

Economic Disadvantage of Being Indigenous: An Analysis of Mexico's First *National Survey of Employment in Indigenous Areas*.

Yoshiaki Hisamatsu* yoshiaki.hisamatsu@nifty.com

Hiroyuki Ukeda hukeda@sta.att.ne.jp

University of Tokyo

August , 2002

Summary.- We estimate the economic disadvantage of being indigenous in Mexico. Using microdata of the 1997 National Survey of Employment in the Indigenous Areas, we show that 'to be indigenous' defined in terms of language usage, has a negative impact on income after controlling for various income-related variables such as education, work experience, geographic location and land assets. And our decomposition analysis measures the probable weight of the indigenous disadvantage in explaining income gap, compared with other income-related variables. Our finding is that the indigenous disadvantage explains less than 20% of income gap. (92 words)

Keywords - indigenous people, poverty, income, decomposition analysis, Latin America, Mexico

* Correspondence: 4-3-12, Sendagi, Bunkyo-ku, Tokyo, 113-0022, Japan.

Prepared for delivery at the 2003 meeting of the Latin American Studies Association, Dallas, Texas, March 27-29, 2003

1. INTRODUCTION

Indigenous peoples are faced with economic and social difficulties in many parts of the world.¹ A notable case is Mexico.² One of the two central themes of indigenous question in the contemporary Mexico is that more than 6 million of the indigenous population remain disproportionately poor (INEGI, 1995; INI, 2000; Panagides, 1994).³ Indigenous Mexican men have lower income per hour than non-indigenous Mexican men. In the second quarter of 1997 the average non-indigenous males in Mexico earned 9.14 pesos an hour, while indigenous males earned 4.74 pesos an hour.⁴

Several possible reasons for the indigenous people's lower income flash on the mind. Among them are schooling and work experience, geographic location, occupation, land assets, language difficulties, and discrimination and other cultural factors. For example, as shown in table 1, indigenous people have much lower attainments in formal education, on average, than the non-indigenous population, and income has a tendency to rise with education. Despite internal and international migration, many indigenous population still live in rural areas, where price levels are relatively low. Being likely to suffer from fluency in the dominant language, Spanish, indigenous Mexicans are prone to be at disadvantage in the labor market. In addition, there is a widespread belief and claim that indigenous Mexicans are the victim of discriminatory practices, and, if they are discriminated in workplace, they cannot obtain the income that their human capital would command.

How much of the income gap described above can be explained by these factors, *inter alia*, by factors solely attributable to being indigenous? This paper tries to provide answers to this question, which is of critical importance also in many other Latin American countries such as Guatemala and Ecuador.

As far as we know, only two efforts have been made to analyze the relative incomes of indigenous and non-indigenous Mexicans in the framework of quantitative method with large sample. One is the research project organized by the World Bank (Psacharopoulos & Patrinos (Eds.), 1994) on poverty of indigenous people in four Latin American countries

(Mexico, Guatemala, Peru and Bolivia). In this volume, Panagides (1994) investigated why Mexican indigenous men were poorer than non-indigenous fellow countrymen, using the 1989 Household Survey data. He estimated earnings equation including indigenous variable, and applied decomposition method similar to Oaxaca (1973). The other is a comprehensive study about rural economy of de Janvry, Gordillo & Sadoulet (1997). Though de Janvry et al. did not focus exclusively on indigenous economic disadvantage, they estimated Probit equation with ethnic variable (a dummy variable for those living in indigenous communities) in a chapter on the causes of income poverty in rural areas of Mexico. They also reported that being indigenous had a significantly negative impact on income after controlling other relevant variables.

In the present study we basically follow the Panagides' research strategy, but with several modifications. Certainly it was a pioneering study to measure the economic disadvantage of indigenous people in Mexico, but we believe it has four problems we could improve. First, the indigenous variable he chose is each percent of indigenous population within a municipality, which is an indirect variable rather than a direct indigenous variable based on individuals.⁵ This is unsatisfactory because considerable number of non-indigenous persons who lived in the municipality are tagged as indigenous. As a result, it has possibility of underestimation. Using more recent data from the 1997 National Survey of Employment in Indigenous Areas (Encuesta Nacional de Empleo en Zonas Indígenas:1997: ENEZI), we take a direct individual-based indicator of indigenous people on usage of language.⁶ Apart from being more exact, ENEZI data allow us to break up indigenous people into bilingual persons and monolingual persons, and to proceed to check the economic effect of speaking Spanish among the same indigenous-language speakers. Second, other previous studies on rural Mexico (e.g., de Janvry, Gordillo & Sadoulet, 1997) highlights the economic importance of land assets in rural area. While Pangides didn't take it into account, we explicitly consider the land variable in the equation. Third, his Household Survey's sample includes both urban and rural households of the entire Mexico, so that his estimation could have higher chance of being exposed to omitted variable bias than the one which only

covers rural households. Since the majority of indigenous people live in rural areas, his indigenous variable might have in fact represented rural effect. As a result, it has a possibility of overestimation. ENEZI sample only deals with households in those rural areas where the share of indigenous inhabitants are high. Therefore, it enables us to avoid the omitted variable bias better than the nation-wide household survey does. Fourth, still remaining omitted variable bias could result from local price level and local labor market conditions which could differ by region. We adopt dummies for resident state to account for the geographical difference.⁷ Panagides didn't take this factor in consideration, either.

Negative sign attached to the indigenous variable in income regression even after controlling for income-related variables could be interpreted as disadvantage of being indigenous. And theoretically it could further disaggregated in two ways. On the one hand, it could be interpreted as insufficient capacity in the form of lack of adequate skill of Spanish given the same educational attainment.⁸ This could be due to both bilingual difficulties of the indigenous students, and/or to the poor arrangement of formal education in the indigenous areas.⁹ On the other hand, it could be interpreted as other non-capacity-related reasons. Discrimination could also affect the coefficient of the indigenous variable.¹⁰ Several scholars have identified cultural difference toward income (e.g., risk-aversion) and/or geographical isolation in remote areas (historically brought about in the process of concentration of fertile land in the hands of Spanish and mestizo landlords) as the causes of the indigenous low income.¹¹ Both capacity-related and non-capacity-related effects could have jointly affected the coefficient of the indigenous variable. Since we don't have the measure to disaggregate further (e.g., fluency in Spanish by way of test score), we cannot separate out these effects in detail. Having said that, we believe that our exercise highlights quantitatively the economic disadvantage of being indigenous.

The rest of the paper is organized as follows. The following section describes the data and the specification of income equation, and explains the method we apply to decompose the earning differential into the portions due to differences in average characteristics and the portions due to differences attributable to unobserved factors and

indigenous disadvantage. The third section presents and interprets the results. The last section concludes.

2. METHOD, DATA, AND VARIABLES

We adopt two types of estimation methods in this study in order to measure an economic disadvantage of being indigenous. First, we estimate the following income equation by ordinary least squares:

$$\ln E_j = \mathbf{b}X_j + cI_j + \mathbf{e}_j \quad (1)$$

where X_j is a vector of observed characteristics, I_j is indigenous variable (0: non-indigenous, 1: indigenous) and β a vector of coefficients.¹² Our indigenous variable is based on the property to speak some indigenous language, the least-ambiguous and the most-frequently-used operational definition of indigenous people in Mexico.¹³ We could measure the economic disadvantage of being indigenous by the sign and magnitude of its coefficient. To examine if speaking Spanish, other variables being equal, has an impact on the percentage change of the indigenous income, we also estimate equation with two indigenous dummies, bilingual and monolingual.

Alternatively, we use a method similar to that developed by Oaxaca (1973). We estimate the following income equation for indigenous Mexicans and non-indigenous Mexicans by ordinary least squares:

$$\ln E_j = \mathbf{b}X_j + \mathbf{e}_j \quad (2)$$

where X_i is a vector of observed characteristics, and β is a vector of coefficients. Having estimated the income function, we can separate the observed income differential between

indigenous and non-indigenous Mexicans into components due to group characteristics and to parameters. The income differential between indigenous Mexicans and non-indigenous Mexicans can be expressed as:

$$\overline{\ln E_N} - \overline{\ln E_I} = (\bar{X}_N - \bar{X}_I) \{D\hat{\mathbf{b}}_N + (I - D)\hat{\mathbf{b}}_I\} + \{\bar{X}_N(I - D) + \bar{X}_I\}(\hat{\mathbf{b}}_N - \hat{\mathbf{b}}_I) \quad (3)$$

where I represents the indigenous group. The left-hand side of (3) represents the difference between the mean log income of indigenous Mexicans and non-indigenous Mexicans. The first term on the right-hand side represents the difference in income that is attributable to differences in person's characteristics; the second term represents the disadvantage part of the income differential and a residual.

The decomposition provides an unbiased estimate of economic disadvantage of being indigenous only if the error terms of the equations follow a standard normal distribution. If significant variables are omitted from the equations, or variables are measured with error, the coefficients in which the estimate of the economic disadvantage is grounded will be biased. Our results, for example, could not be named as the effect of discrimination because the education variables do not adequately control for the effects of the Spanish skill.

The vector "D" with elements valued 0 or 1, determines the benchmark income structure to observed income. Since we don't know the benchmark, we check the two cases that the benchmark structure is the one for the non-indigenous (D=1) and the one for the indigenous (D=0). Indigenous people in Mexico are not a homogenous group. Certainly, various sub-groups with different historical backgrounds and cultures are distinguished within indigenous people. But, in this study, following previous studies, we simply assume as a first approximation that all the indigenous peoples have common economic disadvantage against non-indigenous people.

The data to which we apply the two above-mentioned methods is microdata from the 1997 National Survey of Employment in Indigenous Areas (ENEZI). ENEZI was carried on by the INEGI (Instituto Nacional Económica, Geográfica, e Informática) during the second

quarter of 1997 over ten regions with significant indigenous presence (Table 2), for complementing the deficient informational base about the economic conditions of indigenous people in Mexico.¹⁴ The income samples we selected consist of male aged 14 and above with positive working hours in the previous week who were both self-employed and not self-employed.¹⁵ Industry and employment give rise to different income in the rural area.¹⁶ Therefore, we control both for industry (agriculture and non-agriculture) and form of employment (occupation) (employer, self-employed, and employee).

We have two types of land variable.¹⁷ On the one hand, we have an indicator of land holding area (both privately-owned and ejido parcel). Ejido is a form of community land holding, in which a member person has a claim to use and rent. On the other hand, we have an indicator of land using area (irrespective of property type). As de Janvry, Gordillo & Sadoulet (1997) described, the land is an important asset of rural households.

The Mexican indigenous people with at least 62 distinct languages are scattered in the country. Table 2 shows the sample variation by region and state. The difference in residence alone could contribute to inter-group differences in income due to the local labor market and price level differences. It is therefore important to adjust income for residence indicators when analyzing indigenous vs. non-indigenous differentials. We simply use the dummy variable of resident states to account for the geographical impact.

The natural logarithm of the income reported is regressed on a set of explanatory variables: year of education, potential work experience (age minus year of education minus five), experience squared, employment (employer, self-employed, or employee), industry agriculture or non-agriculture), residence (state dummy), employment formality (affiliation with social security programs for private sector and public sector, IMSS and ISSSTE, respectively) and land variable (land holding or land use).¹⁸ Table 6 provides the definitions of these variables. In the case of employment and industry, we combined them to make 6 categories.¹⁹ Observations with missing data for any of the variables described were dropped from the analysis. Table 1 shows the calculated mean and standard deviation of these variables for total sample, non-indigenous sub-sample, indigenous sub-sample. There

are 7161 indigenous-language-speaking Mexican males and 1050 non-indigenous-language-speaking Mexican males.²⁰ Indigenous sample has lower income and lower education (about half in schooling years), and it is more experienced, working more in agriculture (especially as self-employed), and less formal than non-indigenous sample. Simple mean of land variables in Table 1 is misleading because many non-indigenous don't have land. Table 3 shows more correct picture. Among indigenous sample, 4517 males out of 7161 (63%) hold private and/or ejido land, while only 215 males out of 1050 (20%) hold land among non-indigenous sample. The same pattern could be noted on land use. Total land held is larger than total land used, which means some land is non-used (in fallow or abandoned) or used for non-agricultural purposes.

3. RESULTS

Table 4 shows the result of equation (1). There are four results because we use two types of land variables and two types of indigenous variables. As for the land variable, the equations (E1) and (E2) use land-holding variable, while the equations (E3) and (E4) use land-use variable. As for the indigenous variables, (E1) and (E3) don't separately count monolingual indigenous persons and bilingual indigenous persons, while (E2) and (E4) count them separately. In other words, we have two indigenous variables: monolingual indigenous (M-IND) and bilingual indigenous (B-IND). We test whether the difference of the two indigenous variables is significantly different from zero (see B-IND – M-IND row).

In all the regressions, all the signs fit with the usual hypothesis made at estimating functional relationships between income and human capital.²¹ Except a few coefficients of state dummies, all the coefficients are statistically significant at 1% level. Concerning industry and employment, non-agricultural employer (selected as base) contributes income most. Among the rest of five industry-cum-employment dummies, agricultural self-employment category affects most negatively income, followed by agricultural employer

category and agricultural employee category. This result appears reasonable given the difficulties and restraints of agricultural activities in indigenous-concentrated areas. Formality of employment, IMSS and ISSSTE dummies, has a positive impact, and remarkably so in the case of public employment (ISSSTE). A fact to be noted here is that the 72.9% of 369 ISSSTE beneficiaries works in public education sector.

Turning to our central issue, viz., to measure the impact of being indigenous, signs of indigenous variables are all significantly negative. In both (E2) and (E4), the equality between monolingual coefficient and bilingual coefficient cannot be rejected. This finding suggests that speaking Spanish does amount to economic benefit among the indigenous people only through influencing other income-enhancing variables, for instance, through facilitating schooling and/or raising the chance of obtaining non-agricultural jobs.

The last result indicates that we could use only two sub-samples (indigenous males and non-indigenous males) in decomposition analysis without considering the difference between the monolingual indigenous males and the bilingual indigenous males. Coefficients derived for indigenous vs. non-indigenous decomposition analysis are on table A1. We used land use data for land variable.

Table 5 shows the result of our decomposition analysis.²² The “unexplained” share is 12.4 to 16.3 percent.²³ In another words, 83.7 to 87.6 percent of the income gap is explained by various income-related variables. A previous study reported 34 to 48 % for the “unexplained” share (Panagides, 1994). Our “unexplained” shares are much smaller than the ones reported in Panagides (1994) probably because its rural factor exceeds indigenous specification factor. Even in our estimation, however, there remains an “unexplained” portion in the income gap, which indicates being indigenous is accompanied with economic disadvantage. The data we drew on, ENEZI, does not admit for identification of what factors constitute the residual in decomposition analysis.

4. CONCLUSION

In poverty alleviation of indigenous people, it is necessary to grasp its unique feature. In the last section we have observed that being indigenous presents an economic disadvantage in Mexico, and estimated how much of the indigenous vs. non-indigenous income differential results from it. We calculate how much of the income difference is explained by various measured characteristics other than indigenous variable. We find that the residual which could be called indigenous economic disadvantage constitutes an important factor in lowering the income of indigenous Mexicans. Owing to the data constraint of employment survey, we cannot specify that indigenous disadvantage, which may be the consequence of discrimination, inadequate Spanish skill, geographical isolation and/or some cultural factors. But, it should be emphasized that differences in measured characteristics are the overwhelming (more than 80 percent) reason for the shortfall in income. One caveat would be that discrimination outside the labor market may affect the acquisition of human capital (e.g., education) or selection of occupation by members of indigenous group and lead to differences in observed characteristics.

What kind of implication could be taken out from our findings? One has to be very cautious about interpreting parameter estimates from a linear regression model. It is because the potential impacts of changing a group's characteristics in reality are complex. Many variables do not change independently of one another and they do interact in determining income. It could be argued, however, that the fact that measured characteristics explain more than 80 percent of income difference strongly suggest that, in the Mexican case, the indigenous policies which enhance conventional characteristics (e.g., labor, education, and land) might do have larger impact on narrowing income gap than indigenous policies which seek to focus on indigenous particularities.

Table 1. Means and Standard Deviations of Variables: Males in Income Sample

Variables ^a	Total Sample	Indigenous-Language-Speaking Mexicans	Non-Indigenous-Language -Speaking Mexicans
Observations	8211	7161	1050
INCOME (E)	5.31 (24.6)	4.74 (25.60)	9.14 (16.03)
LNINCOME (lnE)	6.033 (1.172)	5.929 (1.149)	6.746 (1.079)
ED	4.114 (4.016)	3.602 (3.642)	7.604 (4.656)
TC	.0188 (.3912)	.0142 (.1185)	.0495 (.2171)
EXP	30.39 (17.05)	31.75 (16.93)	21.16 (14.90)
EXPSQ	1215 (1222)	1295 (1245)	669 (870)
LNHW	3.767 (.3873)	3.769 (.3771)	3.753 (.4511)
AEMPLOER	.0281 (.1654)	.0285 (.1664)	.0257 (.1584)
ASELF	.5249 (.4994)	.5731 (.4947)	.1962 (.3973)
AEMPLOEE	.1056 (.3073)	.1091 (.3117)	.0819 (.2744)
NASELF	.1025 (.3034)	.0952 (.2936)	.1524 (.3596)
NAEMPLOEE	.2213 (.4151)	.1808 (.3849)	.4971 (.5002)
IMSS	.0427 (.2023)	.0296 (.1695)	.1324 (.3391)
ISSSTE	.0449 (.2072)	.0369 (.1884)	.1000 (.3001)
LANDHOLD	266.9 (1157)	269.8 (599.2)	247.2 (2833)
LANDUSE	185.4 (364.0)	195.3 (334.8)	118.3 (516.7)
YUC	.1159 (.3202)	.0927 (.2901)	.2743 (.4464)
VER	.0656 (.2477)	.0704 (.2558)	.0333 (.1796)
SLP	.0258 (.1586)	.0274 (.1632)	.0152 (.1226)
OAX	.3205 (.4667)	.3282 (.4696)	.2686 (.4434)
NAY	.0385 (.1924)	.0399 (.1958)	.0286 (.1667)
JAL	.0212 (.1440)	.0242 (.1536)	.0001 (.0309)
HID	.0512 (.2203)	.0437 (.2045)	.1019 (.3027)
GUE	.0939 (.2917)	.1014 (.3019)	.0429 (.2026)
DUR	.0118 (.1081)	.0120 (.1089)	.0105 (.1019)
CHI	.0265 (.1608)	.0244 (.1544)	.0410 (.1983)
CHIA	.0938 (.2915)	.0979 (.2972)	.0657 (.2479)
CAM	.0122 (.1097)	.0120 (.1089)	.0133 (.1148)
PUE	.1230 (.3285)	.1258 (.3317)	.1038 (.3052)
IND-LAN	.8721 (.3340)	1 (0)	0 (0)
M-IND	.0901 (.2864)	.1033 (.3044)	0 (0)
B-IND	.7820 (.4129)	.8967 (.3044)	0 (0)

^a Variables are defined in table 6. Numbers in parenthesis are standard deviations.
Source: Authors' calculation based on INEGI, 1998.

Table 2. Geographical Distribution of Interviewed Indigenous Households

Region (Classification by INI)	State	Households
Península	Yucatán and Campeche	1,196
Istmo	Oaxaca	1,083
Papaloapan	Oaxaca	916
Mixteca	Oaxaca	967
Tarahumara	Chihuahua	874
Huicot	Durango, Jalisco and Nayarit	809
Sierra Norte de Puebla	Puebla	996
Huasteca	Hidalgo, Veracruz, San Luis Potosí	1,238
Náhuatl-Tlapaneco-Mixteco-Amuzgo de Guerrero	Guerrero	1,018
Chiapas	Chiapas	823
Total		9,920

Source. INEGI, 1998.

Table 3. Access to Land

	Total Sample		Non-Indigenous-Speaking Males		Indigenous-Language -Speaking Males	
	No. of Persons	Mean of Land Area (100m ²)	No. of Persons	Mean of Land Area (100m ²)	No. of Persons	Mean of Land Area (100m ²)
Land Holding						
Private and Ejido	4732	463	215	1207	4517	428
Private	1295	374	69	1225	1226	327
Ejido	3460	493	146	1199	3314	462
Land Use	5165	295	264	471	4901	285

Note: Land use is for all property types, which include private holding, ejido, and borrowed land.

Source. Authors' calculation based on INEGI, 1998.

Table 4. Male Income Equations; DEPENDENT VARIABLE: LNINCOME

	(E1)	(E2)	(E3)	(E4)
CONSTANT	4.946 a (.143)	4.947 a (.143)	5.011 a (.141)	5.011 a (.141)
ED	.061 a (.004)	.060 a (.004)	.058 a (.004)	.058 a (.004)
TC	.235 a (.076)	.235 a (.076)	.195 a (.076)	.195 a (.076)
EXP	.026 a (.002)	.026 a (.002)	.023 a (.002)	.023 a (.002)
EXPSQ/1000	-.285 a (.031)	-.283 a (.031)	-.261 a (.031)	-.261 a (.031)
LNHW	.388 a (.026)	.388 a (.026)	.385 a (.026)	.385 a (.026)
AEMPLOER	-1.014 a (.096)	-1.015 a (.096)	-1.218 a (.096)	-1.218 a (.096)
ASELF	-1.731 a (.077)	-1.730 a (.077)	-1.809 a (.076)	-1.809 a (.076)
AEMPLOEE	-.837 a (.083)	-.837 a (.083)	-.832 a (.082)	-.832 a (.082)
NASELF	-.735 a (.081)	-.735 a (.081)	-.731 a (.079)	-.731 a (.080)
NAEMPLOEE	-.558 a (.080)	-.558 a (.080)	-.549 a (.079)	-.549 a (.079)
IMSS	.323 a (.055)	.323 a (.055)	.327 a (.054)	.327 a (.054)
ISSSTE	.773 a (.062)	.775 a (.062)	.800 a (.061)	.800 a (.061)
LANDHOLD/10000	.636 a (.087)	.635 a (.087)		
LANDUSE/1000			.490 a (.029)	.490 a (.029)
YUC	.130 a (.046)	.131 a (.043)	.121 a (.042)	.121 a (.042)
VER	.350 a (.048)	.350 a (.048)	.304 a (.048)	.304 a (.048)
SLP	.603 a (.068)	.602 a (.068)	.599 a (.067)	.599 a (.067)
OAX	.223 a (.034)	.223 a (.034)	.185 a (.034)	.185 a (.034)
NAY	.302 a (.058)	.301 a (.058)	.284 a (.058)	.284 a (.058)
JAL	.477 a (.074)	.475 a (.074)	.425 a (.073)	.425 a (.073)
HID	.190 a (.053)	.191 a (.053)	.186 a (.052)	.186 a (.052)
GUE	.091 b (.044)	.092 b (.044)	.075 c (.043)	.075 c (.043)
DUR	.874 a (.096)	.873 a (.096)	.840 a (.095)	.840 a (.095)
CHI	.808 a (.068)	.807 a (.068)	.748 a (.067)	.748 a (.067)
CHIA	.211 a (.044)	.212 a (.044)	.172 a (.043)	.172 a (.043)
CAM	.234 a (.094)	.235 a (.094)	.207 b (.093)	.207 b (.093)
IND-LAN	-.139 a (.032)		-.136 a (.032)	
M-IND		-.154 a (.048)		-.136 a (.047)
B-IND		-.139 a (.032)		-.136 a (.032)
N	8211	8211	8211	8211
Adjusted R2	.423	.423	.439	.439
B-IND – M-IND		.0151 (.0365)		-.0003 (.0360)

Standard error in parenthesis. a: significantly different from 0 at 1% level b: significantly different from 0 at 5% level c: significantly different from 0 at 10% level In the case of state dummies, we adopt Puebla state variable as base. In the case of industry-cum-employment variable, non-agricultural employer is base.

Table 5. Decomposition Analysis

Description	D = 1	D = 1	D = 0	D = 0
	Absolute Gap	Percentage Gap	Absolute Gap	Percentage Gap
Log Income Gap	0.8174	100%	0.8174	100%
Explained Gap	0.7164	87.6%	0.6843	83.7%
Unexplained Gap	0.1010	12.4%	0.1331	16.3%

Source: Authors' calculation.

Table 6. Definitions of Variables

INCOME	= monthly income in the second quarter in 1997 in pesos / (4 * weekly hours worked).
LNINCOME (ln E)	= natural logarithm of monthly income.
ED	= schooling years completed.
TC	= 1 if received technical training; = 0 otherwise.
EXP	= age - schooling years completed - 5.
EXPSQ/1000	= square of EXP / 1000.
LNHW	= natural logarithm of hours worked.
AEMPLOER	= 1 if agricultural employer; = 0 otherwise.
ASELF	= 1 if agricultural self-employed; = 0 otherwise.
AEMPLOEE	= 1 if agricultural employee; = 0 otherwise.
NASELF	= 1 if non-agricultural self-employed; = 0 otherwise.
NAEMPLOEE	= 1 if non-agricultural employee; = 0 otherwise.
IMSS	= 1 if receives benefit from IMSS; = 0 otherwise.
ISSSTE	= 1 if receives benefit from ISSSTE; = 0 otherwise.
LANDHOLD/10000	= land holding of private and/or ejido land (100 square meters/10000).
LANDUSE/1000	= land utilized by the interviewed irrespective of property type (100 square meters/1000).
YUC	= 1 if residence is in Yucatán state; = 0 otherwise.
VER	= 1 if residence is in Veracruz state; = 0 otherwise.
SLP	= 1 if residence is in San Lúis Potosí state; = 0 otherwise.
OAX	= 1 if residence is in Oaxaca state; = 0 otherwise.
NAY	= 1 if residence is in Nayarit state; = 0 otherwise.
JAL	= 1 if residence is in Jalisco state; = 0 otherwise.
HID	= 1 if residence is in Hidalgo state; = 0 otherwise.
GUE	= 1 if residence is in Guerrero state; = 0 otherwise.
DUR	= 1 if residence is in Durango state; = 0 otherwise.
CHI	= 1 if residence is in Chihuahua state; = 0 otherwise.
CHIA	= 1 if residence is in Chiapas state; = 0 otherwise.
CAM	= 1 if residence is in Campeche state; = 0 otherwise.
PUE	= 1 if residence is in Puebla state; = 0 otherwise.
IND-LAN	= 1 if speaks indigenous language; = 0 otherwise.
M-IND	= 1 if speaks indigenous language but does not speak Spanish; = 0 otherwise.
B-IND	= 1 if speaks both indigenous language and Spanish; = 0 otherwise.

Source, INEGI, 1998.

Table A1. Income Equations; DEPENDENT VARIABLE: LNINCOME

	Indigenous-Language-Speaking Males	Non-Indigenous-Language-Speaking Males
CONSTANT	5.070 a (.1594)	4.434 a (.2518)
ED	.0496 a (.0043)	.0804 a (.0070)
TC	.2593 a (.0956)	.0405 (.1039)
EXP	.0186 a (.0027)	.0330 a (.0053)
EXPSQ/1000	-.2164 a (.0335)	-.3505 a (.0865)
LNHW	.3671 a (.0290)	.4483 a (.0517)
AEMPLOER	-1.315 a (.1137)	-.6736 a (.1792)
ASELF	-1.841 a (.0941)	-1.797 a (.1199)
AEMPLOEE	-.8390 a (.0988)	-.9416 a (.1372)
NASELF	-.8322 a (.0983)	-.3675 a (.1184)
NAEMPLOEE	-.5292 a (.0977)	-.6036 a (.1153)
IMSS	.3283 a (.0680)	.3639 a (.0759)
ISSSTE	.8257 a (.0730)	.7356 a (.0969)
LANDUSE/1000	.6177 a (.0350)	.1298 a (.0462)
YUC	.1530 a (.0475)	.0507 (.0882)
VER	.3109 a (.0506)	.0239 (.1384)
SLP	.6137 a (.0714)	.4569 b (.1904)
OAX	.1608 a (.0365)	.3442 a (.0832)
NAY	.2695 a (.0618)	.3121 b (.1491)
JAL	.3972 a (.0755)	.9286 (.7110)
HID	.1945 a (.0594)	.1312 (.1013)
GUE	.0452 (.0460)	.2398 c (.1278)
DUR	.8007 a (.1029)	.9311 a (.2255)
CHI	.6244 a (.0755)	1.170 a (.1309)
CHIA	.1552 a (.0463)	.1661 (.1128)
CAM	.2180 b (.1021)	.2105 (.2060)
N	7161	1050
Adjusted R2	.3952	.5728

Standard error in parenthesis. a: significantly different from 0 at 1% level b: significantly different from 0 at 5% level c: significantly different from 0 at 10% level In the case of state dummies, we adopt Puebla state variable as base. In the case of industry-cum-employment variable, non-agricultural employer is base.

NOTES

1. The ILO defines indigenous people as follows in the Article 1-1-(b). of its 169 Convention, 1989:

“people in independent countries who are regarded as indigenous on account of their descent from the populations which inhabited the country, or a geographical region to which the country belongs, at the time of conquest or colonization or the establishment of present state boundaries and who, irrespective of their legal status, retain some or all of their own social, economic, cultural and political institutions”.

Mexico ratified the ILO 169 Convention in September 1990.

2. According to the Population Census, which has been almost the unique statistics of national coverage that offers indigenous information, in the year 2000 there are 6,044,547 speakers of indigenous languages, who constitutes 7.1% of the total population of Mexico 5 years and older (INEGI, 2001). The 81.5% of them are bilingual, i.e., speak both Spanish and their native languages.

3. The other is the social and political movements for indigenous autonomy, which have gained impetus after the neo-zapatists' rebellion in 1994 (INI, 2000).

4. Authors' calculations from the 1997 National Survey of Employment in Indigenous Areas, as reported in table 1.

5. The indigenous variable used in Panagides' analysis is the person's residence in municipalities where more than 30% of total population is indigenous. Relying on Household Survey, Panagides could not help but employ this indirect method with the help of 1990 Population Census because the Household Survey does not ask interviewed individuals whether they speak indigenous language or not.

6. de Janvry et al. (1997) resorted to institutional classification of ejido households: ejidos with a majority of mestizo population, ejidos with a majority of indigenous population, and indigenous communities. Since their

ethnicity variable is indigenous communities, they do not capture indigenous households inhabited in ejidos.

7. Mexico is a federal country and consists of thirty-two states.

8. Chiswick et al. (2000) takes this line of argument in the case of Bolivia.

9. Bertely (1998) and Ornelas (1995) noted relative and absolute deficiency in education in the indigenous-concentrated areas.

10. Panagides (1994) suggests the existence of discrimination as a probable cause of the unexplained residual in his decomposition analysis.

11. See Aguirre-Beltrán (1991) and INI (2000). According to the 1990 Population Census, 29.3% of indigenous population lived in small localities (localidades) with less than 500 inhabitants (INEGI, 1993).

12. As is well known, the reasoning of this type of estimation dates back to Becker and Chiswick (1966) and Mincer (1974).

13. For example, population census uses this definition.

14. The following sampling procedure was adopted in the ENEZI: The National Institute for Indigenous People (Instituto Nacional Indigenista: INI) -a decentralized governmental organization established in 1948- had identified 20 regions as indigenous-concentrated area in Mexico, and the INEGI chose from these 20 regions the most indigenous-concentrated 10 ones according to the Population Survey 1995 (Conteo de Población y Vivienda 1995). Of each of them, localities with their indigenous residents share 30% or higher were selected, whose inhabitants summed to 3,709,579. In the final stage, 9,920 households were selected as the sample in proportion with the population size of each locality (Of them, 9,726 (98.04%) households accepted the survey). Table 2 shows INI's classification, states and interviewed households, of the ten regions.

15. In the case of agricultural households, income includes self-consumption (its estimated monetary value) and earnings. Regarding non-agricultural households, only earnings are reported as income. As for earnings, only the earnings from the principal job are counted.

16. Warman (2001) pointed out the occupational factor (indigenous concentration in the primary sector) and land factor (less access of the indigenous to fertile land still after the agrarian reform carried out in the 20th century).

17. We don't take land quality into account.

18. IMSS and ISSSTE are abbreviations of Instituto Mexicano de Seguro Social (Mexican Institute of Social Security), and Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (Institute of Social Security and Services of State Workers), respectively.

19. In the regression, non-agricultural employer category is used as base.

20. Among 7161 indigenous-language speaking Mexican males, there are 6421 bilingual indigenous males and 740 monolingual indigenous males.

21. The coefficients of education (ED) are statistically significant and positive, but lower than other estimates on Mexico. For example, Urciaga-García (2002) reports that the estimated coefficient is 0.136 using 1998 National Household Survey. It could be partly because the latter includes urban sample. But, as Chiswick (1997) argues, our education coefficient is partial in the sense that it does not take account of labor market success which is reflected in employment and formality variables. Thus, our estimate has downward bias as the measure of education return.

22. As Oaxaca & Ransom (1999) show, the decomposition analysis produces arbitrary results when attempting

to estimate the separate contributions of sets of dummy variables to the unexplained portion of the decomposition. We, therefore, report only the overall decomposition.

23. The result of decomposition analysis with land holding variable indicates that the residual explains 12.7 to 14.7%.

REFERENCES

- Aguirre-Beltrán, G. (1991). *Regiones de refugio: El desarrollo de la comunidad y el proceso dominical en Mestizoamérica*. México: Fondo de Cultura Económica.
- Becker, G., & Chiswick, B. (1966). Education and the Distribution of Earnings. *American Economic Review*, 56, 358-69.
- Bertely, M. (1998). Educación indígena del siglo XX en México. in Latapí, P. (Ed.). 74-110.
- Chiswick, B. (1997). Interpreting the Coefficients of Schooling in the Human Capital Earnings Function. World Bank Working Paper 1790.
- Chiswick, B., Patrinos, H., & Hurst, M. (2000). Indigenous Language Skills and the Labor Market in a Developing Economy: Bolivia. *Economic Development and Cultural Change*, 48(2), 349-367.
- de Janvry, A., Gordillo, G., & Sadoulet, E. (1997). *Mexico's Second Agrarian Reform: Household and Community Responses*. San Diego: Center for U.S. – Mexican Studies, University of California, San Diego.
- INEGI (Instituto Nacional Económica, Geográfica, e Informática) (1993). *Hablantes de Lengua Indígena: XI Censo General de Población y Vivienda 1990*. México.
- INEGI (1995). *La población indígena mexicana, Tomo III*. México.
- INEGI (1998). *Encuesta nacional de empleo en zonas indígenas 1997*. México.
- INEGI (2001). *XII Censo General de Población y Vivienda 2000*. México.
- INI (Instituto Nacional Indigenista) (2000). *Estado del desarrollo económico y social de los pueblos indígenas de México, 1996-1997*. México.
- Latapí, P. (Ed.) (1998). *Un siglo de educación en México, Tomo II*. México: Fondo de Cultura Económica.
- Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press.
- Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 9, 693-709.
- Oaxaca, R., & Ransom, M. (1999). Identification in Detailed Wage Decompositions. *The Review of Economics and Statistics*, 81(1), 154-157.
- Ornelas, C. (1995). *El sistema educativo mexicano: La transición de fin de siglo*. México: Fondo de Cultura Económica.

Panagides, A. (1994). Mexico. in Psacharopoulos & Patrinos. (Eds.). 127-163.

Psacharopoulos, G, & Patrinos, H. (Eds). (1994). *Indigenous People and Poverty in Latin America: An Empirical Analysis*. Washington: World Bank.

Urciaga-García, J. (2002). Los Rendimientos Privados de la Escolaridad Formal en México. *Comercio Exterior*, 52(4), 324-330.

Warman, A. (2001). *El campo mexicano en el siglo XX*. México: Fondo de Cultura Económica.