Implicit Contracts, Occupational Choice and the Impact of Illness on the Gender Wage Gap in Mexico*

Federico H. Gutierrez†

Yale University

January 20, 2011

Abstract

Labor market institutions and risk coping mechanisms are fundamental topics to understand developing countries' economies. This paper lies in the intersection of the two. I test the role of labor contracts as an insurance mechanism against idiosyncratic productivity shocks, particularly illnesses. In addition, I show evidence that this characteristic of contracts partially explains the gender wage gap in Mexico.

From an implicit contract model with occupational choice, I predict that i) the probability of getting ill, but not illness, negatively affects hourly earnings in an optimal contractual arrangement, ii) workers with better access to ex-post mechanisms to smooth consumption are less likely to choose the contract market, and iii) employers in the contract market internalize the additional responsibilities of women at home reducing their wages in relation to men.

Using the Mexican Family Life Survey, I find consistent evidence that, on average, wage workers are in a contract market that provides them insurance against illness shocks. This institution interacted with the fact that women, but not men, miss days at work when other members of the family get sick, is responsible for reducing women wages in the order of 6% to 7%. Applying the same analysis to self-employed workers, a group for which this type of insurance is not possible, correctly rejects the contract market equilibrium.

*I thank Mark Rosenzweig, T. Paul Schultz and Christopher Udry for their advice and support. I also thank David Atkin, Nancy Qian, Taisuke Otsu, Mir Salim, Veronica Santarosa, Camilo Dominguez, Rachel Heath, Gharad Bryan, Muthoni Ngatia and seminar participants at Yale University for valuable comments and suggestions. All remaining errors are my own

†Yale University, Dep. of Economics, 27 Hillhouse Ave., New Haven, CT. Email: federico.gutierrez@yale.edu

1 Introduction

The organization of labor markets on the one hand, and the risk coping mechanisms on the other are central topics to understand the economy of developing countries.

With respect to risk coping mechanisms, there is a growing number of studies that focus on *ex-post* institutions. That is, on how households smooth consumption after an income shock. Some of the papers analyze particular mechanisms such as the sale of investment assets in moments of financial distress (Rosenzweig and Wolpin [1993]), migration and marriage (Rosenzweig and Stark [1989]) or informal credit markets (Udry [1994]), while others focus on the overall effectiveness of the institutions (Townsend [1994], Ethan Ligon and Worrall [2002] among others).

On the other hand, *ex-ante* mechanisms - those meant to reduce the volatility of income - have also receive some attention. Rosenzweig and Biswanger [1993] analyze how farmers in India choose the composition of productive assets in a way that the variance of profits is reduced at the expense of lowering its expected value. Kochar [1999] studies how workers in rural India shift labor from farm to off-farm employment when they face crop shocks.

Notably, the use of labor contracts as a mechanism to smooth income has not been thoroughly studied in developing countries, at least in an economy-wide setting. Closely related, the literature has focused on explaining the determinants of long-term contract workers versus short-term casual workers (Bardhan [1984], Mukherjee and Ray [1995]). But, no attempt has been made to test if insurance against productivity shocks is a dominant characteristic of labor arrangements in low and middle-income countries[1]. Furthermore, to my knowledge no paper exists that studies how the insurance component in labor contracts interacted with the division of tasks within the household affects the gender wage gap in developing countries.

In this paper I implement a new test for an implicit contract equilibrium against a Walrasian or spot market equilibrium based on the response of labor outcomes to illness shocks. I also analyze the idea that the insurance component of labor contracts in combination with the additional duties women have at home explains part of the gender wage gap in Mexico.

One of the reasons for the existence of labor contracts in the economy is because, in addition to specifying the exchange of labor services for pay, they provide insurance to workers by isolating, at least partially, earnings from productivity shocks. This idea, also known as the implicit contract

[1] Beaudry and DiNardo [1991] provide a test for implicit contract in the US

theory (Azariadis [1975], Baily [1974], Holmstrom [1981]), is based on the notion that risk averter workers with limited access to credit markets prefer a low-variance over a high-variance stream of earnings, both with the same mean. Firms, on the other hand, assumed to be risk neutral or to have better access to credit markets, are indifferent between the two streams of earnings they have to pay - same expected profit - but they are willing to offer the low-variance stream to attract workers. In this way, the worker remuneration is detached in the short-run from his productivity.

This contract theory moves in opposite direction to the standard principal-agent model where labor contracts make the worker remuneration fluctuate, at least stochastically, with his productivity in order to induce certain level of effort. This means that the existence of written labor contracts *per se* is not evidence that they contain an insurance component.

On the other hand, the lack of written contracts is also not evidence that the insurance component is absent in labor arrangements. Thomas and Worrall [1988] theoretically show that this type of contracts can be self-enforcing in the sense that both the employer and the employee have no incentives to renege. So, under certain conditions there is no need to write contracts to ensure their fulfillment.

Beaudry and DiNardo [1991] test the implicit contract hypothesis in the US using different stages of the business cycle. They argue that in a spot market, only current labor market conditions should explain wages (i.e. unemployment rate) but in an implicit contract equilibrium, the past labor conditions should be the relevant ones (i.e. unemployment rate at the moment of signing the contract and the minimum unemployment rate during the life of the contract). In order to test for an implicit contract equilibrium against a spot market one, I adopt a different approach. I use illness as a measure of productivity fluctuation and study the behavior of labor outcomes in response to it.

As explained before, in an implicit-contract market, the worker's remuneration should be invariant *in the short-run* to productivity fluctuations, in this case illnesses. Nonetheless, *in the long-run*, the frequency the worker gets sick affects his average productivity and so his wage level. Let's take for example an employer hiring a new employee. It will remunerate the worker in the same way if this one stays healthy all the time or if he gets sick and misses some days at work, as the implicit contract theory predicts. But, at the moment of 'signing' the contract, the employer have the incentive to correctly estimate how often the worker will miss days at work to offer the appropriate wage level. So, in an implicit contract equilibrium, the probability of being ill, but not illness, is a determinant of the wage level. It is important to note that the opposite is true in the spot market where wages equal the marginal product at all times. In this case, illness have a short-run impact on wages, but the frequency the worker gets sick is irrelevant to explain the current productivity level and hence the current wage rate.

With the same argument, the model I present predicts that the interaction of labor contracts and the sexual division of labor within the household explains part of the gender wage gap. Women tend to be in charge, more than men, of house duties and this may force them to miss more days at work than men. In particular, if the division of tasks within the household is such that when a child - or any other member of the family - gets sick the mother is the one in charge of staying at home to take care of him, then this additional responsibility of women is internalized by the employer at the moment of offering the contract decreasing the wage rate of women in relation to men. Once again, this characteristic is specific of contracts that smooth income and is not present in the spot market.

In order to test the for the presence of insurance in labor contracts, not only do I need information about illness, but I also need information about the probability of getting sick. I approximate this measure with the prevalence of illness in the locality. Consequently the implementation of the test consists of evaluating if the wage rate is lower and the gender wage gap is higher in localities with high prevalence of diseases.

The group of interest to apply the test of implicit contracts and its impact on the gender wage gap are wage workers. But, in order to validate the methodology I also apply the test to self-employed workers. In this group, the employer and the employee are the same person and so his earnings cannot be disentangled from productivity shocks. Consequently, as I will argue later, those in this group should behave in most dimensions as if they were employed in the spot market.

Both in the model and in the empirical analysis, I take into account the endogenous decision of workers choosing their occupations. Those with relatively limited access to ex-post mechanisms to smooth consumption are expected to choose the wage sector where they can obtain insurance against productivity fluctuations. One the other hand, workers with other means to smooth consumption, such as family members in the US that they can ask for help in moments of financial distress, are more likely to be self-employed.

This paper also relates to the literature on the relation between health and labor outcomes. I argue that the impact of illness on wages cannot be understood correctly without taking into consideration labor markets institutions such as the presence of implicit contracts.

The rest of the paper is structured as follows. In section 2, I develop a model comparing the outcomes of an implicit contract market versus a spot market and derive testable implications with respect to earnings, wages and hours worked. In section 3, I present the data, variable definition and descriptive statistics.

In section 4, I estimate the impact of illness on the time allocation, including the labor supply. I show that, consistent with implicit contracts, the response of hours worked to illness is larger in absolute value for wage workers than for self-employed workers. I also show that women, but not men, miss days at work to take care of other members of the family when they get sick, confirming the additional responsibilities women have at home.

In section 5, I present baseline results. In section 6, I deal with occupational choices of workers. I show that, consistent with the model, the availability of ex-post mechanisms to smooth consumption is negatively associated with workers becoming wage workers. In section 7, I analyze how robust results are with respect to migration, firm location, firm size and the potential endogeneity of illness. In section 8, I focus on measuring the extent to which labor contracts explain the gender wage gap. Finally, in section 9, I conclude.

2 The Model

I assume that workers in the economy may work either in the spot market or in the contract market. In this latter case, firms offer a contract (E_s, h_s) that consists of earnings and hours in each state of nature s; where $s = 0$ if the worker is healthy and $s = 1$ if he is ill.

In a competitive market, in order to attract workers, risk neutral firms have the incentive to offer the best possible contract for the employee given the level of expected profit. Then, the optimal contract for the average representative worker solves the following problem:

$$max_{\{C_0,C_1,E_0,E_1,l_0,l_1,b\}} \quad (1-q)\ U(C_0,l_0,0)+q\ U(C_1,l_1,1) \tag{1}$$

$$s.t \quad C_0 = E_0 + y - qb \tag{2}$$

$$C_1 = E_1 + y + (1-q)b \tag{3}$$

$$V = (1-q)\ \left[f'_{h0}h_0 - E_0\right] + q\ \left[f'_{h1}h_1 - E_1\right] = 0 \tag{4}$$

$$T = l_0 + h_0 \ \ ; \ \ T = l_1 + h_1; \ \ l_0, l_1, h_0, h_1 \geq 0 \ \ ; \ \ b \leq \bar{b} \tag{5}$$

Equation (1) is the expected utility function of the worker. With probability $(1-q)$ he is healthy and with probability q he is sick. In each state of nature s, the utility depends on consumption C_s, leisure l_s and illness s that takes the value 0 if healthy and 1 if sick.

Equations (2) and (3) are the budget constraints in each state. In addition to earnings E_s and non-labor income y, the worker may count on b, an ex-post mechanism to smooth consumption such as, but not restricted to, a network of friends and relatives that he can ask for money when needed. Then, the insurance in (2) and (3) can be thought as follows. Workers belong to a large network. At the beginning of the period each member contributes with an amount qb (i.e. the insurance premium). The resulting pot is subsequently divided among those who get sick. Because this group represents a fraction q of the network, when a worker gets sick he receives an amount b. I assume, however, that b is bounded from above ($b \leq \bar{b}$), at least for some workers, indicating that ex-post mechanisms to smooth consumption are incomplete leaving room for ex-ante ones such as those embedded in labor contracts[2].

Equation (4) imposes the condition that the contract should be such that firms don't lose money. V is the expected value of the worker for the firm. If the worker is healthy, that occurs with probability $(1-q)$, the firm gets the marginal product of labor f'_{h0} times the number of hours worked h_0, and it has to pay the worker E_0. On the other hand, with probability q the worker is sick and a similar condition holds although the marginal product of labor, hours worked and earnings may be different in this state. In a competitive environment where firms are free to enter and exit the market, the equilibrium expected value of the worker for the firm should be zero ($V = 0$). For now on, I assume no heterogeneity among firms. But, in subsequent sections, I will incorporate the size of establishments in the analysis.

[2]In the example, I assume for simplicity that illness is independent among members of the network.

Finally, (5) indicates that the total amount of time (T) is divided between leisure and work in each state $s \in \{0, 1\}$

The first order conditions for the employed worker are:

$$
\begin{aligned}
E_0: \quad & U'_{c_0} - \lambda \quad = 0 & (6)\\
E_1: \quad & U'_{c_1} - \lambda \quad = 0 & (7)\\
l_0: \quad & U'_{l_0} - \lambda f'_{h_0} = 0 & (8)\\
l_1: \quad & U'_{l_1} - \lambda f'_{h_1} \geq 0 \quad ,(= 0 \text{ if } l_1 < T) & (9)\\
b: \quad & - U'_{c_0} + U'_{c_1} \geq 0 \quad ,(= 0 \text{ if } b < \bar{b}) & (10)
\end{aligned}
$$

The insurance component in labor contracts and the ex-post mechanisms to smooth consumption are perfect substitutes. This can be seen in the polar situation where the worker has no binding limit to borrow money from the network ($b < \bar{b}$). In this case, equations (10) and, (6) and (7) yield the same condition. If this is the case, the insurance component in labor contracts provides no additional benefit for the worker. However, if ex-post mechanisms are insufficient, as I assume here, there is value in the insurance embedded in labor contracts.

The inequality in (9) captures the possibility that the optimal number of hours worked in case of illness is zero. Assuming that the utility function is additively separable in consumption and hours worked (i.e $U(C_s, h_s, s) = U^c(C_s) + U^h(h_s, s)$), I obtain from equations (6) and (7) implication 1.

Implication 1 $E_0 = E_1$

The optimal contract fully insure the worker paying equal earnings in both states of nature.

Equations (6) and (7) indicate that the optimal contract guarantees the worker the same level of consumption in each state (i.e. $(U'_{c_0} = U'_{c_1}) \Rightarrow (C_0 = C_1)$). A sufficient condition for this to occur is that the employer pays the worker the same in each state ($E_0 = E_1$). However, this is not a necessary condition unless the worker has no other means to smooth consumption (i.e. $\bar{b} = 0$). But, since the employer is a risk-neutral agent and consequently there is no cost for it in providing full insurance, it will do so in order to offer the best contract independently of the worker's access to ex-post insurance mechanisms[3].

[3]We can think that employers don't observe the worker's ex-post mechanisms to smooth consumption. Then,

With respect to labor supply, equations (8) and (9) give the relation between hours worked in sickness and in health.

$$\frac{U'_{l_1}}{U'_{l_0}} \geq \frac{f'_{h_1}}{f'_{h_0}} \qquad \text{, with equality if } h_1 > 0 \tag{11}$$

I assume that the marginal disutility of work increases with illness (i.e. marginal utility of leisure increases, $U'_l|_{s=1} > U'_l|_{s=0},\ \forall l$). This is a simple way of incorporating the fact that the body uses part of the energy to fight the disease leaving less energy to the rest of the activities. Consistent with Becker [1985], it implies that even if the marginal product remains constant ($f'_{h_1} = f'_{h_0}$), energy intensive activities - work in this case - decrease with illness. If, in addition, the marginal product decreases with illness ($f'_{h_1} < f'_{h_0}$) the response of the labor supply is even larger.

Comparing to the spot market where wages equal the marginal product, the response of labor supply to illness is larger in the contract market.

Implication 2 $(h_0 - h_1)_{\text{contract}} \geq (h_0 - h_1)_{\text{spot}}$

The response of the labor supply to illness is larger in the contract market than in the spot market.

The intuition of implication 2 is that, in the contract market, the worker is able to decrease the hours worked in the event of illness without affecting his earnings as shown in implication 1. On the contrary, in the spot market, where the worker receives his productivity in each state of nature, any reduction in the number of hours worked due to illness impacts his earnings. So, the worker has less incentive to reduce the labor supply in the spot market than in the contract market. This can be shown as follows.

The first order condition in the spot market is the standard one derived from the leisure-consumption model ($U'_{c_s} w_s \leq U'_{l_s},\quad s \in \{0,1\}$, with equality if $l_s < T$). When the worker is employed ($l_0 < T$), it implies:

$$\frac{U'_{l_1}}{U'_{l_0}} \geq \frac{U'_{c_1}}{U'_{c_0}} \frac{f'_{h_1}}{f'_{h_0}} \qquad \text{, with equality if } l_1 < T \tag{12}$$

In (12) I use the fact that in the spot market wages equal the marginal product. Comparing

offering a contract that provides full insurance guarantees that it is the optimal one even for those that are well insured by other means.

conditions (12) and (11), and assuming for simplicity an interior solution:

$$\underbrace{\frac{U'_{l_1}}{U'_{l_0}}}_{spot} > \underbrace{\frac{U'_{l_1}}{U'_{l_0}}}_{contract} \tag{13}$$

The inequality in (13) comes from the fact that in the spot market the ratio U'_{c_1}/U'_{c_0} is larger than one because the worker is not insured against illness (so $C_1 < C_0$). This, together with the assumption that the utility function is homothetic, implies that $(l_0/l_1)_{\text{spot}} \geq (l_0/l_1)_{\text{contract}}$ and so a greater response of the labor supply to illness if the worker is in the contract market.

Implication 3 $\quad w = f'_{h_0}(1 - q(1-\gamma))\delta;\quad$ (in contract market)

In the contract market, but not in the spot market, the wage rate depends on the *probability* of getting sick. On the other hand, illness decreases the wage rate only in the spot market.

I derive implication 3 from equation (4) and implication 1. Since earnings in both states are the same $E_0 = E_1$, equation (4) can be written as:

$$E_0 = (1-q)\ \left[f'_{h0}h_0\right] + q\ \left[f'_{h1}h_1\right] \tag{14}$$

Define $\gamma \equiv \frac{f'_{h1}h_1}{f'_{h0}h_0}$, then (14) becomes:

$$E_0 = f'_{h0}h_0(1 - q(1-\gamma)) \tag{15}$$

Then the wage rate (i.e. hourly earning) is:

$$w_s = \frac{E}{h_s} = f'_{h0}(1 - q(1-\gamma))\delta_s \text{ where } \delta_1 = \frac{h_0}{h_1}, \quad \delta_0 = 1 \tag{16}$$

It is important to note the different predictions for the probability of getting sick and for illness itself. The probability of getting sick (variable q) is a determinant of wages only in the contract market (equation (16)) but not in the spot market where wages equal the marginal product at all times ($w_s = f'_{h_s}$). For this reason, illness reduces wages only in the spot market. In the contract market, wages should mechanically increase as a consequence of earnings being constant and hours worked decreasing during illness.

The intuition of implication 3 is as follows. In the contract market, because firms insure workers against idiosyncratic productivity shocks, they pay wages based on the average productivity (across time) and not on the current level of productivity. Consequently, at the moment of offering the contract, firms have the incentive to predict the fraction of the time the worker will miss days at work due to illness (variable q). But, once the contract is set, if the worker becomes ill his remuneration will not decrease, precisely because he is insured against this type of events. In the spot market the opposite is true, the future health of the worker is irrelevant. Wages only depend on the current level of productivity. So, when the worker gets sick his wage rate decreases.

Implication 4 (insurance in contracts) and (women as caretakers) $\Rightarrow$ ($w^{women} < w^{men}$)
In the contract market, but not in the spot market, if women miss more days at work than men because they act as caretakers when other members of the family get ill then the gender wage gap should increase.

Implication 4 is an extension of implication 3. The variable q in (16) is the probability of reducing the labor supply as a consequence of illness. Until now, I have only considered the illness of the worker. Nonetheless, the concept can be extended to include illness of other members in the family. Women usually act as caretakers of family members when these fall sick, which may induce female workers to miss additional days at work. Employers in the contract market, taking this into account, are expected to offer women a contract with a lower wage rate in comparison to men.

Analytically, the probability of missing days at work q in (16) is expected to be larger for women ($q^{women} > q^{men}$). This can be shown as follows. Assume for simplicity that the probability of getting sick is p and is the same for all members in the family. If men only miss days at work when they get sick, then $q^{men} = p$. However, if women stay at home not only when they get sick but when other members of the family get sick, then their probability of missing days at work is given by:

$$q^{women} = 1 - (1 - p)^{n+1} \tag{17}$$

In equation (17), $(1 - p)^{n+1}$ is the joint probability that the female worker and the n members of the family she takes care of (e.g. her children) are healthy. So, (17) simply says that women miss days at work if they or any other family member they are responsible for, gets sick. In this equation, I assume for simplicity that the probabilities of getting ill of family members are independent. Nonetheless, implication 4 is also valid when there is (less than perfect) correlation of illness among members of the household.

Implication 4 links the interaction of labor contracts and the sexual division of tasks in the family with the gender wage gap. This relation is unique to contracts with an insurance component. In contrast, in a spot market the variable q is not a determinant of wages. So, the gender wage gap should not be affected by the different frequencies that women and men miss days at work.

2.1 Occupational Choice

I conclude the model with workers selecting either the contract market or the spot market. In order to do this, I introduce heterogeneity in health. I assume that each worker has a probability q_i of getting sick. The employer does not observe q_i directly. Instead he estimates it based on observable characteristics of the worker ($\bar{q} = E[q_i|x_i]$), and offer a contract based on this information (i.e. a contract that solves problem (1)-(5) for $q = \bar{q}$).

On the other hand, the worker compares the maximum utility he would obtain in the contract market and the maximum utility he would obtain in the spot market. Proposition 1 states that relatively healthy workers tend to choose the spot market and workers who often become sick tend to prefer the contract market.

Proposition 1 $\exists\, q^*$ s.t. if $q_i \geq q^* \Rightarrow U(contract) \geq U(spot)$
if $q_i < q^* \Rightarrow U(contract) < U(spot)$

Proof: If $q_i = 0$ then $U(contract) \leq U(spot)$. If $q_i = \bar{q}$ then $U(contract) \geq U(spot)$. Because $G(q_i, .) = U(contract) - U(spot)$ is continuous and monotonic in q_i, by the intermediate value theorem $\exists\, q^* \in [0, \bar{q}]\ s.t\ G(q_i) = 0$. Q.E.D.

In words, in the spot market the worker maximizes the utility in each state. If q_i is zero, then only one state exists and the utility of the spot market is maximum. If the worker has a probability of getting sick equal to the average for a group with his observable characteristics, then the contract maximizes problem (1)-(5) and so maximizes the utility of the worker. Because the difference in utilities is continuous in q_i, then there exists a value that makes the worker indifferent between choosing the spot market or the contract market.

In this context, the assumption about the employer not observing the probability of getting sick is necessary for some workers choosing the spot market. Otherwise, the employer would offer a contract customized to each worker such that his utility is maximum. However, if the probability of getting sick is private information, as I assume here, other variables influence the occupational choice. For example, the level of risk aversion affects the threshold q^* in proposition 1 and so two workers with the same probability of getting sick may take different decisions, one choosing the spot market and the other one the contract market.

Understanding the occupational selection process is important for two reasons. Firstly, when testing implications 3 and 4, I will empirically analyze wage determinants for groups of workers with different occupations. This may introduce econometric problems if selection is not considered. Secondly and most importantly, workers are expected to choose their occupation based on how much they value the insurance component in labor contracts which gives additional testable implications. For example, those with relatively less access to ex-post mechanisms to smooth consumption are expected to choose the contract market.

Formally the occupational decision is based on the difference between the indirect utility functions of the worker choosing either the contract market or the spot market.

$$G = V^{\text{cont}}(q_i, \bar{q}, f'_{h0}, f'_{h1}, y, \bar{b}, \psi) - V^{\text{spot}}(q_i, f'_{h0}, f'_{h1}, y, \bar{b}, \psi) \tag{18}$$

The worker will choose the contract market if $G \geq 0$, and the spot market otherwise. In addition to the individual probability of getting sick, the decision will depend on the marginal productivities when healthy and sick, non-labor income y, preferences ψ including the risk aversion, the probability of getting sick of the rest of the workers ($\bar{q}$) and the availability of ex-post mechanism to smooth consumption ($\bar{b}$). Of all these elements, only the productivities and $\bar{q}$ (in the contract market) affect wages. The individual probability of getting sick, although affecting the average productivity of the worker, is not included in the contract since it is not observed by the employer.

In the following sections, I test implications 1 through 4 with special emphasis on the last two. They provide a test for the existence of insurance in labor contracts (i.e. implicit contract equilibrium against a spot market one) and the connection between labor contracts and the gender wage gap. Although reasonable, the condition that $q^{women} > q^{men}$ in implication 4 should not be taken for granted. I will provide evidence of it before testing this implication.

3 Data and descriptive statistics

The data I use is the Mexican Family Life Survey (MxFLS)(Rubalcava and Teruel [2007a,b]) . It is a longitudinal multi-thematic survey representative of the Mexican population. So far, there are two rounds available. The first one took place in 2002 and the second one in 2005 ending in 2006.

The sample size for the first round consists on 8,440 households and 35,679 individuals distributed in 150 urban and rural localities throughout Mexico. The second round tracks first round individuals regardless of change of residence, household division or household formation. The recontact rate is over 90%.

Table 1 shows the most important variables I use in the analysis together with some descriptive statistics. The measure of illness I use for adults is an indicator that takes the value one if the person stopped doing any of her daily activities or work due to illness during the four weeks previous to the interview. For children, it is similar. It takes the value one if the child was at least one day inactive because of illness for the same interval or time.

There are some concerns about the interpretation of this measure (see Strauss and Thomas [1998, 2008] for a discussion). First, it is not a perfect measure of health. It is related with the type of activity of the person and the opportunity cost of stopping doing it. Self-employed workers may have less incentive to miss days at work compared to wage workers. For them, any productivity loss has a direct impact on earnings while for wage workers it is not the case if they are in an implicit contract equilibrium. Then, it is possible that self-employed workers do not report illness, according to the definition I use, in situations where wage workers generally do. If this is the case, the consequence for the empirical analysis is that the response of labor supply to illness is upwardly biased (in absolute terms) for self-employed workers in relation to wage workers. This goes in opposite direction to implication 2. Hence, if implication 2 is verified despite this bias, the evidence in favor of an implicit contract is even stronger.

A second concern about this measure of illness is that it is subjective and so what one person considers illness from a clinical point of view, another person may not. In table 2, we see that the prevalence of illness is significantly higher for women than for men. It could be that they tend to get sick more frequently but it could also be that women are more sensitive to symptoms. This is not a concern as long as the additional sensitivity of illness impacts on labor supply. Employers only care if the worker misses days at work or not. Even more importantly, the subjective component in this illness definition does not affect the cross-effect of other members' illness. For example, it does not affect if children's illness affects differently the labor supply of the mother and the father.

The survey provides information about the time spent in different activities. To study the time allocation I group some of these activities into five categories: hours in market work, leisure, taking care of other members, housework and other work. Leisure includes the activities watching TV and reading. Housework includes cooking/preparing food and washing clothes/cleaning the house. Taking care of other members is the number of hours taking care of elder and/or sick members in the family. The number of hours in market work does not require explanation. The only caveat is that, contrary to the other activities, the information is taken from the employment section of the survey and not from the time allocation one. Finally, the category 'other work' includes collecting firewood, water and participating in agricultural activities.

4 Earnings, labor supply and time allocation

In this section I estimate the impact of illness on earnings and on the time allocation, including the labor supply. In the model presented in section 2, for a worker either in the spot market or in the implicit contract market, the hours worked h_{it} in each period t depend on the marginal productivity $f_i'(s_{it})$ (which depends on whether the worker is ill or healthy), illness itself s_{it}, the probability of getting sick q_i (in the case of implicit contract market), ex-post insurance $\bar{b}$, non-labor income y_{it} and time invariant characteristics of the worker ψ_i such as preferences and health endowment.

$$h_{it} = h(f_i'(s_{it}), s_{it}, q_i, \bar{q}, \bar{b}_i, y_{it}, \psi_i) \tag{19}$$

To be more realistic I extend the model in two dimensions. First, workers live in households with several members where decisions are likely to be coordinated and outcomes for one member may effect others. Hence, hours worked should also depend on the marginal productivity, illness, non-labor income and the rest of the determinants in (19) of all members in the household. Second, I include characteristics of the community X_{jt} that may affect the time allocation such as the level of infrastructure or weather conditions.

Considering these extensions, I take a linear approximation and the first difference across time of (19) and obtain the following estimating equation:

$$\Delta h_{ij} = \beta_0^j + \beta_1 \Delta s_{ij} + \beta_2 \Delta s_{-ij} + \beta_3 \Delta y_{ij} + \Delta \epsilon_{ij} \tag{20}$$

The first difference eliminates the impact of preferences and other time invariant characteristics in ψ. It also eliminates the individual probability of getting sick q_i and the availability of ex-post mechanisms to smooth consumption $\bar{b}$ for all members in the family, also assumed to be time invariant. The marginal productivity changes over time due to illness and experience. Since experience is not directly observed, a proxy would be the length of time between the two periods the first difference operates. But, this period is the same for all individuals and so its effect is absorbed in the constant. The parameter β_1 captures the direct impact of illness as well as the indirect impact that it exerts through changes in the productivity. Since the survey provides information about non-labor income and assets at the household level, I assume that resources are pooled in the household and its effect is captured in β_3. Finally, changes in variables at the community level X_{jt} are absorbed by the community fixed effects (β_0^j).

Implication 2 of the model presented in section 2 states that the response of labor supply to illness is larger for workers in the implicit contract market than for workers in the spot market. In table 4, I estimate equation (20) for wage workers and self-employed workers separately. As previously explained, those in this last group should behave as if they were employed in the spot market in the sense that they cannot separate earnings from productivity shocks (i.e. the insurance component of labor contracts is not possible for them). Consequently, I use these workers as a comparison group.

There may be some concern with respect to selection bias in estimating (20) for different groups of workers. Nonetheless, taking the first difference eliminates any unobserved time invariant characteristic of the worker that would induce selection bias if the regression was estimated in levels (for example, health endowment of workers). Furthermore, in subsequent sections I show that self-selection is not an important source of bias.

In table 4, I present three different specifications depending on whose illness status is contained among regressors. The first specification includes only the worker's own illness. The second adds the health status of his children, measured as the fraction of sons and daughters that are ill, and the third includes the health condition of the spouse. As we can observe, on average, wage workers decrease their labor supply seven hours per week when they become sick. On the other hand, self-employed workers show no decrease in the labor supply. The two groups may be performing different tasks that affect differently the productivity when they miss days at work. Nonetheless, the differential response between the two groups is consistent with the idea that wage workers are in an implicit contract equilibrium as implication 2 in the model suggests.

The larger elasticity of labor supply to illness for workers in the implicit contract market is a consequence that for them, but not for workers in the spot market, earnings are invariant if they miss days at work due to illness. The decrease in productivity is absorbed by the employer who acts as insurer.

In order to verify that wage workers are insured against health shocks, I estimate the impact of illness on earnings. I use the same specification as before but instead of hours I use earnings as the dependent variable.

$$\Delta E_{ij} = \beta_0^j + \beta_1 \Delta s_{ij} + \beta_2 \Delta s_{-ij} + \beta_3 \Delta y_{ij} + \Delta \epsilon_{ij} \tag{21}$$

Table 5 shows the results of estimating equation (21) for wage workers. As we can see, neither own illness nor illness of other members in the family affects earnings despite the fact that labor supply is significantly reduced as a consequence of these events. This lack of response of earnings to productivity shocks is consistent with the implicit contract hypothesis as implication 1 in section 2 states

4.1 Time allocation and the division of labor in the family

Implication 4 of the model states that *if* women miss days at work more frequently than men, then this exacerbates the gender wage gap in the implicit contract market. However, the precondition that women miss more days at work than men should not be taken for granted. So, in this section, I present evidence of it. Specifically, I show that when other members of the family get sick (e.g. children), women but not men stay at home to take care of them.

Before moving to regression analysis, in figure 1, I show the identity of the person who takes care of each child under 12 years old. Not surprisingly, in the vast majority of the cases it is the mother. This is true even considering only children whose mothers are employed, suggesting the possibility that women stay at home and miss days at work if children require special care, for instance, in the event of illness. In order to investigate this, I estimate the impact of own and other member's illness on the time allocation of men and women separately by estimating equation (20) for different activities.

Table 6 shows results for women, we see that when children get sick the mother decreases the number of hours worked and increases the number of hours taking care of other members (i.e. the sick child). She also increase s the number of hours doing housework activities that are complementary to care. Cooking and cleaning are tasks that can be done simultaneously while taking care of the child.

When husbands get sick, women also seems to reallocate hours from market activities to home activities. The number of hours worked shows a reduction when the husband gets sick although it is not statistically significant. Nonetheless, the hours taking care of other members, doing home chores and doing leisure activities at home significantly increase suggesting the women also miss days at work not only when children are sick but when their husbands get sick.

Table 7 repeats the analyses of table 6 although for men. It can be observed that own illness decreases market work and increases the time at home, in particular leisure. Contrary to women, men seems not to change their schedule when another member of the family gets sick, either one the children or the wife.

Using the fact that women but not men miss days at work when other members of the family get sick, in the following section I estimate a wage equation that tests for the presence of insurance in labor contracts and I show how this characteristic affects the gender wage gap.

5 Test for insurance in labor contracts

According to the model presented in section 2, the wage rate for workers in the implicit contract market is:

$$w_s = f'_{h0}(1 - q(1-\gamma))\delta_s \qquad s.t. \;\; \delta_1 = \frac{h_0}{h_1}, \quad \delta_0 = 1 \tag{22}$$

Equation (22) indicates that the wage rate when the worker is sick (s=1) or healthy (s=0) equals the value of the productivity when he is healthy (f'_{h0}) adjusted by the probability of being absent at work due to illness (q) that the employer observes. If the worker is fully insured as the model states (i.e same earnings in both states of the world), then illness is expected to increase hourly earnings, as δ_s indicates, and not to decrease them as would be the case in the spot market .

Equation (22) is also valid for women in the implicit contract market with the caveat that, on average, the probability of missing days at work is higher for them than for men. Theoretically, the equivalence between q^{women} and q^{men} is given by equation (17). Empirically and in consistently with the theory, I showed in the previous section that women tend to stay at home and take care of other members in the family when they get sick.

I obtain an estimating equation by taking the natural log of equation (22) and using the approximation ($ln(1-x) \approx -x$).

$$log(w) = log(f'_{h0}) + \beta_1 \; male + \beta_2 \; sick + \beta_3 \; psick + \beta_4 \; psick * male + \mu \tag{23}$$

The log marginal product (f'_L) is approximated with years of formal education and potential experience. I include the binary variable *male* to capture standard discrimination and unobserved productivity difference in gender. The variable *sick* is an indicator of the current health condition. The measure of the probability of getting sick (*psick*) I use is the prevalence of illness among wage workers in the community. The reason for using only wage workers and not all the workers or all the members in the community, is because employers offer contracts based on the conditional expectation of those asking for jobs (section 2.1). Since workers choose occupation based on their probability of getting sick we can expect differences between wage workers and the rest.

The test for implicit contracts is based on the coefficients of illness and the probability of getting sick. If workers are in the spot market, the expected value of the parameters are $\beta_3 = 0$, $\beta_4 = 0$ and $\beta_2 < 0$. That is, wages equal the current marginal product which decreases when the worker is sick. On the other hand, if workers are in the implicit contract market the expected value of the parameters are $\beta_3 < 0$, $\beta_4 > 0$ and $\beta_2 \geq 0$. That is, the higher the probability of getting sick the lower the wage rate. Additionally in communities with relatively high prevalence of illness the gender gap should be higher since women spend on average more time at home taking care of other members.

As a baseline, I estimate equation (23) ignoring workers self-selecting into occupations and potential econometric problems such as migration and firm location, firm heterogeneity and the possible endogeneity of illness. I will address these topics subsequently.

Table 8 presents the results of estimating equation (23) pooling the 2002 and 2005 surveys. Columns 1 and 2 are the standard Mincer equations for wage workers and self-employed workers respectively. As usual, log hourly earnings increases with education, shows an inverted U-shape in potential experience and a positive male wage gap.

In columns 3 and 4, I include the prevalence of illness in the locality, the prevalence of illness interacted with a gender dummy and the current health condition of the worker. Columns 5 and 6 additionally control for the size of the community and for the fraction of the locality labor force in agricultural activities.

Results for wage workers are consistent with the implicit contract hypothesis: higher probability of getting sick (i.e. higher prevalence of illness in the community) implies lower wages. Additionally, we observe that the higher the probability of getting sick, the wider the gender wage gap because women miss days at work more frequently in relation to men to take care of other members in the family when sick.

For self-employed workers, results correctly fail to rejects the spot market equilibrium. Neither locality illness nor locality illness interacted with the gender dummy are significant. Moreover, when community controls are included (column 6), the current health condition of the worker (i.e. own illness) affects negatively the wage rate as the model predicts[4].

6 Occupational Choice

In this section I incorporate in the analysis the worker's decision to become self-employed or wage worker. This is important for two reasons. Firstly, the estimation of the hourly earnings equation ignoring this decision may yield biased results if there are unobservable characteristics of the worker that affect simultaneously the occupational choice and the potential wage rate in each of the occupations. Secondly and most importantly, the model presented in section 2 gives unambiguous predictions about occupational choice determinants that should be empirically verified. Specifically, those workers with relatively less access to ex-post mechanisms to smooth consumption are expected to be more likely to become wage workers since they value more the insurance embedded in labor contracts.

Section 2.1 gives the determinants of the occupational choice. Equation (18) simply indicates that the worker chooses to be a wage worker if the indirect utility function in this occupation is higher than the indirect utility function of being self-employed. Taking a log linear approximation of (18) I obtain the following latent variable equation.

[4] In the regression for self-employed workers the prevalence of illness is computed including only self-employed. However, results are robust to including all workers in the locality

$$G^{*} = log(f'_{h0}) + \beta_1 \ male + \beta_2 \ sick + \beta_3 \ psick + \beta_4 \ psick * male + \beta_5 \ log(y) + \beta_6 \ log(\bar{b}) + \eta \quad (24)$$

$$G = \begin{cases} 1 & \text{if} \quad G^{*} \geq 0 \\ 0 & \text{if} \quad G^{*} < 0 \end{cases} \quad (25)$$

The marginal product is again approximated with education and potential experience as in (23). The variable *sick* is introduced to capture the difference between the marginal product when the worker is sick and when he is healthy. It does not have a direct effect on the occupational choice. As before, *psick* is the average illness among wage workers in the locality. It is used by firms to determine the wage rate in contracts. The interaction with the gender dummy captures the fact that $\bar{q}$ in (18) differs by gender. The individual probability of getting sick q_i and preferences ψ are not observed and so they are contained in the error term.

There are two observable or at least partially observable variables suggested by the model that affect the occupational choice but not the hourly earnings equation. These are non-labor income (y) and the availability of ex-post mechanisms to smooth consumption ($\bar{b}$).

With respect to non-labor income, I only use inheritance for being the one excludable from the hourly earnings equation. The model predicts that workers with higher non-labor income are more likely to become self-employed because they can use it as a safety net in times of meager labor income. So, wealthy workers are more willing to choose a risky activity[5]. Another reason, not included in the model but recognized in the literature, is investment indivisibilities. Becoming self-employed usually involves buying machinery or making another type of lump sum investment that requires the wealth of the worker to be above a certain threshold. Either for self-insurance or for investment indivisibilities motive, a positive income shock such as inheritance is likely to nudge the worker to switch from being an employee to a self-employed worker.

As the model predicts, another factor that increases the probability of becoming self-employed is the availability of ex-post mechanisms to smooth consumption. In developing countries, a very important one is insurance networks. Consumption fluctuations are mitigated with the use of loans and gifts from friends and relatives. Although it is impossible to know how many people the worker can ask for money when needed, the MxFLS survey provides information on whether the worker has a relative in the US. We can expect that households connected with residents in the US use

[5]This is true if the utility of the worker exhibits decreasing absolute risk aversion

these contacts to smooth consumption. I show evidence of this in table 9.

Table 9, column 1, shows the regression of remittances, computed as a binary variable, on illness, on whether any member of the household has a relative in the US and on the interaction of the two. The regression includes time-locality fixed effects to compare workers in the same community at the same year. Results show that workers with relatives in the US are more likely to receive transfers in the event of illness confirming that this is a mechanism to smooth consumption. Columns 2 of the same table shows the same specification with log total expenditure excluding health expenditure as the dependent variable. Consumption does not decrease in the event of illness showing that, on average, workers are insured against this event. Finally, column 3 shows a regression with health expenditure as a dependent variable.

Back to hourly earnings, equations (23) and (24) constitutes a model with endogenous selection. The potential bias from estimating (23) without considering the occupational choice, as I did in the previous section, comes from the possible correlation between the error terms in the two equations ($corr(\eta, \mu) \neq 0$). An element that may induce this correlation is the unobservable part of the worker's productivity not captured by education and potential experience. It is important to emphasize that the unobserved individual probability of getting sick q_i cannot be responsible for the correlation between the error terms μ and η. It is a determinant of the occupational choice but it is not in the error term of the hourly earnings equation. This variable is unknown to the employer so it is not internalized in labor contracts.

In table 10, columns 1 and 2, I estimate simultaneously equation (23) and (24) for wage workers using a maximum likelihood approach assuming that the error terms are jointly normally distributed. As the model predicts, inheritance and having a relative in the US negatively affect the probability of being a wage worker. With respect to the wage equation, results are almost unaffected by making occupational choice endogenous. This is verified with the likelihood ratio test. This test cannot reject the null hypothesis that the selection equation and the wage equation are independent. Columns 3 and 4 include locality characteristics among regressors with no significant impact on results.

In the last two columns of table 10, I relax the parametric assumption of the errors and estimate the model semiparametrically. The method consists of two steps. In the first step, I estimate the selection equation (24) with a flexible functional form.

$$G = \Phi(X\beta) + \gamma_1\Phi^2(X\beta) + \gamma_2\Phi^3(X\beta) + ... + \gamma_{j-1}\Phi^1(X\beta) + u \tag{26}$$

I select the number of terms in equation (26) using generalized cross validation. $\Phi(.)$ is the normal distribution, so if only one term is selected the model reduces to a standard probit model. In this way, the functional form is selected such that a normality test is nested in it.

In the second step I estimate the wage equation (23) using a control function approach that consists of including the single index $X\beta$ of the first stage in a flexible form. In particular I use a polynomial approximation (see Vella [1998])

The last columns of table 10 shows the results of estimating the model semiparametrically. Cross validation indicates that the best fit of (26) is with two terms. The parameter γ_1 is significantly different from zero rejecting normality of unobservables in the selection equation. Nonetheless, the wage equation shows almost identical results compared to the maximum likelihood approach. The control function is statistically significant but it does not impact the estimations.

7 Other results

7.1 Endogeneity of illness

The causality between illness and wages may go in the opposite direction introducing bias to regression (23). That is, very low wages may affect the intake of food debilitating the immune system and increasing illness. Nonetheless, this relation is expected to be highly non-linear. For poor people, an increase of wages is likely to generate an improvement on nutrition and consequently a reduction on illness. But, after certain threshold is reached, an increase in wages is likely to have no further improvement on nutrition.

Figures 2 and 3 show the density of the body mass index (BMI) for wage workers and self-employed workers respectively. Only a very small fraction of them are undernourished suggesting that reverse causality is probably not important for the population of interest. Nonetheless, in table 11 I run the wage equation dropping 10% of the sample with lower wages. In case of reverse causality these workers are the most likely to be responsible for it. Since, I am truncating the sample based on the dependent variables, instead of OLS I use truncated regression assuming that the error term is normally distributed.

Results in table 11 shows that conclusions are unaltered. Locality illness and locality illness interacted with gender continues to explain hourly wages for wage workers but not for self-employed workers.

Another potential source of endogeneity is the possibility that illness is correlated with health endowment and other unobservable determinants of the worker productivity. For example, workers who are currently ill may be less energetic all the time which introduces correlation between the error term μ and *illness* in (23). Nonetheless, for workers in the implicit contract market these elements are determinants of wages only if they are observable by the employer. In section 2.1 I used the model to show that the individual probability of getting sick does not affect wages since it is unknown for the employer.

In order to control for the part of the productivity that is unobservable for the econometrician but not for the employer, I estimate equation (23) using individual fixed effects. With this specification I can still identify β_2 (and only β_2). If the unobservable elements in the error term are not significantly correlated with illness, then the estimated value of this coefficient should be similar in the fixed effect and in the OLS specification.

Table 12 shows the result of comparing the estimated impact of illness on wages using OLS and individual fixed effects. The Hausman test fails to reject the hypothesis that they are equal suggesting that endogeneity is not a concern. This is true for both wage workers and self-employed workers, although in this later case the precision of the estimation with fixed effect is low as a consequence of the reduction in the sample size.

7.2 Migration and firm location

In regression (23) the probability of getting sick, which is computed as the prevalence of illness in the locality, may be correlated with other characteristics in the community affecting wages.

Roback (1982) shows that with perfect mobility of capital and labor, wages in local labor markets are determined by the level of amenities in the community. Some amenities enter the utility function (e.g. good climate conditions) generating immigration, increasing the supply of workers and consequently decreasing wages. On the other hand, other amenities enter the production function (e.g. a river to drain residuals) attracting firms, increasing the capital per worker ratio and hence increasing wages.

A concern in regression (23) is that these amenities may be correlated with the prevalence of illness in the community. For example, in localities where floods occur frequently, the prevalence of illness may be relatively high but outmigration may also be high.

However, if these unobserved locality characteristics have the same impact on all workers in the community, then they cannot explain the difference between wage workers and self-employed workers found in table 8. Nonetheless, it is possible that some amenities affect these two groups of workers differentially since they are likely to perform different tasks (e.g. the tasks of one group may be more physical). For this reason I perform the following robustness check.

Formally, the error term in (23) can be decomposed into locality characteristics or 'amenities' S and an iid shock ϵ ($\mu = S + \epsilon$). As explained above, S may affect wages through migration or firm location and also be correlated with *psick*.

$$log(w) = log(f'_L) + \beta_1 \; male + \beta_2 \; sick + \beta_3 \; psick + \beta_4 \; psick * male + S + \epsilon \qquad (27)$$

The solution is the estimation of (27) with community fixed effects. It eliminates S and the potential endogeneity for omitting this variable. However, the cost of doing this is that β_3 is not identified anymore. But, this methodology still allows the estimation of β_4 which is also a key parameter to identify an implicit contract equilibrium from a spot market one.

In table 13, I estimate equation (27) including locality fixed effects that eliminate S and *psick*. The results confirm that hourly earnings of wage workers behave consistently with the implicit contract hypothesis but not hourly earnings of self-employed workers.

7.3 Firm size

This section introduces heterogeneity in firms. In particular, how the size of the establishment affects labor contracts.

We can expect that in large firms, the absence of a worker has a minor impact on production. Some of the tasks assigned to him can be temporarily performed by other workers within the establishment. But, in small firms, each worker represents an important part of the workforce. His absence may significantly hurt the production process.

I modify the model presented in section 2 by making the productivity of the worker when sick increasing in firm size. With this variation and the maintained hypothesis that employers are risk neutral, the value of the worker for the firm becomes:

$$V = (1-q)\ \left[f'_{h0}h_0 - E_0\right] + q\ \left[f'_{h1}g(z)h_1 - E_1\right] = 0 \tag{28}$$

The difference between (28) and the original equation in model (1)-(5) is the introduction of the increasing function $g(z)$. When the worker is sick he produces per hour the normalized productivity f'_{h1} adjusted by the firm size z. Hence, wages becomes positively associated with firm size, consistent with the empirical evidence in the literature (Oi and Idson [1999]), and also positively associated with the interaction of the probability of getting sick and firm size as (29) and (30) show.

$$\frac{\partial ln(w)}{\partial z} = q\gamma g'(z) > 0 \tag{29}$$

$$\frac{\partial^2 ln(w)}{\partial z \partial q} = \gamma g'(z) > 0 \tag{30}$$

The intuition in equation (29) is simple. Large firms suffer less the absence of a worker and so the contracts they offer have higher wages. In equation (30) the intuition is similar. As argued above, the probability of getting sick affects the future average productivity of the worker and so his wage rate. However, the output of large firms is less sensitive to these shocks.

In table 14, I estimate equation (23) including firm size in levels and interacted with illness and with the prevalence of illness in the community. Results are consistent with the theory, wages increases with firm size and with the interaction of firm size with the locality illness.

Another possible effect of firm size is on the risk aversion of employers. The assumption about risk neutral firms is incorrect if entrepreneurs care not only about the level but also about the volatility of profits. This is likely to be true for owners of small firms. They may be less wealthy and may have less access to credit markets than owners of large firms. If this is the case, both employers and employees in small firms share part of the risk. Thus, wages should fluctuate with productivity shocks. In table 14, the interaction of firm size with *sick* is statistically insignificantly suggesting that even small firms fully insure workers.

8 Implicit contracts and the gender wage gap

In previous sections I showed evidence that, on average, wage workers are in contractual arrangements that insure them against idiosyncratic productivity shocks. Additionally, I showed that these type of contracts are responsible for part of the gender wage gap. But, how important are labor contracts in explaining this gap? This is the question I address in this section.

In the first column of table 15, I estimate the wage equation including locality fixed effects. Assuming that the return to all the characteristics is the same for men and women, the *male* dummy variable simply measures the average gender wage gap. The second column has exactly the same specification but including the gender dummy variable interacted with the prevalence of illness in the locality. This interaction captures the effect of the insurance component of labor contracts on the gender gap. So, in this regression the *male* dummy variable is the residual gender wage gap after purging the effect of labor contracts. In other words, it is what the gender gap would be if employers did not adjust women's contracts by their additional probability of missing days at work due to own illness or illness of other members in the family.

From column i to column ii, the coefficient of the *male* variable decreases from 0.0824 to 0.0148 indicating that the existence of labor contracts that insure workers against productivity shocks are responsible for women wages being approximately 7% lower.

In column iii and iv, I repeat the analysis but considering endogenous the occupational decision of workers. The maximum likelihood specification includes locality fixed effects and assumes normality of the errors. Because this is a non-linear model, there may be concerns about a potential bias from the inclusion of fixed effects (i.e. incidental parameter problem). However, the average number of observations per locality is 57 which is sufficiently large to ignore the problem (Heckman [1979], Green [2003]). Not surprisingly, results in columns iii and iv are almost identical to those in column i and ii indicating once again that selection into occupations is not an important source of bias.

Assuming that the return to all characteristics is the same for men and women is probably inadequate and may bias the estimator of the gender wage gap. For example, the return to potential experience may be different for women taken into account that they tend move in and out of the labor force more frequently than men. In columns v of table 15, I estimate the model interacting all variables but the fixed effects as equation (31) shows.

$$ln(w) = X\beta^f + male * X\beta^I + \alpha_j + \mu \tag{31}$$

The estimation of (31) is numerically equivalent to estimating the wage equation separately for men and women (equations (32) and (33)) imposing the condition that the locality fixed effects α_j are the same in both regressions[6], where the equivalence of parameters is $\beta^m = \beta^f + \beta^I$.

$$ln(w^m) = X^m\beta^m + \alpha_j + \mu^m \quad \text{(men)} \tag{32}$$

$$ln(w^f) = X^f\beta^f + \alpha_j + \mu^f \quad \text{(women)} \tag{33}$$

With the estimation of equations (32) and (33) (or the equivalent in terms of equation (31)), I use the Blinder(1971) and Oaxaca(1971) methodology to estimate the gender gap reduction derived from labor contracts.

$$(\overline{ln(w^m)} - \overline{ln(w^f)})^* = \overbrace{(\bar{X}^m - \bar{X}^f)\beta^m}^{characteristics} + \overbrace{\bar{X}^f(\beta^m - \beta^f)}^{returns+intercept} \tag{34}$$

As equation (34) shows, the methodology consists of decomposing the difference in wages between men and women in two parts. One is the difference in characteristics and the other is the difference in the return to these characteristics and the intercept. This last component is the gender gap for the average female worker compared to a male worker with the same observable characteristics. Since I estimate the model with fixed effects, the right hand side of (34) is not the raw wage gap but the wage gap controlled for locality fixed effects (i.e. the one that would result from regressing log wages on a gender dummy and locality fixed effects)

In order to compute the gender wage gap reduction derived from labor contracts, I estimate the gender gap using the Blinder-Oaxaca decomposition to the model specified with and without locality illness (i.e. columns v and vi in table 15). The difference is the gender reduction due to labor contracts[7].

[6]The condition that the locality fixed effects is the same in both equation is necessary for identification. It would be impossible to recover the coefficient on locality illness interacted with the gender dummy in column vi of table 15 if the locality fixed effects varied independently for men and women. Conceptually, this condition assumes that all elements in the community (e.g. infrastructure, weather shocks, prices, etc) affects men and women in the same magnitude.

[7]The parameters γ^m and γ^f cannot be identified separately in (36). Only the difference is identified and the value is the estimated parameter of the gender dummy interacted with locality illness in table 15

$$\overline{ln(w^m)} - \overline{ln(w^f)} = (\bar{X}^m - \bar{X}^f)\beta_o^m + \overbrace{\bar{X}^f(\beta_o^m - \beta_o^f)}^{A} \tag{35}$$

$$\overline{ln(w^m)} - \overline{ln(w^f)} = (\bar{X}^m - \bar{X}^f)\beta^m + \overbrace{\bar{X}^f(\beta^m - \beta^f)}^{B} + loc\ illness(\gamma^m - \gamma^f) \tag{36}$$

$$\tag{37}$$

A: Gender wage gap not controlled for the impact of locality illness

B: Gender wage gap controlled for the impact of locality illness

A-B = Gender gap attributed to labor contracts

Table 16 shows a summary with the results of the gender gap reduction attributed to the existence of labor contracts. The Blinder-Oaxaca decomposition yields very similar results to the previous estimations. The insurance component of labor contracts lowers women wages approximately 7%.

9 Conclusions

In this paper I explored the role of labor contracts as a mechanism of insurance against idiosyncratic productivity shocks. All results indicate that labor outcomes of wage workers in Mexico are consistent with the theory of implicit contracts and contradicts the spot or Walrasian market hypothesis.

Earnings are invariant to illness shocks suggesting the presence of insurance against productivity shocks. The response of hours worked to illness is higher for wage workers than for self-employed workers consistent with the idea that the insurance component in labor contract reduces the incentive to work in the event of illness. Finally, hourly earnings decreases with the probability of getting ill but not with illness indicating that workers compensations are disentangled from short-run productivity fluctuations and respond to the average productivity across time.

I also explored occupational choices of workers. The predictions of the model indicates that workers who get sick relatively more often and have worse access to mechanisms to smooth consumption are more likely to choose the contract market (i.e. to become a wage worker). Empirically, I showed that remittances from relatives in the US are used to smooth consumption. Consistent with this, results show that having a relative in the US is negatively associated with the worker becoming a wage worker.

Another prediction of the model is that the insurance in labor contracts explain part of the gender gap if women miss more days at work than men. I show evidence that only women stay at home to take care of other members in the family when they get sick. Employers internalize this additional responsibilities of women offering them contracts with lower wages in comparison to the contracts they offer to men. Different methodologies, including a modified Blinder-Oaxaca decomposition, yield the same result: implicit contracts are responsible for women having 6% to 7% lower wages.

References

C. Azariadis. Implicit contracts and underemployment equilibria. *The Journal of Political Economy*, 83(6):pp. 1183–1202, Dec. 1975.

M. N. Baily. Wages and employment under uncertain demand. *The Review of Economic Studies*, 41(1):pp. 37–50, Jan. 1974.

P. Bardhan. *Land, Labor and Rural Poverty*, chapter 4. Columbia University Press, 1984.

P. Beaudry and J. DiNardo. The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data. *The Journal of Political Economy*, 99(4):pp. 665–688, Aug. 1991.

G. S. Becker. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 3(1):pp. S33–S58, 1985.

J. P. T. Ethan Ligon and T. Worrall. Informal insurance arrangements with limited commitment: Theory and evidence from village economies. *Review of Economic Studies*, 69(1):pp. 209–244, 2002.

W. Green. Fixed effects and bias due to the incidental parameters problem in the tobin model. *Manuscript. Department of Economics, Stern School of Business, New York University*, 2003.

J. J. Heckman. The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process. *Structural Analysis of Discrete Data with Econometric Applications*, 47(1):pp. 153–162, Jan. 1979.

B. Holmstrom. Contractual models of the labor market. *The American Economic Review. Papers and Proceedings of the Ninety-Third Annual Meeting of the American Economic Association*, 71 (2):pp. 308–313, May 1981.

A. Kochar. Smoothing consumption by smoothing income: Hours-of-work responses to idiosyncratic agricultural shocks in rural india. *Review of Economics and Statistics*, 81:pp. 50–61, Feb. 1999.

A. Mukherjee and D. Ray. labor tying. *Journal of Development Economics*, 47, 1995.

W. Oi and T. Idson. Firm size and wages. In *Handbook of Labor Economics*, volume 3B, chapter 33. Orley Ashenfelter and David Card, 1999.

M. R. Rosenzweig and H. Biswanger. Wealth, weather risk and the composition and profitability of agricultural investments. *Economic Journal*, Jan. 1993.

M. R. Rosenzweig and O. Stark. Consumption smoothing, migration and marriage: Evidence from rural india. *The Journal of Political Economy*, 97(4):pp. 905–926, Aug. 1989.

M. R. Rosenzweig and K. Wolpin. Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investment in bullocks in india. *The Journal of Political Economy*, 101(2):pp. 223–244, Apr. 1993.

L. Rubalcava and G. Teruel. *The Mexican Family Life Survey: First Wave*, Feb. 2007a. www.mxfls.uia.mx.

L. Rubalcava and G. Teruel. *Users Guide for The Mexican Family Life Survey: Second Round*, Oct. 2007b. www.mxfls.uia.mx.

J. Strauss and D. Thomas. Health, nutrition, and economic development. *Journal of Economic Literature*, 36(2):pp. 766–817, Jun. 1998.

J. Strauss and D. Thomas. Health over the life course. In *Handbook of Development Economics*, volume 4, chapter 54. T. Paul Schultz and John Strauss editors, 2008.

J. Thomas and T. Worrall. Self-enforcing wage contracts. *The Review of Economic Studies*, 55(4): pp. 541–553, Oct. 1988.

R. Townsend. Risk and insurance in village india. *Econometrica*, 62:pp. 539–592, 1994.

C. Udry. Risk and insurance in a rural credit market: An empirical investigation in northern nigeria. *Review of Economic Studies*, 63(3)(no. 208):pp. 495–526, 1994.

F. Vella. Estimating models with sample selection bias: A survey. *The Journal of Human Resources*, 33(1):pp. 127–169, 1998.

Table 1:
Table 1: Variable identification and summary statistics

Variable	Definition	2002		2005	
		Mean	s.d.	Mean	s.d.
Illness					
Adult illness	In the last 4 weeks, she stop doing any of her daily activities or work, due to any illness?	0.088	0.283	0.071	0.258
Child illness	In the last 4 weeks, at least one day the boy/girl was inactive because of any illness	0.079	0.270	0.044	0.205
Time allocation(+)					
Hours worked	During the past week, total number of hours worked	22.178	25.527	19.954	25.283
Care	During the past week, number of hours spent o Taking care of an elderly or sick person and/or any children	10.645	19.827	8.019	17.402
Leisure	During the past week, number of hours spent o Watching TV o Reading	12.949	10.950	12.997	10.780
Housework	During the past week, number of hours spent o Cooking/preparing food o Washing clothes and/or cleaning the house	14.471	16.179	12.900	15.123
Firewood, water and agricultural activ.	During the past week, number of hours spent o Collecting firewood o Collecting water o Participating in agricultural activities	2.440	9.394	1.970	8.233

(+) 21 to 55 years old

Table 2: Descriptive statistics

Illness

2002			2005		
Men	Women	Children	Men	Women	Children
0.055	0.111	0.078	0.052	0.086	0.044
(0.227)	*(0.314)*	*(0.268)*	*(0.221)*	*(0.280)*	*(0.204)*

Composition of the population (21 to 55 years old)

	2002		2005	
	Men	Women	Men	Women
Wage workers	63.92	25.48	63.78	24.73
Self-employed	25.68	11.33	21.39	9.43
without compensatioı	3.06	3.57	4.68	3.47
non-labor force	7.34	59.63	10.15	62.38

Table 3: Descriptive statistics

Fraction in agriculture

	2002		2005	
	Men	Women	Men	Women
Wage workers	0.21	0.05	0.22	0.04
Self-employed	0.35	0.03	0.36	0.04

hours worked

	2002		2005	
	Men	Women	Men	Women
Wage workers	45.8	39.0	46.6	39.5
Self-employed	44.7	30.3	44.2	31.8

locality size

	2002	2005
more than 100,000	42.77	41.03
15,000 - 100,000	9.24	10.27
2,500 - 15,000	10.52	12.23
less than 2,500	37.48	36.47

Table 4: Impact of illness on hours worked (zeros included)

method: first difference, locality fixed effects

Dep Var: hours worked

	Wage workers			Self-employed workers		
VARIABLES	i	ii	iii	iv	v	vi
own illness	-6.951***	-6.935***	-6.932***	0.0229	0.172	0.138
	(1.684)	(1.684)	(1.686)	(3.576)	(3.582)	(3.616)
children illness		-1.584	-1.573		-2.741	-3.145
		(2.171)	(2.173)		(5.427)	(5.460)
spouse illness			-1.103			-1.189
			(1.674)			(4.084)
no child		-1.419	-1.486		4.160	3.913
		(1.389)	(1.399)		(3.647)	(3.667)
no spouse			0.435			2.809
			(1.399)			(3.437)
non-labor income	4.764	5.685	6.106	16.95	17.35	16.49
	(21.11)	(21.12)	(21.19)	(20.41)	(20.43)	(20.50)
assets	1.288*	1.320*	1.316*	0.713	0.687	0.771
	(0.724)	(0.724)	(0.725)	(2.755)	(2.758)	(2.798)
Constant	0.378	0.610	0.530	1.583	0.828	0.303
	(0.516)	(0.581)	(0.632)	(1.142)	(1.304)	(1.456)
Observations	1618	1618	1618	463	463	463
R-squared	0.013	0.014	0.015	0.002	0.007	0.009
Number of localities	143	143	143	126	126	126
mean hours 2002	44.34	44.34	44.34	41.59	41.59	41.59

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Figure 1: Person taking care of the child

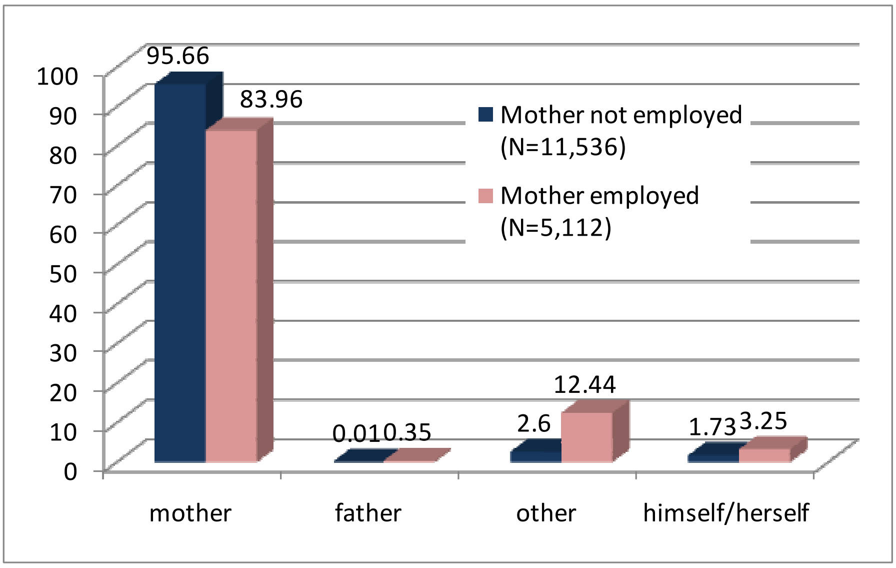

Table 5: Impact of illness on earnings (wage workers only)

method: first difference, locality fixed effects
Dep Var: log earnings

	Wage workers		
VARIABLES	i	ii	ii
own illness	-0.0137	-0.0138	-0.0101
	(0.0695)	(0.0695)	(0.0697)
children illness		-0.113	-0.116
		(0.0929)	(0.0930)
spouse illness			-0.0241
			(0.0705)
no child		-0.0209	-0.0292
		(0.0586)	(0.0590)
no spouse			0.0684
			(0.0593)
non-labor income	0.256	0.280	0.347
	(0.768)	(0.769)	(0.771)
assets	0.0307	0.0322	0.0324
	(0.0297)	(0.0297)	(0.0297)
Constant	0.292***	0.293***	0.281***
	(0.0215)	(0.0242)	(0.0264)
Observations	1189	1189	1189
R-squared	0.001	0.003	0.004
Number of localities	138	138	138

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Figure 2: BMI wage workers

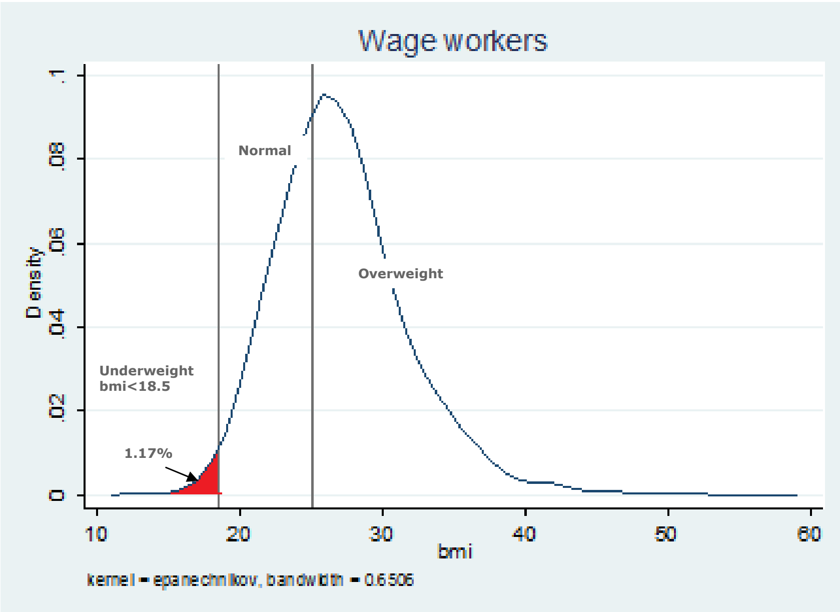

Table 6: Women's time allocation

method: first difference, locality fixed effects

	Hours worked			Taking care of others			Leisure at home			Home Chores		
VARIABLES	i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii
own illness	-2.027**	-1.959**	-1.850**	0.968	0.888	0.797	0.787	0.783	0.748	-0.642	-0.703	-0.747
	(0.909)	(0.910)	(0.908)	(1.269)	(1.266)	(1.266)	(0.535)	(0.535)	(0.535)	(0.802)	(0.802)	(0.803)
children illness		-2.392*	-2.470*		3.691*	3.730*		0.346	0.363		2.399*	2.410*
		(1.437)	(1.434)		(1.990)	(1.989)		(0.844)	(0.844)		(1.267)	(1.266)
spouse illness			-2.399			5.944***			1.466			2.621**
			(1.518)			(2.086)			(0.895)			(1.329)
no child		0.170	0.0993		5.743***	5.724***		1.120**	1.125**		1.086	1.055
		(0.966)	(0.964)		(1.325)	(1.324)		(0.567)	(0.567)		(0.849)	(0.849)
no spouse			3.043***			0.185			-0.443			0.458
			(0.755)			(1.047)			(0.445)			(0.666)
non-labor income	17.52	17.32	18.58*	19.68	19.14	18.27	-23.22***	-23.42***	-23.78***	-3.280	-3.306	-3.590
	(10.83)	(10.83)	(10.81)	(14.53)	(14.48)	(14.48)	(6.310)	(6.309)	(6.311)	(9.468)	(9.465)	(9.466)
assets	-0.222	-0.197	-0.204	0.102	0.0397	0.0411	0.0533	0.0452	0.0464	-0.0255	-0.0561	-0.0559
	(0.280)	(0.280)	(0.280)	(0.371)	(0.370)	(0.370)	(0.164)	(0.165)	(0.165)	(0.245)	(0.245)	(0.245)
Constant	0.831**	0.734*	-0.524	-8.728***	-9.653***	-9.733***	-0.0621	-0.248	-0.0676	-2.614***	-2.735***	-2.925***
	(0.358)	(0.398)	(0.505)	(0.496)	(0.551)	(0.701)	(0.210)	(0.234)	(0.297)	(0.315)	(0.351)	(0.445)
Observations	3723	3723	3723	3387	3387	3387	3649	3649	3649	3654	3654	3654
R-squared	0.002	0.003	0.008	0.001	0.008	0.010	0.004	0.006	0.007	0.000	0.002	0.003
Number of nloc	150	150	150	150	150	150	150	150	150	150	150	150
mean dep var in 2002	11.43	11.43	11.43	22.02	22.02	22.02	12.92	12.92	12.92	26.26	26.26	26.26

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Men's time allocation

method: first difference, locality fixed effects

	Hours worked			Taking care of others			Leisure at home			Home Chores		
VARIABLES	i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii
own illness	-7.365***	-7.358***	-7.338***	1.538*	1.529*	1.497*	1.707**	1.707**	1.757**	0.0105	0.00987	-0.0369
	(1.840)	(1.840)	(1.840)	(0.816)	(0.816)	(0.818)	(0.846)	(0.846)	(0.848)	(0.426)	(0.426)	(0.426)
children illness		-1.973	-2.138		1.370	1.330		0.0253	0.0110		0.130	0.144
		(2.307)	(2.305)		(1.022)	(1.023)		(1.056)	(1.057)		(0.536)	(0.535)
spouse illness			2.294			0.767			-0.0785			0.0993
			(1.532)			(0.682)			(0.705)			(0.354)
no child		-0.717	-0.0393		0.278	0.337		1.151*	1.313*		-0.439	-0.577
		(1.499)	(1.519)		(0.670)	(0.679)		(0.691)	(0.700)		(0.349)	(0.353)
no spouse			-6.275**			-0.538			-1.776			1.579***
			(2.484)			(1.126)			(1.161)			(0.586)
non-labor income	0.643	0.952	-0.165	-0.439	-0.569	-0.719	-1.189	-1.624	-1.848	0.457	0.620	0.813
	(16.43)	(16.44)	(16.42)	(7.235)	(7.240)	(7.242)	(7.479)	(7.482)	(7.484)	(3.750)	(3.753)	(3.749)
assets	1.467	1.510	1.512	-0.0716	-0.0965	-0.107	-0.0693	-0.0925	-0.0762	0.663**	0.670***	0.655**
	(1.132)	(1.133)	(1.132)	(0.499)	(0.499)	(0.500)	(0.516)	(0.516)	(0.517)	(0.259)	(0.259)	(0.259)
Constant	-0.315	-0.245	0.0127	-1.413***	-1.421***	-1.399***	-0.416	-0.617**	-0.545*	-0.153	-0.0723	-0.137
	(0.550)	(0.616)	(0.623)	(0.244)	(0.273)	(0.277)	(0.253)	(0.283)	(0.287)	(0.127)	(0.142)	(0.144)
Observations	2231	2231	2231	2196	2196	2196	2187	2187	2187	2174	2174	2174
R-squared	0.008	0.009	0.013	0.002	0.003	0.003	0.002	0.003	0.005	0.003	0.004	0.008
Number of localities	150	150	150	150	150	150	150	150	150	150	150	150
mean dep var in 2002	43.99	43.99	43.99	3.722	3.722	3.722	11.79	11.79	11.79	1.800	1.800	1.800

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 8: Baseline results

Dep Var: log hourly earnings

VARIABLES	wage worker i	self-employed ii	wage worker iii	self-employed iv	wage worker v	self-employed vi
secondary	0.466***	0.554***	0.465***	0.550***	0.383***	0.406***
	(0.0238)	(0.0666)	(0.0238)	(0.0669)	(0.0239)	(0.0658)
college	1.172***	1.280***	1.166***	1.274***	1.045***	1.054***
	(0.0342)	(0.0987)	(0.0343)	(0.0992)	(0.0357)	(0.0990)
potential experience	0.0412***	0.0292**	0.0409***	0.0301**	0.0404***	0.0286**
	(0.00353)	(0.0120)	(0.00355)	(0.0121)	(0.00357)	(0.0122)
potential experience sq/100	-0.0695***	-0.0414	-0.0688***	-0.0433*	-0.0700***	-0.0401
	(0.00838)	(0.0253)	(0.00841)	(0.0255)	(0.00838)	(0.0255)
male	0.0336*	0.203***	-0.0624	0.264***	-0.000937	0.327***
	(0.0196)	(0.0575)	(0.0417)	(0.0855)	(0.0410)	(0.0831)
year 2005	0.191***	0.306***	0.192***	0.310***	0.208***	0.282***
	(0.0169)	(0.0547)	(0.0170)	(0.0550)	(0.0169)	(0.0543)
locality illness			-1.413**	0.485	-1.492***	0.445
			(0.581)	(0.725)	(0.567)	(0.715)
male * locality illness			1.830**	-0.860	1.328*	-0.657
			(0.710)	(0.861)	(0.685)	(0.852)
own illness			0.000119	-0.169	0.00929	-0.220**
			(0.0446)	(0.119)	(0.0448)	(0.104)
Constant	1.914***	1.722***	1.990***	1.688***	2.162***	2.003***
	(0.0419)	(0.147)	(0.0514)	(0.159)	(0.0534)	(0.163)
locality characteristics	N	N	N	N	Y	Y
Observations	8602	1869	8523	1846	8245	1812
R-squared	0.170	0.132	0.170	0.134	0.195	0.175

Clustered (in worker) standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

locality characteristics include size and fraction of labor force in agriculture. A broader set of controls yields similar res

Table 9: Relative in the US as mechanism to smooth consumption

method: locality-year fixed effect

VARIABLES	remittances i	log cons ii	health expend. ii
log permanent labor income++	-0.0362***	0.747***	0.182***
	(0.00941)	(0.0371)	(0.0235)
relative in US	0.0278***	0.0997***	0.0487***
	(0.00523)	(0.0208)	(0.0132)
relative in US * hhold head sick	0.0304*	-0.0314	0.00293
	(0.0179)	(0.0711)	(0.0450)
hhold head sick	0.00606	0.0746	0.111***
	(0.0135)	(0.0537)	(0.0340)
log household size	-0.00510	0.184***	0.0399***
	(0.00523)	(0.0208)	(0.0132)
Constant	0.322***	1.965***	-1.199***
	(0.0741)	(0.292)	(0.185)
Observations	7084	7173	7176
Number of year-localities	300	300	300

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
++ computed as the prediction of log labor income on education, experience and gend

Table 10: Endogenous occupational choice

Dep. Var.: log hourly earnings	ML estimation		ML estimation		Semiparametric	
	wage equation+	Selection equation	wage equation++	Selection equation	wage equation	Selection equation
VARIABLES	i	ii	iii	iv	v	vi
secondary	0.474***	0.143***	0.386***	0.108***	0.555***	0.195***
	(0.0230)	(0.0372)	(0.0233)	(0.0385)	(0.0381)	(0.0496)
college	1.173***	0.155***	1.049***	0.102*	1.284***	0.266***
	(0.0308)	(0.0531)	(0.0314)	(0.0551)	(0.0513)	(0.0775)
potential experience	0.0401***	-0.0433***	0.0405***	-0.0433***	0.00967	-0.0687***
	(0.00372)	(0.00657)	(0.00368)	(0.00658)	(0.0117)	(0.0139)
potential experience sq/100	-0.0673***	0.0487***	-0.0705***	0.0480***	-0.0270	0.088***
	(0.00865)	(0.0146)	(0.00854)	(0.0147)	(0.0172)	(0.0259)
male	-0.0437	0.0344	0.00181	0.0508	-0.0184	0.058
	(0.0362)	(0.0588)	(0.0359)	(0.0588)	(0.0375)	(0.0778)
year 2005	0.191***	0.408***	0.208***	0.414***	0.429***	0.557***
	(0.0216)	(0.0314)	(0.0216)	(0.0314)	(0.0910)	(0.0633)
locality illness	-1.317***	-0.272	-1.481***	-0.420	-1.448***	-0.3184026
	(0.482)	(0.763)	(0.477)	(0.760)	(0.484)	(0.9941)
male * locality illness	1.576***	0.774	1.289**	0.769	2.057***	1.125418
	(0.578)	(0.919)	(0.571)	(0.913)	(0.605)	(1.209)
own illness	0.00349	-0.150**	0.00658	-0.152**	-0.0730	-0.177**
	(0.0427)	(0.0652)	(0.0421)	(0.0653)	(0.0515)	(0.0815)
inheritance/10,000		-0.0145*		-0.0146*		-0.017*
		(0.00852)		(0.00858)		(0.0092)
relative in US		-0.0951***		-0.0881***		-0.113***
		(0.0328)		(0.0332)		(0.0434)
xb					-0.386**	
					(0.180)	
xb2					-0.0185	
					(0.0354)	
gamma						-0.049***
						(0.0122)
Constant	1.978***	1.317***	2.155***	1.408***	2.694***	1.712***
	(0.0531)	(0.0892)	(0.0531)	(0.0920)	(0.271)	(0.203)
locality characteristics	N	N	Y	Y	N	N
Observations	9720	9720	9720	9720	9720	9720

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
+ LR test of indep. eqns. (rho = 0): chi2(1) = 0.16 Prob > chi2 = 0.6868
++ LR test of indep. eqns. (rho = 0): chi2(1) = 0.10 Prob > chi2 = 0.7525

Table 11: Truncated regression

method: truncated regression; lower 10% of sample dropped
Dep Var: log hourly earnings

VARIABLES	wage worker i	self-employed ii	wage worker i	self-employed ii
secondary	0.610***	0.579***	0.549***	0.437***
	(0.0372)	(0.0861)	(0.0378)	(0.0808)
college	1.482***	1.316***	1.385***	1.110***
	(0.0467)	(0.118)	(0.0474)	(0.112)
potential experience	0.0523***	0.0354**	0.0513***	0.0366**
	(0.00491)	(0.0149)	(0.00489)	(0.0149)
potential experience sq/100	-0.0799***	-0.0569*	-0.0790***	-0.0588*
	(0.0121)	(0.0324)	(0.0120)	(0.0318)
male	-0.125**	0.183*	-0.0760	0.255**
	(0.0541)	(0.107)	(0.0563)	(0.102)
year 2005	0.175***	0.272***	0.187***	0.242***
	(0.0234)	(0.0690)	(0.0231)	(0.0659)
locality illness	-1.736**	1.134	-2.192***	1.229
	(0.735)	(0.844)	(0.784)	(0.790)
male * locality illness	3.189***	-1.307	2.871***	-1.268
	(0.899)	(1.025)	(0.948)	(0.982)
own illness	0.0462	-0.165	0.0658	-0.218
	(0.0602)	(0.162)	(0.0607)	(0.134)
Constant	1.589***	1.658***	1.759***	1.927***
	(0.0775)	(0.199)	(0.0781)	(0.201)
locality characteristics	N	N	Y	Y
Observations	7673	1666	7415	1634

Clustered (in worker) standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
locality characteristics include size and fraction of labor force in agriculture.

Table 12: Hausman test

wage workers

	FE	OLS	Difference	Hausman Test	
own illness	-0.08504	0.000119	-0.08516	chi2(1)	1.6
s.e.	0.078876	0.040947	0.067415	Prob>chi2	0.2065

self employed

	FE	OLS	Difference	Hausman Test	
own illness	0.170532	-0.1688	0.339332	chi2(1)	0.89
s.e.	0.374254	0.106168	0.358879	Prob>chi2	0.3444

Table 13: Locality fixed effect

method: locality fixed effect
Dep Var: log hourly earnings

VARIABLES	wage worker i	self-employed ii
secondary	0.344***	0.262***
	(0.0307)	(0.0725)
college	0.998***	0.892***
	(0.0468)	(0.115)
potential experience	0.0401***	0.0219*
	(0.00317)	(0.0126)
potential experience sq/100	-0.0712***	-0.0320
	(0.00725)	(0.0262)
male	0.0148	0.232***
	(0.0431)	(0.0602)
year 2005	0.220***	0.349***
	(0.0230)	(0.0671)
male * locality illness	1.272*	0.0727
	(0.708)	(0.581)
own illness	-0.000921	-0.172*
	(0.0427)	(0.0949)
Constant	1.984***	1.968***
	(0.0462)	(0.146)
Observations	8523	1846
R-squared	0.131	0.091
Number of localities	150	150

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Impact of firm size

Dep Var: log hourly earnings

	wage worker	
VARIABLES	i	ii
secondary	0.450***	0.368***
	(0.0252)	(0.0249)
college	1.133***	1.018***
	(0.0366)	(0.0373)
potential experience	0.0405***	0.0411***
	(0.00374)	(0.00369)
potential experience sq/100	-0.0679***	-0.0717***
	(0.00881)	(0.00863)
male	-0.0639	-0.0144
	(0.0436)	(0.0423)
year 2005	0.201***	0.218***
	(0.0177)	(0.0174)
locality illness	-2.183***	-2.592***
	(0.740)	(0.694)
male * locality illness	1.638*	1.779**
	(0.840)	(0.789)
own illness	0.00412	-0.00859
	(0.0733)	(0.0719)
log firm size	0.0266**	0.0194*
	(0.0108)	(0.0103)
log firm size * locality illness	0.318	0.424**
	(0.217)	(0.208)
log firm size * male * locality illness	0.0630	-0.130
	(0.193)	(0.186)
log firm size * own illness	0.00799	0.0121
	(0.0298)	(0.0293)
Constant	1.939***	2.125***
	(0.0580)	(0.0585)
locality characteristics	N	Y
Observations	7782	7782
R-squared	0.180	0.202

Clustered (in worker) standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

locality characteristics include size and fraction of labor force in agriculture.

Table 15: Gender gap reduction

Dep. Var.: log hourly earnings

	Fixed effects		ML selet + fixed effects		Fixed effects	
VARIABLES	i	ii	iii	iv	v	vi
secondary	0.344***	0.344***	0.346***	0.346***	0.484***	0.482***
	(0.0229)	(0.0229)	(0.0232)	(0.0231)	(0.0391)	(0.0391)
college	0.999***	0.998***	1.002***	1.001***	1.168***	1.164***
	(0.0319)	(0.0318)	(0.0321)	(0.0321)	(0.0507)	(0.0508)
potential experience	0.0401***	0.0401***	0.0409***	0.0409***	0.0462***	0.0462***
	(0.00343)	(0.00343)	(0.00356)	(0.00355)	(0.00579)	(0.00579)
potential experience sq/100	-0.0711***	-0.0712***	-0.0720***	-0.0722***	-0.0801***	-0.0803***
	(0.00818)	(0.00818)	(0.00825)	(0.00824)	(0.0144)	(0.0144)
male	0.0824***	0.0148	0.0804***	0.0150	0.323***	0.255***
	(0.0185)	(0.0354)	(0.0187)	(0.0354)	(0.0794)	(0.0857)
year 2005	0.219***	0.220***	0.212***	0.213***	0.193***	0.192***
	(0.0176)	(0.0176)	(0.0216)	(0.0214)	(0.0295)	(0.0295)
own illness	-0.00441	-0.000921	-0.00345	-0.000104	-0.00120	0.0150
	(0.0393)	(0.0393)	(0.0400)	(0.0400)	(0.0558)	(0.0563)
male * locality illness		1.272**		1.234**		1.211**
		(0.568)		(0.569)		(0.575)
male * secondary					-0.204***	-0.201***
					(0.0464)	(0.0464)
male * college					-0.250***	-0.245***
					(0.0610)	(0.0611)
male * potential experience					-0.00932	-0.00921
					(0.00716)	(0.00716)
male * potential experience sq/100					0.0139	0.0139
					(0.0175)	(0.0175)
male * year 2005					0.0385	0.0399
					(0.0363)	(0.0363)
male * own illness					-0.00538	-0.0309
					(0.0781)	(0.0791)
Constant	1.983***	1.984***	1.987***	1.987***	1.816***	1.820***
	(0.0408)	(0.0408)	(0.0447)	(0.0445)	(0.0645)	(0.0645)
Observations	8523	8523	8523	8523	8523	8523
R-squared	0.130	0.131			0.133	0.133
Number of localities	150	150	150	150	150	150

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 16: Gender gap reduction

	Fe	Heckman Fe	Blinder-Oaxaca
Gender gap	8.24	8.09	7.76
controlling for labor contracts	1.48	1.44	1.29
Gender gap due to labor contracts	6.76	6.65	6.47

Figure 3: BMI self-employed workers