

# Income Differences on Mexico's Northern Border: A Perspective on Formal and Informal Employment

## Diferencias de ingreso en la Frontera Norte de México: Una perspectiva de empleo formal e informal

*Rogelio VARELA LLAMAS*

Universidad Autónoma de Baja California  
rvll\_uclm@hotmail.com

### ABSTRACT

Using information from the “Encuesta nacional de ocupación y empleo (ENOE)”, income differences between formal and informal workers in the northern border states of Mexico are analyzed. The ANOVA methodology suggests that the income of workers differs according to the type of employment (formal/informal), type of work, and interaction effect. The logit model shows that the likelihood of informal employment is associated with lower levels of education and the job search process.

*Keywords:* 1. informal employment, 2. ANOVA, 3. logit model, 4. regional analysis 5. northern states.

### RESUMEN

Con información de la “Encuesta nacional de ocupación y empleo (ENOE)” se analizan las diferencias de ingreso entre trabajadores con empleo formal e informal en los estados de la frontera norte de México. La metodología ANOVA sugiere que los ingresos de los trabajadores difieren de acuerdo con el tipo de empleo (formal/informal), categoría de ocupación y su efecto de interacción. El modelo logit revela que la probabilidad de tener un empleo informal se asocia a un menor nivel de escolaridad y a un proceso de búsqueda de empleo.

*Palabras clave:* 1. empleo informal, 2. ANOVA, 3. modelo logit, 4. análisis regional, 5. estados del norte.

Date of receipt: April 2, 2014.

Date of acceptance: June 24, 2014.

## *INTRODUCTION*

One of the more important discussions in the area of labor markets has to do with the income differentials between workers in the formal and informal sectors of the Mexican economy. The gaps between one segment and the other are not only pronounced on a large scale, but are also a phenomenon inherent to the different regions that make up the national economy. Some elements empirically associated with these differentials are the various levels of education attained in schools, as well as personal socioeconomic factors and issues related to labor demand such as occupational sector and the size of the workplace. In the specialized literature, a prevailing idea is that one of the structural causes explaining the increase in informal employment is the economy's loss of its potential to generate enough formal, quality jobs. Authors such as Huesca (2010) have indicated that informality persists because the economy does not have the capacity to generate enough formal jobs. In other analyses, such as that of Freije (2002), it is found that the average wages in the formal sector are higher than those of the informal sector in most Latin American countries; nevertheless, there is no consensus on the causes of the wage gap between the two sectors. One would assume, then, that the economically active population does not generally join the informal sector due to an expectation of greater income.

From the perspective of economic growth and the battle against poverty, it is thought to be essential to improve the labor productivity of strategic sectors that can detonate a greater dynamism in the labor market, also considering that active job promotion policies must be consistent and monitored in the medium and long term so that they have favorable effects on strategic indicators such as formal employment and indices of well-being.

This paper does not try to analyze the determinants of formal or informal employment in isolated form or from the optic of the quality or precariousness of employment in a context of implementing specific public policies; it also does not approach the problem from a macroeconomic or aggregate view because that is a perspective that would not help understand the phenomenon at the level of the homes and individuals of the northern border. It seeks to identify those factors and covariates that permit the understanding of the changes in income in each of the groups of workers, based on contrast statistics of equality of means and variances, but using more robust contrasts than provided by multivariate statistical analysis. The ANOVA methodology of analysis of variance in an economic post-

crisis environment is applied, taking the third quarter of 2013 as a point of reference. Factors studied include having formal or informal employment as well as the type of occupation of both groups of workers. In addition, the interaction effect of both factors is evaluated, controlling covariates that could affect the impact generated by individual factors of the target variable.

This paper is limited to studying states on the northern border of Mexico: Baja California, Sonora, Coahuila, Chihuahua, Nuevo León, and Tamaulipas. The data correlated to the strategic indicators of the labor market as well as the micro-data allowing the ANOVA to be carried out were taken from the *Encuesta nacional de ocupación y empleo* (ENOE) of the Instituto Nacional de Estadística, Geografía e Informática (INEGI). The quarterly numbers are consistent with a scenario of relative recovery of the national economy, after the outbreak of the economic crisis of 2008. The data structure is correlated to a cross-sectional estimate where the unit of observation is the head of household.

The literature that discusses the gaps in income between formal and informal workers identifies various explanatory elements; nevertheless, there is still much that can be explored and enriched from a methodological perspective because there are different approaches and tools for investigating the subject of study from various angles. Research such as that of Gallardo (2013) has studied the gross flows of workers in Mexico, examining informal employment in the context of seasonal and cyclical analysis. Other studies such as that of Rodríguez-Oreggia (2007) analyze the comparative dynamics of the informal sector in Mexico using the estimation of a multinomial logit model, and in general the work of Castro and Rodríguez (2014) offers a series of analyses and methodologies of particular interest that address, among other topics, informal work and wage inequality. The contribution of research work comes from applying a statistical methodology that permits—in a disaggregated form—the evaluation of the relevance of factors, interaction effects, and covariates in income performance. The nature of the model and the subject of investigation call for use of proxy education variables and factors associated with the segmentation of the market and type of occupation by category. A binary logit model is estimated to explain the likelihood that a head of household has an informal job. Within the framework of an empirical discussion, it is argued that the incomes of formal workers are on average higher to those registered by the grouping of informal workers, a finding consistent with that raised by Freije (2002). The analysis is centered on the monthly income variable reported by ENOE; the monetary wage

is bypassed because some informal workers may be operating on their own behalf without having a strict employee-employer relationship.

This article has four sections. The first reviews the empirical literature related to the phenomenon of income inequality. The second comes up with a statistical analysis that focuses on strategic indicators of the labor market of Mexico's northern border states. The third section develops an exploratory data analysis. The fourth section deals with a methodology of statistical contrasts as well as a bivariate logit model that complements the preliminary statistical analysis. Finally, the general conclusions resulting from the research are presented.

### *REVIEW OF EMPIRICAL LITERATURE*

The debate about income differences between workers with formal and informal employment has grown significantly in recent years as a result of the relevance it has in terms of working poverty. The discussion also has occurred because informal employment has experienced an important increase that was accentuated as a result of the 2008 crisis, which has caused a greater preoccupation for national economies and international organizations such as the International Labour Organization (ILO) and the Organisation for Economic Co-operation and Development (OECD). One feature of the various studies is that in general they focus attention on the income gaps, but with variants in terms of the spatial and temporal dimension of the data. A large number of the studies recognize two predominant approaches in the literature. The first maintains that a competitive and integrated market exists where workers are thought to make rational decisions, based on a set of incentives, to be part of the informal market. The second says that market segmentation exists where the informal worker is considered to be part of a sector that is assumed to be the residue of what the formal sector cannot generate in terms of employment (Perry *et al.*, 2007).

Numerous works exist that examine the experience of Mexico and whose findings permit a better understanding of the phenomenon of informality from various angles. For example, the work of Alcaraz Chiquiar, and Ramos-Francia (2008) estimates wage differentials in diverse sectors of the economy and posits that the differential between the formal and informal sectors exceeds the differential in industry and services. It says an increase in informal employment within total employment could have adverse effects on aggregate productivity; this has important implications. Once wage distribution functions are estimated

based on ENEU (INEGI, 2004) information, it is found that average formal wages are greater than informal ones.

On the other hand, Esquivel and Ordaz (2008) say that workers receive a wage premium in the formal sector and that workers with similar characteristics receive higher salaries when they work in the formal sector than when they work in the informal one; they also argue that the Mexican economy's labor market, rather than being integrated, is characterized by being segmented, which indicates that the implementation of social programs is not necessarily a cause of the rise in informality. This comes from implementing a methodology based on a probit model with a treatment effect using ENOE data. This contrasts with the position of Levy (2007), who says social programs generate incentives that distort the dynamics of productivity and growth and can cause an increase in informal work. It is argued that social programs are essential in a country where profound social inequality and poverty exist. Nevertheless, it was noted that the programs should be redesigned, because, while well-meant, they are poorly carried out, given that this sector provides a series of social benefits that can operate against formality.

In the same analysis track there are those who clarify that important wage differences exist between workers with formal and informal employment, noting that schooling has a positive relationship with formality, within the framework of a dynamic multinomial logit model that considers a panel structure for men and women for the years 1992 and 1999 (Gong, Van Soest and Villagómez, 2004). It is argued that informal markets work as a last resource for those formal workers who have had difficulty obtaining suitable work. The work of Sojo and Villarreal (2006) uses ENEU (INEGI, 2004) data to argue that informal employment functions as a substitute for formal work for those who have the possibility of returning to the formal sector in the future, which supposes that a dynamic work force mobility lies between both segments.

Using a counterfactual semiparametric technique, Huesca and Camberos (2009) obtained results between 1992 and 2002 that show that self-employed men with an informal job are paid better than such women, whereas women are better paid than men in the wage-earning formal sector. Using a counterfactual scenario, the authors see a better situation for both sexes in the framework of the formal sector. The work of Moreno (2007) also analyzes the differences between formal and informal sector workers in Mexico's labor market. It finds that in the 2000-2003 period, men and women with higher levels of education experience a net gain if they switch to the formal sector, but that there is a negative wage effect in the case

of those with less education, who also face barriers in switching to the formal sector given their level of labor productivity.

There are other studies such as that of Cervantes and Kumar (2013) that approach the formal and informal economy of Mexico during the 1995-2012 period and emphasize that the level of precariousness that has been reached in employment is significant, noting that it was below the well-being line established by Coneval. They add that the growth in the survival economy is important. Using panel data from 1995 to 2001 based on information from the ENOE, Alcaraz (2009) controls for observable characteristics of the workers and regional heterogeneity. The results indicate a negative effect of the unemployment rate on informal sector wages and a positive relation between wage differentials of both sectors with the unemployment rate.

Other studies such as that of Gallardo (2013) maintain that the risk of being an informal worker is greater for men and older workers, whereas the risk of inactivity is greater for young people and women. Ovando and Rodríguez (2013) emphasize that since the implementation of flexibility measures, wage inequality has increased in the Mexican manufacturing industry. A significant and positive relation of various indicators of flexibility was found with wage inequality, using a multiple regression with hierarchical design that associates the Gini coefficient with diverse measures of labor flexibility for the 32 federal organizations.

Understandably, studies focusing on analyzing wages and income of the formal and informal workers vary. They differ as a result of using different methodologies and time and spatial horizons. Standing out among the most-used tools of measurement are the use of panel data, discrete probabilistic modeling, and various approaches that resort to conventional regressions. A common characteristic of most such research about Mexico is that it uses microdata from regular home surveys.

Research in the international arena also has made contributions. For example, Guataquí, García and Rodríguez (2009) show that in the case of Colombia, the income of wage earners, in contrast with that of the self-employed, responds more to educational attainment, which suggests that different functional forms would have to be established to estimate income by sex. In the case of Turkey, Aysit and Kan (2012) conducted an interesting study in which the gap in real pay per hour in the formal and informal sector is analyzed using a methodology of panel data for the period 2006-2009. It indicates that there is labor market segmentation and that there is evidence that informal sector workers receive lower pay than those in the formal sector who have the same characteristics.

In the specific case of Uribe, Ortiz and Correa (2006), it was found that education has a negative marginal effect on informal work, indicating that it is a variable that is usually utilized due to its relevance in the labor literature. It also emphasizes that workers with more education would want to be employed in the formal sector; if the economy is in recession, they would have a greater probability of emigrating. In the case of Spain, Iriondo (2004) estimates the relative influence of various sources of wage dispersion in certain industrial sectors. Using the ANOVA technique, wage variance is broken into four factors: general education, specialized education, discrimination, and establishment. It finds results indicating that the establishment factor contributes the most in explaining remuneration, where the existence of noncompetitive factors can be inferred in the setting of wages.

#### *UNEMPLOYMENT AND INFORMAL WORK AT MEXICO'S NORTHERN BORDER*

Among the indicators that show the performance of the job market in the states are the unemployment rate and the informal work rate. The first measures the proportion of the jobless population relative to the economically active population. The second is the proportion of informal employment in the employed population. A characteristic of the six states on the northern border is that before the economic crisis of 2008 hit, they reported minimal rates of unemployment, but in distinct quarters. Examining the period from 2005 to 2013 by quarter, it was found that Baja California reported the lowest rate during 2005-2Q, which was 1.2 percent. Sonora had a rate of 2.5 percent in 2007-1Q, Coahuila 4.2 percent in 2005-4Q, Chihuahua 2.1 percent in 2005-2Q, Nuevo León 3.9 percent in 2007-4Q, and Tamaulipas 2.9 percent in 2005-4Q. A unique feature is that between 2005 and 2007, various states see rates rise above the minimal level but not increases considerable enough to indicate a strong imbalance in local employment markets, with the adverse effects that could engender in the economically active population or in the rate of critical employment conditions.

During the most acute period of the economic crisis, the states that reach their highest levels of unemployment, all in the third quarter of 2009, are Baja California, Coahuila, Chihuahua, and Nuevo León, with percentages of 7.6, 9.7, 9.7, and 8.5 respectively. Sonora saw its highest rate occur in third quarter of 2010, equivalent to 7.9 percent, and Tamaulipas' highest rate occurred in the second quarter of 2011

(8.4%). Through the third quarter of 2013, Baja California has an unemployment rate of 6.4 percent, Sonora 6.2 percent, Coahuila 5.7 percent, Chihuahua 5.3 percent, Nuevo León 5.8 percent, and Tamaulipas 7.7 percent. During this last quarter, while the level of unemployment in the six cases under study is lower than their maximum levels, it is also reasonable to assume that they continue to be above the minimal rates they were at before the crisis. The preceding reflects how the process of recovery of formal employment has not occurred quickly and consistently in Mexico's northern border. A deficit in new-job generation opportunities counteracts the necessity for work in the region, above all in the segment of young people facing the dilemma of entering the market by either taking the first job offered or continuing their job search, motivated by higher income expectations. It is imperative to face the challenge of generating higher-quality employment in order to reduce the poverty labor trend index (ITLP, or *índice de la tendencia laboral de la pobreza*) in some states such as Baja California and Nuevo León<sup>1</sup>; Coneval reports the index systematically.

Drawing attention is that Chihuahua, Coahuila, and Nuevo León, with the highest unemployment rates during the crisis, were the three states with the lowest unemployment rate in the third quarter of 2013, with 5.3 percent, 5.7 percent, and 5.8 percent respectively. ENOE numbers indicate that the state with the greatest trouble in reducing unemployment has been Tamaulipas; its highest rate was 8.4 percent in the second quarter of 2011 and it was 7.7 percent in the third quarter of 2013. That last figure was the highest among the six states, followed by Baja California with 6.4 percent and Sonora with 6.2 percent.

Baja California and Sonora distinguished themselves during the crisis by registering the lowest unemployment levels of the border states (Table 1). Nevertheless, during the third quarter of 2013 they had the second- and third-highest unemployment in the region. This might show that the recovery process has been slow because of the aftershocks of the crisis, or that the active and passive job-generation policies implemented during the downturn did not bear fruit due to a possible lack of continuity and depth or perhaps because they were not overhauled properly.

The "Encuesta nacional de ocupación y empleo" has been set up to provide important information about informal employment. An indicator that has been

<sup>1</sup> According to Coneval, the ITLP shows the trend in the percentage of people who cannot purchase the basic food basket with their work income each quarter.

TABLE 1. Unemployment Rate in Northern Border States

Period	Baja California	Sonora	Coahuila	Chihuahua	Nuevo León	Tamaulipas
2005-01	1.7	3.7	5	3.4	5.4	5.1
2005-02	1.2	3.4	4.6	2.1	4.7	4.1
2005-03	1.3	4.1	4.3	2.4	6	4.8
2005-04	1.5	3	4.1	2.2	4.5	2.9
2006-01	1.6	2.8	5.9	2.9	4.6	3.9
2006-02	1.3	2.9	5	2.7	4.7	4
2006-03	2.2	3.9	5.6	3	6	5.5
2006-04	2.1	3.5	5	2.9	4.6	4.9
2007-01	2.6	2.5	6.5	4	5	4.7
2007-02	1.7	2.9	4.9	2.6	4.5	4.1
2007-03	1.7	3.2	5.5	2.9	4.9	5.6
2007-04	2.1	2.6	4.3	3.7	3.9	3.6
2008-01	2.6	3.9	5.6	3.7	4.6	4.2
2008-02	2.8	4	4.9	4.2	4.3	4.5
2008-03	3.8	4.3	6.1	4.8	5.1	5.1
2008-04	4.4	4.4	5.7	6.9	4.6	4.9
2009-01	6.1	5.4	8.6	8.4	7.1	5.8
2009-02	5.8	5.6	7.8	7.2	7.2	6.4
2009-03	7.6	6.9	9.7	9.7	8.5	6.9
2009-04	6.7	6.4	7.4	8.5	6.8	6.6
2010-01	5.9	7.6	7.9	8.7	6.9	6.3
2010-02	5.6	6.3	8.1	6.9	6.8	6.9
2010-03	5.9	7.9	7.7	7.6	6.9	8.1
2010-04	5.6	7.6	6.9	5.7	5.9	7.8
2011-01	5.4	6.8	5.6	7.5	6.4	6.7
2011-02	5.6	5.7	6.2	6.4	6	8.4
2011-03	6.1	7.3	6.4	7.6	6.5	7.9
2011-04	7.1	6.1	5.6	6	5.5	6.7
2012-01	5.7	6.9	5.6	7	5.6	6.9
2012-02	6.3	5.2	5.5	6.6	6.3	5.9
2012-03	7	7.2	5.3	6.5	5.8	7.2
2012-04	5.9	6.2	5.6	5.8	5.4	6.1
2013-01	4.4	4.6	5.5	5.4	6.4	7.3
2013-02	5.3	5.5	5.8	5.8	5.7	6.3
2013-03	6.4	6.2	5.7	5.3	5.8	7.7

Source: "Encuesta nacional de ocupación y empleo" (Inegi).

published with regularity is the informal sector employment rate (TOSI, or *tasa de ocupación en el sector informal*) that measures the proportion of the population employed in the informal sector in relation to the total employed population. According to INEGI, the informal sector brings together a set of economic activities that take place based on household resources; they are not constituted as companies with an identifiable and independent status independent of the homes (INEGI, 2005). The ILO (2013) has raised an extended concept of informality that INEGI has adopted in its measurements. It represents a new methodology that not only considers the use of non-registered economic units but also includes a work spectrum that is not registered with the social security system, independently of whether it operates in typical informal sector economic units.

The Inegi now uses the informal sector concept but as a subgroup of a more extensive phenomenon of informality in labor relations that surrounds farming activities and paid domestic work, as well as workers who participate in the production of completely formal economic units, without there being recognition of a labor contract and the obligations that entails. Therefore, it should be pointed out that, methodologically, TOSI is part of the labor informality rate, also known as TIL (*tasa de informalidad laboral*). Also, it is seen that informality covers two dimensions, one tied to the nature of the economic unit and the other having a labor component having to do with jobs carried out without the support of the legal or institutional framework, whether the economic unit that uses their services is a non-registered business or company based out of a home or a formal company (Inegi, 2012). Although both indicators are important in analytical terms, it is important to note that the TIL is a much more effective variable in determining the amount of labor informality in the country and its regions. Samaniego and Murayama (2012) say informality is not just a sector, but a labor condition that shows up in various areas of the Mexican economy and that involves micro and small businesses as well as medium-sized and large companies. Consequently, the TIL is a better reflection of the subject under study and provides a base for better-informed economic policy actions.

A preliminary analysis of the labor informality rate (TIL) indicates that informality has gained ground in the economy following the downturn in formal employment. The extent of informality can be seen by contrasting the TOSI with the TIL. It can be said that in Mexico's six northern border states, this phenomenon represents a challenge for those developing public policy to generate high-quality jobs.

TABLE 2. Informal Employment Rate in Northern Border States

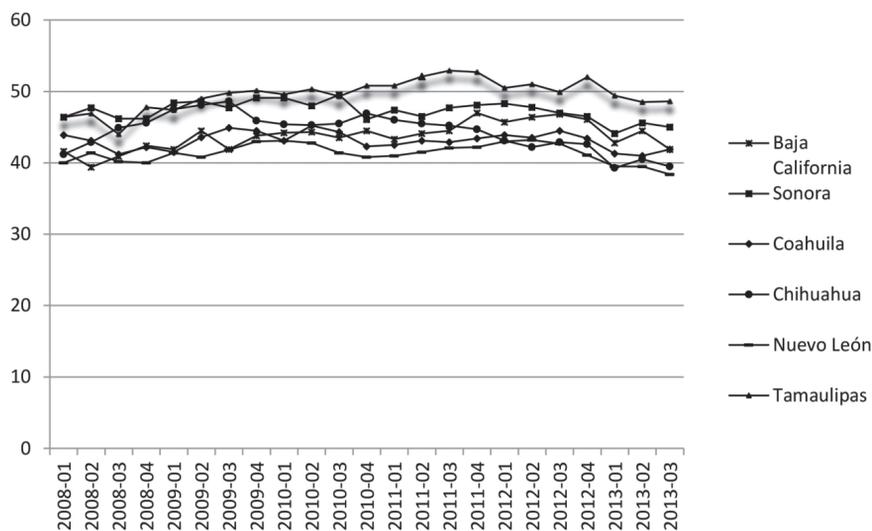
Period	Baja California	Sonora	Coahuila	Chihuahua	Nuevo León	Tamaulipas
2008-01	41.6	46.4	43.9	41.2	40	46.4
2008-02	39.4	47.7	43.1	42.9	41.4	46.9
2008-03	40.9	46.2	41.2	44.9	40.2	44
2008-04	42.4	46.2	42.2	45.6	40	47.8
2009-01	41.9	48.4	41.5	47.5	41.4	47.4
2009-02	44.5	48.6	43.6	48.1	40.8	49
2009-03	41.9	47.7	44.9	48.6	41.8	49.8
2009-04	43.8	49.1	44.5	45.9	43	50.1
2010-01	44.2	49.1	43.1	45.4	43.1	49.6
2010-02	44.3	48	45.2	45.3	42.8	50.3
2010-03	43.5	49.5	44.3	45.5	41.4	49.3
2010-04	44.5	46.1	42.3	46.9	40.8	50.8
2011-01	43.3	47.4	42.5	46	41	50.8
2011-02	44.1	46.5	43.1	45.5	41.5	52.1
2011-03	44.5	47.7	42.9	45.2	42.1	52.9
2011-04	47	48.1	43.4	44.7	42.2	52.7
2012-01	45.7	48.3	43.9	43.1	43	50.5
2012-02	46.4	47.8	43.5	42.2	43.2	51
2012-03	46.8	47	44.5	42.9	42.7	49.9
2012-04	46.1	46.5	43.4	42.6	41.1	52
2013-01	42.8	44.1	41.3	39.3	39.5	49.4
2013-02	44.5	45.6	41	40.5	39.5	48.5
2013-03	41.9	45	41.9	39.5	38.4	48.6

Source: "Encuesta nacional de ocupación y empleo" (Inegi).

It is found that Tamaulipas has the greatest rate of labor informality at 52.9 percent. Next highest are Sonora and Chihuahua, with 49.5 and 48.6 percent respectively. The states with the lowest rates are Baja California with 47 percent, Coahuila with 45.2 percent, and Nuevo León at 43.2 percent. One feature exhibited by the states, with the exception of Baja California and Tamaulipas, is that in the third quarter of 2013 they report a rate of labor informality lower than that of the first quarter of 2008, which shows an effort to deal with the problem. During the period of study, Nuevo León reports the lowest rate of informality

at 41.34 percent, on average, followed by Coahuila and Baja California with 43.09 percent and 43.73 percent respectively. On the other hand, Tamaulipas and Sonora are the states with the highest labor informality, at 49.55 percent and 47.26 percent respectively (Table 2).

In terms of the labor informality rate trend from the first quarter of 2008 through the third quarter of 2013, it can be seen that in the worst years of the economic crisis, 2008 to 2009, Tamaulipas and Sonora are the two states with the highest level of informality; this includes Chihuahua until the beginning of 2009. Beginning in 2010, it is interesting to note that Sonora and Chihuahua begin to see a drop in their rate, while the trend in Tamaulipas does not change; on the contrary, its gap with respect to the other states widens. The case of Nuevo León attracts attention because it is the state with the lowest rate of labor informality, along with Coahuila. It can be seen that by mid-2011 the six states begin to have a lower rate, which could culminate in a falling trend as a result of the process of what could be called a slow or gradual recovery, in the sense that there are no indicators of a new recession. The case of Chihuahua stands out in this scenario; the drop in its rate begins to show in the middle of 2010, whereas the reduction in informality in the other states is more delayed (Graph 1).



Source: "Encuesta nacional de ocupación y empleo" (Inegi).

GRAPH 1: Informal Employment Rate

### *EXPLORATORY DATA ANALYSIS*

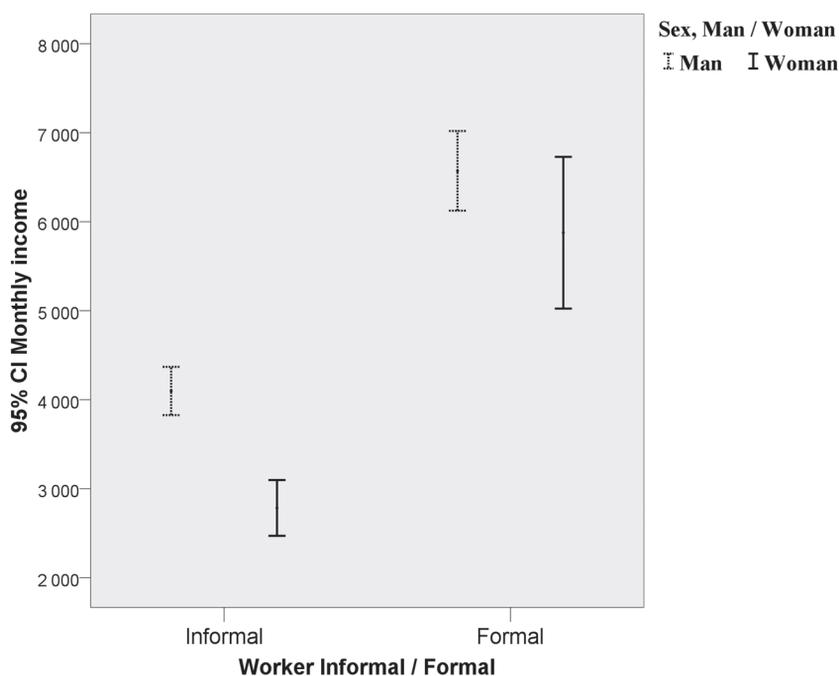
In this section, an exploratory data analysis takes place in order to examine, in preliminary form, whether differences in income exist between formal and informal workers. Once the existence of differentials is confirmed in the study variable, a variance-covariance analysis is developed to primarily examine the statistical significance of factors<sup>2</sup> and variables that could explain their variability. The variable of interest under study corresponds to the monthly income captured in sociodemographic indicators from the ENOE.

A sample of 2 703 heads of households was used, with some adjustments to limit the sample. For example, the only cases considered are those that had full interviews and that correspond to the fifth interview in a rotating panel of the survey of habitual residents or new residents between the ages of 14 and 65. Only the heads of households in the six northern Mexican states are studied. The mean income of the segment of informal workers is 3 794, while that of formal workers is 6 449. The standard deviation of the first group is 3 675 and of the second group is 8 252. The central tendency and dispersion measures suggest that both segments are characterized by the income levels they attain. As a result of working with a microdata sample, and due to its size, first an exploration of the data is made through recording instruments, considering the level of income by sex, state, and other relevant socioeconomic attributes. An error bar chart is used as a tool to help explore how the study variable in each market group behaves, taking under consideration a comparison-group design.

The error bar is a tool that takes the average income of each worker classification (formal and informal) and involves a 95 percent confidence interval, permitting the valuation of statistical relevance of the groups with different distributions. It was found that the average income of the women and the men with formal employment is far greater than the average income of men and women with informal work. When carrying out the analysis for formal and informal workers with the same level of schooling, it was corroborated the average income of formal workers exceeds that of informal ones in the case of men and women. This exercise is particularly interesting because it represents a pairwise comparison on a more disaggregated scale. The empirical regularity in these exercises is similar to that carried

<sup>2</sup> When reference to a factor is made, it is being spoken of as a numerical categorical variable with different levels of study that are not necessarily ordinal variables, but rather nominal ones.

out for the general case (Graph 2). The study levels that are analyzed are not having completed primary school, having completed primary school, having completed middle school, and having completed high school and higher education<sup>3</sup>.



Source: Own calculations based on the “Encuesta nacional de ocupación y empleo” (Inegi).

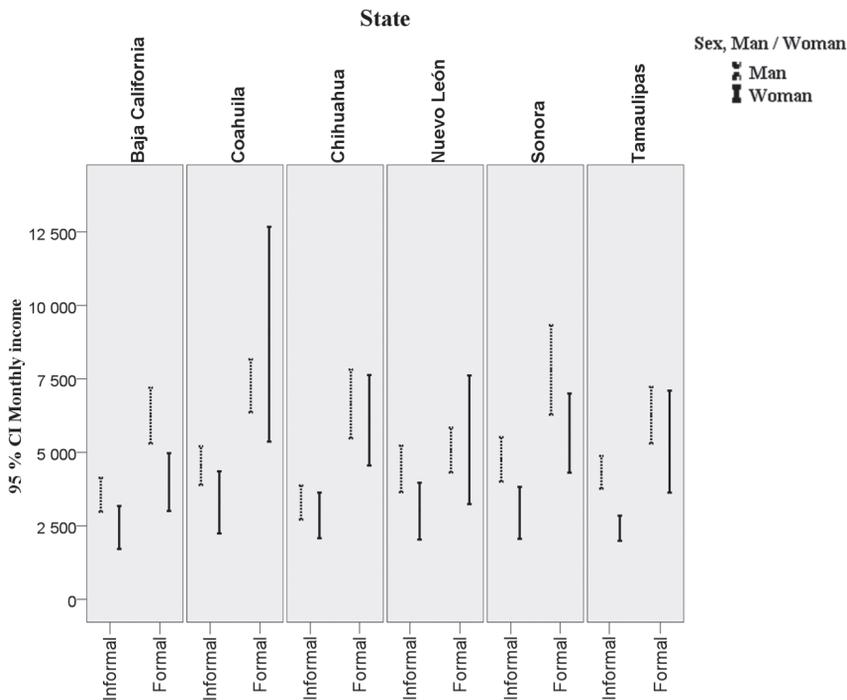
GRAPH 2. Income Level Error Bars by Sex

It should be underlined that the confidence intervals are very different in the various cases. In this sense, Pardo and Ruiz (2005) outline that when these intervals do not overlap, it can indicate that average income could be significantly different, which suggests the implementation of the analysis of variance and covariance. The exploratory analysis also identified, through box-plots, that the inter-

<sup>3</sup> The error graphs that compare formal and informal workers with the same level of schooling and that constitute a pairwise analysis are not incorporated for space reasons; nevertheless, they are at the disposition of readers for their analysis.

quartile range and its whisker do not overlap, which implies income differences by type of employment.

Graph 3 represents an extension of what was just described; it is unique in that the error bars take account of the northern Mexico border states separately and allow the data exploration to be more precise, in an extended panel. It shows that in the six states the average income of men and woman with formal employment is greater than the income of those with informal work. It is clear that the confidence intervals for the case of the men and women with formal employment are greater than those established for the average income of men and women in the informal sector. The confidence intervals in terms of the average income of women with formal employment are greater than those of men when they work in the same field.



Source: Own calculations based on the “Encuesta nacional de ocupación y empleo” (Inegi).

GRAPH 3. Error Bars of Income Level, Sex, and State

TABLE 3. Results of One-Factor ANOVA and Contrast Statistics

Total sample	Statistic	Significance
<i>F</i>	96.31	0.000
Levene	74.71	0.000
Welch	130.5	0.000
Brown-Forsythe	130.5	0.000
<i>Did not finish primary school = 252, informal = 175, formal = 77</i> <i>men = 199, women = 53</i>		
<i>F</i>	29.54	0.000
Welch	26.74	0.000
Brown-Forsythe	26.74	0.000
<i>Average informal income = 2 901.22 Average formal income = 4 773.53</i> <i>Completed primary school = 513, informal = 291, formal = 222</i> <i>men = 396, women = 117</i>		
<i>F</i>	28.87	0.000
Welch	27.09	0.000
Brown-Forsythe	27.09	0.000
<i>Average informal income = 3 341.66. Average formal income = 4 982.86</i> <i>Completed middle school = 1 035, informal = 396, formal = 639, men = 831, women = 204</i>		
<i>F</i>	21.37	0.000
Welch	22.63	0.000
Brown-Forsythe	22.63	0.000
<i>Average informal income = 4 156.47 Average formal income = 5 265.58</i> <i>High school and college=903, informal=185, formal=718, men=744, women=159</i>		
<i>F</i>	16.69	0.000
Welch	37.12	0.000
Brown-Forsythe	37.12	0.000
<i>Average informal income = 4 575.85. Average formal income = 8 136.91</i>		

Source: Own calculations based on the “Encuesta nacional de ocupación y empleo” (Inegi).

To deepen the statistical analysis, an ANOVA model with one factor was estimated to answer the primary question of the study: Is there an income difference between workers with formal and informal employment? In order to be able to find an answer to this concern and later specify an extended factor model, first a

specification restricted with one factor is considered, whose contrasts are reported in Table 3 for the general sample as well as by level of schooling.

The statistic  $F$  is obtained as a quotient of the inter-group and intra-group quadratic mean. Its level of significance leads to the rejection of the hypothesis of the equality of means where the average income of the workers of one group and another are equal; it is a contrast that is consistent with that observed in the bar charts. The Levene test statistic allows the clear rejection of the hypothesis of equality of variances and covariates of income in the two samples analyzed. Therefore, it should be taken into account that the Welch and Brown-Forsythe<sup>4</sup> statistics are an alternative to the  $F$  contrast when it is not possible to assume that the population variations are equal, as happened in this case. Because the critical level of both statistics is less than 0.05, the hypothesis of equality of means is rejected, and it is concluded that the average incomes of the formal and informal workers differ significantly. These results show the relevance of carrying out an extended analysis of variance and covariance.

### STATISTICAL METHODOLOGY AND RESULTS

This empirical work, in order to arrive at a deeper understanding, uses a univariate general linear model to evaluate the relevance of factors related to the type of work (factor A: formal and informal employment) and employment status (factor B: labor activity) in the income of the heads of households. This extended estimation makes use of covariates that could be interfering with the effect the factors have on the objective variable. The model specification is the following:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk} \quad (1)$$

<sup>4</sup> The *Brown-Forsythe* statistic for an ANOVA of a factor it is given by the following expression:

$$f = \frac{N - p}{p - 1} \frac{\sum_j^p = 1 \times n_j (\bar{z}_{.j} - \bar{z}_{..})^2}{\sum_j^p = 1 \sum_{i=1}^{n_j} (z_{ij} - \bar{z}_{.j})^2}$$

where  $p$  is the number of groups,  $n_j$  the number of observations in the group  $j$  and  $N$  is the total of observations. The term  $\bar{z}_{.j}$  corresponds to the mean of  $z_{ij}$  and  $\bar{z}_{..}$  is the total of the average of  $z_{ij}$ .

What is of primary interest is to focus attention on the statistical evidence that helps verify the hypothesis of equality of means in the groups studied. Three components related to the effects taken into account were found; the term  $\alpha_j$  relates to the first factor of the study and  $\beta_k$  represents the second factor mentioned. The third element  $\alpha\beta_{jk}$  makes reference to an interaction effect of the two factors examined and  $\varepsilon_{ijk}$  symbolizes those factors that have been omitted. The subindex  $j = 1, 2 \dots J$ , shows the levels on the first factor, while  $k = 1, 2 \dots K$ , corresponds to the categories of the second factor<sup>5</sup>.

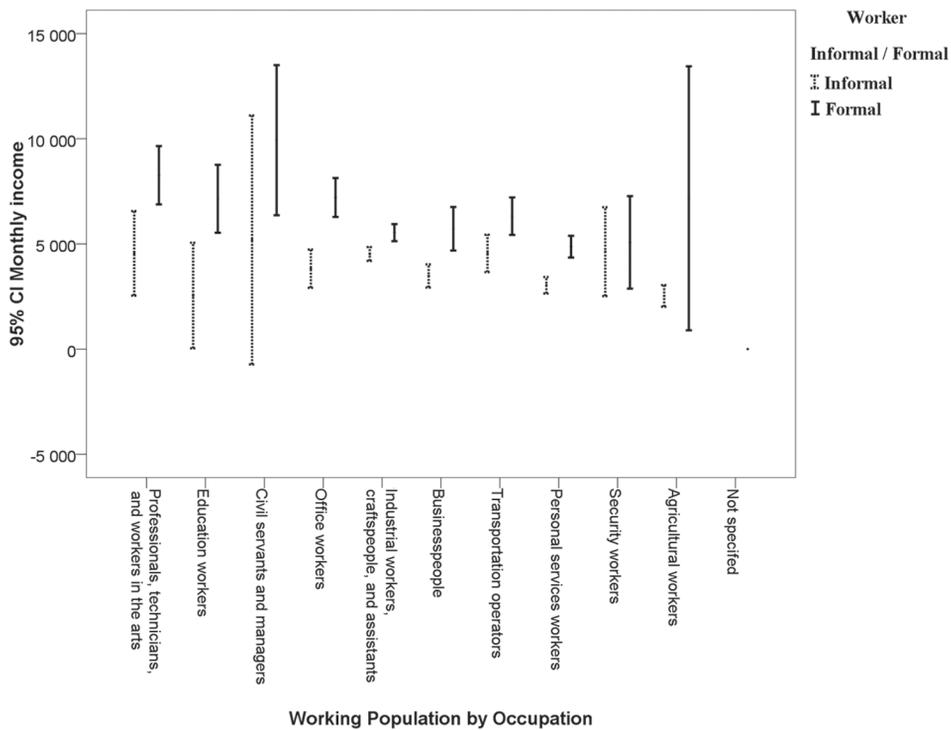
In the first instance, a brief graphical analysis based on error bars is developed. In Graph 4, it is seen that in each occupation type, there is a contrast between the average income of formal and informal workers. It is evident that the average income of formal workers exceeds that of informal ones in each type of occupation. Average income varies considerably within the groups of formal and informal workers. Therefore, it is stipulated that an average income inequality exists inter as well as intra group.

In the case of those with formal employment, it is the civil servants and managers who report a higher average income, and to a lesser extent, professional and technical wage earners and those who labor in the arts. Below them are education and office workers. All these categories also report average incomes higher than those received by informal workers in the same areas. The area reporting a smaller income within the formal employment segment is personal services workers. It is interesting to note that the categories of civil servants, managers, and agricultural workers have the widest confidence intervals. In the case of informal workers, this only applies in the case of civil servants and managers.

After considering diverse factors of socioeconomic character, it was found that factor A, factor B, and the interaction effect turned out to be significant, according to the  $F$  test for the sample set. This means that the incomes received by heads of households who live and work in northern border states are different depending on the type of work they have (formal or informal) and their occupation (Graph 4). The two covariates introduced in the analysis are potential experience and hours worked per week. In the case of work hours, it is a variable that indicates a certain degree of labor flexibility and therefore should not be seen, in a strict sense, as an indicator of preset work day hours. Age is traditionally con-

<sup>5</sup> With this methodology, the discussion can center on the results of an ANOVA with covariates, highlighting the *inter-subject contrasts*.

sidered in the empirical literature as a proxy variable of potential work experience. Data illustrating the dimension of the first covariate show that, on average, an informal worker labors 38 hours per week, as opposed to 45 for a formal worker. On the other hand, the average age of a formal worker is 42.5 and that of an informal worker 43.8.



Source: Own calculations based on the “Encuesta nacional de ocupación y empleo” (Inegi).

GRAPH 4. Error Bars of Income by Type of Occupation and Informal/Formal Work

When estimating the model (1) in disaggregated form by level of study attained, it also is found that differences in income exist pairwise at the level of both factors studied for the case of not having finished primary school, having finished primary school, and having finished middle school. In the case of high

TABLE 4. ANOVA model contrasts with covariates

Factors/Covariates	F	Significance
Hours worked per week	12.230	0.000
Worker age	4.346	0.037
Factor A [informal work/formal work]	21.340	0.000
Factor B [Working population by employment status]	2.477	0.006
Interaction effect (Factor A * Factor B)	1.901	0.048
<i>Did not complete primary school</i>		
Hours worked per week	15.710	0.000
Worker age	2.087	0.150
Factor A [informal work/formal work]	8.626	0.004
Factor B [Working population by employment status]	4.890	0.000
Interaction effect (Factor A * Factor B)	1.881	0.085
<i>Completed primary school</i>		
Hours worked per week	6.730	0.010
Worker age	0.212	0.645
Factor A [informal work/formal work]	4.220	0.040
Factor B [Working population by employment status]	2.074	0.025
Interaction effect (Factor A * Factor B)	0.173	0.984
<i>Completed middle school</i>		
Hours worked per week	19.290	0.000
Worker age	3.439	0.064
Factor A [informal work/formal work]	13.670	0.000
Factor B [Working population by employment status]	1.865	0.054
Interaction effect (Factor A * Factor B)	2.118	0.026
<i>High school and higher education</i>		
Hours worked per week	1.343	0.247
Worker age	1.097	0.295
Factor A [informal work/formal work]	13.330	0.000
Factor B [Working population by employment status]	0.981	0.454
Interaction effect (Factor A * Factor B)	0.738	0.658

Source: Own calculations based on the “Encuesta nacional de ocupación y empleo (ENOE)” (Inegi).

school and higher education, the differences are only seen for factor A (formal and informal work). As far as the covariates, it is observed that hours worked per week are statistically significant for the first three levels of instruction, and turn out to be irrelevant in terms of age. There also is the interaction effect of factors, which plays a preponderant role only at the level of having completed primary and middle schools. In these two levels, income differences depend on the type of employment and the type of work done. In any case, what was found is that income differences in general and by level of schooling depend on whether the work is informal or formal, as illustrated in Graph 4, suggesting that these inequalities operate in favor of formal employment.

In order to deepen understanding of the matter, a discrete choice model was estimated based on a function of logistic distribution. The primary objective is to analyze some elements that help explain the likelihood that a head of household has informal employment. In the previous analysis it was found that differences in income exist in both segments of the market and that income is higher for workers with formal employment in general and pairwise. This analytical framework establishes that greater income is not an incentive to participate in the informal sector, because it is a sector that normally is tied to low levels of productivity and schooling; while it can have skilled workers who have not found formal work, this evidently does not constitute a generality. In this context, other socioeconomic indicators are taken into consideration.

Among the variables involved in binary logistic regression are monthly income, level of schooling, hours worked per week, and a dichotomous variable that has the value of 1 if the worker looked for work and 0 if not. The income variable is introduced to evaluate whether it is a factor that explains the decision to become an informal worker, much as in the previous methodological discussion. Hours worked are a proxy variable of the flexibility of the job market while schooling is assumed to be a variable of human capital. The incorporation of the dummy variable associated with the search process is considered relevant because if individuals have looked for work, it could be that they are out of work and this favors their entry into the informal market on a temporary basis, or, if they are employed, this simply shows dissatisfaction with the work they have and that they are looking for a better job fit. The estimated logit model uses the following function:

$$Y_i = \frac{1}{1 + \ell^{-(\beta_0 + \beta_1 bus_i + \beta_2 ht_i + \beta_3 ae_i + \beta_4 il_i)}} \quad (2)$$

When linearizing (2), the following specification is obtained:

$$L_i = \ln\left(\frac{P_i}{1-p}\right) = z_i = \beta_0 + \beta_1 bus_i + \beta_2 ht_i + \beta_3 ae_i + \beta_4 li_i \quad (3)$$

The  $bus_i$  variable is a dummy that captures the job search process,  $ht_i$  quantifies hours worked per week,  $ae_i$  captures the years of schooling of the head of household, and  $li_i$  represents monthly income. The estimation results demonstrate that each variable is statistically significant at a standard confidence level. Standard errors in general are smaller and the  $LR^6$  statistic that follows a *chi* squared distribution shows that the set of variables is important. The goodness-of-fit statistic *Pseudo R*<sup>2</sup>, equal to 0.14, allows it to be said that there is a reasonable adjustment of the model where these types of econometric specifications are more important than the statistical significance of the variables and the sign of the coefficients (Greene, 2012).

The worker's monthly income has a negative logit correlation, that is to say, the natural logarithm of the likelihood ratio rate favors the worker having informal employment versus not having it. Nevertheless, due to the practical difficulty faced in interpreting the estimated coefficient, the marginal effect of the head of the household's income is calculated using the probability of having an informal job. It is estimated that if income rises one unit, the probability of being an informal worker diminishes 0.0018 percent, which means that, in fact, the decision to participate in the informal market is not appreciably affected by received income or the worker's expectation for himself or herself. The hours worked per week are another determinant of whether a head of household has informal employment, because it is in a certain sense a *proxy* variable of labor flexibility and affects the decision-making of the worker, while not being a determinant of the phenomenon of informality but rather of the decision to be a part of it.

It is found that when more hours are worked per week, the probability of having informal employment falls 0.498 percent. On the other hand, the years of schooling of the head of household  $ae_i$  show an inverse relation exists, because the marginal effect shows that for every additional year of study, the probability of having informal employment falls 4.18 percent, an important relationship that

<sup>6</sup> The statistic  $LR = -2\ln(\lambda) = -2\ln(LCR/LSR)$  or likelihood ratio, allows carrying out joint evaluation of the regressors,  $LCR$  and  $LSR$ , the logarithm of the likelihood function of the model with restrictions and without restrictions respectively.

also helps in compiling the profile of the average informal worker. Finally, it is found that informal employment is strongly correlated with the job-search process. This case has a qualitative variable where if a discrete change from zero to one is recorded, with the value of one showing a worker looking for work, then the probability of being an informal worker increases 31.82 percent.

TABLE 5. Results of the Logit Model Estimation

$X_0, X_{1i} \dots X_{4i}$	$\beta_0 \dots \beta_{4i}$	Robust standard error	$z$	Prob.	Odds ratio	dr/dx
$bus_i$	1.3195	0.4876	2.72	0.007	3.7416	0.3182*
$ht_i$	-0.0217	0.0024	-8.70	0.000	0.9784	-0.0049
$ae_i$	-0.1826	0.0116	-15.66	0.000	0.8331	-0.0418
$il_i$	-0.00008	0.0000	-6.45	0.000	0.9992	-0.0000
c	2.604	0.1711	15.22	0.000	13.525	
Likelihood log = -1546.07				Prob. $chi(2)$ = 0.0000		
Interaction = 4				Wald $chi(2)$ = 341.46		
Correctly classified predictions = 71.7 %				$Pseudo R^2$ = 0.1432		
Estimated average probability = 0.3549				Correlation = $ht_i, il_i$ = 0.088		
*For a discrete change of the variable dummy				Correlation = $il_i, ae_i$ = 0.227		
N = 2 703						

Source: Stata estimates based on the "Encuesta nacional de ocupación y empleo" (Inegi).

The previous results show that the income level of workers in northern border states is a significant variable from a statistical point of view; nevertheless, its marginal effect on the probability of being an informal worker is limited, which indicates that income is not necessarily an incentive to enter the informal labor market. Schooling, understood as a variable *proxy* of education representing a store of knowledge, abilities, and skills, has a greater weight because it is evident that the more schooling people have, the less likely they are to seek informal work, as formal jobs are not limited in a considerable way. It is very clear that the variables having the greatest predictive power, due to the magnitude of their marginal effect, are the dummy variable that defines the search process and having the same level of formal education.

## CONCLUSIONS

The results of the analysis of variance and covariance indicate that there are noticeable differences in income levels that different groups of workers have in Mexico's northern border region. On average, those who have formal work have higher incomes than those with informal work. In study sample set, the model shows that once potential experience and hours worked per week are controlled for, factor A: formal employment/informal employment, and factor B: working population by work status, help the income gaps be understood. The estimation also shows that income differs between formal and informal workers in each one of the labor categories studied, showing that the interaction effect between both factors is quite important. When carrying out a disaggregated analysis considering formal and informal workers with the same amount of education or the equivalent, it is found that income differences persist in each factor category. The exception is factor B, which is not significant at the educational level of high school and higher education. With respect to the estimation of the binary logistic regression, it is found that the probability of having informal employment is explained by the educational level of the head of household, but that it is more likely that those engaged in a relatively long job search have such employment. It is inferred that the search process can lead to informal work because of the serious difficulties involved in getting a job within a reasonable time in the formal sector. The costs of a long-term search could eventually bring about the decision to join the formal workforce if, on the other hand, there are limited offers of formal work. Against this background, it is necessary to differentiate that one feature is the determinants of informality in terms of their structural and short-term causes and another feature is the individual decision to be part of this sector. In this sense, the probit model offers elements that help bring about the understanding of informality from a perspective of socioeconomic indicators. One of the implications for economic policy that can be seen in relation to the reduction of informality is to implement timely educational policies to fortify labor productivity levels as well to establish incentives favoring the search for formal work. On the macroeconomic level, it is necessary to generate sustained growth and to incentivize the industrial structure so that it reaches its potential in terms of the generation of formal, quality jobs.

## REFERENCES

- ALCARAZ, Carlo, 2009, "Informal and Formal Labour Flexibility in México," *Desarrollo y Sociedad*, no. 63, 1<sup>st</sup> semester, pp. 115-143.

- ALCARAZ, Carlo; Daniel CHIQUIAR, and Manuel RAMOS-FRANCIA, 2008, "Diferenciales salariales intersectoriales y el cambio en la composición del empleo urbano de la economía mexicana en 2001-2004," México, D. F., Banco de México, July, pp. 1-39 (Documento de Investigación, no. 2008-06).
- AYSIT, Tansel, and Elif Oznur KAN, 2012, "The Formal/Informal Employment Earnings Gap: Evidence from Turkey," Ankara, Middle East Technical University/Cankaya University, May, pp. 1-45 (Munich Personal RePEc Archive, Paper, no. 38498).
- CASTRO LUGO, David, and E. Reyna RODRÍGUEZ PÉREZ, coords., 2014, *El mercado laboral frente a las transformaciones económicas en México*, México, CISE-Universidad Autónoma de Coahuila/Plaza y Valdés.
- CERVANTES, José Juan, and Arun Kumar ACHARYA, 2013, "La economía formal e informal en México 1995-2012: Implicaciones para los mercados laborales," *Gaceta Laboral*, vol. 19, no. 2, pp. 175-199.
- ESQUIVEL, Gerardo, and Juan Luis ORDAZ DÍAZ, 2008, *¿Es correcto vincular la política social a la informalidad en México? Una prueba simple de las premisas de esta hipótesis*, Mexico, D. F., CEPAL, December, pp. 1-32 (Serie Estudios y Perspectivas, no. 104).
- FREIJE, Samuel, 2002, "El empleo informal en América Latina y el Caribe: Causas, consecuencias y recomendaciones de política," n. p., Banco Interamericano de Desarrollo, pp. 1-49 (Serie Documentos de Trabajo Mercado Laboral).
- GALLARDO del Ángel, Roberto, 2013, "Gross Flows of Formal and Informal Workers in the Mexican Labor Market," *Estudios Económicos*, vol. 28, no. 2, July-December, pp. 299-324.
- GONG, Xiaodong; Arthur van SOEST, and Elizabeth VILLAGÓMEZ, 2004, "Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico," *Economic Development and Cultural Change*, University of Chicago Press, vol. 53, no. 1.
- GREENE, William H., 2012, *Econometric Analysis*, 7<sup>th</sup> ed., United Kingdom, Pearson.
- GUATAQUÍ, Juan Carlos; Andrés Felipe GARCÍA, and Mauricio RODRÍGUEZ, 2009, "Estimaciones de los determinantes de los ingresos laborales en Colombia con consideraciones diferenciales para asalariados y cuenta propia," n. p., Universidad del Rosario, August, pp. 1-22 (Documento de Trabajo, no. 70).
- HUESCA REYNOSO, Luis, 2010, "El empleo informal en la frontera norte de México y el caso de Sonora: Un análisis de expectativas de los ingresos," *Región y Sociedad*, vol. XXII, no. 49, pp. 17-43.

- HUESCA REYNOSO, Luis, and Mario CAMBEROS CASTRO, 2009, "El mercado laboral mexicano 1992-2002: Un análisis contrafactual de los cambios en la informalidad," *Economía Mexicana*, vol. XVII, no. 1, pp. 5-40.
- INSTITUTO NACIONAL DE ESTADÍSTICA, GEOGRAFÍA E INFORMÁTICA (INEGI), 2004, "Encuesta nacional de empleo urbano. Bases de microdatos", Aguascalientes, México, Inegi.
- INSTITUTO NACIONAL DE ESTADÍSTICA, GEOGRAFÍA E INFORMÁTICA (INEGI) [methodological paper], 2005, "Encuesta nacional de ocupación y empleo: Una nueva encuesta para México", Aguascalientes, México, Inegi, pp. 1-56.
- INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA (INEGI), 2012, "Boletín de prensa 449/12", Aguascalientes, México, pp. 1-4.
- INTERNATIONAL LABOUR ORGANIZATION (ILO), 2013, *Measuring Informality: A Statistical Manual on the Informal Sector and Informal Employment*, Geneva, ILO.
- IRIONDO MÚGICA, Iñaki, 2004, "Diferencias salariales en la industria manufacturera española: Estimación y factores explicativos," *ICE. Tribuna de Economía*, no. 815, May-June, pp. 261-278.
- LEVY, Santiago, 2007, "¿Pueden los programas sociales disminuir la productividad y el crecimiento económico? Una hipótesis para México," *El Trimestre Económico*, vol. LXXIV (3), no. 295, July-September, pp. 491-540.
- MORENO TREVIÑO, Jorge O., 2007, "Los salarios del sector formal e informal en México: Análisis de ganancias y pérdidas por formalización," *Ensayos. Revista de Economía*, vol. XXVI, no. 1, May, pp. 1-44.
- OVANDO ALDANA, Wendy, and Óscar M. RODRÍGUEZ PICHARDO, 2013, "Flexibilidad laboral y desigualdad salarial. La industria manufacturera mexicana como evidencia, 2005-2010," *Análisis Económico*, vol. XXVIII, no. 67, 1<sup>st</sup> quarter, pp. 59-76.
- PARDO MERINO, Antonio, and Miguel Ángel RUIZ DÍAZ, 2005, *Análisis de datos con SPSS*, Madrid, McGraw-Hill.
- PERRY, Guillermo E.; William F. MALONEY, Omar S. ARIAS, Pablo FAJNZYLBER, Andrew D. MASON, and Jaime SAAVEDRA-CHANDUVI, 2007, *Informality: Exit and Exclusion*, Washington, D.C., The World Bank (World Bank Latin American and Caribbean Studies, no. 40008).
- RODRÍGUEZ-OREGGIA, Eduardo, 2007, "La dinámica comparativa del sector informal en México," México, D. F., Instituto de Investigaciones sobre Desarrollo Sustentable y Equidad Social-Universidad Iberoamericana (Documentos de Investigación).

- SOJO, Eduardo, and Roberto VILLARREAL, 2006, "Public Policies to Promote Productive Occupation and Increase Formality among the Moderate Poor: The Mexican Agenda," Basudeb Guha-Khasnobis, and Ravi Kanbur, eds., *Informal Labour Markets and Development*, Hampshire, Palgrave Macmillan.
- SAMANIEGO, Norma, and Ciro MURAYAMA, 2012, "¿Qué tan informales somos hablando de informalidad?," *Nexos*, no. 419, November, pp. 20-23.
- URIBE, José Ignacio; Carlos Humberto ORTIZ, and Juan Byron CORREA, 2006, "¿Cómo deciden los individuos en el mercado laboral? Modelos y estimaciones para Colombia," *Lecturas de Economía*, Universidad de Antioquia, no. 64, January-June, pp. 59-89.