Who is Coming to the Experiment? A Cautionary Tale from China

Paul Frijters
School of Economics
University of Queensland

Tao Sherry Kong
Institute of Social Science Survey
Peking University

Elaine M. Liu*
Department of Economics
University of Houston

* Corresponding author. Email: emliu@uh.edu. Tel: 713-743-3861/Fax 713-743-3798.

Abstract

We compare the characteristics and regression coefficients between the participants in a field experiment in China and the survey population from which they were recruited. The experimental participants were more educated, younger, more likely to be male, more risk-loving and work fewer hours than the more general population. The estimates of their regression coefficients in the standard analyses of wages, happiness and entrepreneurship differed significantly from non-participants, indicating that inferences drawn from experimental samples may not hold for more representative groups of the population.

I. Introduction

Choice experiments are increasingly relied on in economics to glean information about subjects' preferences and behavioural traits. A key question about choice experiments, whether in the lab or in the field, is whether there is external validity of the results. This hinges on whether the behaviour displayed is natural[1] and on whether the subjects who participate in the experiments are typical. If not, then inferences drawn from experiments might not be generalizable.[2] Regarding potential selectivity, most lab experiments conducted in Universities are done with university students who are not representative of the entire population, and it is even unclear whether the selected students who decide to participate in lab experiment are representative of the student population itself. For example, Camerer and Lovallo (1999) find that students' entry into competition game depends on their level of overconfidence. In contrast, Falk et al (forthcoming) find that students behave similarly to the general population in social preferences including prosocial behaviors and trust.

'Artefactual field experiments'[3], which recruit real world subjects, are a partial response to the criticism that students are not normal people. Yet, artefactual field experiments could possibly face worse selection biases in recruitment, because the pool of potential subjects is then more heterogenous in things like their time constraint and education level than students;

[1] This paper will not examine whether the behaviour of people in an experiment differs from the behaviour of the same people in a real world situation (see Ashraf et al, 2006; Liu, *forthcoming*, and Schecters,2007).

2 See Harrison and List (2004) for more elaborate details of this criticism and Deaton (2009) for a wider discussion of the limits of what can be learned by experiments. Harrison et al (2009) show that providing non-stochastic show-up fees for the experiment participates can lead to a selection of more risk averse subjects. Lazear et al (2012) find that giving experimental subjects a choice of opt-out from an experiment on sharing can significantly affect the measure of social preference. Levitt and List (2007) suggest that experiments using university students as subjects may understate the importance of social preference in the real world. von Gaudecker et al. (2011) find that sampling from university student population lead to lower estimates of risk aversion and loss aversion compared to sampling from general population. In contrast, there are also several papers suggesting that the issue of self-selection into experiments may be negligible (Cleave et al. 2011; Falk et al. forthcoming; Andersen et al 2012).

[3] In the taxonomy of Harrison and List (2004), artefactual field experiments include a conventional lab experiment but with a nonstandard subject pool.

therefore, the underlying characteristics of experiment participants (EP) in artefactual field experiment could differ from that of non experiment participants (NEP) more strongly than one students will differ from another. It is therefore an open question just how representative participation in artefactual field experiments is. The objectives of this paper are then twofold: to investigate the sample selection bias of an artefactual field experiment, and to ascertain the bias in standard analyses from ignoring this selectivity.

In most experimental settings, one has no control group to say something about the characteristics of those not in the experiment. The unique design of our dataset allows us to address this issue. The dataset we use as our base is the Urban Migrant Survey (UMS) that includes 5,240 migrant households across 15 cities in China in 2009. This survey contains detailed information on many variables, including education, income, wealth, birth order, family size, a general risk preference measure, and general trust measures.

Within months after the general survey, we invited all household heads from surveyed households to participate in an artefactual field experiment.[4] This experiment, which lasted about 2.5 hours, elicited particular traits—such as risk preferences, time preferences and trust measures, as well as their cognitive ability. The fixed show-up fee was the equivalent of more than a half day's average wage (50 Yuan), and participants earned additional income depending on the outcome of the experiments. Despite our efforts to make the experiment accessible by scheduling multiple sessions in the evenings and on the weekends, and along with the non-

[4] To our knowledge, this is the first large-scale panel dataset to include such an elaborate version of an experimental module with real lottery. Most other large, representative samples including the Italian Survey of Household Income and Wealth (SHIW) (Guiso et al. ,2002; Diaz-Serrano and O'Neill, 2004), Health and Retirement Survey (Barsky et al, 1997) and CentER Savings Survey in Netherlands (Donkers et al, 2001) measure risk preferences with abstractly-framed, hypothetical lotteries. German Socioeconomic Panel (SEOP) is unique since 450 subjects also answer a set of lottery choices with real payoff. Unlike our experiment, SEOP experiment module was done with trained interviewers visiting subjects' house (Dohmen et al, 2011).

negligible compensation, only one third of the initial survey respondents decided to participate in the experiments.

Using the UMS dataset, we examine the differences in characteristics between experiment participants (EP) and non experiment participants (NEP). This directly allows us to see to what extent there is selectivity in terms of whom shows up for field experiments, with the main theoretical prediction being that those with a higher opportunity cost of time and those who are less trusting are less likely to show up.

We look at whether selectivity matters by checking several of the most commonly-examined relations in the field of labor economics, where we look for parameter heterogeneity in the coefficient estimates across EP / NEP groups. If the coefficient estimates for the two groups are drastically different, then we interpret this as selectivity on unobservable differences in relationships. It would mean that the relations found between observed characteristics of lab participants do not provide an inference for the same relationship in the underlying population.

We find that the EPs are, on average, more likely to be male, younger, less likely to be self-employed, more educated, have higher risk tolerance and work fewer hours. This is in line with what one would expect from the point of view of experimental subjects being more risk taking. Whilst we expected a lack of trust in the survey agency to play a role, we do not find any difference in the self-assessed trust measure between EP and NEPs. Nor do we find any difference in hourly wages and wealth, which means an opportunity cost interpretation of the selectivity is imperfect. If anything, we mainly seem to select on whether someone is young and adventurous.

We then run standard Mincerian wage-regressions, as well as standard analyses of subjective wellbeing and entrepreneurship. We find statistically robust evidence that the results

from regressions we often see in the literature differ for the NEP versus the EPs, though in absolute terms the differences are small.

The paper proceeds as follows: Section 2 describes the survey methodology and the experimental methodology. Section 3 gives summary statistics and analyses the participation decision. Section 4 provides evidence of selection on underlying relations. Section 5 concludes.

2. The Survey and Recruitment into the Experiment

The dataset used in this paper is from the 2009 Urban Migrant Survey (UMS) in China.[5] This survey is an on-going 5-year project with the first wave collected in 2008. The Urban Migrant Survey covers 15 cities across nine provinces or metropolitan areas, namely Shanghai, Guangdong, Jiangsu, Zhejiang, Anhui, Hubei, Sichuan, Chongqing and Henan. The survey team developed a listing-sampling scheme, selecting 5,240 migrant households randomly from a mini-census of representative enumeration areas in each city. The face-to-face interviews elicited detailed information about every family member living in the surveyed household, and included questions about individual characteristics, work, family expenditures and income, health, social networks, and subjective risk preference and trust measures. The 2009 wave took place between March/April to July of 2009.[6] All participating households in the 2009 UMS survey form the *population* from which we recruited participants for artefactual-field experiments.

Each household head of the 2009 survey received multiple calls/texts starting in mid-June.[7] They were reminded that they had participated in the UMS survey and were invited to participate in an individual preference experiment (*geren pianhao diaocha).* They would receive

[5] This is a part of the Rural–Urban Migration in China and Indonesia (RUMiCI) research project, which includes a longitudinal survey in both countries. For more details about RUMiC sample scheme, see http://rumici.anu.edu.au/ or Chapter 7 of Meng et al. (2011).
[6] Within each city, all data collection occurred within a one-month period.
[7] Some cities started the experiment as late as Sep 2009.

50 Yuan as a show-up fee to compensate for transportation costs and time. They were told the experiment would take approximately 2.5 hour and there would be an additional payoff depending on the outcomes of strategic games and lotteries. If they could not participate in person, we asked if their spouses would be able to come. If they agreed to participate, they were given a list of available sessions to come to the designated venue, which were usually in rented school classrooms or conference rooms at hotels in the center of each city.

The lab sessions were mostly on the weekend with some during the week days, to suit a variety of working schedules. Out of the 5240 households we contacted, 1745 households sent a participant. There are 52 household members who showed up for the experiments but were themselves not in the survey so were not included in the EP sample. Therefore, our EP sample consists of 1693 subjects-- 1601 household heads (94.5%), 74 spouses (4.4%) and 18 (1.1%) other household members.[8]

Each of the experiment participants would complete experimental modules consisting of 4 parts--- including a set of lotteries eliciting risk preferences[9], a set of lotteries eliciting time preferences, a game eliciting their attitudes on trust and trustworthiness[10], and a 30-min Raven's Progressive Matrices test measuring their cognitive ability. Since this paper focuses on selection, we omit a detailed discussion of these experiments in this paper, and simply mention that they were fairly standard economic experiments.

Since the EP sample contained a particular proportion of household heads, spouses and others family members, we had to construct an appropriate NEP sample as the relevant baseline population. We first restricted our NEP sample to all those household members who were present at the survey, meaning that we deleted the proxy-respondents. Since our primary

[8] A further 40 participants were dropped due to missing data.
[9] The design of the lottery is similar to the lottery design by Holt and Laury (2002) and Dohmen et al (2011).
[10] We use an experimental protocol similar to Burks et al's (2003).

experimental targets were the household heads, we included all 3,322 household heads who did not show up for the experiments but have participated in the survey in the NEP sample.[11] Given that 4.4% of all EP are spouses, we randomly selected 154 spouses from the primary NEP population to keep the proportion of spouse to household head the same in both EP and NEP sample. We performed the same procedure with respect to the selection of other relatives from NEP participants. In total, the NEP sample includes 3513 individuals.[12]

3. Who Comes to the Experiment?

3.1 Differences in descriptive

To investigate whether there is a difference in the characteristics between the EP and NEP, Table 1 presents the summary statistics of the UMS 2009 sample by whether they participated in the experiments or not.

[Insert Table 1 About Here]

Since the objective of the experiment was to elicit attitudes of risk and trust, it would be especially worrisome if the EP and NEP differs in this dimension. In the survey, the risk-taking question was self-assessed: "*Some people in a society are more likely to take risk while others are less likely. On a scale from zero to ten, where zero indicates do not take any risks and ten indicates you like to take risks very much, could you please rank your own risk taking level"*. Table 1 shows that experimental participants report a significantly higher degree of average risk-taking (4.351 versus 4.096).

There were two measures of trust in the survey, one self-assessed and one derived from revealed behaviour. The trust question based on self assessment was: *"In general, do you think*

[11] The number does not equal to 5240-1693= 3547 because some heads of household were not present at the time of the survey.
[12] We do not expect NEP + EP to be 5240 for the same reason as in Footnote 11.

most people can be trusted or one should be careful when he/she deals with others. Select either 'Yes, most people can be trusted'; 'No, one should be careful'; or 'Do not know'." The revealed choice measure went: *"Do you lend money or objects to your friends or acquaintances? Select either 'usually'; 'sometimes'; 'seldom'; 'never'."* We find that experiment subjects are significantly more likely to have lent money to friends, but did not self-assess as having a higher or lower level of general trust than non-participants.[13]

The EP sample also differs from NEP in other dimensions: the experiment participants are younger, more likely to be migrants who are from the same province where the experiment took place, have more years of schooling, have slightly bigger social networks[14], are less likely to be self-employed, more likely to be men, and more likely to be single. Surprisingly, they are less healthy and less happy compared to the NEP participants.

Whilst these differences are highly significant, the size of these differences is small in a statistical sense: the difference in risk preferences, self-employment levels, age, and whether the respondent comes from the same province as the city in which (s)he lives, is about $1/10^{th}$ of a standard deviation. The biggest difference is for education with almost $2/10^{th}$ of a standard deviation.

[Insert Table 1 Panel A and Panel B About Here]

3.2 Participation regressions

In Table 2 we examine the difference in underlying characteristics for those who participated in the experiment using a probit model with the dependent variable being the participation in the experiments.

[13] A similar set of risk and trust questions are also used in German Socio-Economic Panel (GSOEP).
[14] The size of the social network is proxied by the answer to the question "how many people who reside in urban area have you contacted/phoned/visited/greeted during Chinese New Year?"

Results in Column 1 show that selection based on trust and risk preferences is an issue if those are the only variables looked at: those who are more risk loving and more likely to lend money to their friends are significantly more likely to show up for the experiments at the 1% and 5% significance level, respectively. These results are robust when we control for other individual characteristics in Column 2. There, we also see that men, more highly educated individuals and individuals who work fewer hours are more likely to participate.

One fear is that we are picking up differences in the labor market conditions in the area where participants live, and that areas differ in terms of who lives there; therefore, we control in Column 3 for the city-level labor market conditions, including average work hours per week and average wages. Whilst this has little effect on the other findings, we do find that in places where average wages are higher or hours of work are longer, people are more likely to attend the experiment.

Another worry is related to the city-level differences associated with the organization of the experiments. The experiments were organized by different local survey agencies in their respective cities.[15] It is possible that some particular survey agencies may work harder than others on persuading their respondents to come to the experiment agencies. In addition, migrants may also have some unobserved characteristics that differ across areas. Therefore, we control for city-level fixed-effects in Column 4. The results show little difference.

Throughout the specifications, when we test for the joint significance of risk and trust measures, they are consistently significant at the 10% level. We also consistently find that male, more educated respondents who work fewer hours are more likely to participate.

[Insert Table 2 About Here]

[15] We hired a private survey company with a main office in Beijing and several branches in China to manage the survey collection and experiment process. In different cities, different persons would be in charge, potentially allowing for a divergence in recruitment effort.

4. Possible impact of the experimental selection on outcome variables

In the previous section, we found some underlying differences in the characteristics of the EP and NEP samples. In this section we examine whether the self- selection bias in the experimental sample affects the representativeness of regression coefficients of interest. In a statistical sense this means we look at whether the selectivity is on the unobservables whose effect includes an inter-action with observables.

4.1 The correlates of wages, entrepreneurship, and life satisfaction

In this sub-section, we want to test differences in coefficients between the EP and NEP samples for some of the most commonly studied outcomes in the field of applied economics, including earnings, entrepreneurship and subjective well-being (happiness). The idea is to see whether regression coefficients in the experimental sample differ from the non-experimental sample.

Table 3 presents the results from a standard earnings equation. In the first column we control only for basic individual characteristics and key city level outcomes to wash out city level variation in economic conditions, including the average monthly wage and hours worked. All we look for in Column 1 in terms of selectivity is whether the intercept term differs by sample. The result shows that those who participated in the experiment on average earn 2.9 percent more than the non-participants, which is a statistically significant but modest amount.

If we look at the actual coefficients, our estimate of the returns to schooling is about 3.4%. This is slightly below to Liu and Zhang estimates (2012) of 6% return to schooling using China Health and Nutrition Survey (CHNS) among urban Chinese population, but it is in line with the

finding that migrants stay in the city for less long than urban residents and thus are less likely to have equal benefits from their education.

Consistent with the existing literature, we also find a concave wage-experience relation (Li, 2003; Zhang, 2011); we find that females earn lower hourly wages than men (Li, 2003; Zhang, 2011); and we find that self-assessed height is positively correlated with hourly wage (Gao and Smyth, 2010; Case and Paxson, 2008). Hence, in terms of overall relations, this data concurs with that of others.

In the second column we add an additional city variable which measures the share of the experimental participants in the full sample for each city. The variable intends to capture the effort level of the survey company and other unobserved city-level differences. While the additional variable is statistically significant, including it does not change the intercept for EPs.

In Columns 3 to 5 we look at whether the EP sample differs in terms of general correlates of earnings by having interaction terms for each of the individual-level variables. The most significant interaction variables are age and marriage, showing that older experimental participants earn about 1.8% less per additional year than non-experimental participants who are the same age, whilst the marriage premium on wages are 9.3% higher for experimental participants in the specification of column 3 and 6.9% in the richer specification of column 4. Compared to the coefficients amongst the NEP sample, these two interactions are 50% to 100% of the magnitude, showing that the relation between age/marriage and earnings differs both statistically and substantially between the two samples.

We then test the joint significance level for the participation dummy and all the interaction terms (see F-test at the bottom of Table 2). The joint significant test shows that the null-hypothesis that the probability of two sub-samples being the same in terms of regression

coefficients is rejected at the 1% or even 0.1% significance levels. Column 3 also includes city fixed-effects to take out any unobserved difference between cities, but this does not change the finding that the interactions are jointly significant.

Column 4 adds the three subjective measures of risk and trust to Column 3's specification. This reduces the significance level of F-test somewhat, but it is still significant at the 1 percent level. Finally in Column 5 we threw in a kitchen sink of other individual characteristics, including type of employment contract (long versus short term), employment type (SOE, Private, Foreign, etc), hukou status and size of the social network. This further reduced the F-test slightly, but it remains significant at the 1 percent level.

In total, the results presented in Table 3 seem to indicate that even with a very rich set of observable variables from normal household survey data as well as some "unobservable" variables from normal household survey data, we are unable to wash out the difference in earnings between the participant and non-participant sub-samples, both in terms of levels and in terms of relations between earnings and other characteristics.

[Insert Table 3 About Here]

In Table 4, the dependent variable is whether one is self-employed or not. We use linear probability model instead of probit or logit models since we are interested in the interaction terms between the experiment participation and several covariates. The interpretation of interaction terms in a non-linear model such as probit and logits are less intuitive (Norton et al., 2004). In Columns 1 and 2, we control for basic individual characteristics and city level characteristics, just as with Table 3. In the first 2 columns we find that EPs are 3.7 to 5 percentage points less likely to be self-employed.

In Columns 3, 4 & 5, city fixed effect and a set of experiment interaction terms and city-interaction terms are included, and we again test for the joint significance of all the interaction terms. The null hypothesis that interaction terms are jointly equal to zero is rejected at the 1% level. This holds across specifications, even when we include the subjective trust measures or a large set of additional controls.

[Insert Table 4 About Here]

In Table 5, we examine the relationship between subjective wellbeing and individual characteristics using ordinary least square (OLS). Again even though the dependent variable ranges from 0 (least happy) to 3 (most happy), we use OLS instead of an ordered-probit model for the ease of interpreting the interaction terms. Column 1 reveals that, once we control for individual characteristics, EPs are on average just as happy as NEPs (Column 1). Consistent with numerous previous studies on the determinants of self-reported happiness, we find education (Frey and Stutzer 1999), health (Diener et al. 1999), wealth (Ferrer-i-Carbonell and Frijters 2004), and being married (Argyle, 1999; Gerlach and Stephan, 1996) are all positively associated with higher level of happiness. As with other cross-sectional studies involving many covariates, we find a U-shape in age (eg. Gerdthama and Johannesson, 2001).

When we test for the joint significance of the EP dummy and all EP interaction terms in Column 1, the null hypothesis cannot be rejected at the 10% significance level. In subsequent analyses, we include subjective measure of risk and trust, and additional controls, where the joint significant test for interaction terms means they are significantly different from 0 at the 5% significance level (column 3). Yet, this is not due to a major difference in the coefficient of these observables, but rather due to the importance of the interaction between the EP sample and the

city-intercepts. It is thus the influence of unobserved city characteristics that differs between these two samples.

[Insert Table 5 About Here]

5. Conclusion

Our study compared the characteristics and regression coefficients between participants in a field experiment in China and the survey population they were recruited from. The participants who showed up were more educated, younger, more likely to be male, more risk-loving, and less hard working than the more general population. The regression coefficients in standard analyses of wages, happiness, and entrepreneurship differed significantly, indicating that inferences drawn from experimental samples may not in general hold for more representative groups of the population. Though the difference was usually relatively small, some particular differences in coefficients were substantial. Most notably, we found that amongst the experimental sample, married and younger people had relatively higher wages compared to equivalent individuals who did not show up for experiments. This suggests that those who show up for experiments are risk-takers with more than average returns to their social capital and health capital.

This study thus casts additional doubt on the generalizability of field experiments. More in general, it brings into focus the enormous importance of selectivity in socio-economic research based on voluntary cooperation.

References

Anderson, J., Burks, S., Carpenter, J., Götte, L., Maurer, K., Nosenzo, D., et al. (2012) Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: Evidence from one college student and two adult samples. *Experimental Economics*, , 1-20.

Adams, P., Hurd, M. D., McFadden, D., Merrill, A., & Ribeiro, T. (2003). Healthy, wealthy, and wise? tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics, 112*(1), 3-56. doi:10.1016/S0304-4076(02)00145-8

Argyle, M. (1999). Causes and correlates of happiness. *In Kahneman Et Al.(1999), Chapter 18.*,

Ashraf, N., Karlan, D., & Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines*. *Quarterly Journal of Economics, 121*(2), 635-672.

Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. The Quarterly Journal of Economics, 112(2), 537-579.

Burks, S. V., Carpenter, J. P., & Verhoogen, E. (2003). Playing both roles in the trust game. *Journal of Economic Behavior & Organization, 51*(2), 195-216.

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review, 89*(1), pp. 306-318.

Case, A., & Paxson, C. (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of Political Economy, 116*(3), 499-532.

Cleave, Blair Llewellyn, Nikiforakis, Nikos and Slonim, Robert, Is There Selection Bias in Laboratory Experiments? The Case of Social and Risk Preferences. IZA Discussion Paper No. 5488. Available at SSRN: http://ssrn.com/abstract=176565

Deaton, A. (2010). *Instruments, randomization, and learning about development*

Deaton, A. S. (2009). *Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development*National Bureau of Economic Research.

Demakakos, P., Nazroo, J., Breeze, E., & Marmot, M. (2008). Socioeconomic status and health: The role of subjective social status. *Social Science & Medicine, 67*(2), 330-340. doi:10.1016/j.socscimed.2008.03.038

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (Mar 1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*(2), 276-302.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). *Are risk aversion and impatience related to cognitive ability?*

Dohmen, T. J., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). INDIVIDUAL RISK ATTITUDES: MEASUREMENT, DETERMINANTS, AND BEHAVIORAL CONSEQUENCES. Journal of the European Economic Association, 9(3), 522-550.

Donkers, B., B. Melenberg, and A. V. Soest (2001): "Estimating Risk Attitudes Using Lotteries: A Large Sample Approach," Journal of Risk and Uncertainty, 22(2), 165–195

Falk, Armin, Stephan Meier, and Christian Zehnder. (forthcoming) "Do lab experiments misrepresent social preferences? The case of self-selected student samples." *Journal of the European Economic Association*

Ferrer-i-Carbonell, A., & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness?*. *The Economic Journal, 114*(497), 641-659. doi:10.1111/j.1468-0297.2004.00235.x

Frederick, S. (2005). *Cognitive reflection and decision making*

Frey, B. S., & Stutzer, A. (2000). Happiness, economy and institutions. *The Economic Journal, 110*(466), 918-938. doi:10.1111/1468-0297.00570

von Gaudecker, H., van Soest, A. and Wengstrom E., (2011) Experts in Experiments: How Selection Matters for Estimated Distributions of Risk Preferences. Unpublished Manuscript

Gao, W., & Smyth, R. (2010). Health human capital, height and wages in china. *Journal of Development Studies, 46*(3), 466-484. doi:10.1080/00220380903318863

Gerdtham, U., & Johannesson, M. (2001). The relationship between happiness, health, and socio-economic factors: Results based on swedish microdata. *Journal of Socio-Economics, 30*(6), 553-557. doi:10.1016/S1053-5357(01)00118-4

Gerlach, K., & Stephan, G. (1996). A paper on unhappiness and unemployment in germany. *Economics Letters, 52*(3), 325-330. doi:10.1016/S0165-1765(96)00858-0

Giulietti, C., Ning, G., & Zimmermann, K. F. (2011). Self-employment of rural-to-urban migrants in china. *CEPR Discussion Papers, C.E.P.R. Discussion Papers no. 8473.*(London, Centre for Economic Policy Research. http://www.cepr.org/pubs/dps/DP8473.asp.)

Guiso, L., & Paiella, M. (2008). Risk aversion, wealth, and background risk. Journal of the European Economic Association, 6(6), 1109-1150.

Haizheng, L. (2003). Economic transition and returns to education in china. *Economics of Education Review, 22*(3), 317-328. doi:10.1016/S0272-7757(02)00056-0

Harrison, G. W., & List, J. A. (December 2004). Field experiments. *Journal of Economic Literature, 42*, 1009-1055(47).

Harrison, G. W., Lau, M. I., & Elisabet Rutström, E. (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. Journal of Economic Behavior & Organization, 70(3), 498-507.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review, 92*(5), 1644-1655.

Lazear, E. P., Malmendier, U., & Weber, R. A. (2012). Sorting in experiments with application to social preferences. American Economic Journal: Applied Economics, 4(1), 136-163.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? The Journal of Economic Perspectives, 21(2), 153-174.

Liu, E. (Forthcoming). Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in china. *The Review of Economics and Statistics.*

Liu, E. M. and Zhang, S.(2012). *A meta-analysis of the estimates of returns to schooling in china.* Unpublished manuscript.

Norton, Edward C., Hua Wang, and Chunrong Ai. (2004) "Computing interaction effects and standard errors in logit and probit models." *Stata Journal* 4 (2004): 154-167.

Rong, Z. (2011). Individual heterogeneity in returns to education in urban china during 1995–2002. *Economics Letters, 113*(1), 84-87. doi:10.1016/j.econlet.2011.05.028

Ross, C. E., & Wu, C. (1995). The links between education and health. *American Sociological Review, 60*(5), pp. 719-745. Retrieved from http://www.jstor.org/stable/2096319

Schechter, L. (2007). Theft, gift-giving, and trustworthiness: Honesty is its own reward in rural paraguay. *The American Economic Review, 97*(5), 1560-1582.

Smith, J. P. (1998). Socioeconomic status and health. *The American Economic Review, 88*(2, Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association), pp. 192-196. Retrieved from http://www.jstor.org/stable/116917

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *The Journal of Economic Perspectives, 13*(2), pp. 145-166. Retrieved from http://www.jstor.org/stable/2647123

Smith, J. P., & Kington, R. (1997). Demographic and economic correlates of health in old age. *Demography, 34*(1, The Demography of Aging), pp. 159-170. Retrieved from http://www.jstor.org/stable/2061665

Wyke, S., & Ford, G. (1992). Competing explanations for associations between marital status and health. *Social Science & Medicine, 34*(5), 523-532. doi:10.1016/0277-9536(92)90208-8

Yuen, L. (2009). Self-employment in urban china: Networking in a transition economy. *China Economic Review, 20*(3), 471-484. doi:10.1016/j.chieco.2009.01.002

Zhang, X. (2011). The rate of returns to schooling: A case study of urban china. *International Journal of Humanities and Social Science, 1*(18), 173-180.

Table 1 Panel A: Summary Statistics by Experiment Participation

	(1) Experiment Participants	(2) Not In Experiments	(3) p-value (difference!=0)
Trust and Risk Taking Measures			
Subjective Risk Preference[a]	4.351	4.096	0.0011***
(0=Risk Averse..10=Risk Seeking)	(2.571)	(2.656)	
Often/sometimes lend money to	0.713	0.673	0.0040***
acquaintances and friends[b]	(0.452)	(0.469)	
Trusting People (Yes=1)[c]	0.573	0.572	0.9431
	(0.495)	(0.495)	
Labor Market Outcome			
Self-Employed (yes=1)	0.197	0.237	0.0011***
	(0.397)	(0.425)	
Total monthly wage (1000 Yuan)	1.770	1.817	0.3928
	(1.183)	(2.103)	
Weekly work hours	59.136	61.713	0.0000***
	(19.326)	(19.898)	

Note: Standard deviation is reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

a. Question from the survey: *Some people in a society are more likely to take risk while others are less likely. On a scale from zero to ten, where zero indicates do not take any risks and ten indicates you like to take risks very much, could you please rank your own risk taking level*

b. Question from the survey: *Do you lend money or objects to your friends or acquaintances?* The 4 choices were 'usually'; 'sometimes'; 'seldom'; 'never'. Those who answered usually or sometimes would be assigned 1, otherwise 0.

c. Question from the survey: *In general, do you think most people can be trusted or one should be careful when he/she deals with others.* The choices were "Yes, most people can be trusted"; "No, one should be careful" and "Do not know." Those who answered "Yes, most people can be trusted" would be assigned 1, otherwise 0.

Table 1 Panel B: Summary Statistics by Experiment Participation

	(1) Experiment Participants	(2) Not In Experiments	(3) p-value (difference!=0)
Individual Characteristics			
Female	0.350	0.367	0.2250
	(0.477)	(0.482)	
Age	30.901	31.781	0.0042***
	(9.803)	(10.631)	
Ethnic Minority (yes=1)	0.0171	0.020	0.4890
	(0.130)	(0.140)	
Married (yes=1)	0.537	0.560	0.1134
	(0.499)	(0.496)	
From the same province	0.634	0.561	0.0000***
	(0.482)	(0.496)	
network in cities ('100 persons)[d]	0.0363	0.0365	0.9379
	(0.0732)	(0.092)	
Years of edu	9.710	9.288	0.0000***
	(2.613)	(2.521)	
Log(Wealth)	7.370	7.318	0.3683
	(1.896)	(1.956)	
Time elapsed since migrating (Years)	8.183	8.297	0.5583
	(6.442)	(6.665)	
2009 New Households (Yes=)	0.652	0.655	0.8199
	(0.477)	(0.476)	
Having any child (age<16) resides	0.145	0.146	0.9664
in the same city	(0.353)	(0.353)	
Subjective Wellbeing	2.107	2.142	0.0601*
(0=very unhappy,..,3=very happy)	(0.614)	(0.624)	
Self-Rated Health	4.093	4.140	0.0393**
(1=very bad,..5=very healthy)	(0.777)	(0.740)	
Observations	1693	3513	

Note: Standard deviation is reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

d. Question from the survey: *how many people who reside in urban area have you contacted/phoned/visited/greeted during Chinese New Year?*

Table 2: Probit Model of Experiment Participation

	(1)	(2)	(3)	(4)
Subjective Risk Preference	0.021***	0.017**	0.018**	0.010
	(0.007)	(0.007)	(0.007)	(0.007)
Often lend money to friends	0.104**	0.080**	0.074*	0.091**
	(0.042)	(0.038)	(0.044)	(0.040)
Trusting People	-0.024	-0.049	-0.064*	-0.062*
	(0.040)	(0.040)	(0.039)	(0.037)
Female		-0.049	-0.046	-0.078*
		(0.039)	(0.042)	(0.040)
Age		-0.001	-0.000	0.001
		(0.003)	(0.002)	(0.003)
Ethnic Minority (yes=1)		-0.032	-0.016	0.058
		(0.146)	(0.135)	(0.146)
Married		0.063	0.050	0.035
		(0.050)	(0.051)	(0.049)
Self-Employed		-0.038	-0.081	-0.090
		(0.057)	(0.056)	(0.065)
year of edu		0.032***	0.035***	0.025***
		(0.008)	(0.008)	(0.009)
Total monthly wage in 1000 Yuan		-0.019	-0.012	-0.011
		(0.013)	(0.012)	(0.012)
Weekly work hours		-0.003**	-0.004***	-0.003**
		(0.001)	(0.001)	(0.001)
Average hours work in the city			0.034***	
			(0.007)	
Average Monthly Wage in the City			0.168**	
			(0.076)	
Size of Work Unit Dummy[a]				X
City FE				X
Chi-Test	14.88	11.86	10.98	9.35
Prob > Chi2	0.0019	0.0079	0.0118	0.025
Number of Observations	5,157	4,957	4,957	4,952

Standard errors (in parentheses) are based on 200 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.1. Results of Chi-Test for the joint significance of trust, lend money and self-reported risk preference are reported.

a. dummies breakdown by work unit size: 1-20 persons, 20-50 persons, 50-99 persons, 100-999 persons, 1000 above)

Table 3: Labor Market Outcome between EP and NEP

	Dep Variable: Log(Hourly Wage)				
	(1)	(2)	(3)	(4)	(5)
Experiment Participants (EP)	0.029**	0.035***	0.708	0.720*	0.626
	(0.012)	(0.012)	(0.438)	(0.421)	(0.422)
Age	0.030***	0.029***	0.037***	0.037***	0.036***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
(Age)2	-0.000***	-0.000***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.060***	-0.059***	-0.063***	-0.060***	-0.055***
	(0.016)	(0.016)	(0.020)	(0.020)	(0.020)
Ethnic Minority (yes=1)	0.005	0.003	0.036	0.031	0.025
	(0.040)	(0.040)	(0.050)	(0.049)	(0.048)
Years of Education	0.034***	0.034***	0.035***	0.031***	0.025***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Married	-0.012	-0.011	-0.059***	-0.033*	-0.016
	(0.016)	(0.016)	(0.021)	(0.020)	(0.022)
self-assessed height	0.004***	0.004***	0.005***	0.004***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
EP*Age			-0.018**	-0.017**	-0.015*
			(0.009)	(0.008)	(0.008)
EP*$(Age)^2$			0.000**	0.000*	0.000*
			(0.000)	(0.000)	(0.000)
EP*Female			-0.006	-0.007	0.001
			(0.036)	(0.035)	(0.035)
EP*Non-Han			-0.042	-0.026	-0.043
			(0.091)	(0.087)	(0.088)
EP*Education			0.004	0.006	0.005
			(0.005)	(0.005)	(0.005)
EP*Married			0.093**	0.069**	0.050
			(0.036)	(0.035)	(0.035)
EP*Height			-0.003	-0.002	-0.002
			(0.002)	(0.002)	(0.002)
Subjective Risk Preference				0.007***	0.005**
(0=Risk Averse..10=Risk Seeking)				(0.002)	(0.002)
Trusting People (Yes=1)				0.015	0.014
				(0.011)	(0.011)
Often/sometimes lend money				0.033***	0.030**
to acquaintances and friends				(0.012)	(0.012)
Size of Network					0.301***
					(0.065)
Average Monthly Wage in the City	0.290***	0.298***			
	(0.022)	(0.023)			
Average hours work in the city	-0.014***	-0.013***			
	(0.002)	(0.002)			
Share of survey participants		-0.128**			

		(0.057)			
City Fixed Effect (FE)			X	X	X
EP * City Dummies			X	X	X
Additional Controls[a]					X
Observations	4,948	4,948	4,948	4,901	4,746
Adjusted R-squared	0.194	0.195	0.141	0.207	0.232
F-test			10.73	2.351	1.728
Prob > F			0.0000	0.000360	0.0186

*** p<0.01, ** p<0.05, * p<0.1. In Columns 3, 4 & 5, results of F-test for the joint significance of all in-experiment interaction terms and in-experiment dummy are reported.

a. Additional controls include type of employment contract (long versus short term), employment type (SOE, Private, Foreign, etc), hukou status & size of network

Table 4: Self-Employment between EP and NEP

	Dep Variable: Self Employed				
	(1)	(2)	(3)	(4)	(5)
Experiment Participants (EP)	-0.037***	-0.050***	0.018	0.010	0.016
	(0.011)	(0.011)	(0.096)	(0.096)	(0.096)
Age	0.002***	0.002***	0.000	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.029***	0.024**	0.024*	0.023*	0.014
	(0.011)	(0.011)	(0.014)	(0.014)	(0.017)
Ethnic Minority (yes=1)	-0.023	-0.018	-0.056	-0.053	-0.063
	(0.038)	(0.038)	(0.046)	(0.046)	(0.046)
year of edu	-0.018***	-0.019***	-0.017***	-0.017***	-0.015***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Married	0.139***	0.137***	0.160***	0.159***	0.130***
	(0.014)	(0.014)	(0.017)	(0.017)	(0.019)
Log(Wealth)	0.065***	0.065***	0.067***	0.067***	0.069***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Years since Migrating to City	0.004***	0.005***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
EP*Age			0.005***	0.005***	0.005***
			(0.002)	(0.002)	(0.002)
EP*Female			-0.010	-0.011	-0.012
			(0.024)	(0.024)	(0.024)
EP*Non-Han			0.147*	0.144*	0.155*
			(0.082)	(0.082)	(0.082)
EP*Education			-0.003	-0.003	-0.003
			(0.005)	(0.005)	(0.005)
EP*Married			-0.069**	-0.067**	-0.066**
			(0.031)	(0.031)	(0.031)
EP*Subjective Risk Preference			0.002	0.002	0.003
			(0.005)	(0.005)	(0.005)
EP*Log(Wealth)			-0.015**	-0.015**	-0.017**
			(0.007)	(0.007)	(0.007)
EP*Years resides in Urban Area			-0.004**	-0.004**	-0.004*
			(0.002)	(0.002)	(0.002)
Subjective Risk Preference	0.003	0.002	0.002	0.002	0.002
(0=Risk Averse..10=Risk Seeking)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Trusting People (Yes=1)				-0.037***	-0.038***
				(0.011)	(0.011)
Often/sometimes lend money				-0.001	0.000
to acquaintances and friends				(0.012)	(0.012)
network in cities					-0.067
					(0.062)
Number of Children					0.033***
					(0.010)
unhealthy					-0.004

					(0.039)
self-assessed height					-0.001
					(0.001)
Average Monthly Wage in the City	0.054**	0.036*			
	(0.022)	(0.022)			
Average hours work in the city	0.019***	0.016***			
	(0.002)	(0.002)			
Share of survey participants		0.312***			
		(0.055)			
City Fixed Effect (FE)			X	X	X
In-Exp * City Dummies			X	X	X
Additional Controls[a]					X
F-test			3.83	3.82	3.87
Prob > F			0.000	0.000	0.000
Observations	4,939	4,939	4,939	4,939	4,932

*** p<0.01, ** p<0.05, * p<0.1. Coefficients from linear probability model are reported. In Columns 3, 4 & 5, results of F test for the joint significance of all in-experiment interaction terms and in-experiment dummy are reported. All regression includes a dummy indicating whether wealth is reported as zero or not.

a. additional controls include size of network in urban area, number of children, hukou status, self-reported health & height. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Determinants of Subjective Wellbeing

	dependent Variable: happiness (0 least happy ..3 most happy)		
	(1)	(2)	(3)
Experiment Participants (EP)	-0.001	0.036	0.061
	(0.158)	(0.157)	(0.158)
Age	-0.006***	-0.005***	-0.006***
	(0.001)	(0.001)	(0.002)
Female	-0.013	-0.008	-0.002
	(0.022)	(0.022)	(0.029)
Ethnic Minority (yes=1)	0.020	0.017	-0.008
	(0.077)	(0.077)	(0.077)
year of edu	0.012**	0.010**	0.009*
	(0.005)	(0.005)	(0.005)
Married	0.132***	0.133***	0.142***
	(0.029)	(0.028)	(0.032)
Log(Wealth)	0.022***	0.021***	0.015**
	(0.007)	(0.007)	(0.007)
Total monthly wage in 1000 Yuan	0.007	0.007	0.007
	(0.005)	(0.005)	(0.005)
unhealthy	-0.438***	-0.439***	-0.435***
	(0.083)	(0.082)	(0.082)
network in cities	0.136	0.123	0.115
	(0.120)	(0.120)	(0.120)
EP*Age	0.003	0.003	0.003
	(0.003)	(0.002)	(0.003)
EP*Female	0.029	0.029	0.044
	(0.040)	(0.040)	(0.040)
EP*Non-Han	-0.233*	-0.222	-0.154
	(0.138)	(0.138)	(0.141)
EP*Education	0.007	0.006	0.002
	(0.008)	(0.008)	(0.008)
EP*Married	-0.033	-0.038	-0.055
	(0.050)	(0.050)	(0.051)
EP*Log(wealth)	-0.001	-0.002	-0.001
	(0.012)	(0.012)	(0.013)
EP*Total Monthly Wage	0.012	0.013	0.015
	(0.015)	(0.015)	(0.015)
EP*Unhealthy	0.014	0.016	0.016
	(0.135)	(0.134)	(0.133)
EP*Network in City	-0.026	-0.032	-0.034
	(0.246)	(0.244)	(0.245)
Subjective Risk Preference		-0.002	-0.002
(0=Risk Averse..10=Risk Seeking)		(0.004)	(0.004)
Trusting People (Yes=1)		0.123***	0.115***
		(0.018)	(0.018)
Often/sometimes lend money		0.068***	0.072***
to acquaintances and friends		(0.019)	(0.020)

self-assessed height			0.001
			(0.002)
City Fixed Effect (FE)	X	X	X
In-Exp * City Dummies	X	X	X
Additional Controls[a]			X
F-test	1.31	1.44	1.59
Prob > F	0.13	0.07	0.03
Observations	4,965	4,965	4,798

*** p<0.01, ** p<0.05, * p<0.1. Estimates from ordinary least square. The dependent variable is subjective measure of wellbeing (0=not happy,..3=very happy). In Columns 1, 2 &3, results of F-test for the joint significance of all in-experiment interaction terms and in-experiment dummy are reported.

a. additional control includes height, type of contracts, employment types, hukou status, size of network, number of children & hours worked last week.

Appendix Table: Labor Market Outcome between EP and NEP including Only Wage Earner

	Dep Variable: Log(Hourly Wage)				
	(1)	(2)	(3)	(4)	(5)
Experiment Participants (EP)	0.049***	0.056***	0.560	0.641	0.539
	(0.012)	(0.012)	(0.414)	(0.398)	(0.393)
Age	-0.002**	-0.002**	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.031*	-0.030*	-0.021	-0.016	-0.005
	(0.016)	(0.016)	(0.021)	(0.020)	(0.020)
Ethnic Minority (yes=1)	0.024	0.021	0.070	0.055	0.057
	(0.040)	(0.040)	(0.048)	(0.046)	(0.045)
Years of Education	0.036***	0.037***	0.039***	0.035***	0.028***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Married	0.073***	0.073***	0.039**	0.050***	0.061***
	(0.015)	(0.014)	(0.019)	(0.018)	(0.019)
self-assessed height	0.006***	0.006***	0.007***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
EP*Age			-0.002	-0.003*	-0.003*
			(0.002)	(0.002)	(0.002)
EP*Female			-0.034	-0.040	-0.034
			(0.036)	(0.035)	(0.034)
EP*Non-Han			-0.081	-0.060	-0.110
			(0.093)	(0.088)	(0.089)
EP*Education			-0.001	0.002	0.002
			(0.005)	(0.005)	(0.005)
EP*Married			0.079**	0.072**	0.064**
			(0.032)	(0.030)	(0.030)
EP*Height			-0.003	-0.003	-0.002
			(0.002)	(0.002)	(0.002)
Subjective Risk Preference				0.004*	0.001
(0=Risk Averse..10=Risk Seeking)				(0.002)	(0.002)
Trusting People (Yes=1)				-0.002	-0.001
				(0.011)	(0.011)
Often/sometimes lend money to acquaintances a				0.027**	0.024**
				(0.012)	(0.012)
Size of Network					0.264***
					(0.064)
Average Monthly Wage in the City	0.282***	0.296***			
	(0.023)	(0.023)			
Average Work Hours in the City	-0.013***	-0.011***			
	(0.002)	(0.002)			
Share of survey participants as EP		-0.163***			
		(0.057)			
City Fixed Effect (FE)					
EP * City Dummies					
Observations	3,888	3,888	3,888	3,846	3,702

R-squared	0.205	0.206	0.150	0.218	0.264
F-test			11.89	3.395	3.128
Prob > F			0.000	0.000	0.000

Standard errors (in parentheses) are based on 200 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.1. In Columns 3, 4 & 5, results of F-test for the joint significance of all in-experiment interaction terms and in-experiment dummy are reported. Column 5 also control for type of employment contract (long versus short term), employment type (SOE, Private, Foreign, etc), hukou status & size of network