comm	25.0	25.3	20.3	22.2	20.0
Trans	4.6	1.0	0.5	0.6	0.5
F. Serv	1.6	1.0	0.9	1.2	1.1
S. Serv	41.0	47.0	49.0	48.4	47.3
Other	--	--	--	--	--

Source: Encuesta Nacional de Empleo Urbano (several years)

Figure A1: Workers in the Informal Sector



Source: own estimation using ENEU 1987:01-1999:04

Figure A2: Mexico: Evolution of Gini (16 cities)



Appendix B

Variables used to estimate the wage premium in the informal sector

The wage equation was estimated for each quarter during the period 1987:01-1999:04. The equation is for workers between ages 14 and 65. We excluded unemployed and people working but not getting a monetary payment. Income is monthly labor income. Recalling that the equation used is:

$$\begin{aligned} \ln W_i = & \alpha_1 G_i + \alpha_2 A_i + \alpha_3 A_i^2 + \sum_j^J \beta_j e_{j i} + \sum_k^K \delta_k c_{k i} + \sum_m^M \gamma_m S_{m i} + \sum_n^N \theta_n F_{n i} \\ & + \sum_p^P \vartheta_p P_{p i} + \sum_q^Q \pi_q I_{q i} + \mu_i \end{aligned}$$

Where

G_i is the gender of the i th individual. It takes value 1 if male and 0 otherwise

A_i is age of i -th individual

$e_{j i}$ is the educational level of individual i th. We define five educational categories e_0 = No Formal Education, e_1 = Primary Education, e_2 = Secondary Education, e_3 = High School and e_4 =College Education or higher

$c_{k i}$ is a set of dummy variables that controls for the city where the i th individual lives.

Until 1991:04 ENEU only included 16 cities. Between 1992:01 and 1994:02 the number of cities included in the survey increased to 32. The number of cities included in the

survey again increased in 1994:03 to 39. From then onwards the number of cities continuously increased so that by 1999:01 it was 45.

S_{mi} is a set of dummy variables that controls for economic sectors where the individual works. We define 10 economic sectors i.e., s1=Agriculture, s2=Mining, s3=Manufactures, s4=Construction, s5=Utilities (Electricity, Gas, Water), s6=Commerce and Restaurants, s7=Transportation and Communications, s8=Financial Services, s9=Social and Personal Services and s10=Other.

F_n is a set of dummy variables that controls for firms' size. We define 6 sizes: F0 when is self-employed, F1=firms that employ between 2 and 5 workers, F2 for firms that employ between 6 and 50 workers, F3 = firms that employ between 51 and 100 workers, F4 for firms that employ between 101 and 250 workers, and F5 for firms that employ more than 250 workers.

P_{pi} is a set of dummy variables that control for the type of labor contract the worker has. We define 7 different categories: P1 is fixed salary, P2 is by hour or by day, P3 by piece produced, P4 is by percentage or commission, P5 by Tips, P6 by a percentage of utilities, P7 other form of payment.

I_{qi} is a dummy variable that controls for informality. We define two types of informalities. I1 is informal boss and I2 is informal employee.

Appendix C

Cities included in the analysis

Mexico City
Guadalajara
Monterrey
Puebla

León
Torreon
San Luis Potosí
Mérida

Chihuahua
Tampico
Orizaba
Veracruz

Ciudad Juarez
Tijuana
Matamoros
Nuevo Laredo

Appendix D

Table D1: Unit Root Test (ADF)

Variable	τ	τ_{μ}	τ_t
<i>g</i>		-2.3501 (2)	
<i>wfi</i>		-1.6264 (0)	
<i>y</i>		-0.5027 (0)	-2.5372 (7)
<i>wef40</i>		-2.3999 (1)	
<i>wef31</i>		-2.6320 (5)	-0.6383 (0)
<i>wei31</i>	-0.8634 (2)		
<i>p</i>		-0.6116 (3)	-3.0848 (3)
Δg	-2.9938 (7)		
Δwfi	-7.7894 (0)		
Δy		-5.6432* (0)	
$\Delta wef40$	-8.9701* (0)		
$\Delta wef31$	-3.5135* (4)		
$\Delta wei31$	-3.5392* (4)		
Δp		-2.3022 (3)	-2.2966 (3)
$\Delta^2 p$	-8.5491* (1)		

Source: Own estimates

Notes: (1) Number of lags in parenthesis; (2) (*) Reject the Null at 1%

Table D2: Cointegration Test

Date: 11/13/12 Time: 15:53
 Sample: 1989Q1 1999Q4
 Included observations: 44
 Trend assumption: Linear deterministic trend
 Series: LG LFOR_INF FWE40 IWE40 FWE31 IWE31 LINFL
 Lags interval (in first differences): 1 to 1

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.600776	149.2377	125.6154	0.0008
At most 1 *	0.540980	108.8355	95.75366	0.0046
At most 2 *	0.448671	74.57433	69.81889	0.0198
At most 3 *	0.364235	48.37572	47.85613	0.0446
At most 4	0.278249	28.44698	29.79707	0.0709
At most 5	0.165328	14.09968	15.49471	0.0802
At most 6 *	0.130407	6.148135	3.841466	0.0132

Trace test indicates 4 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

There was also cointegration in the bi-variate, tri-variate models.