Estimating the Effects of Formality on Mexican Informal Microfirms
A Joint Multivariate Approach

Sonia Di Giannatale, Gibrán Ramírez-Abarca and Ricardo Smith*
Fecha de recepción: 2 de febrero de 2012; fecha de aceptación: 20 de mayo de 2013.

Abstract: In this article we use a multivariate framework to estimate net profits and levels of capital and labor for informal Mexican microfirms had they been formal. We estimate a Roy model to simultaneously model three different microfirm responses conditional on the sector (formal or informal) choice. Responses include profits and levels of labor and capital to be used in each sector. Our results indicate that the formal counterfactuals of currently informal microfirms employ higher levels of labor and physical capital; yet, they do not earn higher profits.
Keywords: microfirm behavior, informal sector, Roy model, developing economies.

Estimación de los efectos de la formalidad sobre microempresas informales mexicanas: Un enfoque multivariado conjunto

Resumen: En este artículo usamos un análisis multivariado para estimar los beneficios netos y los niveles de capital y trabajo de microempresas informales mexicanas de haber sido éstas formales. Estimamos un modelo de Roy para modelar simultáneamente las respuestas de las microempresas condicionales a la selección del sector de operación, formal e informal. Las variables de respuesta incluyen los beneficios netos y los niveles de capital y trabajo usados en cada sector. Nuestros resultados indican que los contrafactuales formales de microempresas informales usan mayores niveles de capital y trabajo, pero no obtienen mayores beneficios netos.
Palabras clave: comportamiento de microempresas, sector informal, modelo de Roy, economías en desarrollo.

JEL *Classification:* C3, D2, O12, O17.

*Sonia Di Giannatale, sonia.digiannatale@cide.edu, associate professor of Economics, Department of Economics, CIDE. Gibrán Ramírez-Abarca, gibran_ramirez@hacienda.gob.mx, director, Department of Public Policies for Productivity, Secretaría de Hacienda y Crédito Público. Ricardo Smith, rsmithramirez@gmail.com, adjunct professor of Economics, Department of Economics, CIDE. Mexico City. The authors have benefited from comments and suggestions by Robert Duval and participants at the CIDE Economics Seminars. The authors thank three anonymous referees for enriching comments. The authors also thank Rodolfo Cermeño, Itza Curiel-Cabral and Alejandro López-Feldman for technical support. Remaining errors are the authors' responsibility.

Introduction

According to Fajnzylber *et al.* (2009), microfirm owners and their workers accounted for approximately 50 per cent of the Latin American labor force during the last decade. In all Latin American countries, and especially in Mexico,[1] a significant proportion of these microfirms operates in the informal sector of the economy; that is, they neither pay taxes nor fulfill their legal obligations toward their workers. Yet, they make use of many public goods, such as the public health system, the sidewalks, streets and parks where they conduct business, and, in many cases, do not pay for the electricity or water they use as production inputs. However, the quality of those public goods might be compromised because of the resulting loss of tax revenues. The discussion over the determinants of informality is still ongoing. A country's regulatory burden and its capacity to enforce the law have both been identified as determinants of informality, but there is empirical evidence that suggests that as the latter grows stronger the former becomes less important (Dabla-Norris *et al.*, 2008).

At the individual microfirm level, the owner's choice of operating in the formal or the informal sector of the economy should be part of a profit maximization decision, as Levenson and Maloney (1998) discuss. Under this perspective, Fajnzylber *et al.* (2009) analyze whether formalization has an impact on the performance of Mexican microfirms and which treatment (credit supply, entrepreneurial training, and access to government services) is more effective on business performance. They find that formalization has potential positive effects in profits and the survival rates of microfirms.

In this article we concur with the idea that formality is an input of the productive process, where its marginal value increases as the size of the firm increases. For larger firms, meeting institutional requirements becomes more important because a larger operation is more likely to draw the attention of governmental agencies. For some firms, however, either because of specific characteristics of their type of activity and capital restrictions, or their owner's entrepreneurial ability, the optimal decision is to remain in the informal sector and to maintain a small scale of production. For instance, Dabla-Norris *et al.* (2008) empirically establish a positive association between small firms and informality. On the other hand, Cunningham and Maloney (2001) show that, since the informal sector in

[1] According to Levy (2008), 58 per cent of the Mexican labor force and 60 per cent of the employed population worked in the informal sector in 2006.

Mexico is highly heterogeneous, a reasonable explanation for microfirm owners' decision to remain in the informal sector can be found in differences in their entrepreneurial abilities and heterogeneous preferences, and not in distortions in the labor or credit markets, as it is usually assumed.

In this paper, we focus on the decision-making process of informal Mexican microfirms trying to answer two questions: *a)* are, *ceteris paribus*, formal firms more profitable than informal ones?, and *b)* is, as suggested by the model of Rauch (1991), the informal sector the optimal choice for small firms and the formal sector the preferred choice for larger firms? To answer these questions, we perform quantitative comparisons on three measures of firm size: net profit, investment in physical capital, and number of workers. Our study explicitly assumes the endogeneity of the firms' decisions. That is, we assume that the decision of whether or not to participate in formal institutions is part of the entrepreneur's profit maximization decision problem (Levenson and Maloney, 1998). Accordingly, assuming profit-maximizing owners, a microfirm operates in the formal sector only if the net profits (monetary or not) obtained from operating in this sector exceed those obtained in the informal sector.

We perform a joint simultaneous analysis of a vector of interrelated microfirm responses resulting from the sector choice by using a Roy model with multiple response variables (Heckman and Vytlacil, 2007). We estimate all the parameters in the model using full information maximum likelihood by means of a Montecarlo EM algorithm. This methodology allows us to control for the endogeneity of the responses, account for the correlation among the different responses, and, therefore, obtain more efficient estimates of the treatment effect. Additionally, we include a heteroskedastic error covariance matrix in the estimation. This multivariate approach distinguishes our analysis from other related empirical papers that analize microfirm responses separately using either standard selection models or matching methods. Differently from most of the related literature on treatment effects, in which the effect of treatment (formality in this case) on the treated (formal) sub-population is the focus of the estimation, we estimate the effect that treatment would have on the untreated (informal) firms. We study informal microfirms for three reasons. First, previous studies have focused either on the formal sub-population or on the whole population. The proportion of the informal sector in the economy has increased steadily in the last decade in Mexico (Duval and Orraca, 2011). This fact suggests that in Mexico there are incentives for operating

in the informal sector, contradicting the results of Fajnzylber *et al.* (2009), who, by using both matching and standard control function techniques, found positive effects of formality on the profits and survival rates of formal Mexican microfirms. McKenzie and Sakho (2010), on the other hand, by using instrumental variables and matching techniques on Bolivian data, found that effects of formality on profits depend on firm size. In particular, they found that, for small firms, lower profits are associated with registering for taxes; while middle-size firms might benefit from formalization. Second, Latin American governments have shown an increasing interest in reducing the size of the informal sector by providing incentives to informal microfirms to move into formality; therefore, it is interesting to explore how well informal firms, as they are, would perform under the rules of the formal sector. Our third reason is methodological. Although we use data from more recent waves of the survey used by Fajnzylber *et al*. (2009), the data still have the same problem that those authors encountered, which is the difficulty to satisfy the common support assumption for the treated sub-population; an issue that is less problematic for the untreated sub-population. In simple words, for each informal entrepreneur it is possible to find a similar (in observable) formal counterpart; the opposite, however, is not always true as there are types of formal entrepreneurs (*e.g.* very highly trained or educated ones) that cannot be found in the informal sector.

Our empirical results indicate that the main determinants of sector choice are the microfirm owner's level of formal education and access to credit, and the microfirm's type of economic activity. Even though the nature of our estimation is static, our results show that the right incentives for a microfirm's growth are only found in the formal sector of the economy, and that this growth occurs if the entrepreneur's education level and ability to obtain outside funds are adequate, as well as if the microfirm operates in the sectors of commerce and services. Moreover, when comparing the behavior of currently informal microfirms and their expected behavior were they to operate in the formal sector, we find that these microfirms would obtain an annual net profit in the formal sector that is lower than what they currently obtain in the informal sector, even after adjusting their capital and labor levels. Hence, those entrepreneurs would optimally choose to stay in the informal sector.

This article is organized as follows: in section I we describe the data we use and present some descriptive statistics. In section II, we present the econometric model and the methodology we use to quantify the im-

pact of formality on net profits and the levels of physical capital and labor. In section III, we discuss our results. Finally, in the last section, we offer our concluding remarks.

I. Data

We employ data from the Mexican Microfirm National Survey (Encuesta Nacional de Micronegocios, Enamin) for the years 1998 and 2002. This survey is part of an additional questionnaire that comes with the National Survey of Urban Employment (Encuesta Nacional de Empleo Urbano, ENEU), where people who say they are owners of a firm with ten or fewer employees must answer the Enamin questionnaire, which gathers specific information about the characteristics of the microfirm.

The numbers of registered questionnaires in Enamin-1998 and Enamin-2002 were 10 738 and 11 106, respectively. We use data from both surveys in our estimations and, because our tests of the existence of structural change between 1998 and 2002 were statistically rejected, in this article we present and discuss only the results of the joint sample, which comprises a total of 21 844 microfirms.

Many informal microfirm owners, however, use the informal market as a source of subsistence while seeking a wage-earning job. Mexican informal firms are heterogeneous in the sense that a fraction of their owners are true entrepreneurs, while some owners are trying to earn a living while searching for a job with a paid salary and benefits, (Bruhn, 2012). This means that many microfirms have a very short life span, a situation that is not representative of microfirms owned by individuals who have actively decided to be their own bosses. In order to avoid this bias, in the final estimation we use only microfirms whose owners are 20 years old or more, who invest more than US $100 in capital, and who have been operating for at least 4 years. We decided to distinguish the true entrepreneurs from the others by imposing the above requirements, that are based on the owners' behavior, as opposed to directly using the answers to questions, included in the questionnaire, that explicitly ask the microfirm owners about their motivations to open their businesses and their plans for the future. Our intention is to avoid the possibility that while answering those questions, the owners could be misreporting their real motivations and business plans.

Because we use individual dummies to control for heterogeneity among different municipalities, in order to correctly identify municipal effects, we consider only those municipalities with 15 observations or more.

Table 1. Descriptive statistics

Variable	*Average*	*Std. Dev.*	*Median*	*Minimum*	*Maximum*
Sub-sample informal sector					
Monthly net profit*	3742	4785	2614	16	91579
Capital stock*	31095	48415	10769	1098	434434
Labor	1.32	0.71	1.00	1.00	7.00
Gender	0.26	0.44	0.00	0.00	1.00
Marital status	0.71	0.46	1.00	0.00	1.00
Age	43.05	11.50	43.00	20.00	70.00
Primary	0.42	0.49	0.00	0.00	1.00
EHS	0.29	0.45	0.00	0.00	1.00
High school	0.12	0.32	0.00	0.00	1.00
University	0.08	0.28	0.00	0.00	1.00
Time (years)	10.84	8.32	8.00	3.00	50.00
Formal credit	0.03	0.16	0.00	0.00	1.00
Informal credit	0.16	0.37	0.00	0.00	1.00
Manufacture	0.19	0.39	0.00	0.00	1.00
Commerce	0.26	0.44	0.00	0.00	1.00
Services	0.46	0.50	0.00	0.00	1.00
Sub-sample formal sector					
Monthly net profit*	6990	9670	4579	6	176384
Capital stock*	90435	92115	57506	1098	441282
Labor	1.87	1.14	1.00	1.00	7.00
Gender	0.26	0.44	0.00	0.00	1.00
Marital status	0.78	0.41	1.00	0.00	1.00
Age	44.45	10.77	44.00	20.00	70.00
Primary	0.32	0.47	0.00	0.00	1.00
EHS	0.22	0.42	0.00	0.00	1.00
High school	0.13	0.34	0.00	0.00	1.00
University	0.26	0.44	0.00	0.00	1.00
Time (years)	11.01	8.12	8.00	3.00	50.00

Table 1. Descriptive statistics (Cont.)

Variable	*Average*	*Std. Dev.*	*Median*	*Minimum*	*Maximum*
Formal credit	0.06	0.24	0.00	0.00	1.00
Informal credit	0.19	0.39	0.00	0.00	1.00
Manufacture	0.11	0.31	0.00	0.00	1.00
Commerce	0.37	0.48	0.00	0.00	1.00
Services	0.46	0.50	0.00	0.00	1.00

Source: Authors' own elaboration using data from Enamin 1998 and 2002. *Notes:* *Monthly net profit and capital stock are expressed in 2002 Mexican pesos. 1US$=10.2MX$ in 2002. EHS: Elementary High School. Sample size: N = 8800. Observations informal sector: 4315 4414 (weighed). Observations formal sector: 4485 4386 (weighed).

All these restrictions reduce the final sample to 8 800 observations. Finally, we assume a microfirm to be formal if it is registered with the Ministry of Finance and Public Credit (Secretaría de Hacienda y Crédito Público, SHCP). According to table 1, 49.8 per cent (weighted) of the microfirms in our sample are formal.

About 26 per cent of all microfirms in the sample are run by women. The microfirm owners' average age is 44 years, and around 70 per cent of them are married. It is observed that most of the microfirm owners have only finished primary or elementary high school education, 71 per cent in the sub-sample of informal microfirm owners, and 54 per cent in the sub-sample of formal microfirm owners. Only 8 per cent of informal microfirm owners have completed a college education, while 26 per cent of formal microfirm owners have done so.

Regarding the characteristics of the microfirms, the average time of operation of the microfirms in the sample is 11.9 years. On the other hand, the average monthly net profit of informal microfirms is 3 742 Mexican pesos with a standard deviation of 4 785 Mexican pesos, these monetary amounts (and all monetary amounts from now on) are expressed in 2002 Mexican pesos and the average exchange rate in 2002 was 1 US dollar per 10.2 Mexican pesos. Also, the average monthly net profit of formal microfirms is 6 990 Mexican pesos with a standard deviation of 9 670 Mexican pesos. The average value of the capital stock of informal microfirms, expressed in 2002 Mexican pesos, is 31 095 Mexican pesos with a standard deviation of 48 415 Mexican pesos, while that of formal microfirms is 90 435 Mexican pesos with a standard deviation of 92 115 Mexican pesos. The average number of workers in informal microfirms, including the owner, is 1.32 while that of formal microfirms is 1.87. No microfirm in the sample employs more than 7 workers.

Finally, 3 per cent of the informal microfirms in the sample used a formal credit source to start the business, while 15 per cent of them used an informal credit source. Interestingly, 6 per cent of the formal microfirms in the sample used a formal credit source to start the business, while 19 per cent of them used an informal credit source. The rest of the microfirms, formal and informal, financed their businesses by using personal savings or benefits obtained from the owner's previous job.

II. Econometric Model

To motivate our empirical approach let us consider N profit-maximizing microfirms that use capital (K) and labor (L) as inputs. Each microfirm i ($i = 1, \ldots, N$) must choose whether to operate in the formal sector ($s = 1$) or the informal sector ($s = 0$). Simultaneously, and conditional on the chosen sector, the microfirm must determine the optimal amounts of capital and labor, as well as the optimal production level (Y). Thus, the problem microfirm i solves is:

$$(\tilde{S}_i, \tilde{Y}_i, \tilde{K}_i, \tilde{L}_i) = argmax \{s_i \cdot \pi^1(Y_i, K_i, L_i) + (1 - s_i) \cdot \pi^0(Y_i, K_i, L_i)\}$$

$$s.t. \quad \begin{array}{l} s_i = 0, 1 \\ Y_i \leq g(K_i, L_i; a_i) \\ K_i \geq 0 \\ L_i \geq 1 \end{array}$$

where, $g(K_i, L_i)$ is a production function that satisfies $g_1 > 0, g_2 > 0, g_{11} < 0$, $g_{22} < 0$, and a_i is an indicator of the owner's productivity depending on his/her characteristics such as education and entrepreneurial capacity. The profit functions π^s are given by $\pi^s, (Y_i, K_i, L_i) = p^s Y_i - r^s K_i - w^s L_i - C^s (b_i, Y_i, K_i, L_i)$, where p^s, r^s and w^s are the prices of the product, capital, and labor, respectively, in sector s. The functions C^s represent other operating costs in sector s. These costs depend on the characteristics and regulatory environment, b_i, of the municipality in which the microfirm is located. The microfirms that operate in the formal sector must pay municipal legal fees, taxes (some of which are proportional to the firm size, Y_i and K_i), and the social security payments of its workers in proportion to the amount of labor employed, L_i. The components of the costs C^0 for microfirms in the informal sector include fines for failure to pay permits, taxes, and/or social-security payments to its workers. The probability of being fined will

depend on local characteristics, b_i, such as the efficiency of local governments in curbing the informal economy, and also on the firm size, Y_i, K_i, L_i, (larger firms are easier to detect). An informal microfirm faces the risks of having its capital and merchandise confiscated and having to pay bribes to inspectors to avoid legal conflicts with the local authority. Thus, there is an optimal response triplet $(\tilde{Y}_i, \tilde{K}_i, \tilde{L}_i)$ for each sector, in which production level and optimal input levels will depend on characteristics of the microfirm and owner, a_i, and those of the market, b_i, p^s, r^s, w^s.

A profit-maximizing microfirm will operate in the formal sector if $\pi^1 - \pi^0 - T_i > 0$, where, T_i denotes the transaction costs of moving from one sector to the other. Because the regimes to which a microfirm may belong (informal or formal) are mutually exclusive, our econometric model must allow us to construct a counterfactual against which we can compare each response variable (net profits or firm size), and, thus, to quantify the expected impact of moving from one regime to the other. The most frequently used methods for constructing a counterfactual are the matching methods (Rosenbaum and Rubin, 1983) and the Roy model (Roy, 1951; Heckman and Vytlacil, 2007). Instead of imposing the assumption of conditional independence required by matching methods, here we use a generalized Roy model with a dichotomous selection equation and multiple responses. Econometrically, the microfirm's problem is equivalent to the following reduced model:

$$
\begin{aligned}
&S_i^* = X_{1i}\beta_1 + \varepsilon_{1i} && \text{selection equation}\\
&\left.\begin{aligned}
\pi_i^0 &= X_i\beta_2^0 + \varepsilon_{2i}^0\\
K_i^0 &= X_i\beta_3^0 + \varepsilon_{3i}^0\\
L_i^0 &= X_i\beta_4^0 + \varepsilon_{4i}^0
\end{aligned}\right\} && \text{response equations, informal sector } S_i^0\\
&\left.\begin{aligned}
\pi_i^1 &= X_i\beta_2^1 + \varepsilon_{2i}^1\\
K_i^1 &= X_i\beta_3^1 + \varepsilon_{3i}^1\\
L_i^1 &= X_i\beta_4^1 + \varepsilon_{4i}^1
\end{aligned}\right\} && \text{response equations, formal sector } S_i^1
\end{aligned}
\qquad (1)
$$

In our analysis, S_i^* is a latent variable representing the net benefit of moving from the informal to the formal sector of the economy. The observable counterpart of S_i^* is the dichotomous optimal response S_i^0, which takes the value 0 if S_i^* is nonpositive and it takes the value 1 otherwise. The response variable π_i^0 is the net profit of the microfirm i when operating in

sector S_i. We decided to use this variable instead of the production level because we are interested in comparing the level of net profits between the two sectors (Note that π_i^0 is a function of $\tilde{Y}_i^s, \tilde{K}_i^s$ and $\tilde{L}_i^s$). The other variables that we use to construct the vector of dependent variables are the investment in physical capital, $\tilde{K}_i$, and the number of workers (including the owner)[2], $\tilde{L}_i$, respectively. All these variables are used in their logarithmic form. The observable variable for number of workers is censored at 1 since no microfirm can operate with less than one worker (in its logarithmic form, the censorship occurs at zero). Thus, the equation system in (1) includes three latent variables, which are denoted by the symbol (*).

The vectors X_i contain exogenous variables associated with characteristics of the microfirm, its owner and the local market where it operates. The vector X_{1i} is also exogenous and contains the determinants of the choice of sector of operation, but some of the components of X_{1i} do not influence the optimal response vector $(\tilde{\pi}_i^s, \tilde{K}_i^s \tilde{L}_i^s)$.

In order to control for differences in behavior among microfirms operating in different sectors, the model allows that $\beta_j^0 \neq \beta_j^1 (j = 2, 3, 4)$. The difference between the slope vectors can be evaluated statistically. The disturbances ε_{1i} and ε_{ji}^s are assumed to be heteroskedastic and normally distributed with zero average and covariance matrices $\Omega_{s,i}$. In order to control for nonobserved heterogeneity, we assume multiplicative heteroskedasticity. Thus, the elements in $\Omega_{s,i}$ are

$$\begin{aligned} \operatorname{var}[\varepsilon_{1i}] &= \sigma_{11} \exp(H_i \delta_1) = \exp(H_i \delta_1) \\ \operatorname{cov}[\varepsilon_{1i}, \varepsilon_{ji}^s] &= \sigma_{1j}^s \exp\left(\frac{H_i \delta_1 + H_i \delta_j^s}{2}\right) \\ \operatorname{var}[\varepsilon_{ji}^s] &= \sigma_{jj}^s \exp(H_i \delta_j^s) \\ \operatorname{cov}[\varepsilon_{ui}, \varepsilon_{vi}^s] &= \sigma_{uv}^s \exp\left(\frac{H_i \delta_u^s + H_i \delta_v^s}{2}\right) u,v = 2, 3, 4 \end{aligned} \tag{2}$$

[2] A better variable to represent labor would have been total wages paid by the microfirm. Many microfirms, however, use household labor (e.g. the owner's spouse and children comprise all or part of the microfirm labor), which receives neither paid market wages nor monetary payment at all. Therefore, we were unable to construct that variable. However, splitting the labor variable in two components: household and non-household labor, our model can be extended to partially tackle this issue. This implies replacing the labor equation in each regime of the empirical model by two new equations modelling the two types of labor. Such a model would allow us to study the relationship between the two types of labor and determine, for instance, whether it is true that informal microfirms make more intesive use of household labor than formal ones.

where $\sigma_{11} = 1$, according to the usual reparametrization in models with dichotomous variables. All the parameters in the model are estimated simultaneously using full information maximum likelihood by means of a Monte Carlo EM algorithm (Wei and Tanner, 1990). This algorithm allows us to efficiently handle the presence of multiple latent variables without resorting to the calculation of multiple integrals (see, for instance, Natarajan *et al.*, 2000).

II.1. The Effect of Formality on Microfirm net Profits and Size

The expected "effect" of operating in the formal sector, conditional on currently operating informally, i.e. the treatment effect on the untreated (TEUT), is estimated according to:

$$\begin{aligned} \text{TEUT}_i &= E[y_{ji}^1 \mid s_i = 0] - E[y_{ji}^0 \mid s_i = 0] \qquad j = 2, 3, 4 \\ &= x_i\,(\beta_j^1 - \beta_j^0) \\ &\quad - \left\{ \sigma_{1j}^s \exp\left(\frac{H_i \delta_j^1}{2}\right) \right. \\ &\quad \left. - \sigma_{1j}^0 \exp\left(\frac{H_i \delta_j^0}{2}\right) \right\} \exp\left(\frac{H_i \delta_1}{2}\right) \frac{\Phi\left(\dfrac{x_i \beta_1 s}{\exp\left(\dfrac{H_i \delta_1}{2}\right)}\right)}{\Phi\left(-\dfrac{x_i \beta_1 s}{\exp\left(\dfrac{H_i \delta_1}{2}\right)}\right)} \end{aligned} \tag{3}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative function and the density function of the standard normal distribution, respectively. Variables y_{ji}^s ($j = 2, 3, 4$) correspond to $\tilde{\pi}_i^s$, $\tilde{K}_i^s$, and $\tilde{L}_i^s$, respectively. The sample average of this "effect" for each of the three response variables was obtained by calculating $TEUT_i$ for each observation in the informal subsample and then calculating the corresponding sample average. The standard error for the average effect was estimated by means of the delta method (Greene, 2007). According to Roy's hypothesis that comparative advantages drive microfirm behavior, we expect the average estimated value of $TEUT_i$ for the microfirm net benefit to be statistically non-positive.

II.2. Specification of the Model

Next, we discuss the regressors (in small capitals) used in our estimates. We consider socioeconomic characteristics like GENDER and SCHOOLING as determinants of the sector of operation. The variable GENDER is dichotomic, and the motivation for including it as a determinant of the sector of operation is the evidence that in Mexico there exists an important proportion of women who decide to work as entrepreneurs in the informal sector, since they consider that in this way they could have a better work-family balance, (Maloney, 2004). On the other hand, educated microfirm owners are expected to extract greater benefits from operating in the formal sector. In the model, we consider five dummies for SCHOOLING: ILLITERATE, PRIMARY (between 1 and 6 years of schooling, not including preschool years of education), ELEMENTARY HIGH SCHOOL (between 7 and 10 years of schooling, including the previous schooling level), HIGH SCHOOL (between 11 and 13 years of schooling, including the previous levels), and UNIVERSITY (more than 13 years of schooling). The first dummy is excluded from the model for the usual reasons of identification. The motivation for including SCHOOLING as a determinant of the sector of operation is to use it as a proxy of entrepreneurial ability. There are theoretical models that obtain that the size of a firm is determined by its owner's entrepreneurial ability, (Lucas, 1978; Jovanovic, 1982; Rauch, 1991; and De Paula and Scheinkman, 2011).

The second variable we consider is the TIME the microfirm has been operating. We expect that firms with more time in the market would tend to be in the formal sector (Jovanovic, 1982). In order to capture nonlinearities in the effect of this variable, it is included in linear and quadratic forms.

The third variable is the microfirm owner's access to credit. Because formal microfirm owners tend to have broader access to credit from the formal market, we use the type of credit the individual used to start the microfirm.[3] The variable includes three dummies: OWN CAPITAL; FORMAL CREDIT, if the credit was provided by the suppliers of the microfirm or by a formal credit institution; and INFORMAL CREDIT, if the credit came from a

[3] Certainly, it is possible that history of credit accessibility is correlated with unobserved components of entrepreneurial ability, which may generate inconsistency of some specific coefficients in our model. Our main interest, however, is to fit the expected differences between responses of informal microfirms and their counterfactuals, not to identify specific coefficients. Regressors can fail to be exogeneous, but the treatment effect can still be identified as discussed in Heckman *et al*. (1998).

source such as family, friends, or another informal lender. The dummy OWN CAPITAL is excluded from the regression. Given the possibility that access to credit might be endogenous to the sector decision, we estimated two additional specifications as robustness checks: one in which we exclude all the credit dummies, and another in which we only use one dummy variable that takes the value one if the entrepreneur started the microfirm with a credit from any source (formal or informal) and zero if he/she used funds of his/her own.

The fifth variable is the productive sector to which the microfirm belongs. We consider four dummies: AGRICULTURE, MANUFACTURING, COMMERCE, and SERVICES. The construction sector is excluded from the analysis because this sector is under-represented in the sample. The dummy AGRICULTURE is excluded from the regression to avoid the problem of multicollinearity.

In our core specification, we include AGE and MARITAL STATUS as regressors in the selection equation only. In the response equations we exclude these regressors for econometrical reasons, as we explain below. Our intention is to take into account evidence that supports the idea that younger individuals tend to be less risk averse (Jovanovic, 1979), while individuals who have family responsibilities tend to be more risk averse (Carrasco, 1999). Because operating in the informal sector entails higher risks for the microfirm owner of being discovered and punished by governmental officials, we expect that older and married microfirm owners would prefer to operate in the formal sector. Therefore, by including these variables in the selection equation, we intend to model the effect of the entrepreneur's risk aversion on the decision of sector of operation. Regarding to the response equations, the variable AGE is highly correlated with the entrepreneur's experience, which is already captured in this model by the linear and quadratic effects of TIME. So, having AGE in the response equations entails the risk of multicollinearity.[4] Also, we do not expect any significant effect of MARITAL STATUS on the microfirm's size, conditional on the sector in which the entrepreneur has decided to operate. Nonetheless, we tested an additional specification in wich we added MARITAL STATUS while excluding AGE in the response equations. The variable AGE is measured in years, while the variable MARITAL STATUS is a dichotomic taking the value one for married individuals.

[4] We performed the experiment of adding age while excluding marital status in the response equations. The model, however, did not converge, an outcome that supports our hypothesis of multicollinearity when adding age to the response equations.

In order to control for the characteristics of local markets, specific dummies for municipalities are considered in each equation. Also, to control for heteroskedasticity, we consider time of operation (in its log form) as the variable H_i. We expect that younger microfirms (i.e. in the early stages of adaptation to the market conditions) would display greater variability in the response variables than those that have been operating for a longer time.

Finally, the dependent variables MONTHLY NET PROFITS, CAPITAL STOCK and LABOR are used in their log forms in the estimation.

III. Results

In this section we test our model and discuss the effect that formality may have on informal microfirms. Since direct discussion of slope estimates can lead to erroneous conclusions in nonlinear models, in addition to the coefficient estimates, table 2 reports the marginal effects for the selection equation, which are suitable estimators for quantifying the impact of each regressor on the expected value of a dichotomous dependent variable (Greene, 2007).

We evaluate three hypotheses regarding our model specification. In each case, we use a Wald (Chi-2) test, based on the likelihood of the unrestricted model. The first test simultaneously evaluates the hypotheses H_0: $\beta_j^0 = \beta_j^1$ $j = 2, 3, 4$; that is, it evaluates whether net monetary profits, the degree of investment in capital, and labor levels are determined in the same way in both the formal and informal sectors. The test rejected the null hypothesis (p< 0,0001), indicating that microfirms react differently depending on the sector in which they operate.

The second test evaluates the existence of heteroskedasticity; that is, it simultaneously evaluates the hypotheses $H_0: \delta_1 = \delta_j^0 = \delta_j^1 = 0 \;\; j = 2, 3, 4$. In this case, the null hypothesis was rejected (p=0.179), although according to table 2, δ_3^0 is significant individually.

The third test evaluates whether there are efficiency gains when estimating a unique Roy model with three responses, instead of estimating three separate Roy models with only one response. If the covariances between the error terms of the different responses for each regime are zero, then a joint estimation will not be more efficient than estimating equations for the three responses separately. The corresponding null hypothesis is $H_0 : \sigma_{uv}^s = 0 \;\; u, \;\; v = 2, 3, 4 \;\; s = 0, 1$. The Wald test rejected H_0 (p< 0,001), an outcome that supports our joint specification.

Table 2. Estimates and marginal effects for the selection equation

		Informal sector		*Formal sector*	
Equation	*Variable*	*Estimate*	*Std. error*	*Marg. effect*	*Std. error*
Sector (selection)	Constant	-1.172 [a]	0.156		
	Gender	-0.015	0.035	-0.005	0.012
	Marital status	0.158 [a]	0.037	0.055 [a]	0.012
	Age	0.016 [a]	0.002	0.006 [a]	0.000
	Primary	-0.003	0.055	-0.001	0.019
	EHS*	0.154 [a]	0.058	0.056 [a]	0.021
	High school	0.399 [a]	0.070	0.145 [a]	0.023
	University	1.008 [a]	0.094	0.348 [a]	0.021
	Time	0.003	0.006	-3.1E-4	9.7E-4
	$Time^2$	-1.9E-4	1.5E-3		
	Formal credit	0.529 [a]	0.083	0.182 [a]	0.024
	Informal cred	0.186 [a]	0.041	0.066 [a]	0.013
	Manufacture	-0.125 [b]	0.063	-0.044 [b]	0.022
	Commerce	0.423 [a]	0.068	0.150 [a]	0.020
	Services	0.120 [b]	0.057	0.042 [b]	0.019
Monthly net profit	Constant	8.345 [a]	0.119	7.970 [a]	0.145
	Gender	-0.686 [a]	0.037	-0.463 [a]	0.035
	Primary	0.091	0.056	-0.231 [a]	0.062
	EHS*	0.202 [a]	0.059	-0.052	0.064
	High school	0.498 [a]	0.069	0.228 [a]	0.071
	University	1.086 [a]	0.077	0.632 [a]	0.077
	Time	0.009	0.006	0.013 [b]	0.006
	$Time^2$	-1.9E-4 [b]	1.5E-4	-3.1E-4 [c]	1.6E-4
	Formal credit	0.402 [a]	0.093	0.125 [c]	0.067
	Informal credit	0.156 [a]	0.042	0.137 [a]	0.039
	Manufacture	-0.397 [a]	0.064	-0.009	0.074
	Commerce	0.190 [a]	0.062	-0.142 [b]	0.067
	Services	-0.134 [b]	0.057	-0.159 [b]	0.063
Capital stock	Constant	10.70 [a]	0.149	11.36 [a]	0.191
	Gender	-0.319 [a]	0.045	-0.359 [a]	0.043
	Primary	0.067	0.068	-0.322 [a]	0.077

Table 2. Estimates and marginal effects for the selection equation (Cont.)

		Informal sector		*Formal sector*	
Equation	*Variable*	*Estimate*	*Std. error*	*Marg. effect*	*Std. error*
	EHS*	0.258[a]	0.071	-0.098	0.079
	High school	0.606[a]	0.084	0.087	0.088
	University	0.886[a]	0.104	0.396[a]	0.099
	Time	0.008	0.007	0.015[b]	0.007
	Time^2	-4.2E-4[b]	1.8E-4	-1.9E-4	1.9E-4
	Formal credit	0.330[a]	0.120	0.142[c]	0.083
	Informal credit	0.095[c]	0.052	0.143[a]	0.048
	Manufacture	-1.634[a]	0.077	-0.477[a]	0.091
	Commerce	-1.242[a]	0.077	-0.671[a]	0.084
	Services	-1.866[a]	0.069	-0.810[a]	0.077
Labor	Constant	-1.696[a]	0.210	-0.403[b]	0.189
	Gender	-0.169[a]	0.057	-0.137[a]	0.037
	Primary	0.281[a]	0.087	0.138[b]	0.066
	EHS	0.042	0.092	0.057	0.068
	High school	0.358[a]	0.106	0.101	0.077
	University	-0.114	0.145	0.076	0.100
	Time	0.005	0.012	0.001	0.006
	Time^2	-5.9E-4[b]	3.0E-4	-7.0E-5	1.7E-4
	Formal credit	0.310[b]	0.142	0.253[a]	0.074
	Informal credit	0.141[b]	0.064	0.072[c]	0.041
	Manufacture	0.790[a]	0.123	0.787[a]	0.081
	Commerce	1.178[a]	0.126	0.492[a]	0.079
	Services	0.916[a]	0.116	0.436[a]	0.070
Covariance matrix	σ_{12}	0.867[a]	0.052	0.412[a]	0.082
	σ_{13}	-0.173[c]	0.092	0.127	0.120
	σ_{14}	0.129	0.115	0.063	0.135
	σ_{22}	1.300[a]	0.099	0.910[a]	0.074
	σ_{23}	0.201[a]	0.052	0.326[a]	0.038
	σ_{24}	0.356[a]	0.078	0.213[a]	0.038
	σ_{33}	1.741[a]	0.121	1.395[a]	0.098
	σ_{34}	0.331[a]	0.045	0.283[a]	0.027

Table 2. Estimates and marginal effects for the selection equation (Cont.)

		Informal sector		*Formal sector*	
Equation	*Variable*	*Estimate*	*Std. error*	*Marg. effect*	*Std. error*
	σ_{44}	1.248 [a]	0.229	0.722 [a]	0.082
	δ_1	0.028	0.067	0.028	0.067
	δ_2	-0.033	0.031	0.032	0.031
	δ_3	-0.076 [b]	0.031	-2.3E-04	0.031
	δ_4	0.027	0.082	0.064	0.051

Source: Authors' Own elaboration using data from Enamin, 1998 and 2002. *Notes:* *EHS: elementary high school. [a] Significant at 0.01; [b] Significant at 0.05; [c] Significant at 0.10.

III.1. Determinants of Sector Choice

According to table 2, the main determinants of sector choice (formal or informal) are: the owner's schooling, the owner's access to credit to start the microfirm, and the microfirm's type of economic activity.

The probability that an individual with university studies will operate in the formal sector is 34 percentage points higher than one without formal studies and 20 points higher than one who has achieved a high school education. An individual with a high school education has 9 more probability points than an individual who achieved an elementary high school education to operate in the formal sector. Moreover, a microfirm owner with an elementary high school education has 5 more probability points than an individual with primary education to be a formal entrepreneur. Finally, our analysis indicates that the probability of operating in the formal sector for an individual that has obtained only a primary school education is not greater than that of an individual with no formal education, and that the probability of operating in the formal sector is increasing with respect to the achievement of higher education levels.

A microfirm owner that had access to either a formal or informal credit source to open his business has 18 and 6, respectively, more probability points to choose to operate in the formal sector than an individual who used his own money to do so. This indicates that an individual who was able to obtain seed capital from a formal or informal credit source has more probability points of choosing the formal sector than an individual who used his own money. Moreover, an entrepreneur that has obtained outside financing to start his microfirm, either formal or informal, tends to show higher monthly net profits, capital investment and labor than an

entrepreneur that used his own money to do so. This last result is statistically significant both for informal and formal entrepreneurs. However, this tendency seems to be stronger for informal entrepreneurs that have obtained outside funds from a formal source to start his business (3 per cent of the informal entrepreneurs sub-sample).

Also, an entrepreneur who works in the sector of commerce or services has 15 and 4, respectively, more probability points to work in the formal sector than an individual who works in agriculture.

On the other hand, a married individual has 5 probability points more than an unmarried individual to work in the formal sector, and this result is statistically significant. The owner's age, albeit statistically significant, seems to explain little of the owner's choice of sector of operation. The owner's gender has not statistically significant effect on entrepreneur's choice of sector of operation.

Some studies, like Jovanovic (1982), conclude that larger and older firms are more efficient. This conclusion leads us to expect that older firms would be more inclined to move from the informal sector to the formal sector after attaining a certain size and level of abilities. However, our estimates indicate that the time a company has been operating is not in itself a factor in the sector choice (the marginal effect of time is not significant, see table 2). On the other hand, a longer operation time implies greater size (in terms of a more intensive use of capital and labor), which suggests a more elaborated relation between the age of the microfirm and its likelihood to remain in the informal sector. As table 2 shows, microfirms that have been in the market longer tend to show greater net profits and investment in capital than those of younger microfirms. Note that this tendency is statistically significant only among formal microfirms, indicating that only the formal sector provides the necessary incentives for the growth of a microfirm. Our empirical estimation is static; but our results suggest that a microfirm born in the informal sector may eventually face the decision to remain at a certain size or move to the formal sector to extend its scale of production.

III.2. Effect of the Sector Choice on net Profits and Microfirm Size

In order to compare each microfirm against its counterfactual, we use expression (3) for microfirms currently operating in the informal sector. Table 3 provides a means of comparing the behavior of microfirms currently in the informal sector and their expected behavior had they been operating in the formal sector.

Table 3. Effects of formality on the net profits and size of Mexican microfirms

	*Expected log difference between operating formally minus operating informally, conditional on operating informally**	
Response	*Effect*	*Standard error*
Net annual profit	-0.168	0.135
Capital stock	1.135 [a]	0.195
Labor	0.216 [b]	0.118
Adding marital status		
Net annual profit	-0.237 [c]	0.121
Capital stock	0.954 [a]	0.151
Labor	0.128 [b]	0.062
Excluding all credit dummies		
Net annual profit	-0.255 [b]	0.123
Capital stock	0.937 [a]	0.149
Labor	0.125 [c]	0.067
Including a single credit dummy		
Net annual profit	-0.237 [c]	0.140
Capital stock	0.948 [a]	0.154
Labor	0.124 [c]	0.063

Source: Authors' own elaboration using data from Enamin, 1998 and 2002. *Estimation carried out on the subsample of informal microfirms. [a] Significant at 0.01; [b] significant at 0.05; [c] significant at 0.10.

The estimations indicate that in the formal sector, these same microfirms would obtain an annual net profit that is 6 per cent lower than what they currently make in the informal sector.[5] The difference is not statistically significant, however. This is true regardless of their scale of operation, since, according to the values in the same table, when operating in the formal sector these microfirms would have to raise their investment in capital in 250 per cent and increase the number of workers in 80 per cent.

[5] Note that the dependent variables in the model are in log form; thus, figures displayed in table 3 are the difference between the expected values of two logarithms. The percentages presented in the text for the potential transition from the informal to the formal sector were obtained from the ratio

$$\frac{E[e^{y^1 ji} \mid s_i = 0] - E[e^{y^0 ji} \mid s_i = 0]}{E[e^{y^0 ji} \mid s_i = 0]}, \; j = 2, 3, 4.$$

On the other hand, as we mentioned before, we performed several robustness checks in order to test the suitability of several of the decisions we made in our econometric specification; that is: *a)* Adding MARITAL STATUS while excluding AGE in the response equations, *b)* Excluding CREDIT dummies, and *c)* Including only one CREDIT dummy. In all the aforementioned specifications, see table 3, we obtained results that show that currently informal microfirms would obtain (were they in the formal sector) a profit that is at least 20 per cent lower (statistically significant only in one specification) than what they currently obtain, after increasing their capital investment in at least 155 per cent and the number of workers in at least 13 per cent. Even though we observe numerical differences, these additional results are of the same nature as our core results, mentioned in the previous paragraph.

In summary, our results indicate that in order to compete in the formal sector, a microfirm must increase its size. Despite these adjustments, microfirms that are currently informal would not perform better in the formal sector in terms of profits than they currently do operating informally. An explanation to this result is that, probably, those microfirm owners do not have the ability of exploiting the advantages that the formal sector offers. Therefore, those owners' optimal decision is to remain working in the informal sector.

Our results are along the lines of Levenson and Maloney (1998), who consider formalization to be part of the microfirms' growth process in developing countries. Thus, above a certain size, a microfirm is better off in the formal sector, because it can continue growing, pay more qualified workers, and earn enough profits to compensate for the additional cost of formalization.

Conclusions

In this article, we use a multivariate framework to study the effect of formality on three response variables of informal microfirms, including net profits and levels of capital stock and labor. We are able to control for the correlation between microfirm responses and heteroskedasticity. Using a Montecarlo EM algorithm allows the estimation of a seven-equation model using full information maximum likelihood.

Our results indicate that the main determinants of sector choice are the microfirm owner's level of formal education, access to credit and type of economic activity. In addition, microfirms with more time in the market

tend to have higher net profits and capital than younger companies; but this tendency is statistically significant only for formal microfirms. All of this allows us to suggest, even taking into account the static nature of our estimation, that the necessary incentives for the growth of a microfirm exist only in the formal sector, and that this growth takes place only if the microfirm owner's education level and ability to obtain outside funds are adequate, as well as if the microfirm works in commerce and services.

Regarding the effect of formality on microfirm net profits and size, our results indicate that, when comparing the behavior of currently informal microfirms and their expected behavior were they to operate in the formal sector, the microfirms would not obtain a higher annual net profit in the formal sector than what they currently obtain in the informal sector. In addition, if operating in the formal sector, the currently informal microfirms would need to considerably increase their investment in capital and their number of workers. It is interesting to note that, despite these adjustments in capital and labor, informal microfirms would not do better in the formal sector in terms of profits than they currently do operating informally. This result suggests that the level of formal education, ability to obtain outside funds and type of activity of those individuals that choose to operate in the informal sector limit the profits they can obtain in the formal sector. Therefore, the optimal decision of such entrepreneurs is to remain informal.

Given our empirical results, a direction for further research could be studying the effect of increasing the level of formal and financial education of the informal microfirm owners by means of an effective qualification of middle-aged entrepreneurs and incentives for young entrepreneurs to resume their formal education. Fajnzylber *et al.* (2009) find empirical evidence of the positive impact of training on the profits of microfirms.

Even though formal education does not seem to be relevant in some international measures of entrepreneurial ability based on personal characteristics —see, for instance, Ardagna and Lussardi (2010)— our empirical results seem to suggest that, given the low level of education and lack of outside funding observed in the sample of entrepreneurs of our study, investment in formal and financial education might improve the entrepreneurial ability of those individuals.

Finally, this analysis provides a framework that might lead to further discussion of whether a more educated informal entrepreneur would be able to adjust his microfirm's structure were it to operate in the formal sector and to exploit the advantages of the formal sector of the economy, considering different policy scenarios.

References

Ardagna, S. and A. Lussardi (2010), "Explaining International Differences in Entrepreneurship: The Role of Individual Characteristics and Regulatory Constraints", in J. Lerner and A. Schoar (eds.), *International Differences in Entrepreneurship,* Chicago, The University of Chicago Press.

Bruhn, M. (2012), "A Tale of Two Species: Revisiting the Effect of Registration Reform on Informal Business Owners in Mexico", Policy Research Working Paper, 5971, World Bank.

Carrasco, R. (1999), "Transitions to and from Self-Employment in Spain: An Empirical Analysis", *Oxford Bulletin of Economics and Statistics*, 61, pp. 315-341.

Cunningham, W. and W. Maloney (2001), "Heterogeneity among Mexico's Microenterprises: An Application of Factor and Cluster Analysis", *Economic Development and Cultural Change*, 50, pp. 131-156.

Dabla-Norris, E., M. Gradstein, and G. Inchauste (2008), "What Causes Firms to Hide Output? The Determinants of Informality", *Journal of Development Economics*, 85, pp. 1-27.

De Paula, A. and J. Scheinkman (2011), "The Informal Sector: An Equilibrium Model and Some Empirical Evidence from Brazil", *Review of Income and Wealth, International Association for Research in Income and Wealth*, 57, pp. S8-S26,05.

Duval, R. and P. Orraca (2011), "Análisis por cohortes de la participación laboral en México (1987-2009)", *El Trimestre Económico,* LXXVIII (2), 310, pp. 343-375.

Fajnzylber, P., W. Maloney and G. Montes-Rojas (2009), "Releasing Constraints to Growth or Pushing on a String? Policies and Performance of Mexican Micro-Firms", *Journal of Development Studies*, 45, pp. 1027-1047.

Greene, W. (2007), *Econometric Analysis* sixth edition, Pearson Prentice Hall.

Heckman, J., H. Ichimura, J. Smith and P. Todd (1998), "Characterizing Selection Bias using Experimental Data", *Econometrica*, 66, pp. 1017-1098.

Heckman, J. and E. Vytlacil (2007), "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation", *Handbook of Econometrics,* 6, pp. 4779-4874.

Jovanovic, B. (1979), "Job Matching and Theory of Turnover", *Journal of Political Economy*, 87, pp. 972-990.

_____ (1982), "Selection and Evolution of Industry", *Econometrica*, 50, pp. 649-670.

Levenson, A. and W. Maloney (1998), "The Informal Sector, Firm Dynamics, and Institutional Participation", Policy Research Working Paper, 1988, World Bank.

Levy, S. (2008), *Good Intentions, Bad Outcomes: Social Policy, Informality and Economic Growth in Mexico,* Washington, D. C., Brookings Institution Press.

Lucas, R. (1978), "On the Size Distribution of Business Firms", *Bell Journal of Economics*, 9, pp. 508-823.

Maloney, W. (2004), "Informality Revisited", *World Development,* 32, pp. 1159-1178.

McKenzie, D. and Y. Sakho (2010), "Does it Pay Firms to Register for Taxes? The Impact of Formality on Firm Profitability", *Journal of Development Economics*, 91, pp. 15-24.

Natarajan, R., C. McCulloch, and N. Kiefer (2000), "A Monte Carlo EM Method for Estimating Multinomial Probit Models", *Computational Statistics and Data Analysis*, 34, pp. 33-50.

Rauch, J. (1991), "Modelling the Informal Sector Formally", *Journal of Development Economics*, 35, pp. 33-47.

Rosenbaum, P. and D. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, pp. 41-55.

Roy, A. (1951), "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 3, pp. 135-146.

Wei, C. and M. Tanner (1990), "A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms", *Journal of the American Statistical Association*, 85, pp. 699-704.