Labor Risk Sharing

by

Lucas Manuelli

A.B., Princeton University (2012)

Submitted to the Department of Economics in Partial Fulfillment of the requirements for the Degree of

Masters of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

Signature of Author

Signature redacted

Department of Economics

Certified by

Signature redacted

Ivan Werning

Professor Of Economics

Thesis Supervisor

Accepted by

Signature redacted

Ricardo Caballero

Ford International Professor of Economics

Chairman, Committee For Graduate Students

Labor Risk Sharing

by

Lucas Manuelli

Submitted to the Department of Economics in Partial Fulfillment of the requirements for the Degree of Masters of Science in Economics

ABSTRACT:

In this paper we aim to test the extent of labor risk sharing exists in thai village economies. Specifically we test the null hypothesis of full risk sharing at the village level. We outline a simple planner's problem that motivates our empirical specification. Our empirical specification consists of two equations, a labor supply equation that determines how many hours you work conditional on participating in the labor market, and a selection equation which determines the probability of working positive hours. Our empirical specification allows for fixed effects that correspond to different Pareto weights for the agents. Our dataset, an unusually long panel survey spanning over 160 months conducted in 16 villages in Thailand, allows us to deal with these fixed effects. Our results lead us to reject the null of full risk sharing since non-labor income has a significant negative effect on participation. In most specifications it also has a significant but small negative effect on hours worked conditional on participation. In light of these results we reject the null of full risk sharing.

Thesis Supervisor: Ivan Werning

Title: Professor of Economics

1 Introduction

In this paper we aim to test the extent of labor risk sharing exists in thai village economies. Specifically we test the null hypothesis of full risk sharing at the village level. We outline a simple planner's problem that motivates our empirical specification. Our empirical specification consists of two equations, a labor supply equation that determines how many hours you work conditional on participating in the labor market, and a selection equation which determines the probability of working positive hours. Our empirical specification allows for fixed effects that correspond to different Pareto weights for the agents. Our dataset, an unusually long panel survey spanning over 160 months conducted in 16 villages in Thailand, allows us to deal with these fixed effects. The paper follows the general outline of the working paper of Bonhomme, Chiappori, Townsend and Yamada (2013) henceforth Chiappori et. al (2013) [1].

Our hypothesis of full labor risk-sharing is motivated by the fact that there seems to be significant amounts of consumption risk sharing in the same Thai villages. A recent paper by Chiapporti, Samphantharak, Schulhofer-Wohl and Townsend (2013) [2] is not able to reject the null of full risk-sharing in consumption among households in Thai villages. They also find significant heterogeneity in risk-aversion among households. One limitation of our current empirical formulation is that it cannot account for heterogeneous utility for leisure or work. While we do think that this may be an important feature of the data we leave it for future work.

The paper is organized as follows. In section 2 we present a model that motivates our empirical specification. In section 3 present our empirical specification and estimation strategy. Section 4 describes the data. Section 5 discusses the results and some robustness checks and section 6 concludes.

2 Model

In this section I outline a model that motivates our empirical specification in section 3. Our goal is to test null hypothesis of full risk sharing particularly in regards to labor supply and wage risk. With this goal we consider the following planner's problem. The planner solves

$$\max \sum_t \sum_i \alpha_i \beta^t U(c_{it}, h_{it})$$

subject to the period by period budget constraint given by

$$\sum_i c_{it} \leq \sum_i w_{it} h_{it} + K_t$$

The α_i is the Pareto weight for individual i. In the budget constraint K_t represents other terms that could be borrowing or saving by the planner. These won't matter for our purposes because we only care about the planner's first order conditions on consumption c and labor h. One important thing to note about the budget constraint is that we take the production technology available to the planner to be linear in individual i's wage times his hours. This is a simplification that allows us to bypass the village level production technology. Although this is certainly not realistic I believe that it serves as a good first approximation. Of course this also means that we ignore capital accumulation. In practice it is not completely unreasonable to suppose that for individuals who work for a paid wage their contribution to the village level budget constraint is $w_{it}h_{it}$, i.e. linear in their wage. Since we want to test the null hypothesis of full risk sharing in the model we abstract away from moral hazard and adverse selection although it would interesting to build these in a full structural model to study how they interact with labor risk sharing. The empirical specification will essentially come from the linear relations that correspond to the first order conditions of the planner's problem. In order for the first order conditions to correspond to linear regressions we make the following strong parametric assumption that the utility function separable in consumption and leisure. Thus we can write $U(c,h) = U(c) + \frac{(T-h)^{1-\gamma}}{1-\gamma}$ where c is consumption, h is labor supply and T is total hours available for leisure. With this parametric specification Now we can write down the first order conditions of the planner's problem. Since we assumed for simplicity that the utility function is separable between consumption and leisure we only need to look at the first order conditions for labor. The first order condition for hours is

$$\alpha_i L_{it}^{-\gamma} \geq \lambda_t w_{it}$$

with equality if $h_{it} > 0$ where $L_{it} = T - h_{it}$. Now we can take a log of the above first order condition and we have

$$\log(\alpha_i) - \gamma \log(L_{it}) = \log(\lambda_t) + \log w_{it}$$

Thus if hours are positive we have

$$\log L_{it} = \frac{1}{\gamma}\log\alpha_i - \frac{1}{\gamma}\log\lambda_t - \frac{1}{\gamma}\log w_{it} \tag{1}$$

In addition an agents works positive hours when

$$\log w_{it} \geq \log(\alpha_i) - \gamma\log T - \log\lambda_t \tag{2}$$

Our empirical specification will follow Chiappori et. al (2013) and uses hours h_{it} instead of leisure hours but the intuition is the same. Now we describe the effect of several different factors on hours worked. If an agent has a high Pareto weight α_i, then we expect that agent to have lower work hours since the planner cares more about the utility from leisure of this agent. Alternatively if the risk-sharing unit Lagrange multiplier is low then hours will decrease because the extra utility that can be gotten by relaxing the budget constraint is relatively low. Finally changes in the wage rate can also affect hours and this is captured by the parameter γ, which is the elasticity of labor supply. These are the main forces affecting labor supply. On the participation margin we see that equation (2) tells us that an agent participates if his wage is above a given reservation wage. In the empirical specification we add demographic information that helps to account for heterogeneity that is not present in the model. Hence the model is really a guide to help us interpret our empirical specification.

3 Empirical Specification

In this section I describe the empirical specification that will be estimated. The econometric model uses the theoretical model as a guide.

3.1 Econometric Model

The empirical specification follows that in Chiappori et. al (2013). Let H_{it} be the hours of work supplied by individual i at time t. In the equations below the subscript i will be used to refer to individuals while subscript j will denote a village. For most of our estimations we take the village as the basic risk-sharing unit. In other words the planner is solving the problem at the village level using the village level budget constraint. This implies that the Lagrange multipliers will also be at the village level. The hours equation is based on equation (1) from the planning problem. In particular we specify

$$\log H_{it} = e_i + f_i D_{jt} + g_i \log W_{it} + h' X_{it} + u_{it} \tag{3}$$

where W_{it} is the wage of individual i at time t. Now we go through each of the terms in the above equation. The term e_i corresponds to the Pareto weight of individual i. The D_{jt} corresponds to the logarithm of the Lagrange multiplier on the village budget constraint $\log(\lambda_{jt})$. The coefficient of the wage term reflects the elasticity of labor supply in the utility function of the individual $\log W_{it}$. The X_{it} are determinants of hours that contain demographic information (age, household size, gender, education etc.). Although this is not explicitly in the model we include it in the econometric specification in order to account for observable heterogeneity.

One thing that we must take into account is that about 30% of our sample of individual-month observations have zero labor supply. Thus we potentially have a selection problem as in Heckman (78) [3]. Equation (2) from the model suggests modeling the participation equation as

$$P_{it} = \mathbf{1}\left\{\log W_{it} \geq a_{it} + b_i D_{jt} + c' X_{it} + d' Q_{it} + \epsilon_{it}\right\} \tag{4}$$

Here $P_{it} = 1$ indicates that individual i is participating in the wage labor market in month t. In order for the two-step estimator that we will use to be well identified we want to have some covariates Q_{it} which appear in the participation equation but do not appear in the hours equation. Exclusion restrictions are not completely uncontroversial but following Chiappori et. al (2013) we will use lagged participation as our Q_{it}. One possible reason for including lagged participation in the model is the existence of sunk costs. We don't include these in the model because it would become significantly more complicated due to dynamic considerations.

Following CT we close the model by writing down a Mincer-type equation for log-hourly wages

$$\log W_{it} = l_i + m' X_{it} + v_{it} \tag{5}$$

so the wage can depend on observable characteristics X_{it} and unobservable individual characteristics l_i. One important consideration in the empirical analysis is that wages are very likely recorded with error. Measurement error tends to be a common feature of survey data, especially in developing countries such as Thailand. Given the fact that our wages are probably observed with error we model the measurement error as

$$\log \tilde{W}_{it} = \log W_{it} + \eta_{it} \tag{6}$$

where $\tilde{W}_{it}$ is the observed hourly wage. We suppose that η_{it} is classical measurement error, i.e. it is uncorrelated with the other regressors and error terms. Note that if we simply ran our regressions without accounting for measurement error this would cause a downward bias in our coefficient estimates on the log wage.

The equations (3) - (6) together form a joint model of hours worked, participation, true and observed wages. We assume that the shocks (ϵ, u, v, η) follow a multivariate normal distribution. Our exclusion restrictions are that (ϵ, u, v) are uncorrelated with D_{jt}, X_{it} and Q_{it}. We also assume as is typical with measurement error that it is independent of all shocks and regressors.

Now we want to test for the null of full risk sharing. In order to do this we include in the regressors X_{it} a household-level indicator of income shocks. The risk-sharing hypothesis implies that this should have no effect on participation or hours worked. This is because the model tells us that participation and hours worked should only respond to changes in wages and the aggregate shock which is captured by the Lagrange multiplier D_{jt}. The variables we will use are household non-labor income and household non-labor income including cash flows from production.

In order to estimate the model we impose several restrictions on the coefficients in the equations (3) -(6). We restrict the coefficients b_i and f_i to depend only on the risk sharing unit, in this case the village. Thus everyone responds the same to the aggregate shocks. Secondly we assume that the g in the hours equation is constant across individuals and villages within the same province. This is a particularly strong assumption as it implies that all individuals in the same province have the same elasticity of labor supply. However as the Townsend et. al show in their 2013 paper "Heterogeneity and Risk Sharing in Village Economies" there is considerable heterogeneity among households in their aversion to consumption risk. This may be an indication that there is may also be heterogeneity in their elasticity of labor supply. We will attempt to address this concern in the results section.

3.2 Estimation Strategy

In order to estimate the model we treat the individual-specific fixed effects as parameters to be estimated. In other words we need to estimate

$$\{a_i, e_i, l_i\}$$

This approach is justified by the fact that the number of time periods $T = 160$ is large and thus these parameters can be precisely estimated. Now we present our method for estimating the parameters of the hours and participation equations. The approach to dealing with the selection problem is similar to Heckman's two step estimator. Substituting the wage equation into the participation equation gives us

$$P_{it} = \mathbf{1}\{l_i + mX_{it} - a_i - bD_{jt} - cX_{it} - dQ_{it} \geq \epsilon_{it} - v_{it}\}$$

Now we let $\chi_{it} = \{D_j, X_{it}, Q_{it}, \delta_{it}\}$ be the set of regressors. The Pareto weights D_{jt} will simply be time-village dummies. The δ_{it} are the individual specific dummies. Now, assuming normality of the error $\epsilon_{it} - v_{it}$ the above equation can be estimated by a probit regression. In particular the probit delivers estimates $\hat{\beta}$. We will use these later in computing the selection correction factor.
Now we turn to the hours of work equation. Since there is measurement error in $\log \tilde{W}_{it}$ we see that the hours equation can be written as

$$\log H_{it} = e_i + fD_{jt} + g \log \tilde{W}_{it} + h' X_{it} + u_{it} - g\eta_{it}$$

There are several issues in the above regression. The first problem is that $\log \tilde{W}_{it}$ is endogenous since it is correlated with the measurement error η. The second issue is that due to the fact that u_{it} can be correlated with ϵ_{it} we can have $E[u_{it}|P_{it} = 1] \neq 0$. Now let $\tau_{it} = v_{it} - \epsilon_{it}$ be the error in the participation equation. Then we know that

$$E[\tau_{it}|P_{it} = 1] = E[\tau_{it}|\tau_{it} > -\chi_{it}\beta] = \frac{\phi(-\chi_{it}\beta)}{1 - \Phi(-\chi_{it}\beta)} = \frac{\phi(\chi_{it}\beta)}{\Phi(\chi_{it}\beta)}$$

We can estimate the above by $\hat{\lambda} = \frac{\phi(\chi_{it}\hat{\beta})}{\Phi(\chi_{it}\hat{\beta})}$, the inverse mills ratio. Let ν be the covariance between τ_{it} and u_{it}. Then we see that $E[u_{it}|P_{it} = 1] = \nu\lambda$. Then if we let $\tilde{u}_{it} = u_{it} - \nu\lambda$ then $E[\tilde{u}_{it}|P_{it} = 1, \chi_{it}] = 0$. Also note that since the measurement error was assumed to be independent of all regressors and shocks we see that $E[\eta_{it}|P_{it} = 1, \chi_{it}] = 0$. The only remaining problem is the endogeneity of the observed wage due to the measurement error. This can be solved if we can find an instrument for the observed wage. If we had instruments Z_{it} for $\log \tilde{W}_{it}$ then we could consistently estimation the log hours equation by performing a fixed effect regression combined with

2sls for instrumenting the log wage.

$$\log H_{it} = e_i + fD_{jt} + g_i \log \tilde{W}_{it} + h'X_{it} + \nu\hat{\lambda} + (\tilde{u}_{it} - g\eta_{it})$$

Hence we run a fixed effect regression on the above equation where we instrument the observed log wage with Z and include the estimated inverse mill's ratio $\hat{\lambda}$ as a regressor. To instrument for $\log \tilde{W}$ we use the job form wage interacted with the job form payment type. As was explained in the data section for each individual and each job that that individual has they fill out a job form questionnaire. This contains information about the type of job, the employer etc. In particular the individual must report their wage, what we call the "job form wage," and the "job form payment type" which tells us whether this is a daily, weekly or monthly wage. Then by interacting the job form wage with the job form payment type we can instrument for the observed (noisy) wages $\log \tilde{W}$. This approach relies on the fact that Z (the job form wage) is uncorrelated with u conditional on $P = 1$ and (X, D, Q).

4 Data

The data we use comes from the Townsend Thai Monthly survey. This intensive monthly survey began in August 1998 and is still ongoing. As such it provides us an unusually long time series, $T = 160$ months at the time of this writing. Now we give a brief overview of the dataset and the variables that we will use in our estimation.

The Townsend Thai monthly survey is a comprehensive monthly survey that began in 1998 in four provinces of Thailand and is still ongoing. The four provinces are Chachoengsao, Lopburi, Buriram and Sisaket. Chachoengsao and Lopburi are semi-urban provinces in the fairly well developed central region. They are within about 160 kilometers of the capital city Bangkok. On the other hand Buriram and Sisaket are rural and in the less developed northeastern region of Thailand which borders Cambodia. Within each province there are four villages that are part of the survey making for a total of 16 villages. In the first month of the survey, August 1998 there was an initial census in each village. Each structure was recorded and grouped into "household" units depending on sleeping and eating patterns. This means that all individuals and households in each of the 16 villages can then be identified in later months of the survey. There are roughly households 45 in each village during the months of the survey. The monthly updates track a wide range of variables of interest for each family (and household member within the family when applicable). In particular there are 24 modules each focusing on a different set of variables, i.e. consumption, labor supply, savings, borrowing, household business etc. Since the focus of our paper is on labor supply we restrict our sample to individuals who are of age $18 - 60$. The reason for this restriction is that individuals who are older than 60 and younger than 18 don't really participate in the labor market. We also exclude individuals who are involved in full time schooling. The empirical specification requires us to construct the hours of work and wage rate of each individual and the non-labor income of the household.

4.1 Wages

As mentioned before our empirical specification requires that we know the wage rate of each individual for each month that they participate in the labor market. This is relatively straightforward for individuals who are working in the wage labor market in a given month. We have data on earnings and hours and thus we can compute the wage rate as earnings divided by hours. Note that in general this will produce a noisy estimate of the wage rate since hours are reported with error. For individuals who are not working on the wage labor market it is not clear what the correct measure of the hourly wage would be. For example many individuals work on cultivating their own plots or taking care of livestock. These are activities with a large payment at harvest time but little income the rest of the year. Thus it isn't clear what the analogue of the hourly wage would be. In light of this we focus exclusively on labor supply in the paid labor market. By doing this we observe the wage rate of each individual in each month during which they work on the paid labor market.

4.2 Labor Supply

Computing labor supply for an individual is quite a complicated and time consuming task with the dataset. The difficulty comes from the fact that individuals spend time doing a variety of different economic activities. Economic activities include cultivating own polots, taking care of livestock, business, fish/shrimp and paid labor. Hours must be compiled from each of the monthly update forms separately and merged. The monthly surveys record the number of days spent on a each economic activity and the average number of hours in a day that were spent on that activity on a day when that activity was performed. From this we can compute total hours spent on each economic activity in the given month. Since the number of days between interviews varies (mostly between 27 and 35) we renormalize the working hours to reflect a 30 day month. Following Chiappori et. al (2013) I will focus on labor supply in the wage labor market. The reason is that in order to estimate our empirical specification we need to know the wage that the individual is earning in a given month. In order to precisely estimate the individual fixed effects terms in the regressions following Chiappori et. al (2013) we restrict our sample to individuals who work at least 40 months (out of 160) on the wage labor market. We also restrict our sample to individuals who were present in at least 90 of the 160 months of the sample. Wages are deflated using the regional CPI from the Thai chamber of commerce.

4.3 Non-labor income

In order to test our hypothesis of full risk sharing we include in the regressors X_{it} an measure of household income shocks. Under the full risk sharing hypothesis the theory tells us that participation and hours worked should not react to idiosyncratic shocks, rather it should respond only to aggregate shocks, which appear as changes in the Lagrange multipliers λ_{jt}. We use two variables that capture household income shocks. The first is household non-labor income. The other variable that we use has household non-labor income but also includes cash flows from production (i.e. livestock, fish/shrimp etc.). Again we deflate these using the regional CPI from the Thai chamber of commerce.

5 Results and Discussion

Recall that the sample includes four regions or changwats and we've sampled four villages within each changwat. Since there seems to be considerable heterogeneity across the changwats we run our all our estimations at the level of the changwat. We exclude Srisaket (province 53 in the data) because after trimming our sample in the above manner the number of remaining observations, i.e. individuals working on the wage labor market, is quite small.. As was discussed before, it is difficult to accurately assign a wage to individuals who aren't participating on the wage labor market and thus we chose to focus on labor supply in the paid labor market. Given this choice Srisaket doesn't provide a sufficient sample size for meaningful analysis.

Firstly we focus on labor force participation. Looking at tables 1 and 2 shows us that the effects of non-labor income on labor force participation are negative and significant (at the one percent level) in almost all cases. This means that increased non-labor income results in decreased participation in the wage-labor market. This is an indication of rejection of the null hypothesis of full risk sharing since the under the null hypothesis non-labor income should not have an effect on labor force participation. We normalized the magnitude of non-labor income by its standard deviation in order to facilitate interpretation of the results. For example in province 7 (Chachoengsao) a one standard deviation increase in non-labor income causes the probit to decrease by about 0.08. This implies that the probability of labor force participation decreases by about 2%. In the other provinces the effect is even larger. In provinces 27 and 49 a one standard deviation increase in non-labor income causes participation probability to decrease by around 8%. In the specification without individual fixed effects the coefficient of non-labor income has a larger magnitude across the three provinces. This could arise from the fact that non-labor income is negatively correlated with the fixed effect in the probit. This would imply that individuals with larger non-labor income are also those whose unobserved taste for work is lower. One way to rationalize this is that individuals have different abilities for earning income from sources other than wage labor. Thus individuals who have high non-labor income are also those who have higher ability to earn from sources other than wage labor and thus they participate less in the labor market.

Across all the specifications lagged participation, especially in the previous month and the average of the previous $2-6$ months, is an important predictor of participation probability in the current month. In particular the coefficients on lagged participation are relatively stable across different provinces and specifications. Comparing the probit with and without individual fixed effects also gives fairly similar coefficients on the lagged participation indicators.

We recognize that including fixed effects in the selection equation can lead to inconsistent estimates due to an incidental parameters problem. However, following Chiappori et. al (2013), since we have an unusually long panel, $T = 160$, this problem is mitigated to a certain extent. In addition to the individual fixed effects we have month-village dummies which represent the Lagrange multipliers. Estimating these parameters consistently can also be a challenge since the Lagrange multipliers are similar to incidental parameters since we have a limited number of observations per village and each time we add a village we add another Lagrange multiplier. However at each time t we have observations for each member of the village (generally around 40). A Hausman test of individual fixed effects and village-time effects versus only village-time effects accepts the null of no fixed effects. I believe that this is largely due to the fact that the degree of the chi-squared is very large since there is a large number of estimated parameters due to the fact that we have village-month dummies. In order to test this further I considered a specification where we only include regional dummies in the probit equation. That is we have region-month dummies instead of village-month dummies. See tables 3 and 4 for these estimates. This reduces the number of estimated parameters significantly, almost by a factor of four. Then we can run the probit estimation with the fixed effects and compare it with the results from the probit without fixed effects. In this case a Hausman test of these two alternatives *does* reject the random effects model. This suggests that there are individual fixed effects in the selection equation which, according to the model, would come from non-constant Pareto weights.

Now we turn to the hours equations. The relevant estimates are in tables 5 and 6. A Hausman test of random effects vs first effects in the hours equations rejects the null of fixed effects. Thus we focus on the fixed effects estimates but report both estimates for consistency. The first thing to look at is the coefficients on non-labor income. In the model with full fixed effects (in both the selection and the hours equation) (see 5) we see that the coefficient on non-labor income is significant in provinces 7 and 27. In province 7 the coefficient is actually positive taking the value .018. Since we normalized the non-labor income variable by its standard deviation this means that a one standard deviation increase in non-labor income would actually *increase* labor hours by about 1.8% in province 7. Thus even though the coefficient is significant its magnitude is quite small. In province 49 on the other hand the coefficient is negative and slightly larger at -0.063. Thus a one standard deviation increase in non-labor income decreases hours by 6.3%. This means that we reject the hypothesis of full risk sharing in provinces 7 and 49. On the other hand since the magnitude of the coefficient is relatively small the actual deviation of hours from the full risk-sharing model is quite small. In province 49 we can't reject the null of full-risk sharing in the labor hours equation since the coefficient isn't significantly different from zero at the 10% level. [1]

One important thing to notice is that in the specifications with fixed effects the coefficient on the log wage in the log hours equation is negative for provinces 7 and 49. This is a surprising result since the theory suggests that this coefficient should be positive. It is however positive in province 27.

[1]Since the regressor $\hat{\lambda}$ in the hours equation is itself the projection from a first stage estimation we have an issue with generated regressors. This implies that the usual standard errors need to be corrected for this. I attempted to bootstrap the the standard errors but was not able to get the code to converge as the probit estimation is quite computationally intensive. Thus the I report the usual standard errors.

Thinking back to the planner's problem log hours takes the form

$$\log h_{it} = -\frac{1}{\gamma}\log\alpha_i + \frac{1}{\gamma}\log w_{it} + \frac{1}{\gamma}\ln\lambda_{jt}$$

Thus after accounting for individual fixed effects (which correspond to Pareto weights in the planner's problem) the planner should assign more hours of work to those who individuals who are more productive, i.e. the agents with higher wages. This would lead us to expect that the coefficient on the log wage in the hours equation should be positive. Thus it is somewhat surprising to see that the coefficient is actually negative and significantly so in all the specifications for provinces 7 and 49. One possible explanation are income effects. In the absence of full risk sharing if an individual becomes wealthier, say by having a higher wage, he could reduce labor supply due to the dominance of the income effect but this seems unlikely.

Additionally note that in all specifications the coefficient on the mills ratio is significant. This tells us that there is indeed a selection bias in the hours equation so that the mean of u_{it} conditional on participation is not zero. It is not obvious how to interpret the sign of this coefficient. The coefficient on the mills ratio is negative. Going back to the estimation section this tells us that the correlation between τ_{it} and u_{it} is negative. In other words individuals with a high τ in the selection equation will actually have a low u_{it} in the hours equation. This is somewhat confusing as it says that individuals that choose to participate in the wage labor market have lower shocks u_{it} in the hours equation than the population average. On the other hand this is consistent with the results of the Chiappori et. al. (2013) paper.

Recall that we instrumented for the log wage in the hours equation because we were worried about endogeneity because the observed wage data were most likely contaminated by measurement error. Performing Wu-Hausman and Durbin test for endogeneity gives a p-value of around 0.2 which means that we can't reject the null that log wages are exogenous in the hours equation. This is surprising given that we think that our wage observations are contaminated by measurement error. A possible explanation is that the instruments aren't exogenous. A Sargan chi-squared test of overidentifying rejects the exogeneity of the instruments, so if we don't have exogenous instruments that could explain why the Wu-Hausman test didn't reject exogeneity of log wages since that test is effectively just comparing the IV and OLS estimates and they may be quite similar (even if log wage is endogenous) if the instruments don't satisfy exogeneity. On the weak instruments front we don't have a problem since our first stage R-squareds are around 0.7 with an F statistic above 30. thus we continue with reporting our iv specification even though OLS would be more efficient. This doesn't come at a huge efficiency loss since our instruments have a quite high first stage R^2. We don't report the results of using our other measure of non-labor income (which includes cash flows from production), as the results are quite similar.s

As was discussed before we imposed the strong assumption that the elasticity of labor supply is the same across all villages in a given province. We can attempt to test this hypothesis by re estimating the equation for log hours but at the village level thus allowing each village to have its own elasticity of labor supply. Table 7 reports the result for province 7. The disparity in the coefficient estimates of the log wage indicate would seem to indicate that there is considerable heterogeneity even within province. In some cases the coefficient even changes sign as in the case of village 4. It would be surprising if there was truly such a large degree of heterogeneity within the villages. Thus these results are probably a reflection that some portion of the empirical model is misspecified.

6 Conclusion

This paper tests the null of full labor risk sharing in Thai villages. Our results lead us to reject the null of full risk sharing since non-labor income has a significant negative effect on participation. In most specifications it also has a significant but small negative effect on hours worked conditional on participation. In light of these results we reject the null of full risk sharing. However our analysis was limited by the nature of the data. In particular we only considered paid wage labor. Thus some of what we classified as non-participation could be individuals participating in non wage labor activities. If this is also correlated with an increase in household non-wage income this produce a significant negative coefficient in the participation probit while still being consistent with full labor risk sharing. On the other hand the estimation did produce some worrying results. Firstly in provinces 7 and 49 the coefficient of log wages in the log hours regression was negative, which runs counter to what we would expect. In addition the robustness check of running the hours equation at the village level for province 7 showed significant heterogeneity in the log wage coefficient across villages in the same province. This suggests that the model is most likely misspecified. Given this possible misspecification it is best to refrain from drawing definitive conclusions.

References

[1] S Bonhomme, P A Chiappori, R Townsend, and H Yamada. Sharing wage risk. 2008.

[2] PA Chiappori, K Samphantharak, Robert M Townsend, and Sam Schulhofer-Wohl. Heterogeneity and Risk Sharing in Village Economies. *Federal Reserve Bank of Minneapolis Staff Report*, 2013.

[3] JJ Heckman. Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46(4):931–959, 1978.

7 Tables

Table 1: Labor Market Participation

	(1) Province 7	(2) Province 27	(3) Province 49
Age (-20)	0.116	-2.752	0.905
	(0.449)	(100.8)	(0.710)
Age (21-40)	0.596	-2.226	1.076
	(0.449)	(100.8)	(0.705)
Age (41-60)	0.561	-2.129	0.894
	(0.457)	(100.8)	(0.709)
Household size	-0.0185	0.00778	0.0244
	(0.0164)	(0.0202)	(0.0182)
Sick days	-0.0567***	-0.0162	-0.0820***
	(0.0106)	(0.0195)	(0.0225)
Nonlabor income	-0.0863***	-0.225***	-0.173***
	(0.0194)	(0.0381)	(0.0191)
Lagged participation (1 month)	1.563***	0.863***	1.067***
	(0.0517)	(0.0484)	(0.0544)
Lagged participation (2-6 months)	0.260***	0.201***	0.259***
	(0.0161)	(0.0162)	(0.0181)
Lagged participation (7-12 months)	0.0804***	0.108***	0.110***
	(0.0127)	(0.0136)	(0.0149)
N	15059	8412	11936

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes individual fixed effects and time-village dummies

Table 2: Labor Market Participation

	(1) Province 7	(2) Province 27	(3) Province 49
Age (-20)	-0.0397	-2.851	0.602
	(0.399)	(112.6)	(0.556)
Age (21-40)	0.200	-3.232	-0.239
	(0.394)	(112.6)	(0.543)
Age (41-60)	0.128	-3.254	-0.277
	(0.395)	(112.6)	(0.543)
Household size	0.0800***	-0.00804	-0.0215**
	(0.00680)	(0.0127)	(0.00849)
gender	0.0455*	0.0142	-0.0104
	(0.0249)	(0.0329)	(0.0264)
Education 2	-0.727***	-0.174***	-0.763***
	(0.111)	(0.0550)	(0.111)
Education 3	-0.533***	-0.0101	-0.430***
	(0.116)	(0.0655)	(0.115)
Education 4	-0.726***	-0.958***	-0.624***
	(0.117)	(0.131)	(0.121)
Education 5	-0.123	0.629***	
	(0.125)	(0.113)	
Sick days	-0.0424***	0.00733	-0.0435**
	(0.00910)	(0.0173)	(0.0192)
Nonlabor income	-0.216***	-0.204***	-0.270***
	(0.0144)	(0.0286)	(0.0149)
Lagged participation (1 month)	1.419***	0.873***	0.942***
	(0.0458)	(0.0455)	(0.0501)
Lagged participation (2-6 months)	0.245***	0.218***	0.235***
	(0.0141)	(0.0148)	(0.0161)
Lagged participation (7-12 months)	0.0969***	0.124***	0.104***
N	18781	9148	13348

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes time-village dummies

Table 3: Labor Market Participation

	(1) Province 7	(2) Province 27	(3) Province 49
Age (-20)	0.118	-3.094	0.698
	(0.426)	(110.7)	(0.721)
Age (21-40)	0.520	-2.544	0.735
	(0.425)	(110.7)	(0.716)
Age (41-60)	0.394	-2.487	0.675
	(0.432)	(110.7)	(0.719)
Household size	-0.0131	-0.00663	0.0535***
	(0.0152)	(0.0187)	(0.0166)
Sick days	-0.0505***	-0.0157	-0.0780***
	(0.0101)	(0.0186)	(0.0208)
Nonlabor income	-0.0711***	-0.179***	-0.154***
	(0.0180)	(0.0347)	(0.0176)
Lagged participation (1 month)	1.503***	0.797***	1.035***
	(0.0487)	(0.0450)	(0.0509)
Lagged participation (2-6 months)	0.241***	0.196***	0.236***
	(0.0151)	(0.0152)	(0.0168)
Lagged participation (7-12 months)	0.0814***	0.0973***	0.106***
	(0.0119)	(0.0127)	(0.0139)
N	15069	8429	11960

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes individual fixed effects and time-province dummies

Table 4: Labor Market Participation

	(1) Province 7	(2) Province 27	(3) Province 49
Age (-20)	-0.0586	-3.230	0.428
	(0.386)	(194.8)	(0.577)
Age (21-40)	0.130	-3.644	-0.401
	(0.381)	(194.8)	(0.564)
Age (41-60)	0.0401	-3.614	-0.398
	(0.381)	(194.8)	(0.564)
Household size	0.0744***	-0.0173	-0.0101
	(0.00640)	(0.0120)	(0.00818)
gender	0.0580**	0.0105	0.0127
	(0.0243)	(0.0311)	(0.0256)
Education 2	-0.687***	-0.255***	-0.790***
	(0.107)	(0.0519)	(0.106)
Education 3	-0.518***	-0.0204	-0.469***
	(0.111)	(0.0623)	(0.111)
Education 4	-0.665***	-0.935***	-0.656***
	(0.112)	(0.128)	(0.115)
Education 5	-0.0954	0.766***	
	(0.121)	(0.111)	
Sick days	-0.0385***	0.000265	-0.0451**
	(0.00886)	(0.0166)	(0.0183)
Nonlabor income	-0.195***	-0.208***	-0.252***
	(0.0137)	(0.0270)	(0.0139)
Lagged participation (1 month)	1.389***	0.809***	0.932***
	(0.0441)	(0.0427)	(0.0476)
Lagged participation (2-6 months)	0.234***	0.210***	0.219***
	(0.0135)	(0.0140)	(0.0151)
Lagged participation (7-12 months)	0.0956***	0.111***	0.100***
	(0.0100)	(0.0115)	(0.0123)
N	18795	9168	13372
R^2			

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes time-province dummies

Table 5: Log-hours

	(1) Province 7	(2) Province 27	(3) Province 49
log wage	-0.478***	0.400**	-0.877***
	(0.0419)	(0.159)	(0.0753)
Age (-20)	-0.166	0.839	-0.303
	(0.130)	(0.725)	(0.301)
Age (21-40)	-0.0664	-0.415	-0.210
	(0.130)	(0.670)	(0.299)
Age (41-60)	-0.0697	-0.629	-0.315
	(0.131)	(0.674)	(0.301)
Household size	-0.00750*	-0.0198	-0.0117
	(0.00431)	(0.0160)	(0.0101)
Sick days	-0.0373***	-0.0370***	-0.0123
	(0.00497)	(0.0139)	(0.0123)
Nonlabor income	0.0178***	-0.0637**	-0.00193
	(0.00576)	(0.0305)	(0.0118)
Selection Factor	-0.413***	-0.417***	-0.397***
	(0.0215)	(0.0523)	(0.0440)
N	10290	3357	4951
R^2	0.591	0.399	0.505

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes individual fixed effects and time-village dummies

Table 6: Log-hours

	(1) Province 7	(2) Province 27	(3) Province 49
log wage	-0.114***	0.499***	-0.142***
	(0.0135)	(0.0522)	(0.0198)
Age (-20)	-0.0560	-0.190	-0.207
	(0.148)	(0.843)	(0.228)
Age (21-40)	-0.0150	0.0467	-0.0699
	(0.146)	(0.821)	(0.225)
Age (41-60)	-0.0292	0.111	-0.214
	(0.146)	(0.821)	(0.225)
Household size	-0.00727***	-0.00588	0.0659***
	(0.00244)	(0.0114)	(0.00467)
gender	-0.0731***	-0.237***	0.0890***
	(0.00884)	(0.0286)	(0.0154)
Education 2	0.124***	-0.0235	0.0221
	(0.0396)	(0.0448)	(0.0617)
Education 3	0.161***	-0.0409	0.188***
	(0.0413)	(0.0589)	(0.0611)
Education 4	0.251***		0.190***
	(0.0416)		(0.0657)
Education 5	-0.0616	-0.251***	
	(0.0453)	(0.0898)	
Sick days	-0.0267***	-0.0466***	-0.0272**
	(0.00553)	(0.0133)	(0.0129)
Nonlabor income	0.0402***	0.0184	0.0920***
	(0.00609)	(0.0242)	(0.0118)
Selection Factor	-0.823***	-0.804***	-0.838***
	(0.0233)	(0.0477)	(0.0463)
N	13734	4018	6042
R^2	0.229	.	0.303

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Includes village month-dummies but no individual fixed effects

Table 7: Log_hours

	(1) Village 2	(2) Village 4	(3) Village 7	(4) Village 8
log wage	-0.760***	0.405**	-0.281*	-0.549***
	(0.0586)	(0.166)	(0.145)	(0.0477)
Age (-20)	-0.368*	0.196	0.0690	0.0557
	(0.199)	(0.368)	(0.344)	(0.253)
Age (21-40)	-0.212	0.151	0.0299	0.00450
	(0.197)	(0.367)	(0.349)	(0.251)
Age (41-60)	-0.0487	0.0302	0.350	-0.210
	(0.201)	(0.370)	(0.351)	(0.253)
Household size	-0.0401***	0.0228**	0.0251	-0.00190
	(0.00716)	(0.0113)	(0.0173)	(0.00787)
Sick days	-0.0409***	-0.0617***	-0.0363***	-0.000118
	(0.0119)	(0.0118)	(0.00620)	(0.0167)
Nonlabor income	0.0535***	-0.00145	-0.0197	0.0310***
	(0.0104)	(0.0127)	(0.0159)	(0.0108)
N	3702	3150	1351	2087
R^2	0.583	0.352	0.607	0.650

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, ***$p < 0.01$

Includes individual fixed effects and time-village dummies

Regressions are at the village level to allow each village

to have its own elasticity of labor supply