Testing Efficient Risk Sharing with Heterogeneous Risk Preferences
Maurizio Mazzocco and Shiv Saini
Web Appendix

Statistical Appendix

A1. Estimation of the Risk-sharing Functions

We will briefly describe how the estimators proposed by Whitney K. Newey, James L. Powell, and Francis Vella (1999) and Hidehiko Ichimura (1993) can be employed to recover the risk-sharing functions. Suppose first that the parameters defining the heterogeneity term $d^{i,j}$ are known. Observe that the risk-sharing functions cannot be estimated using standard non-parametric methods because $E\left[m^k|\rho^{i,j}\right] \neq 0$. To address this issue, following Newey et al. (1999) let q be a set of instruments in the sense that the following conditions are satisfied:

$$\rho^{i,j} = h\left(q^{i,j}\right) + \varphi^{i,j}, \quad E\left[\varphi^{i,j}|q^{i,j}\right] = 0, \quad \text{and} \quad E\left[m^k|\varphi^{i,j}, q^{i,j}\right] = E\left[m^k|\varphi^{i,j}\right].$$

Then, we have that

$$\begin{aligned} E\left[\ln\rho^k|\rho^{i,j}, w^i, w^j, d^{i,j}, q^{i,j}\right] &= g^k\left(\rho^{i,j}, w^i, w^j, d^{i,j}\right) + E\left[m^k|\rho^{i,j}, w^i, w^j, d^{i,j}, q^{i,j}\right] \\ &= g^k\left(\rho^{i,j}, w^i, w^j, d^{i,j}\right) + E\left[m^k|\varphi^{i,j}, q^{i,j}\right] = g^k\left(\rho^{i,j}, w^i, w^j, d^{i,j}\right) + \lambda\left(\varphi^{i,j}\right), \end{aligned}$$

where $\lambda(\varphi) = E\left[m^k|\varphi\right]$. Newey et al. (1999) propose to estimate the function g^k in two steps. In the first step the error term φ is estimated non-parametrically as $\hat{\varphi}^{i,j} = \rho^{i,j} - \hat{h}\left(q^{i,j}\right)$. In the second step, the function $g^k\left(\rho^{i,j}, w^i, w^j, d^{i,j}\right) + \lambda\left(\varphi^{i,j}\right)$ is estimated using the estimated residuals in place of the true ones. An estimator of the function g^k can then be recovered by isolating the components that do not depend on the residuals φ. The estimation will be performed using the series estimator proposed by Newey et al. (1999) with polynomials.

The parameters of the heterogeneity term $d^{i,j}$ are not known, but they can be estimated using one of the semi-parametric methods developed for the estimation of single-index models. We use the semi-parametric least square approach proposed by Ichimura (1993). Specifically, each risk-sharing function is estimated by iterating the following two steps. First, for a given value of the parameters in $d^{i,j}$ we estimate the risk-sharing function using the series estimator proposed by Newey et al. (1999). We then compute the sum of squared errors corresponding to the estimated function. We stop when the sum of squared errors is minimized.

A2. Test of Homogeneous Risk Preferences For a Pair of Households

The test is implemented in three steps. In the first step, the difference between risk-sharing functions is estimated using the method discussed in the previous subsection. Let $\hat{g}^{i,j}$ be the estimated difference. In the second step, the test statistic $\hat{\xi}_1^{i,j}$ is computed as follows:

$$\hat{\xi}_1^{i,j} = -\left(\sum_{l=1}^{n} \mathbf{1}_{\left\{\hat{g}^{i,j}\left(\rho_l^{i,j}\right)\geq 0\right\}} \hat{g}^{i,j}\left(\rho_l^{i,j}\right)\right)\left(\sum_{l=1}^{n} \mathbf{1}_{\left\{\hat{g}^{i,j}\left(\rho_l^{i,j}\right)< 0\right\}} \hat{g}^{i,j}\left(\rho_l^{i,j}\right)\right),$$

where $\mathbf{1}_{\{\}}$ is an indicator function. In the final step, the distribution of $\hat{\xi}_1^{i,j}$ is recovered by bootstrap. To increase the power of the test, we follow Hall and Wilson (1991) and compute the bootstrap distribution by resampling $\hat{\xi}_1^{i,j*} - \hat{\xi}_1^{i,j}$ instead of $\hat{\xi}_1^{i,j*}$, where $\hat{\xi}_1^{i,j*}$ is the estimated test statistic obtained using a bootstrap sample. The null is then rejected for the pair composed of households i and j if $\hat{\xi}_1^{i,j}$ is too large, i.e. if

$$\hat{\xi}_1^{i,j} > q^{i,j*}\left(0.95\right),$$

where $q^{i,j*}\left(0.95\right)$ is the 95-th percentile of $\hat{\xi}_1^{i,j*} - \hat{\xi}_1^{i,j}$.

A3. *Test of Efficiency Based on Omitted Variables for a Pair of Households*

This test is implemented using the semi-parametric approach proposed by Yanqin Fan and Qi Li (1996). We will briefly describe the method. Consider the pair composed of households i and j and suppose that the risk-sharing function of household i depends on an omitted variable y^i. Then, Subsection A.1 implies that the function can be written in the following form:

$$\ln \rho^i = g^i\left(\rho^{i,j}, w^i, w^j, d^{i,j}, y^i\right) + \lambda\left(\varphi^{i,j}\right) + \epsilon^i = f^i\left(X^{i,j}, y^i\right) + \epsilon^i,$$

where $E\left[\epsilon^i | X^{i,j}, y^i\right] = 0$. Under the null hypothesis of efficiency, y^i should not enter the risk-sharing function. As a consequence

$$f^i\left(X^{i,j}, y^i\right) = E\left[\ln \rho^i | X^{i,j}, y^i\right] = E\left[\ln \rho^i | X^{i,j}\right] = r^i\left(X^{i,j}\right).$$

But under the alternative hypothesis we have that

$$f^i\left(X^{i,j}, y^i\right) \neq r^i\left(X^{i,j}\right).$$

Let $\nu^i = \ln \rho^i - r^i\left(X^{i,j}\right)$. Then, $E\left[\nu^i | X^{i,j}, y^i\right] = f^i\left(X^{i,j}, y^i\right) - r^i\left(X^{i,j}\right) = 0$ under the null and $E\left[\nu^i | X^{i,j}, y^i\right] \neq 0$ under the alternative. Now observe that

$$E\left[\nu^i E\left[\nu^i | X^{i,j}, y^i\right]\right] = E\left[\left\{E\left[\nu^i | X^{i,j}, y^i\right]\right\}^2\right] \geq 0,$$

where the inequality follows from the previous discussion and the inequality is replaced by an equality if and only if the null hypothesis is correct. Fan an Li propose to test for the omitted variable y^i using this inequality and the following idea. If ν^i and $E\left[\nu^i | X^{i,j}, y^i\right]$ were known, one could estimate $E\left[\nu^i E\left[\nu^i | X^{i,j}, y^i\right]\right]$ using its sample analog $n^{-1}\sum_i \nu^i E\left[\nu^i | X^{i,j}, y^i\right]$. The authors suggest to replace the residuals ν^i with estimated ones and $E\left[\nu^i | X^{i,j}, y^i\right]$ with its kernel estimator. Finally, to overcome the random denominator problem in the kernel estimation, they propose to replace the sample analog with its density weighted version:

$$\frac{1}{n}\sum_i \left[\nu^i f\left(X^{i,j}\right)\right] E\left[\nu^i f\left(X^{i,j}\right) | X^{i,j}, y^i\right] f\left(X^{i,j}, y^i\right),$$

where $f(.)$ denotes the probability density function. Finally, the estimated sample analog is divided by its estimated standard deviation and multiplied by $nh^{d/2}$, where n is the number of observations and h is a smoothing parameter in the kernel estimator. In the present paper we estimate the residuals using the series estimator described in Subsection A.1 and the densities using a standard gaussian kernel estimator.

At this point we have one test statistic for each household in the pair. To compute the test statistic for the pair observe that efficiency is rejected if y^i affects the risk-sharing function of at least one household. We can therefore compute the test statistic for the pair $\hat{\xi}_2^{i,j}$ by taking the maximum of the individual test statistics. Similarly to the homogeneity test, the distribution of the test statistic for the pair is obtained using bootstrap by resampling $\hat{\xi}_2^{i,j*} - \hat{\xi}_2^{i,j}$. The null hypothesis is then rejected for the pair composed of households i and j if $\hat{\xi}_2^{i,j}$ is too large, i.e. if

$$\hat{\xi}_2^{i,j} > q^{i,j*}(0.95).$$

A4. Test of Efficiency Based on Increasing Risk-sharing Functions For a Pair

We will first provide the intuition underlying the test introduced by Peter Hall and Nancy E. Heckman (2000) using a simpler version of the economy considered in this paper. We will then consider the more general case. Suppose that preferences are separable between consumption and leisure, there is no observable and unobservable heterogeneity, and no measurement errors. In this case, household i's expenditure is only a function of $\rho^{i,j}$ and there is no endogeneity issue, i.e.

$$\ln \rho^i = g\left(\rho^{i,j}\right) + \epsilon. \tag{A1}$$

Let $\left\{\left(\ln \rho_t^i, \rho_t^{i,j}\right), 1 \leq t \leq T\right\}$ be data generated by equations (A1) and denote by $\left\{\left(\ln \bar{\rho}_t^i, \bar{\rho}_t^{i,j}\right), 1 \leq t \leq T\right\}$ the same data sorted in increasing order of aggregate resources $\rho^{i,j}$. Consider a subset of the sorted data $\left\{\left(\ln \bar{\rho}_t^i, \bar{\rho}_t^{i,j}\right), r \leq t \leq s\right\}$ and estimate the slope of the linear regression of $\ln \rho^i$ on $\rho^{i,j}$. Repeat the last step for any subset of the sorted data that contains enough information to estimate the slope. Hall and Heckman's idea is that under the hypothesis that the function $g\left(\rho^{i,j}\right)$ is increasing, the minimum over all the estimated slopes should not be negative.

Formally the test is implemented as follows. For a given integer m that will be defined later, let r and s be integers that satisfy $0 \leq r \leq s-m \leq T-m$ and let a and b be scalars. Denote by $h\left(w^i, w^j, d^{i,j}\right)$ a polynomial in the wages and heterogeneity term, and by $\delta\left(\varphi^i\right)$ a polynomial in the first stage residuals in the estimator proposed by Newey et al. (1999). Define

$$S\left(a, b, h, \delta | r, s\right) = \sum_{t=r+1}^{s} \left\{ \ln \rho_t^i - \left[a + b\rho_t^{i,j} + h\left(w_t^i, w_t^j, d_t^{i,j}\right) + \delta\left(\varphi_t^i\right) \right] \right\}^2 .$$

For each choice of r and s, let $\hat{a}\left(r,s\right)$, $\hat{b}\left(r,s\right)$, $\hat{h}\left(r,s\right)$, and $\hat{\delta}\left(r,s\right)$ be the solution of the following least square problem:

$$\left(\hat{a}, \hat{b}, \hat{h}, \hat{\delta}\right) = argmin\ S\left(a, b, h, \delta | r, s\right).$$

The variance matrix of the estimated coefficients is equal to $\sigma^2\left(X'X\right)^{-1}$, where σ^2 is the variance of the residuals in the risk-sharing function and X is the matrix of regressors. This implies that the variance of $\frac{\hat{b}}{\sqrt{\left(X'X\right)_{b,b}^{-1}}}$ is equal to σ^2, where $\left(X'X\right)_{b,b}^{-1}$ is the diagonal element of the inverse matrix that corresponds to $\hat{b}$. The test statistic for each household in the pair can then be defined as

$$\hat{\xi}_3^i = \max\left\{ -\frac{\hat{b}\left(r,s\right)}{\sqrt{\left(X'X\right)_{b,b}^{-1}}} : 0 \leq r \leq s-m \leq T-m \right\}.$$

Note that the integer m plays the role of a smoothing parameter in the sense that larger values of m reduce the effect of outliers. Similarly to the first efficiency test, the test statistic for the pair $\hat{\xi}_3^{i,j}$ can be computed by taking the maximum of the two individual test statistics. The test rejects the null if $\hat{\xi}_3^{i,j}$ is too large.

The distribution of the test statistic is derived using the bootstrap method suggested by Hall and Heckman (2000). According to this method, the bootstrap distribution should be derived under the hypothesis that the function under investigation is constant in $\rho^{i,j}$ because it is the most difficult nondecreasing function for which to test. As in the previous tests, the bootstrap distribution is obtained by resampling $\hat{\xi}_3^{i,j*} - \hat{\xi}_3^{i,j}$. We can then reject the null for the pair composed by households i and j if

$$\hat{\xi}_3^{i,j} > q^{i,j*}\left(0.95\right).$$

A5. Tests at the Economy Level

The tests at the economy level are based on the multiple testing procedure developed in Joseph P. Romano, Azeem M. Shaikh and Michael Wolf (2007). Consider n hypotheses $H_1, ..., H_n$ and let $T_1, ..., T_n$ be the associated test statistics. Suppose that one is interested in a null hypothesis H_0 which is equal to the intersection of $H_1, ..., H_n$, in the sense that H_0 is not rejected only if each individual hypothesis H_k is not rejected. Romano, Shaikh and Wolf (2007) propose a method for testing H_0 that controls the k-FWE rate, i.e. the probability of rejecting at least k true hypotheses. To describe it, let $T_{r_1} \geq T_{r_2} \geq ... \geq T_{r_n}$ be the test statistics ordered from the highest to the lowest. For any subset of individual hypotheses D, denote by $k - max\{T_i\}$ the k-th largest test statistic and by $c_D(1-\alpha, k)$ the $1-\alpha$ percentile of its sampling distribution. In the first step, all the individual hypotheses are considered, i.e. $D = D_1 = \{H_{r_1}, ..., H_{r_n}\}$, and the following generalized confidence region is constructed:

$$[T_{r_1} - c_{D_1}, \infty) \times ... \times [T_{r_n} - c_{D_1}, \infty) .$$

The individual hypothesis T_{r_i} is then rejected if $0 \notin [T_{r_i} - c_{D_1}, \infty)$, or equivalently $T_{r_i} > c_{D_1}$. If in the first step no individual hypothesis is rejected, the null H_0 is also not rejected and the procedure stops. If at least one hypothesis is rejected H_0 is also rejected. If the number of rejections is smaller than k, the procedure stops. Otherwise, Romano, Shaikh and Wolf (2007) show that the power is increased by proceeding to the second step. Let R_1 be the number of individual hypotheses rejected in the first step. In the second step, one considers the individual hypotheses not yet rejected, i.e. $D = D_2 = \left\{H_{r_{R_1+1}}, ..., H_{r_{R_n}}\right\}$, and construct the corresponding generalized confidence region:

$$\left[T_{r_{R_1+1}} - c_{D_2}, \infty\right) \times ... \times [T_{r_{R_n}} - c_{D_2}, \infty) ,$$

where the threshold c_{D_2} is constructed using the following method. Construct all the possible subsets that contain the $n - R_1$ hypotheses that were not rejected plus $k-1$ of the rejected hypotheses. Denote by $D_{2,i}$ the i-th subset. For each subset compute the $1-\alpha$ percentile of the sampling distribution of the k-th largest test statistic $c_{D_{2,i}}(1-\alpha, k)$. The threshold c_{D_2} is the maximum of all $c_{D_{2,i}}$. The hypothesis $T_{r_{R_i}}$ is then rejected if $T_{r_{R_i}} > c_{D_1}$. If the number of rejected hypotheses is smaller than k one should stop. Otherwise, one continues in this stepwise fashion until less then k hypotheses are rejected. The $1-\alpha$ percentile of the sampling distribution of the k-th largest test statistic is computed using the bootstrap method illustrated in Romano, Shaikh, and Wolf (2006).

In the empirical implementation of the tests, k will be set equal to a fraction of the individual hypotheses considered. The fraction is chosen following the results of the simulation study that is discussed in the next appendix. It is important to remark that k will be set equal to the same fraction of hypotheses for every group of households. As a consequence, the k-FWE rate will be fixed at the same percentage of hypotheses for groups that are characterized by 20 hypotheses as well for groups that are characterized by 500 hypotheses. This choice implies that the same definition of size of the test will be used for small and large groups of households. The only exceptions are groups for which the constant fraction of hypotheses implies a k that is smaller than 1. In those situations we will set k equal to 1 and the the test will be more likely to reject.

Simulation Study

In this appendix we will study the performance of the three tests proposed in this paper using simulations. The simulation study is important for three reasons. First, the statistics used to test homogeneity in preferences and full-insurance are not smooth functions. This feature may be problematic if one wants to bootstrap their distributions. The study will enable us to evaluate the performance of the tests when the distributions are bootstrapped. Second, under the assumption made in Section V of the paper, the measurement errors enter non-linearly in aggregate resources, which may be problematic for the estimator proposed by Newey et al. (1999). With the simulation study we will be able to understand what is the effect of the measurement errors on the performance of the tests. Finally, the results of the simulations will enable us to evaluate the power and control of the tests and to improve them when the tests are implemented using the ICRISAT data. Following Romano, Shaikh and Wolf (2007) we will focus on three measures of test performance: the average number of false hypotheses rejected; the average number of true hypotheses rejected; the empirical FWE rate.

All the simulations share the following features. To be consistent with the data, it is assumed that the group under investigation is composed of thirty households.[1] All households have a utility function which is nonseparable between consumption and leisure and has the following form:

$$u^i\left(c, l; z, \eta\right) = \frac{\left(c^{\sigma_i} l^{1-\sigma_i} + a_i\right)^{1-\gamma_i}}{1-\gamma_i} exp\{\theta z + \eta\}.$$

This utility generates risk-sharing functions that satisfy the restrictions discussed in Section V of the paper. The parameters σ_i and a_i are assumed to be identical across households with $\sigma_i = 0.5$ and $a_i = 1$. The risk aversion parameter γ_i is allowed to vary across households. Fifteen households are assumed to have $\gamma_i = 1.2$, whereas the corresponding parameter for the other fifteen is set to 2.5. Households can save using a risk-free asset with no constraint on their borrowing ability. The interest rate is fixed at 0.05 and the discount factor is set equal to

[1] We have also performed the simulation study with 10 households obtaining similar results.

0.95. Each household can draw a high or low daily wage with equal probability. The high and low wages are set equal to 3 and 5 rupees, respectively. We allow for unobservable heterogeneity in the form of a pair fixed effect and for observable heterogeneity using the following three variables: mean adult age, caste ranking, and number of infants. Household decisions are simulated for 160 periods. To approximate the length of the panel in the ICRISAT, the test is then performed using the 120 periods that are between $t = 21$ and $t = 140$. The simulation is repeated 500 times. The distribution of the test statistics is determined using 500 bootstraps.

To evaluate the effect of measurement errors on the outcome of the tests, we add measurement errors to the natural logarithm of household expenditure ρ^i. We consider two types of measurement errors. In the first case, they are drawn from a normal distribution with mean zero and a negligible standard deviation ($\sigma_m = 0.01$). In the second case they are drawn from a normal distribution with the same mean but a standard deviation that is equal to half the standard deviation of the simulated household expenditure. We expect the two standard deviations used in the simulation to be a lower and upper bound for the standard deviation of the measurement errors in the data.[2] In the homogeneity in risk preferences test, the effect of measurement errors depends also on the quality of the instruments. In the simulation of that test we consider two sets of instruments. The first set contains non-labor income, the demographic variables, the wages of each household in the group, and lagged savings. It generates an average R^2 for total expenditure of about 0.97, where the average is computed across pairs. The second set contains the wages of the two households considered in the test, lagged total expenditure, lagged savings, and rain which is an exogenous shock in the simulation. In this case, the average R^2 is about 0.5.

The implementation of tests requires the researcher to choose k in the k-FWE rate and the order of the polynomial in ρ^{ij}. We experimented with several values for k. We report the results for k equal to 2.5 percent, 5 percent, and 10 percent of total hypotheses, since generally for k greater than 10 percent the loss in control more than dominates the gain in power. For the first two tests, we also experimented with polynomials of order 1, 2, 3, and 5. For each one, we will report only the results that are useful to understand the relationship between the order of the polynomial and the test performance. By construction, the efficiency test based on increasing functions does not depend on the order of the polynomial in ρ^{ij}.

In the implement of the homogeneity in risk preferences test one must control for the variation in wages, and in observable and unobservable heterogeneity. We control for this variation in two steps. We first estimate semi-parametrically the risk-sharing functions and their differences. We then fix wages and the heterogeneity term at the household mean and perform the tests.

[2]Martin Ravallion and Shubham Chaudhuri (1997) convincingly argue that there are measurement errors in the ICRISAT. We could not find a paper, however, that estimates their standard deviation.

B1. Simulations for the Test of Homogeneity in Risk Preferences

In this subsection we evaluate the performance of the test of homogeneity in risk preferences. To that end, we simulate the decisions of the thirthy households under the maintained assumption of efficient risk sharing. The Pareto weights are chosen so that the risk-sharing functions of households with heterogenous risk preferences cross. The goal of the simulation study is therefore to evaluate whether different specifications of the test can detect these crossings. The actual data may correspond to Pareto weights for which the household risk-sharing functions do not cross even if preferences are heterogeneous. The results should therefore be interpreted as an upper bound for the power of the test.

The results indicate that the performance of the test depends on four features of the simulated data: the severity of the measurement error problem; the quality of the instruments used to address it; the order of the polynomial in total expenditure; the choice of k in the k-FWE rate. The results are reported in Table B1 for a polynomial in ρ^{ij} of order 2, which gives the best results. To understand them it is important to remember that the risk aversion parameter is equal to 1.2 for fifteen households and to 2.5 for the remaining fifteen. In the simulation there are therefore 225 pairs with heterogeneous preferences and risk-sharing functions that cross, and 210 pairs with homogeneous risk preferences.

Table B1—Simulation Results for the Test of Homogeneity in Risk Preferences

Null Hypothesis	Homogeneity in Risk Preferences		
	Average Number of False Hypotheses Rejected	Average Number of True Hypotheses Rejected	Empirical FWE Rate
Large Instrument Set			
$\sigma_m = 0.0$, k=11	224.4/225 (99.7%)	1.8/210 (0.9%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=11	142.0/225 (63.0%)	1.1/210 (0.5%)	0.01
$\sigma_m = 0.0$, k=22	224.5/225 (99.8%)	5.8/210 (2.8%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=22	162.6/225 (72.2%)	5.3/210 (2.5%)	0.01
$\sigma_m = 0.0$, k=44	224.5/225 (99.8%)	15.2/210 (7.2%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=44	180.1/225 (80.0%)	20.8/210 (9.9%)	0.01
Small Instrument Set			
$\sigma_m = 0.0$, k=11	216.3/225 (96.0%)	1.1/210 (0.5%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=11	52.6/225 (23.4%)	0.38/210 (0.2%)	0.00
$\sigma_m = 0.0$, k=22	217.5/225 (96.7%)	4.4/210 (2.1%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=22	85.3/225 (37.9%)	2.5/210 (1.2%)	0.0
$\sigma_m = 0.0$, k=44	219.1/225 (97.4%)	14.2/210 (6.8%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=44	115.7/225 (51.4%)	12.3/210 (5.8%)	0.0

Note: The results are obtained by simulating an economy with thirty households that share risk efficiently. Fifteen households have a coefficient of risk aversion equal to 1.2 and fifteen households have a coefficient of risk aversion equal to 2.5. There are therefore 210 pairs for which the null of identical risk preferences is satisfied and 225 pairs for which the null is violated.

We start the description of the results by discussing the effect of the measurement errors on the outcome of the test. If the standard deviation of the measurement errors is negligible, the test can easily detect the existence of heterogeneity in risk preferences. The average rejection of false hypotheses is 224.5 out of 225. If the standard deviation is large the average number of false hypotheses rejected by the test drops significantly for some specifications of the test. The lowest average is obtained when we employ the small set of instruments and $k = 11$. For this specification the average number of correct rejections is just 23.4% of the total. The number increases significantly if we use the large set of instruments and we set k equal to 44. In this case the average number of correct rejections is 80%. To understand why measurement errors have a substantial effect on the power of the test, observe that the test statistic is constructed using the difference in risk-sharing functions. In the simulation the difference is generally much smaller than the actual expenditures. When the measurement errors are drawn from a distribution with a standard deviation that equals half the standard deviation of household expenditure, the difference in measurement errors dominates the difference in true expenditures in many instances. It is therefore difficult for the test to detect a crossing unless a strong set of instruments is used and k is increased. The measurement errors have very little effect on the number of true hypotheses rejected by the test. The average number of false rejections is always very small with the highest number being 10% of the total for the large set of instruments, high σ_m, $k = 44$.

The simulation study also indicates that the choice of k has an important effect on the power of the test. When the standard deviation of the measurement errors is large the test is too conservative if k is set equal to 11. When we increase k to 22 or 44 we observe a substantial gain in power with little loss in control. For instance when we use the large set of instruments, an increase in k from 11 to 22 raises the average number of correct rejections from 63% to 72.2% of total rejections. At the same time the average number of false rejections increases only slightly from 0.5% to 2.5%. An additional increase in k to 44 increases the power of the test with 80% of false hypotheses rejected. But at same time, the number of true hypotheses rejected becomes four times as large. When we use the small set of instruments, $k = 44$ appears to be the optimal choice.

The test of homogeneity in risk preferences in the ICRISAT will be set up taking into account the results of the simulation study. We expect the standard deviation of the measurement errors to be between the ones considered in the simulation study. Moreover, the set of instruments that will be used produces an average R^2 of about 0.71. Because of all this, we will implement the test using the following specification. We will set k equal to 5% of the total number of individual hypotheses and the order of the polynomial to 2.

B2. Simulations for the Efficiency Test Based on Non-labor Income

In this subsection we will discuss the performance of the efficiency test based on non-labor income. Its evaluation requires the simulation of an economy in which a first group of households share risk efficiently, whereas efficiency is violated for a second group. We will assume that fifteen households behave efficiently and that the remaining fifteen are in autarky and can insure themselves against income shocks only by using the risk-free asset. We therefore have 435 individual hypotheses, 330 of which are false. It is important to point out that autarky with a risk-free asset is only one possible alternative to efficiency. Other alternatives are autarky without savings, cooperation without commitment, and cooperation with asymmetric information. We have chosen autarky with a risk-free asset because it has been shown in the finance literature that in this environment households can achieve a degree of insurance similar to the degree that can be achieved in an economy with efficient risk sharing.

The ability of households in autarky to insure themselves against income shocks using the risk-free asset depends on the properties of the non-labor income process. In the simulation we consider a process that attempts to replicate non-labor income in the ICRISAT data. It is assumed that the process is distributed according to a normal distribution with a mean that depends on lagged non-labor income, mean adult age, caste, and number of infants. This specification enables us to capture the fact that non-labor income is highly persistent in the data: everything else equal, an increase in lagged non-labor income by 100% increases current non-labor income by about 50%. Using this specification and the ICRISAT data we can estimate the mean and variance of the process. We can then compute the probability of drawing different realizations for non-labor income.

The simulation results for the first efficiency test differ from the results obtained from the homogeneity test in two respects. First, in the efficiency test the measurement errors have a smaller effect on the outcome of the test. To provide the intuition behind this result note that the present test is based on the estimated risk-sharing functions and not on their differences. Consequently, measurement errors drawn from the same distribution have a smaller effect on the test. For this reason we only report the outcome of the simulation study for the large set of instruments. The second difference is that the best performance is obtained with a polynomial in $\rho^{i,j}$ of order 4. To understand why this test requires a polynomial of higher order, observe that households that share risk efficiently have risk-sharing functions that are approximately concave in total expenditure after one takes the natural logarithm of household expenditure. A polynomial of order 2 is therefore the optimal choice for this group. The group of inefficient households, however, have risk-sharing functions with a more complicated functional form. As a consequence they require a polynomial of higher order. A polynomial of order 4 enables us to approximate in the best possible way the risk-sharing functions of both groups of households. We only report the results for this specification.

Table B2 describes the outcome of the simulation study. If the standard deviation of the measurement errors is negligible, the test is able to reject almost all false hypotheses. When $k = 11$ we reject on average 328.5 false hypotheses out of 330. When k is set equal to 22 or 44 we reject all false hypotheses. When we increase the standard deviation, the average number of correct rejections decreases but only slightly with 314.1 average rejections for $k = 11$, 328.7 rejections for $k = 22$, and 329.8 rejections for $k = 44$. The average number of true hypotheses rejected is small for both specifications of the measurement errors.

TABLE B2—SIMULATION RESULTS FOR THE TEST OF EFFICIENCY BASED ON NON-LABOR INCOME

Null Hypothesis	Efficiency Based on Non-labor Income		
	Average Number of False Hypotheses Rejected	Average Number of True Hypotheses Rejected	Empirical FWE Rate
$\sigma_m = 0.01$, k=11	328.5/330.0 (99.4%)	0.4/105.0 (0.4%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=11	314.1/330.0 (95.2%)	1.8/105.0 (1.7%)	0.0
$\sigma_m = 0.01$, k=22	329.6/330.0 (99.9%)	1.7/105.0 (1.7%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=22	328.7/330.0 (99.6%)	1.7/105.0 (1.7%)	0.0
$\sigma_m = 0.01$, k=44	329.9/330.0 (100.0%)	9.7/105.0 (9.0%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=44	329.8/330.0 (99.9%)	5.6/105.0 (5.0%)	0.0

Note: The results are obtained by simulating an economy with thirty households. Fifteen of them share risk efficiently, whereas the remaining fifteen are in autarky. There are therefore 105 pairs for which the null of efficiency is satisfied and 330 pairs for which the null is violated.

The results of the simulation study are used to set up the test that will be used to evaluate efficiency in Indian villages. The test will be implemented using a polynomial of order 4 in ρ^{ij} and a k that corresponds to 5% of total hypotheses.

B3. Simulations for the Efficiency Test Based on Increasing Functions

In this section we will describe the performance of the efficiency test based on increasing functions. Similarly to the non-labor income test, we simulate an economy in which fifteen households share risk efficiently and fifteen households are in autarky.

The computation of the test statistics for the present test requires a choice for the smoothing parameter m. We have experimented with $m = 10$ and $m = 15$. When the measurement errors have a negligible standard deviation the power of the test is maximized without sacrificing control when $m = 10$. If we increase this parameter to 15, however, the test becomes to conservative and we reject too few false hypotheses. When the measurement errors have a large standard deviation, we reject too many true hypotheses if m is set 10 because the impact of the outlying data is too large. With the high standard deviation we obtain the best balance between power and control when we set $m = 15$. In the first part of Table B3, we report the results for $m = 10$ for the low variance case and for $m = 15$ for the high variance case.

As mentioned above, by construction the performance of the test is not affected by the choice of the order of the polynomial, but it is by the choice of k for the k-FWE rate. We report our findings in Table B3 for the two specifications of the measurement errors and for k equal to 11, 22, and 44. The test has good power and control when the variance of the measurement errors is low. Its performance is not as good when the variance of the measurement errors is large. In this case,

Table B3—Simulation Results for the Test of Efficiency Based on Increasing Risk-sharing Functions

Null Hypothesis	Efficiency Based on Increasing Functions		
	Average Number of False Hypotheses Rejected	Average Number of True Hypotheses Rejected	Empirical FWE Rate
$\sigma_m = 0.01$, k=11, m=10	265.8/330.0 (80.6%)	1.7/105.0 (1.6%)	0.01
$\sigma_m = 0.5 * \sigma_{exp}$, k=11, m=15	147.0/330.0 (44.5%)	0.04/105.0 (0.0%)	0.0
$\sigma_m = 0.01$, k=22, m=10	307.1/330.0 (93.1%)	2.0/105.0 (1.9%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=22, m=15	273.4/330.0 (82.9%)	2.8/105.0 (2.6%)	0.02
$\sigma_m = 0.01$, k=44, m=10	319.5/330.0 (96.8%)	2.2/105.0 (2.1%)	0.0
$\sigma_m = 0.5 * \sigma_{exp}$, k=44, m=15	316.1/330.0 (95.8%)	19.8/105.0 (18.9%)	0.09
An Illustration of the Effect of Changing the Smoothing Parameter m			
$\sigma_m = 0.01$, k=22, m=8	313.1/330.0 (94.9%)	16.2/105.0 (15.4%)	0.20
$\sigma_m = 0.01$, k=22, m=9	311.1/330.0 (94.3%)	7.9/105.0 (7.6%)	0.03
$\sigma_m = 0.01$, k=22, m=10	307.1/330.0 (93.1%)	2.0/105.0 (1.9%)	0.0
$\sigma_m = 0.01$, k=22, m=11	299.6/330.0 (89.5%)	1.8/105.0 (1.7%)	0.0
$\sigma_m = 0.01$, k=22, m=12	281.3/330.0 (85.3%)	0.2/105.0 (0.2%)	0.0

See note in Table B2.

it is difficult for the test to detect the non-monotonicity that characterizes the false hypotheses if k is low. When k is large there is loss of control with the rejection of too many true hypotheses. The choice of $k = 22$ appears to achieve the best balanced between power and control.

In the second part of Table B3, we describe the effect of changing the smoothing parameter m. Specifically, we fix the standard deviation of the measurement errors to the lower level and k to 22 and then we vary m from 8 to 12. The results show that an increase in the smoothing parameter reduces the power of the test by decreasing the number of false hypotheses that are rejected, but it has the beneficial effect of of reducing the probability of a type I errors by decreasing the number of true hypotheses rejected by the test.

Given the results obtained with the simulations, we will test efficiency based on increasing functions in the ICRISAT by setting the smoothing parameter to 10 and k equal to 5% of total individual hypotheses.

REFERENCES

Fan, Yanqin, and Qi Li. 1996. "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms." *Econometrica*, 64(4): 865–890.

Hall, Peter, and Nancy E. Heckman. 2000. "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions." *The Annals of Statistics*, 28(1): 20–39.

Ichimura, Hidehiko. 1993. "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models." *Journal of Econometrics*, 58(1-2): 71–120.

Newey, Whitney K., James L. Powell, , and Francis Vella. 1999. "Nonparametric Estimation of Triangular Simultaneous Equations Models." *Econometrica*, 67(3): 565–603.

Ravallion, Martin, and Shubham Chaudhuri. 1997. "Risk and Insurance in Village India: Comment." *Econometrica*, 65(1): 171–184.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2007. "Formalized Data Snooping Based on Generalized Error Rates." *Econometric Theory*, 24(2): 404–447.