UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Applied Microeconomics

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Economics

by

Carolina Mejía Mantilla

2013

ABSTRACT OF THE DISSERTATION

Essays on Applied Microeconomics

by

Carolina Mejía Mantilla

Doctor of Philosophy in Economics

University of California, Los Angeles, 2013

Professor Kathleen McGarry, co-Chair

Professor Adriana Lleras-Muney, co-Chair

Each chapter of this dissertation studies a different question within the field of Applied Microeconomics. The first chapter examines the mid- and long-term effects of the 1998 Asian Crisis on the educational attainment of Indonesian children ages 6 to 18, at the time of the crisis. The effects are identified as deviations from a linear trend for specific age groups using data from the Indonesian Family Life Survey (IFLS) and Census data. Contrary to previous studies, I find that the crisis had negative mid-term effects on the probability of attending school (5 to 7 percentage points) and grade progression (5 to 14 percentage points) of older children, aged 13 to 18 at the time of the crisis. Similarly, these children lost around 0.5 years of education and increased the numbers of hours worked, suggesting that the income effect (lower income) dominated the substitution effect (lower opportunity cost of school). More importantly, the evidence points to large long-term negative effects of around 1.5 years of education for these same children, about twice the increase of the average educational attainment in the last decade. There are also adverse long-term effects

on high-school graduation rates and real wages; where the latter are largely (but not fully) explained by the lower educational attainment.

The second chapter explores the causal effect of peer feedback on the teaching performance of graduate teaching assistants (TAs) using a Randomized Control Trial (RCT).[1] The participants of the intervention were the TAs of the Department of Economics of a large public university, and the duration of the intervention was one academic quarter. We analyzed the students' evaluations of these TAs, both for the quarter in which the intervention took place as well as for the following quarter, and the students' raw grades for the quarter in which the intervention took place. The results show an effect of almost one half of a standard deviation for the students' TA evaluations in the quarter following the intervention. Nonethless, the intervention had no effect on the student evaluations of the concurrent quarter, suggesting that it takes time for TAs to adjust their teaching practices. A detailed analysis of the TA evaluations for the following quarter suggests that the intervention had a large effect on the TAs' communication skills, and a more modest effect on the following aspects: concern with student learning, organization, and interaction with students.

Finally, the third chapter studies risk sharing and heterogeneous risk preferences. More specifically, it introduces a simple test that incorporates risk preference heterogeneity in the traditional test of efficient risk sharing, overcoming a problem previous studies may have encountered: rejecting the efficient risk sharing hypothesis even when it was true. The requirement to implement this test is a household panel data set with considerable waves, that besides expenditure and income recordings contains a measure of risk preferences. To my knowledge, no dataset fulfills all these requirements at the moment, so I develop an alternative way to incorporate risk preference heterogeneity into the analysis: implement the traditional test within groups of households that share the same risk preferences, using the Mexican Family Life Survey (MXFLS). I use a measure of risk aversion to classify households in one of six groups (in which homogenous risk preferences are likely to hold, as required

[1]This chapter is co-authored with Gabriela Rubio, UCLA.

by the traditional test) and implement the traditional test within each of these groups. The results show that within-groups efficient risk sharing is rejected in almost 60% of the cases, mainly when the total household income is considered as the relevant income variable (as opposed to non-labor income). Further refinement of the risk groups result in low power as a result of few observations in the sub-groups.

The dissertation of Carolina Mejía Mantilla is approved.

Paola Giuliano

Maurizio Mazzocco

Adriana Lleras-Muney, Commitee co-Chair

Kathleen McGarry, Commitee co-Chair

University of California, Los Angeles

2013

To my wonderful parents Luis and Elsa, who encouraged me every step of the way and never lost faith in me

Contents

List of Figures

List of Tables

Acknowledgements

I am deeply grateful to my principal advisors Kathleen McGarry and Adriana Lleras-Muney for their invaluable guidance and support throughout this lengthy process. They patiently provided an optimal combination of academic insight and enthusiastic encouragement - and knew precisely when each was needed. I am also very grateful to my committee members, Paola Guilano and Maurizio Mazzocco, for sharing their time and knowledge with me. I would also like to thank Day Manoli, Leah Boustan, Moshe Buchinsky and the participants of the applied microeconomics seminar, for their helpful comments. I am particularly grateful to my close friends, both in Los Angeles and Bogotá, who were a constant source of optimism, inspiration and energy during these five years. In particular, I am grateful to Javier Pérez-Estrada, Elizabeth Gutiérrez, Matt Luzzetti and Seth Neumuller. Finally, I am very thankful for the love and support of my wonderful (and large) family, in particular Camilo Casas, María José Casas and Adriana Mejía.

Curriculum Vitae

Education

- M.A. Economics, University of California Los Angeles, 2009
- M.Sc. Economics (Honors Degree) Universidad de los Andes, Bogotá Colombia, 2005
- B.A. Economics (Honors Degree) Universidad de los Andes, Bogotá Colombia, 2004

Work Experience

- Research Economist, Inter American Development Bank (IADB), Washington D.C. Summer, 2011
- Research Economist, Fedesarrollo, Bogotá Colombia, 2007 – 2008
- Visting Research Economist, Institute for Fiscal Studies (IFS), London UK, 2007
- Head of the Economic Outlook Team and Editor of the bi-annual Journal "Coyuntura Económica", Fedesarrollo, Bogotá Colombia, 2005 – 2006
- Research Assistant, Universidad de los Andes, Bogotá Colombia. 2002 – 2005

Awards and Distinctions

- UCLA Graduate Student Fellowship, 2009 – 2013
- Central Bank of Colombia Study Abroad Fellowship, 2008 – 2012
- Ramón de Zubiría Academic Excellence Scholarship 2004

Chapter 1

1 Crisis and Education: Mid-term and Long-term Effects of the 1998 Asian Crisis in Indonesia[2]

1.1 Introduction

The Asian Crisis was particularly severe in Indonesia, where after almost three decades of sustained growth real, GDP fell by almost 15 percent in 1998. In that same year, the Indonesian Rupiah lost over two thirds of its value in a matter of weeks, inflation reached an overwhelming 80 percent and interest rates remained extremely volatile. The severity of the crisis, illustrated in the context of several recent macroeconomics crisis around the world in Table 1.1, prompted the fall of President Suharto after thirty years of holding power, which in turn led to important changes in the political and social structures.

From a theoretical point of view, the effect of macroeconomic shocks to human capital accumulation is ambiguous due to the tension between the substitution and income effects: a drop in the wage level lowers the opportunity cost of studying (encouraging school attendance), while a drop in the household income increases the marginal utility of the labor income of all members of the household (deterring school attendance). While the relative importance of these opposing effects depends on several observable factors, the overall effect in any given situation remains an empirical question. For the most part, the empirical studies on the subject have shown that recessions are counter-cyclical for education for high-income countries, pro-cyclical for low-income countries and have little or no effect for middle-income

[2]Special thanks to Jack W. Clift for detailed comments and suggestions. Comments from participants of the applied micro group meetings, in particular Aprajit Mahajan, and the labor proseminar at UCLA are also acknowledged. All remaining errors are mine.

countries.[3] In contrast with the previous literature, this paper finds significant and robust negative effects for Indonesia, a middle-income country.[4]

The impact of a large economic crisis on human development is complex and heterogeneous (Lundberg and Wuermli, 2012). Thus, understanding the effects that macroeconomic shocks have on human capital accumulation, and on education in particular, of children has important implications for public policy. Identifying the direction, magnitude and heterogeneity of these effects is informative for the design of the stabilization policies that governments and non-governmental organizations usually implement in response to macroeconomic shocks, and can enhance their effectiveness. As Thomas (1999) points outs, if such policies are to succeed, they must be based on truthful information on who has been affected, how they have been affected and how they are changing their behavior in response to the crisis. Also essential for the effectiveness of the public policy response to this type of crisis is the distinction and comparison between the mid- and long-term effects. It may be the case that the mid-term effects underestimate the magnitude of the long-term effects, as shown by Chetty et al. (2010) in the context of the returns to being in a small size class in early childhood. It could also be the case that effects are not observable in the mid-term, but are observable and large in the long run, as Black, Devereaux and Salvanes (2007) find when looking at *in utero* nutrition. Therefore, it is important, if possible, to quantify and compare the mid- and long-term effects , as pointed by Currie (1997).

An additional motivation to analyze the effects of the crisis on human capital accumulation of children is that the economic literature has widely documented the importance of early environmental conditions, opportunities and human capital investment on shaping the cognitive and non-cognitive skills of individuals (Heckman, 2007). Different types of abilities

[3] As will be discussed in detail in the literature review section.

[4] Note that in the case of severe shocks, such as the one experienced by Indonesia, that there are other possible effects not operating through these mechanisms. For example, if the economic crisis is accompanied by changes in the political regime and organization of the government, the supply of education may also be affected, both in terms of quantity and quality. Similarly, political unrest and violent protests as a result of the economic crisis may affect the children's access to school and education center. At this point, I am unable to separate these effects, even though the results on the labor supply of children are suggestive that the effects of the crisis were transmitted mainly through economic mechanisms at the household level.

seem to be shaped at different ages: early ages seem to be particularly important for cognitive skills while teenage years are a sensitive period for the formation of non-cognitive abilities.[5] Another feature of skill formation characterized in the literature is that of "dynamic complementarities", which suggests that previously acquired skills enable the development of further skills (Cunha and Heckman, 2007). Thus, if there is an interruption or disruption of schooling during the teenage years, this might have long-term consequences for acquiring skills throughout college and/or the professional career.

This paper looks at whether the Asian Crisis had any mid- or long-term effects on the educational attainment of Indonesian children 7 to 18 years old.[6] In the mid-term (two years after the crisis), it considers school attendance and years of education as the main outcomes of interest but it also analyzes grade progression and education quality outcomes (through standardized test scores). For the long-term (nine years after the crisis), it studies the effects on educational attainment, progression outcomes, log wages and self-employment. For the long-term, the analysis disentangles the effects of lower educational attainment from the wider effect that crises have on wages. The paper also explores if these effects vary by gender, birth order, area of residence (urban/rural), and wealth. The effects are identified as deviations from a linear time trend for specific age groups, which takes into account pre- and post-crisis data, coming from the Indonesia Life Family Survey (IFLS) and Census data.

The contribution of this paper is threefold. First, it investigates whether there were long-term effects of an economic crisis on educational attainment, something that no other paper has attempted so far, and, in the particular case of th Asian Crisis and Indonesia, it studies the mid-term effects on a broader set of education outcomes, including progression indicators and education quality outcomes.[7] It also compares the results obtained in the

[5] Along these lines, it will be important to analyze the effect of the crisis on different age groups, to look for critical windows.

[6] The Asian Crisis refers to the macroeconomic crisis that affected much of Asia during 1997 (second half) and 1998. In the case of Indonesia, the crisis took place in the beginning of 1998.

[7] A growing strand of the economic literature has determined that graduating in a recession has substantial, negative and long-term effects on the earnings of college graduates. Oreopoulos et al. (2012) document a persistent earnings loss for large number of representative cohorts of male college graduates in Canada and Kahn (2010) concludes that the effects of graduating in a bad economy in the US are bad, negative and

mid-term versus those obtained in the long-term. Second, it uses a more precise identification strategy to study these phenomena, in which two pre-crisis and two post-crisis waves of the household survey are used. Finally, it controls for idiosyncratic characteristics of the children's household, such as socio-economic status, a very important determinant of educational attainment.

Previous studies on the impact of the Asian Crisis in Indonesia suggest that there were adverse short-term effects in terms of school enrollment and share of household budget devoted to education, but that these effects vanished in the mid-term (Strauss et al. (2004), Thomas and Frenkelberg (2006, 2007)). This paper shows that the crisis had a negative mid-term effect on the probability of attending to school of around 5 to 7 percentage points for children of age 11-16 at the time of the crisis, while it increased the number of hours worked. Similarly, these children lost about 0.5 years of schooling in the mid-term. Effects vary by birth order and wealth status, but not by gender or by rural/urban area. The differences in results stem from the fact that I am able to incorporate more data both before and after the crisis, which allows me to identify linear time trends of the educational outcomes, and thus, to identify the effect of the crisis, at various points of the age profile of children.

More importantly, the effects are long lasting (for the same age group affected in the mid-term) and roughly equal to 1.5 years of education completed. The magnitude is equivalent to twice the increase in the average educational attainment in the last decade. These effects also vary by wealth status but not by gender or by rural/urban area. In the long-term, the probability of graduating from high school and of progressing to college are lower, while the probability of being self-employed is higher. The estimates also indicate a long-term effect on real log wages, which is largely but not fully explained by the lower educational attainment.

My findings have important implications for the design and implementation of public policies aimed at mitigating the effects of macroeconomic recessions. First, they support the idea that in developing countries, educational outcomes are pro-cyclical, particularly for

persistent, among other studies.

low-income households. Therefore, the design of policy responses should take into account the heterogeneity of the consequences. Also, the effects are heterogeneous depending on age, which may suggest distinctive policies for different school levels. Finally and more importantly, the long-lasting effects of the crisis suggest that programs whose objective is to avert school attrition during recessions may have substantial benefits for the future of these children, which may help to justify the existence of these programs in difficult economic times.

The remainder of the paper is organized as follows. Section 1.2 presents a brief review of the literature that studies the effects of macroeconomic crisis on human capital accumulation, and of the studies that have focused on the Asian Crisis and Indonesia. Section 1.3 provides some background on how the crisis impacted Indonesian households and introduces the main dataset used in this study. The identification strategy and results for the mid-term effects are presented in Section 1.4, while the results for the long-term effects are presented in Section 1.5. Section 1.6 presents some robustness checks using Census data and Section 1.7 concludes.

1.2 Literature Review

1.2.1 The Effects of Macroeconomic Crisis

From a theoretical perspective, the effects of macroeconomic crisis on educational outcomes are ambiguous. On the one hand, recessions are accompanied by lower wage levels, which lowers the opportunity cost of attending school and hence promotes schooling. On the other hand, overall declines in household income both reduce the household resources available for educational expenses and increase the marginal utility of income from children's earnings. Ferreira and Schady (2008) formalize this argument and conclude that there are several factors that help to determine which of the two forces dominates. The ini-

tial household income level matters because marginal utility is lower at higher income levels, implying that richer households and richer countries should have a countercyclical demand for schooling. Credit-constrained households (or countries where credit markets are not well developed) should be more affected by the crisis, as household budgets must be balanced by cutting costs and/or raising income, rather than by borrowing. Finally, the magnitude of the shock should also determine the direction and magnitude of the effect, in the sense that deeper and longer crisis should have a pro-cyclical effect on schooling outcomes (in the same direction as the income effect).

For the most part, the empirical evidence has supported this hypothesis: crises are counter-cyclical for schooling in developed countries, pro-cyclical in low-income countries, and have very little or no effect in middle-income countries. For the US, Goldin (1999) and Black and Sokoloff (2006) find a counter-cyclical pattern of the Great Depression in the US, with high-school graduation rates increasing nearly 50 percent.[8] Betts and McFarland (1995) find a similar counter-cyclical pattern at the college level, using more recent data: a one percent increase in the unemployment of adults is correlated with a four percent increase in the enrollment rates of community colleges. A different picture arises from the analysis of the effects of adverse macroeconomic shocks in low-income countries: Jensen (2000) uses data from Cote d'Ivoire to conclude that the 1986-1987 drought lowered school enrollment by almost 20 per cent. The World Bank (2007) shows how a similar rainfall shock in the mid-1990s increased school absenteeism by a similar magnitude in Malawi, particularly for poorer children.

The studies that analyze episodes of economic crisis in Latin American (mainly middle income countries) present mixed results, but the general conclusion is that there is little or no effect on education outcomes of children. McKenzie (2003) uses a difference-in-difference approach to conclude that the Mexican Peso Crisis (1995-1996) had a small but positive

[8]However, the US of the Great Depression resembles more a developing country of now. As a matter of fact, the real GDP per capita was roughly that of Indonesia today, close to $4,500 US dollars of 2000.

effect on the school attendance of children in their late teens.[9] Maluccio (2005) also finds that enrollment of children aged 7-12 increased by around 15 per cent in Nicaragua, using a rural sample (not representative at the national level) employed mainly in coffee plantations, a child labor intensive agricultural activity.

Schady (2004) uses cross-section household surveys from Peru during the late 1980s economic downturn to establish that the probability of school enrollment didn't change throughout the crisis but the probability of working was much lower. Similarly, he observes an increase of 0.25 years of completed education for those individuals that were school-aged during the crisis compared to those that were not school-aged. Duryea and Arends-Kuenning (2003) study two crisis episodes (1981/1983 and 1990/1992) in Brazil and find no evidence of a change in school enrollment rates, both for girls and boys. In the case of Costa Rica, Funkhouser (1999) determines that the economic downturn that took place in the early 1980s resulted in a decline in enrollment rates of around 7 per cent. However, the result is not reflected in educational attainment, which is somewhat puzzling.

Most of these studies focus on the short term effects of the crisis, impose strict assumptions on pre-crisis and post-crisis trends in terms of education outcomes and use cross-sectional household surveys or administrative data. This paper attempts to overcome these flaws by analyzing both the mid-term and long-term effects of a macroeconomic crisis; by introducing a different identification strategy that explores pre- and post-crisis information at different age groups and by using a longitudinal panel household survey, which counteracts the problem of selection bias in the sample over time. Similarly, very little attention (in part because of lack of data) has been given to the effect of macroeconomic crisis on actual measures of skills and competences of schooled aged children, measures of education quality, with a few exceptions. A study of the 1998– 2000 crisis in Ecuador found significant decreases in vocabulary test scores in children 6–11 months and 18–29 months (Hidrobo, 2011) and another study using US data, suggests that lower investment (in terms of time

[9]Of around 5 per cent for boys and 4 per cent for girls.

and resources) in adolescent children may decrease their literacy and math skills, negatively affecting future educational and labor outcomes, Gershoff et al. (2007). The present paper represents a first step towards understanding the effects of macroeconomic crisis on outcomes measuring the quality of education, by studying whether the Asian Crisis had an effect on the standardized test scores that the Indonesian government requires at the end of each education level (primary, middle school and high school).

1.2.2 The Effects of the Asian Crisis in Indonesia

Given the severity with which the Asian Crisis hit Indonesia, several studies have addressed its consequences on different aspects of the welfare of households. Thomas, Frankenberg and Beegle (1999) study the immediate effects of the crisis (one year after) to conclude that there was a significant reduction in household real expenditure (both in essential and non-essential goods), as well as a decline in the share of the budget devoted to education.[10] This was true, particularly for the left tail of the income distribution. Using summary statistics from administrative data Thomas et al. (2004) finds that in 1998 (the year of the crisis), school enrollment declined for 10-14 year olds, mainly those from lower income households.[11] However, a closer look to the IFLS 2 (1997) and IFLS2+ (1998) summary statistics, used in the paper, shows that enrollment decreased significantly for older children, ages 15 to 19, as well. While enrollment decreased by around 3.5 per cent for children aged 13-15, it declined by around 15 per cent for children ages 16 -19. Similarly, the linear regression analysis of the paper is only informative regarding a bigger effect for low-income children, but not regarding heteregeneous effects by age. By matching data on price changes with data on household consumption, Levinsohn et al. (1999) reach the conclusion that the spike in prices affected the real expenditure of poorer households disproportionately and that agricultural and self-production activities attenuated the impact of the crisis.

With respect to mid-term effects, Strauss et. al. (2002) uses the IFLS data to conclude

[10] Average declined from 13.1% to 10% of the total real household expenditure.
[11] SUSENAS data from 1996 through 1998.

that two years after the crisis, Indonesian children do not appear to be worse off, in terms of health and income poverty, than they were just before the crisis (if anything they seem to be a little better off). They discuss, but do not provide any evidence, how it is possible that in the mid-run the crisis did not hit many households hard, that its impacts were very short lived or that households were able to smooth the children's health outcomes. Similarly, various studies conclude that by 2000, most reductions in human capital outcomes, mainly educational outcomes, were reversed, despite to the deterioration right after the crisis (Strauss et al. (2004), Thomas and Frankelberg (2006) and (2007)). This result is mainly derived from the fact that primary school enrollment was higher in 2000 compared to 1997, and secondary school enrollment was more or less the same in both years (the difference is small and statistically insignificant). The latter is interpreted by the authors as the result of the strong social safety net and the effective government response to the crisis.[12]

Nonetheless, if one observes that in a middle-income country like Indonesia, education indicators follow a positive time trend (particularly for middle-school and high-school children), the relevant question is by how much more would enrollment increased had it not been for the crisis. Thus, by using all four waves of the IFLS longitudinal data and assuming a time trend for educational outcomes, this paper provides a more precise identification strategy. More importantly, this paper looks at the long-term effects of the crisis and analyzes a broader range of educational outcomes, which provide a more complete picture of the impact of the crisis.

[12] In terms of other wellbeing indicators, Friedman et al. (2007) determine that the crisis impacted Indonesians' psychological well-being adversely, as there was a substantial increase in the psychological distress indicators among females and male across the income distribution. More importantly, these effects persist in the mid-term, even after the households have partially recovered in terms of other economic indicators.

1.3 Data and Characterization of the Crisis in Indonesia

1.3.1 IFLS Data

The Indonesian Family Life Survey (IFLS) is a longitudinal household survey.[13] It collects extensive information related to individuals, their households and the communities in which they live. The survey is a representative sample from thirteen of Indonesia's twenty-seven provinces (representing 83% of the Indonesian population), and contains more than 30,000 individuals (7,000 households) per wave.[14] The first wave (IFLS 1), was collected in 1993, the second wave (IFLS 2) was collected in 1997, shortly before the onset of the Asian Crisis, the third wave (IFLS 3) was collected throughout 2000 and, finally, the last available wave (IFLS 4) was collected in 2007.[15]

For most of the mid-term analysis, I considered only individuals between the ages of 7 and 18 (schooling age), while for the long-term analysis, I considered individuals between the ages of 18 and 30 (the number of children of each age at every point in time is shown in Table 1.12 of the Appendix). The dataset provides detailed information on each household's consumption, assets, income and family business. In addition, it contains information on individual health, education, labor market, marriage and migration variables, among others. The education module contains detailed information on attendance, grade progression and fees while the labor market module contains questions on wages, participation, occupation and related topics. See Table 1.2 for the summary statistics of the most important education variables and wage that will be used throughout the paper.[16]

With longitudinal surveys, attrition of individuals (and sometimes entire households) can be a problem. Nonetheless, as Thomas et al. (2012) point out, attrition has not been a

[13]It is conducted by the RAND Corporation and all datasets are available in the website: http://www.rand.org/labor/FLS/IFLS.html.

[14]The provinces included are North Sumatra, West Sumatra, South Sumatra, Lampung, Jakarta, West Java, Central Java, Yogyakarta, East Java, Bali, South Kalimantan, South Sulawesi and W. Nusa Tenggara.

[15]There was another wave of the IFLS, the IFLS 2+, collected in 1998, right in the midst of the crisis. This wave is used in most studies that analyze the immediate effects of the crisis, discussed in the previous section. Nonetheless, this survey only follows 25 percent of the full IFLS sample.

[16]See Table 1.13 in the appendix to the summary statistics by age.

major complication in the four waves of the IFLS, with follow up rates of almost 80 percent in every wave. In part, attrition is low because movers are followed.

1.3.2 Characterization of the Asian Crisis in Indonesia

As mentioned above, Indonesia was one of the countries that was more severely affected by the Asian Crisis, along with South Korea and Thailand.[17] Following the collapse of the Thai Baht in July 1997, the Indonesian Rupiah came under increasing pressure from speculative attacks throughout the second half of 1997, which forced the monetary authorities to switch from a managed floating exchange regime to a free-floating exchange regime. Despite the efforts of the Indonesian monetary authorities and the rescue funds that the IMF approved by the end of the year, the Indonesian Rupiah lost two-thirds of its value in the first six weeks of 1998. This rapid devaluation precipitated a broader economic collapse, exacerbated by several characteristics of the Indonesian economy: high external private sector debt, poor financial regulation, and weak institutional development under the 30-year political regime of President Suharto (Hartono and Ehrmann, 2001).

The GDP per capita in Indonesia decreased by 15 percent in 1998 and annual inflation reached 80 percent. Food prices rose by even more than the price index, disproportionately affecting poorer households who spend a larger portion of their budget in food (Thomas and Frankenberg, 2007). In order to appease domestic and international investors, the government took a fiscally conservative stance, cutting food and fuel subsidies, which triggered riots and violent protests. As Hill and Shiraishi (2007) point out, with the bankruptcy of some externally indebted firms owned by government officials and the Suharto family, the crisis destroyed the informal funding mechanism of the state. This helped to fuel popular discontent and ultimately resulted in the resignation of President Suharto (May 1998) and change from a centralized authoritarianism regime to a decentralized democracy. The transition was gradual and took five years in total: it was not until 2004 that Indonesians held a free and popular voteto elect their president. As mentioned by the authors, there were several

[17] Other countries severely affected were Hong Kong, Malaysia, Laos and the Philippines.

constitutional amendments that took place during the transition period, mainly aimed at ensuring the separation of powers and creating a system of checks and balances.

Indonesia implemented a series of institutional reforms, as part of the financial agreement signed with the International Monetary Fund (IMF) during the period 1997 to 2003 (Hartono and Ehrmann, 2001). The main objective of these reforms, and perhaps of macroeconomic policy in general during these years, was to restore macroeconomic and financial stability: monetary and fiscal policy were the main focus of the economic authorities. Hill and Shiraishi (2007) suggest that social investment and national planning (including national education policy) became relevant only after macroeconomic stability was restored, around the year 2003 (almost five years after the onset of the crisis).

Even though the crisis was widespread, there were some groups that were more affected than others, as can be observed in Table 1.3. Given the financial nature of the crisis (in which there were considerable spikes of the exchange rate and of the nominal interest rate) and that food prices increased significantly, rural households were less affected than urban ones; the average decrease in per capital real expenditure is almost half the size for rural households (7.9 percent versus 13.1 percent). Moreover, wealthier households were able to use their assets and savings to smooth consumption more effectively than poorer households. While households in the first asset quintile (poorest households) saw their real per capita expenditure decrease by almost 16 percent, those in the fifth quintile (richest) only experience a 5.22 percent decline.[18] This is also discussed by Friedman and Levinsohn (2001), who use pre-crisis consumption data and accurate information on the change in commodity prices to assess how the crisis impacted households across the income distribution.

1.4 Mid-term Effects of the Crisis on Education Outcomes

This section analyzes the mid-term effects (two years after) of the Asian Crisis in the

[18]These patterns are confirmed when regressing the change in per capita household expenditure between 1997 and 2000 on a series of socioeconomic characteristics of the household, as shown in Table 1.14of the Appendix.

education outcomes of children aged 7 to 18 at the time of the crisis. It focuses on school attendance but also considers years of completed education, grade progression, repetition and, finally, standardized test scores at the elementary, middle-school and high-school level (EBTANAS). I defined grade progression as the difference between the grade the child is currently attending and the grade he is supposed to attend according to the date in which he entered elementary school. Grade repetition is an indicator variable of whether the child repeated a grade.

1.4.1 Education in Indonesia

The Indonesian education system consists of six years of primary school, three years of lower-secondary school or middle school and a further three years of upper-secondary education or high school, followed by various kinds of tertiary education, the most common one being a technical degree (two years) and a university degree (four years). As Jones and Hagul (2001) explain, since 1993, the Indonesian government has focused on providing all children with nine years of education, which has improved enrollment rates at the primary level, and to a lesser extent at the secondary level. Figure 1.1 shows the attendance rates by age for the four waves of the IFLS, which shows a general improvement over the last two decades, as well as a noticeable decrease in attendance between 1997 and 2000, respectively the year before and two years after the crisis. [19]

The efforts to promote education are also reflected in an upward trend in the years of completed education, particularly for older children (see Figure 1.2). Schools in Indonesia can be either public or private, and religious (mainly Islamic) or secular. In the IFLS sample, almost 80 percent of the students attend a public institution, and close to 85 percent attend non-religious schools. Public non-religious schools are the responsibility of the Ministry of Education and the Ministry of Culture, while public religious school are under the supervision of the Ministry of Religious Affairs.

[19] Figure 1.16 of the Appendix shows the average attendance by age when the sample is split between urban and rural areas, where a similar pattern can be observed.

1.4.2 Identification Strategy and Results

Figure 1.1 shows the attendance rates for children aged 6 to 18 at every IFLS wave: 1993, 1997, 2000 and 2007. For primary-school-aged children (6 to 11 years old approximately) attendance has been relatively high, above 93 percent, at every point in time since the 1990s. Also, for this group, it seems as though there was no change in attendance due to the crisis: there are only small differences between the lines representing the 1997 and the 2000 attendance (dotted line and grey line correspondingly). The pattern is different for older children, mainly those aged 12 to 18, for whom it is clear that school enrollment in 2000 is lower than that of 1997 (the grey line is below the dotted line). As an exploratory exercise, separate age-group regressions are run that include all IFLS waves with school attendance as the dependent variable and year dummies as independent variables.[20] The results reveal that the year 2000 coefficients are negative and significantly different from zero for children ages 11 to 18 and are robust to the inclusion of controls (see Figure 1.17 in the Appendix), suggesting that the 1998 crisis dampened school attendance.

In order to fully capture the mid-term effects of the crisis I use the following (main) specification:

$$y_{iht} = \alpha + \sum \delta_k age_k + \gamma t + \sum \psi_k age_k t + Yr2000 + \sum \phi_k age_k Yr2000 + \beta X_{iht} + \varepsilon_{iht} \quad (1.1)$$

where y_{iht} is the education outcome of interest of children i in household h at time t, age_k are the six age group dummies (age at the time of the crisis): ages 5 - 6, ages 7 - 8, ages 9 - 10, ages 11 - 12, ages 13 - 14 and ages 15 - 16, t is a linear time trend (which is allowed

[20] I group children aged 7 to 18 in groups of two, for a total of six groups.

to change for each age group), Yr2000 is a dummy variable for the year 2000, X_{iht} is a set of socio-economic covariates, and ε_{iht} is the error term.[21] The mid-term effects of the crisis are captured by the ϕ_ks, and the identification comes from the deviations from the linear time trend at each specific age group. In this way, I take advantage of the availabilty of pre- and post-crisis data of the IFLS, as well as the fact that it contains important socio-economic information about the children's household.

Figure 1.3 depicts the mid-term effects of the crisis on school attendance: an indicator variable that the child is attending school during that academic year. The effect is negative, but small and not statistically different from zero, for those children who were aged 7 to 12 at the time of the crisis. On the contrary, the effect is sizable and statistically significant for older children: a 5.4 percentage point drop for those aged 13 and 14, and a 7.6 percentage point drop for those aged 15 and 16 at the time of the crisis. The point estimates are robust to the inclusion of socioeconomic controls at the individual and household level.[22][23]

These results suggest that the income effect dominated the substitution effect, as is usually the case in less developed countries where households are more credit constrained. When compared to previous studies, the effects observed are smaller than those found in very poor African countries, but similar to those observed for middle-income countries.

In a similar fashion, Figure 1.4 presents the same specification but uses years of completed education as the dependent variable. When no covariates are included, there is a negative and significant effect of around 0.4 years of education for children aged 9-12, and of around 0.5 for those aged 13-16. However, consistent with the previous estimation only the effect for those children aged 13 to 16 remains significant once the covariates are included. This

[21] The covariates include gender of the children, # of siblings the child has, urban indicator, household asset quintile, adjusted household size, % males in household, % children under 6 in household, average age of household individuals, and an indicator for a female household head.

[22] This will remain true for the rest of the empirical exercises but the results with and without covariates will only presented for the main results on attendance and years of education.

[23] The effects do not change (both in magnitude and significance) when a probit model is used instead of the linear probability model (see Figure 1.18 in the Appendix).

effect is equivalent to that of moving a child from the first asset quintile (poorest) to the fourth asset quintile (second most rich).

These results, summarized in Table 1.4, show that the short-term effects on education outcomes were *not* reversed in the mid-term, as had been suggested by previous studies. One potential reason for this discrepancy is a more efficient identification strategy that takes advantage of all the available waves of the IFLS. Moreover, the effects are important: for attendance rates equivalent to reductions of 7-10% of the average and for years of education equivalent to reduction of 6-7% of the average (See Table 1.13 in the Appendix for the averages of these variables by age).

If the income effect is larger than the substitution effect, as implied by these results, then it is very likely that that the group of affected children adjusted their labor supply, in order to contribute to the household income. Table 1.5 shows the results of using the main specification (Equation 1) using three work-related variables as dependent variables: an indicator variable for whether the individual was participating in the labor market (working or looking for a market job) or working in the family business last week, the number of hours actually worked during the last week and, finally, the number of hours normally worked per week during the last year. Both a linear probability model and a probit model are used for the first indicator, while a linear probability model and a tobit model are used for the other two. It seems that the crisis has no effect on the probability of participating in the job market (the extensive margin), except for those aged 13 - 14 when the probit specification is used. Now, there is a positive and significant effect in the intensive margin, both under the linear probability model and the tobit model. In the first case, two years after the crisis, children aged 13 and 14 at the time of the crisis are likely to work 2.7 hours a week more, while this number is 2 hours for those in the 15 and 16 age group. The marginal effects under the tobit model are slightly larger, around 3.4 hours for the first group and 2.3 for the second.

Did the mid-term effects of the crisis affect all children in a similar way or did these effects

vary for different groups?[24] The position of women in Indonesia is generally considered to be favorable despite its Muslim tradition. As Peacock (1973) points out, "(W)omen enjoy a high status in Indonesia. Females and males are both considered legally competent, and are perceived as equals. Even in santric circles, women are not kept secluded and under wraps as in many Muslim cultures".[25] Nevertheless the effects of the crisis on education outcomes might differ by gender if labor skills (and thus opportunity cost) vary. Figure 1.5 presents the results of the main specification when the sample is separated by gender and it shows no statistical difference between between boys and girls in those age groups that were affected by the crisis.

Nonetheless, the effects are heterogenous by order of birth: the attendance of the oldest child is basically unaffected by the crisis in the mid-term, while the rest of the children were (see Figure 1.6).[26] Since the oldest child will take care of the family once the parents grow old, as is the tradition in Muslim families, contributing to his human capital is considered important, almost an investment.

As discussed in the previous section, the crisis affected urban households more than their rural counterparts, which might in turn suggest a larger effect on children from urban households. Surprisingly, as shown in Figure 1.7, there is basically no difference between urban and rural children. A potential explanation for this is that urban households value education more than rural households, possibly because returns to education are higher in urban areas. Finally, Figure 1.8 shows the differential effect on children on opposite sides of the wealth distribution: those from the poorest 20 percent of the population (Quintile 1) versus those from the richest 20 percent (Quintile 5).[27],[28] As the theoretical framework

[24]This section focuses on the heterogenous effects on school attendance but the results are the same if years of education is considered as the outcome of interest.

[25]Pg. 145.

[26]Figure 1.19 of the Appendix shows the differential effects both by order of birth and gender, and the results do not change.

[27]To minimize the measurement error in income variables, assets are used as a proxy for income.

[28]A similar result is obtained when I compare the effects of the crisis on households whose head is a highschool graduate versus housholds whose head is not a highschool graduate. This is not surprising given the relationship between education and income/wealth.

predicts, wealthier households are able to cope better with an adverse shock since they can dissave their assets to avoid having to cut investments in education. Similarly, wealthier households may also value education more, and may prioritize human capital investment over other expenditures.

In order to obtain a more complete picture of the effects of the crisis on education, other outcomes besides attendance and years of schooling are considered. The first two are a measure of progression: grade differential and whether the child repeated a grade. The IFLS provides information on the age at which individuals entered primary school and thus it is possible to calculate the ideal grade the individual should be attending in the absence of disruptions. The grade differential variable is constructed as the difference between the actual grade the child is attending and that ideal grade. Thus, a negative value for this variable, means the child is lagged behind in the education process. As for the second measure of progression, the indicator variable takes the value of one if the child repeated one or more grades.

The results of using these progression measures as dependent variables in the main specification (Equation 1) are shown in Table 1.6. There is a negative and significant mid-term effect on the grade difference of around -0.12 of a grade for children ages 5 to 8, of around -0.05 for children aged 13 and 14, and of -0.14 for children aged 15 and 16 at the time of the crisis. This means that for the latter two age-specific groups, the crisis not only led to lower attendance and educational attainment, but it also caused a lag for those who returned to school after the crisis passed. There are no statistically significant effects on grade repetition.

1.4.3 Standardized Tests

The third education outcome is a proxy for the quality of education, the test scores from

the standardized tests that the Ministry of Education requires from all students at the end of each educational level EBTANAS (primary, mid-school and high-school) in order to progress to the next one.[29] The IFLS provides information on the date and the overall score for each individual. There is yearly information available on test scores from 1985 until 2001, year in which the implementation of the EBTANAS stopped. Given that the structure of the data is different, the following specification is used to capture the effects of the crisis:

$$y_{iht} = \alpha + \delta_k post + \gamma t + \psi_k post \cdot t + \beta X_{iht} + \varepsilon_{iht} \tag{1.2}$$

where y_{iht} is the test score of individual i in household h at time t, $post_t$ takes the value of one after the year 1998, t is a linear time trend, X_{iht} is a set of socio-economic covariates, and ε_{iht} is the error term.[30] The results of this estimation using the three test scores (primary, mid-school and high-school) are presented in Table 1.7: of main interest is the total effect of *post* (evaluated at the mean of t). There seems to be a negative and significant effect for the primary-school test scores, of about 2 points out of 50 possible (the mean is around 32 and the standard deviation is around 5.5 for all tests). On the contrary, the effect is of the same magnitude but positive for the middle school test scores, which represents an effect equivalent to one third of a standard deviation.

In both cases it is likely to observe some selection bias.[31] In the case of the primary-school test scores, the effect of the variable $post_t$ disappears once selection is accounted for by using the Heckman correction method (Table 1.8). In the case of middle-school, the effect remains unchanged once selection is controlled for, and results are suggestive of a positive selection

[29] EBTANAS is an acronym for National Evaluation Study of the Final Stage and it was implemented without interruption from 1980 to 2001.

[30] The covariates include gender of the children, # of siblings the children has, urban indicator, household asset quintile, adjusted household size, % males in household, % children under 6 in household, average age of household individuals, and an indicator for a female household head.

[31] Table 1.15 in the appendix shows the difference in covariates before and after the crisis for the whole sample and for each EBTANAS test: primary, middle and high school. The means test suggests that there was some recomposition of the children in each group after the crisis.

bias: those children who neither dropped out of school nor lagged behind (most likely those from wealthier and more educated households) took the test and performed better (positive and statistically significant coefficient of the inverse Mills Ratio). It must be noted that these estimations should be interpreted carefully, given that there are few years after the crisis in which students took the test (3 years) and the magnitude of the effects is relatively small. The preferred interpretation of these results is that there is suggestive evidence that the crisis impacted the quality of education in Indonesia, but further research on this particular topic is required.

1.5 Long-term Effects of the Crisis

This section explores the long-term effects (nine years after) of the Asian Crisis on the education and labor market outcomes of those individuals that were of school age when the shock hit. This is the first paper to analyze the long-term effects of the 1998 crisis in Indonesia. More specifically, it focuses on the age groups for which important mid-term effects were observed: children aged 11 to 18 at the time of the crisis. As mentioned in Section 2, there are three main education outcomes of interest in the long run: years of (completed) education, whether the individual is a high school graduate and whether he progressed to college; and three labor market indicators: labor force particpation, self-employment and wages. For all of these, the following specification is implemented:

$$y_{iht} = \alpha + \sum \delta_k age_k + \gamma t + \sum \psi_k age_k t + Yr2000 + \quad \sum \phi_k age_k Yr2000 + \\ Yr2007 + \sum \theta_k age_k Yr2007 + \beta X_{iht} + \varepsilon_{iht} \tag{1.3}$$

Where y_{iht} is the outcome of interest of child i in household h at time t, age_k are four age group dummies (age at the time of the crisis): ages 11 - 12, ages 13 - 14, ages 15 - 16 and ages 17 - 18, t is a linear time trend (which is allowed to change for each age group),

Yr2000 is a dummy variable for the year 2000, Yr2007 is a dummy for the year 2007, X_{iht} is a set of socio-economic covariates, and ε_{iht} is the error term.[32] In this specification, the coefficients of interest are the θ_ts, which identify the long-term effects as deviations from the linear trend (different from the mid-term effects). Figure 1.9 shows the long-term effects of the crisis on the years of education completed, focusing on those individuals who were aged 11-18 in 1998, year of the crisis.

The point estimates are negative for the four age specific groups but statistically different from zero only for those children in the late teenage years at the time of crisis. The effects equals -1.64 years of education completed for those aged 15 and 16, and -1.46 years for those aged 17 and 18. The results are very similar with and without the inclusion of the covariates. How do these results compare to the mid-term effects? The long-term effects are larger than the mid-term effects (0.5 years), which suggests that the individuals affected in the mid-term did not return to school after the year 2000, a likely event since they were already in their late teenage years (or early twenties) and probably already started to participate in the labor market. These long-term effects are sizable, as 1.5 years is twice the increase in the average number of years of education that took place in Indonesia during the last decade. This is equivalent to the effect of moving a child from the first asset quintile (poorest) to the fifth asset quintile (richest).

The existence of heterogenous long-term effects across different dimensions is also tested in the long-term. There is no evidence that the long-term effects on years of education completed varied by gender (Figure 1.10) or by area of residence (urban versus rural children, see Figure 1.11). Nonetheless, and along the lines of the mid-term results, the long-term effects on education attainment were different across socio-economic status. As shown in Figure 1.12, there is no effect for wealthier individuals (asset quintile 5), while the effect is negative and significant for the less wealthy (asset quintile 1).

[32]The covariates include gender of the children, # of siblings the child has, urban indicator, household asset quintile, adjusted household size, % males in household, % children under 6 in household, average age of household individuals, and an indicator for a female household head.

These findings are consistent with the long-term effects on the other two educational outcomes studied in this section and depicted in Figure 1.13: probability of being a high-school graduate and of progressing to college. Individuals who were between the ages of 11 and 18 at the time of the crisis are 25 percentage points less likely to graduate from high school and are 10 percentage points less likely to progress to college.[33] As before, the inclusion of the covariates does not change the results.

A natural question that arises is: how are these long-term effects on educational outcomes translated into wages? As several studies have demonstrated, there are important and long-lasting effects on wages of graduating in a recession (Beaudry and DiNardo (1991), Oreopoulos, von Wachter, and Heisz (2012); Kahn (2010), and Oyer (2006, 2007)). In an attempt to separate the effects of education from the effects of entering the job market during a recession, two different specifications of Equation 3 with log real wages as the dependent variable are used.[34],[35] In the first, the individual's years of education are excluded from the covariates to capture the total effect (or the gross) of the crisis, and in the second, the education of individuals is included as a covariate to control for the effect of the crisis on educational attainment. As presented in Figure 1.14, wages for the individuals who were of age 11-18 at the time of the crisis are around 1 percent lower. Interestingly, for those in the age group 15-16 and 17-18, the effects are largely explained by lower educational attainment, while that is not the case for those in the age group 13-14. These findings, specially the small magnitude of the effects, are consistent with the gradual process of recovery reported in the literature on graduating in a recession (Oreopoulos, von Watcher and Heisz (2012)).

Most likely, there is a selection bias problem since we only observe the wage of the

[33]It must be noted that only 6.75 percent of the individuals older than 18 observed in 2007 progressed to college

[34] It seems likely that some of these kids that didn't graduate, entered the job market permanently during the crisis.

[35]Figure 1.20 of the Appendix presents the average log real wage by age for each of the four waves of the IFLS; real wages in the year 2000 are considerably lower than in the year 1997 for most age groups.

individuals participating in the labor market, which might result in a smaller effect of the crisis.[36] The lower panel of Table 1.9 presents the results once selection bias is controlled for, using the Heckman two-step method. The results suggest that there are factors related to both a lower probability of participating in the labor force and a lower wage (unobserved skill and enthusiasm, for example), as evidenced by the significant and negative coefficient accompanying the Inverse Mills Ratio (lambda). The long-term effects of the crisis seem to be slightly larger for all age groups when no covariates are included, as can be seen by comparing columns 3 and 5. This is no longer true when covariates (including educational attainment) are taken into account, comparing columns 4 and 6.

Finally, Figure 1.15 shows the long-term effects of the crisis with respect to two other labor market indicators: labor force participation and self-employment. There is a negative and significant effect of 20 percentage points on the probability of participating in the labor market for those who were 15 - 16 years old at the time of the crisis. Similarly, there is positive and significant effect of the same magnitude on the probability of being self-employed for the same age group and those who were 13 - 14 years old at the time of the crisis. Usually, self-employment in developing countries is a sign of more flexibility but also of worse working conditions in terms of benefits, insurance and wages. In sum, the crisis had long-term effects not only on the educational attainment outcomes of the individuals of schooling age at the time of the crisis, but also on their job market outcomes.

1.6 Robustness Check: Census Data

A potential weakness of the results presented above is that the data used are not representative of Indonesia at the national level, and that there is a potential for selective attrition, even though Thomas at al. (2012) dismiss this latter element as a minor problem.[37] In this section, I investigate whether the negative effects of the crisis in the mid-term are corrobo-

[36] Since it is more likely that the effect is smaller for those participating in the labor force

[37] Some may potentially argue that the effects observed are a result of the composition of the IFLS sample over time.

rated in nationally-representative Census data. The data come from Integrated Public Use Micro data Series (IPUMS) International, and include Census data for the years 1990, 2000 and 2010 as well as intercensal data (from the Bureau of National Statistics of Indonesia) representative at the national level for the years 1995 and 2005. To avoid an over-representation of the cross-section surveys in the Census years, I obtain a random sub-sample of the data in those years to match the number of individual observations of the intercensal surveys, around 300,000 per year.

The only education outcome available in the IMPUMS data is a categorical variable of educational attainment that takes the following values: 1 if the individual completed less than primary, 2 if he completed primary, 3 if he completed junior high, 4 if he completed high school, 5 if he completed a technical degree and 6 if he completed college education. The more detailed educational attainment information in the IFLS was recoded to match the Census categories. As shown in Table 1.10, the distribution of this variable is very similar in the IFLS and the Census data, the only difference being perhaps that in the Census data the first category (less than primary) is 4 percentage points higher while the second category (primary completed) is 3 percentage points lower.

In order to see if the mid-term effects of the crisis varied across the two different samples, the following specification is used:

$$y_{iht} = \alpha + \sum \delta_k age_k + \gamma t + \sum \psi_k age_k t + \sum \phi_k age_k Yr2000 + \beta X_{iht} + \varepsilon_{iht} \qquad (4)$$

where y_{iht} is the categorical education variable described above of individual i in household h at time t, age_k are the six age group dummies (at the time of the crisis): ages 7 - 8, ages 9 - 10, ages 11 - 12, ages 13 - 14, ages 15 - 16 and ages 17 - 18, t is a linear time trend (which is allowed to change for each age group), Yr2000 is a dummy variable for the year 2000, X_{iht} is a set of socio-economic covariates, and ε_{iht} is the error term. In the case of the Census data, the only available covariates are gender and household size. Table 1.11

depicts the results of the estimation. When the IFLS data are used, the mid-term effects on this constructed measure of educational attainment are negative and significant for all age groups, except for the age 11 - 14 group. When the Census data are used, the effect is negative and significant for all age groups, and the effect is observed to decrease with age. Encouragingly, the point estimates for the different age groups are close if both samples are compared, regardless of whether they are significant or not in the IFLS regressions. As expected, the main difference between the two sets of regressions are the smaller standard errors of the specifications that use Census data, due to the large number of observations.

Unfortunately, the results using this categorical variable of educational attainment are not comparable to the previous mid-term results using years of education (or attendance). Nonetheless, I interpret the results of this section as, first, evidence of a negative and significant effect of the crisis in the mid-term and, second, that the effects found using the IFLS data are not the result of sample composition.

1.7 Conclusions

Understanding and quantifying the mid- and long-term effects of severe macroeconomic shocks on human capital accumulation is important. The estimates of these effects are a valuable input in the planning of public policy responses, particularly if the effects vary across dimensions such as socio-economic status and age. This task is even more relevant when we consider that human capital accumulation (and more specifically, education) at early ages has important effects on the acquisition of skills later on in the professional path, with related long-term effects on outcomes such as labor earnings. From a theoretical perspective, it is not clear whether educational outcomes should improve or worsen as a result of an adverse macroeconomic shock, given that the crisis triggers two opposing forces: the substitution effect (which decreases the opportunity cost of attending to school) and the income effect (which increases the marginal utility of labor income). It remains an empirical question

which effect dominates in any given circumstance.

This paper studies the mid- and long-term effects of the 1998 Asian Crisis on the educational attainment of children aged 7 to 18 (at the time of the crisis) in Indonesia. It contributes to the existing literature by looking at the long-term effects of a recession in a developing country, by broadening the array of outcomes of interest in the mid-term, by improving the identification strategy of previous analysis and by controlling for important socioeconomic variables by exploiting a rich longitudinal data set. Contrary to previous papers that analyze the effects of the Asian Crisis in Indonesia, I find non-trivial negative effects both in the mid- and long-run, mainly for older children in their late teens. In the mid-term (two years after the crisis) these older children are 7 percentage points less likely to attend school, have completed 0.5 fewer years of education, and increased hours of work in 2 to 3 hours a week, compared to what would have been expected given the linear trend over time. These are significant effects comparable to that of moving a child from a household in the first asset quintile (poorest) to one in the fourth asset quintile (second most rich). These children are also more likely to be lagged in their school progression, though the effect is relatively small. I mainly attribute the differences in the results to the more precise identification strategy, but differences may also arise because this paper focuses solely on educational outcomes while most other papers cover a broader set of welfare measures with less detail.

Overall, the mid-term results suggest that the income effect was larger than the substitution effect, and that children aged 13-18 were forced to adjust their labor supply in response to the crisis. The effects vary by order of birth (there is no effect for the oldest child) and by wealth (there is no effect for the richer children). However, the effects do not vary by gender, which confirms the relatively high status of women in Indonesia, nor by region of residence (urban versus rural). This last result is somewhat counterintuitive, since the crisis affected urban more harshly, but may be reconciled by higher returns to education or, alternatively, by a larger substitution effect in the urban areas. In terms of quality of education,I find a

positive effect on middle-school test scores, after controling for selection bias. This suggests that children who were more likely to do better in the test are the ones that remained in school. Nonetheless, these results on test scores should be considered as more suggestive than definite, given the few post-crisis observations available.

More importantly, the estimates show that there are long-term effects of the crisis on educational attainment and labor market outcomes. These are consistent with the previous set of mid-term results, since they are observed for the same group who were affected then and go in the same direction. In summary, they completed 1.5 fewer years of education, they are 25 percentage points less likely to graduate high school, and 10 percentage points less likely to progress to college. In addition, there is a long-term effect negative effect of one percent on the monthly real wage, largely but not fully explained through lower educational attainment.[38] The crisis also had long-term effects on two other labor market outcomes: labor force participation and self-employment. These are important results that point out that temporary macroeconomic shocks may have consequences that affect human capital indicators in the longer run, even after economic activity has returned to pre-shock levels. Similarly, the magnitude of the effect on years of education is worth noting. The loss in years of education of the two cohorts is equivalent to the increase in the national average educational attainment in the last decade.

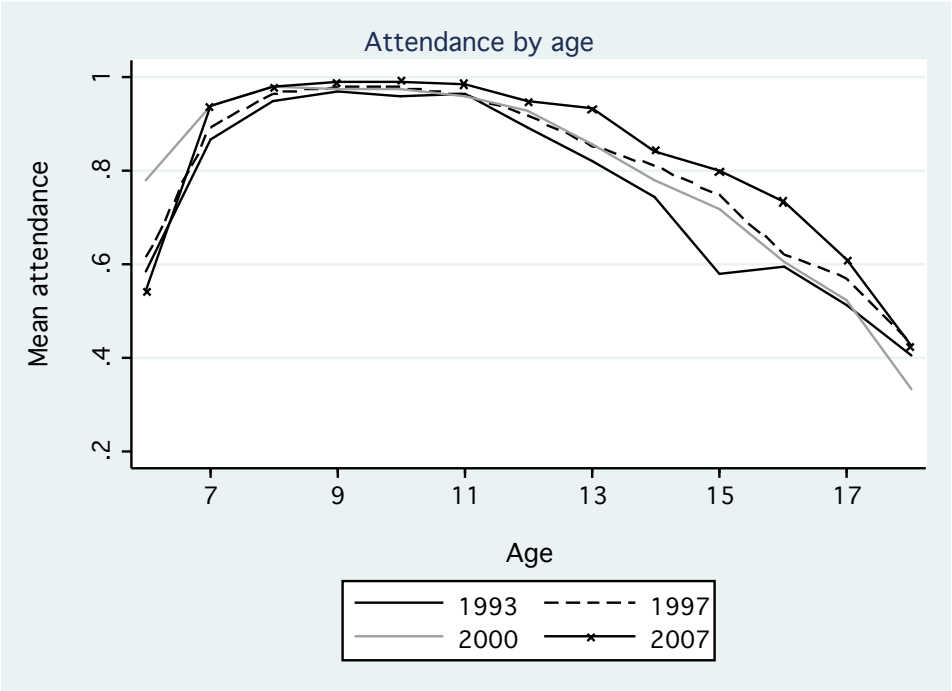
It is important to recognize that the identification of the effects of the Asian Crisis obtained in this paper relies on relatively few pre-crisis observations. Without doubt, it would be ideal to have more frequent and longer panel data in order to better identify the outcome's trends before and after the crisis. Another limitation of this paper is its inability to parameterize the effect of the crisis, mainly because of a lack of a suitable variable that captures the magnitude of the crisis accurately. Regarding possible future related research on this topic, and considering that the effects of a severe crisis may touch other aspects of human capital accumulation, it would be interesting to analyze the mid- and long-term

[38] Even after controlling for selection bias using the Heckman method.

effects on other factors (such as children's health and subjective welfare indicators), using the same framework used in this paper. In addition, potential extensions to this topic involve obtaining better and more comprehensive measures of quality of education; this paper takes a small step towards understanding the effect of macroeconomic crises on educational quality, indicating a promising avenue for future research.

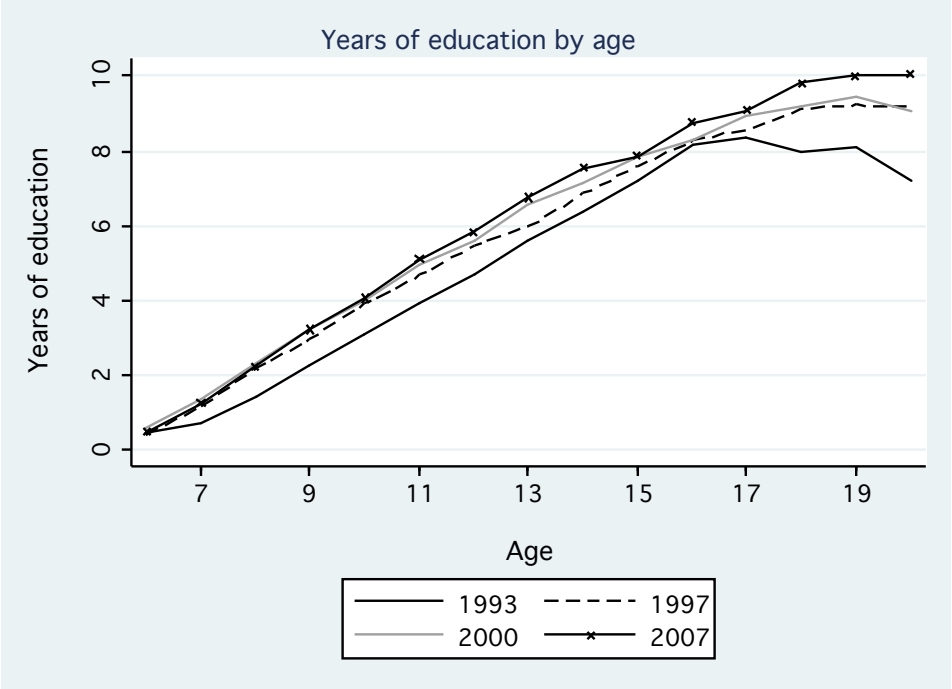
1.8 Main Figures and Tables

Figure 1.1: Attendance By Age



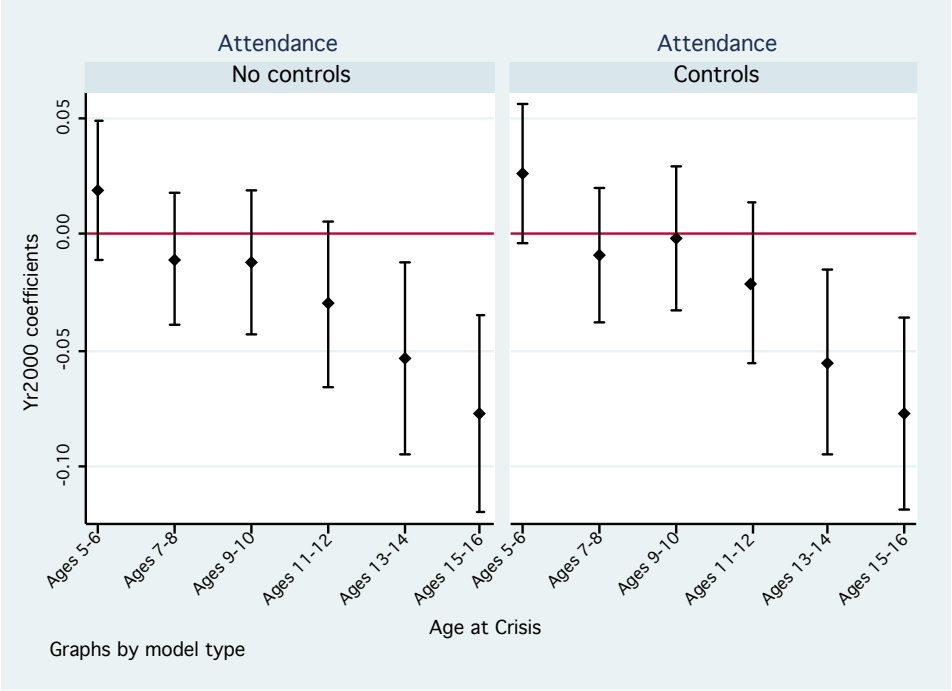
Note: Data from IFLS survey, all waves (1993, 1997, 2000, 2007). Mean attendance for every age in every year using sample of all children from ages 7 to 18. Attendance is an indicator variable that equals one if child is attending to school.

Figure 1.2: Years of Education by Age



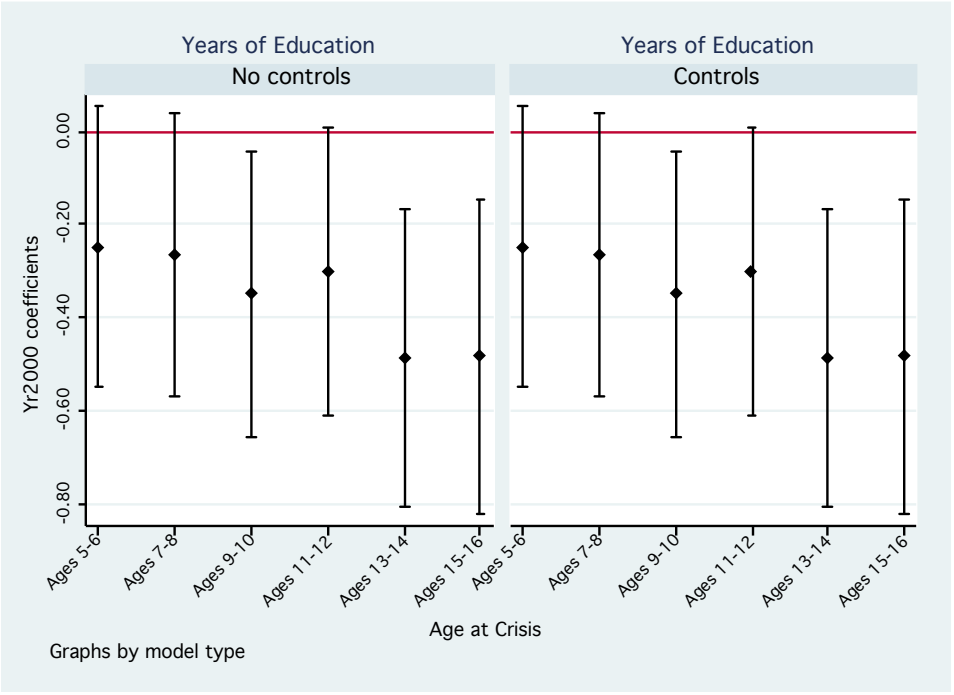
Note: Data from IFLS survey, all waves (1993, 1997, 2000, 2007). Mean completed years of education for every age in every year using sample of all children from ages 7 to 18.

Figure 1.3: Mid-term Effects of the Crisis on Attendance



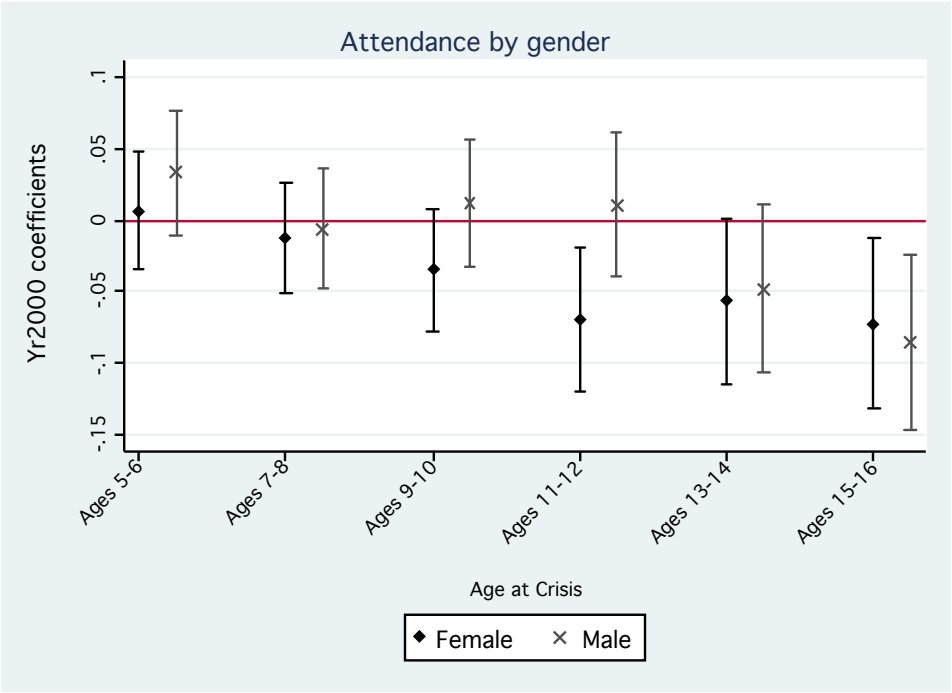
Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18. Left panel doesn't include any covariates while right panel includes: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.4: Mid-term Effects of the Crisis on Years of Education



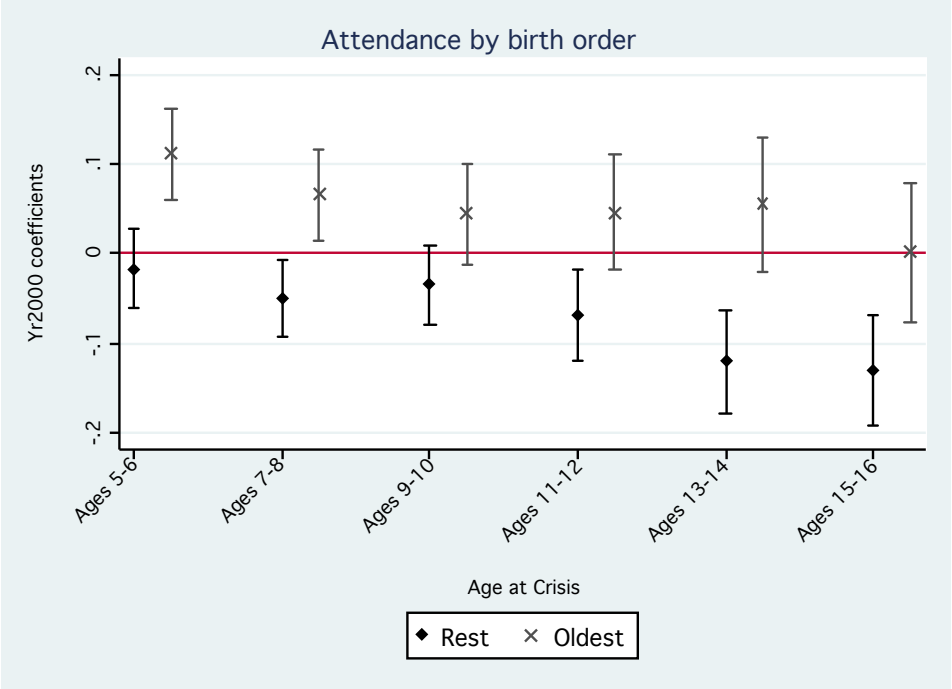
Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is (completed) years of education at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18. Left panel doesn't include any covariates while right panel includes: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.5: Mid-term Effects of the Crisis on Attendance by Gender



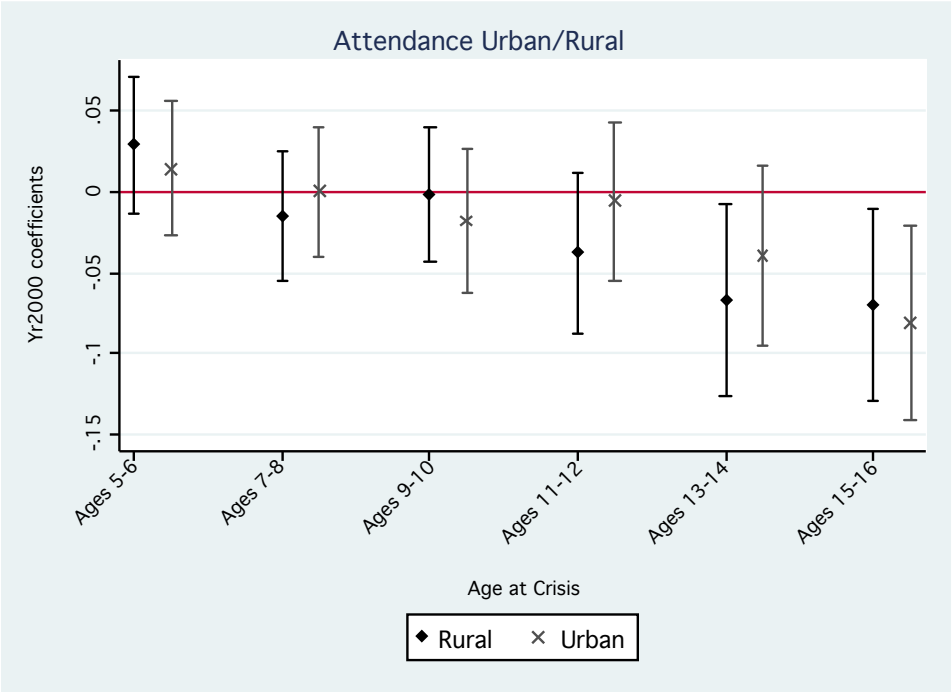
Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18, divided between female and male. Controls include: urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.6: Mid-term Effects of the Crisis on Attendance by Birth Order



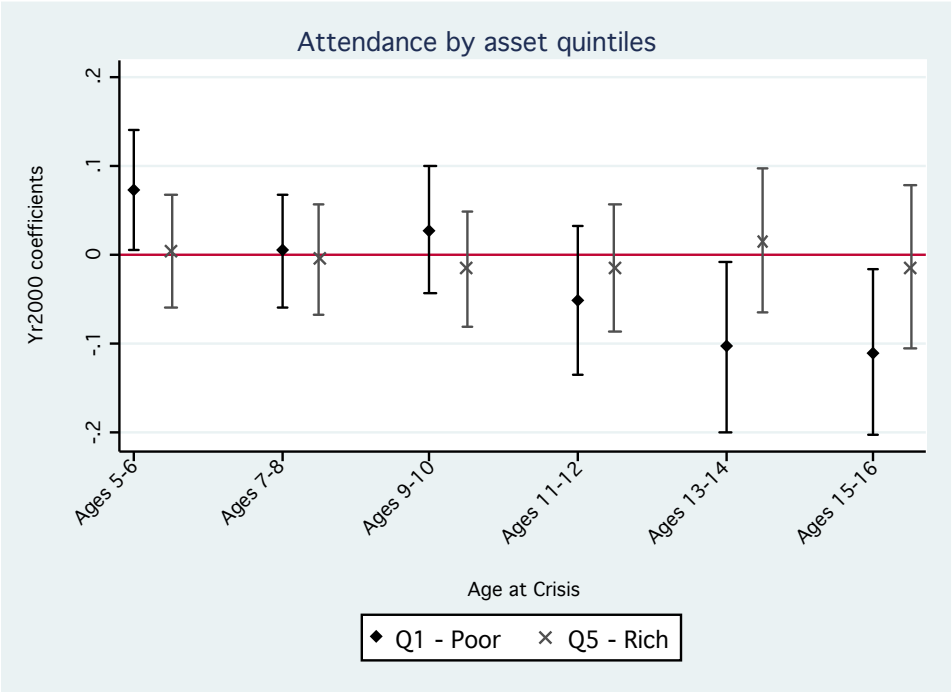
Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18, divided by oldest and rest of the children. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.7: Mid-term Effects on Attendance by Urban/Rural Areas



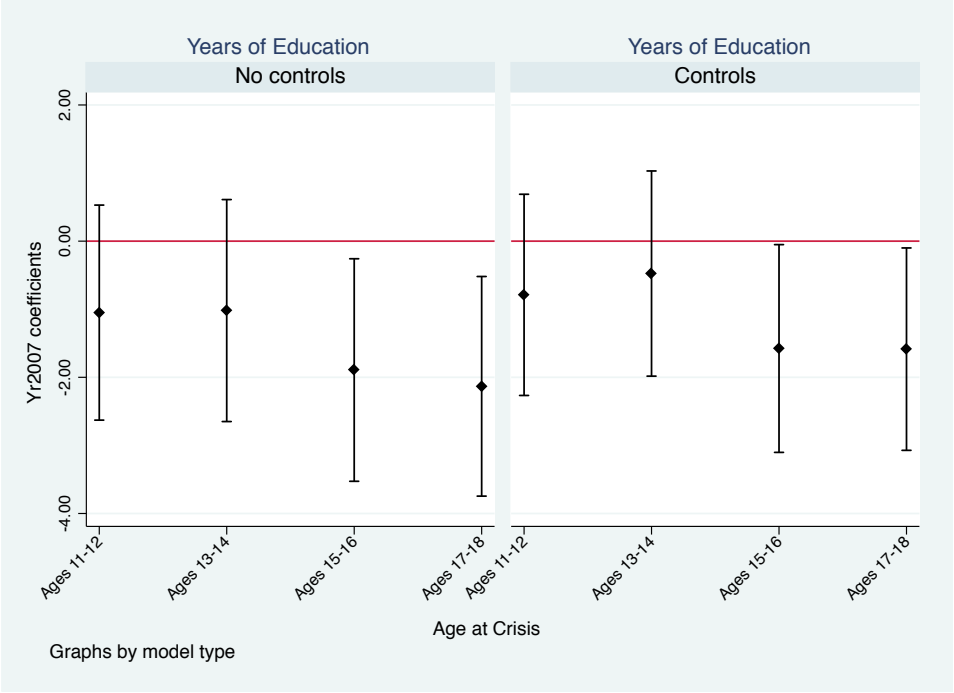
Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18, divided by urban and rural children. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.8: Mid-term Effects of the Crisis on Attendance by Asset Quintiles



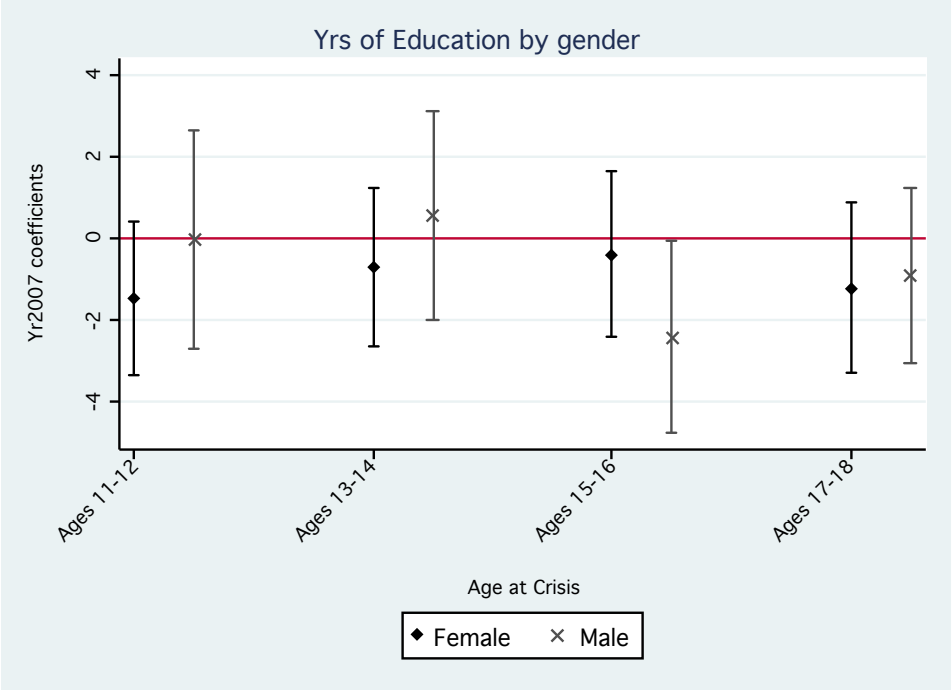
Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18, divided by quintile 1 (poor) and quintile 5 (rich). Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.9: Long-term Effects of the Crisis on Years of Education



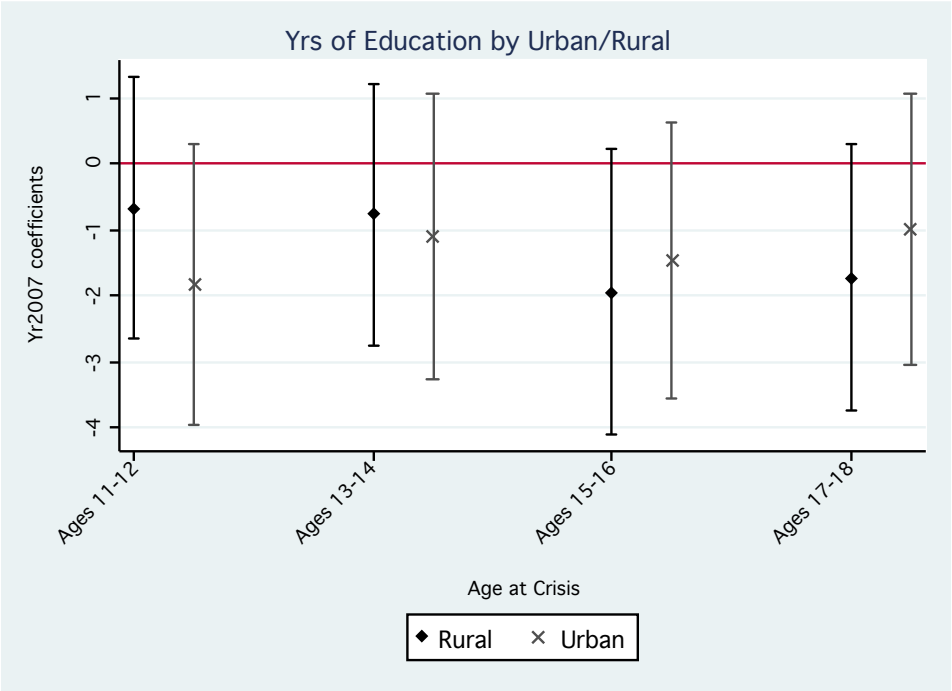
Note: Depicts the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is completed years of education at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all individuals from ages 7 to 26. Left panel doesn't include any covariates while right panel includes: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.10: Long-term Effects of the Crisis on Years of Education by Gender



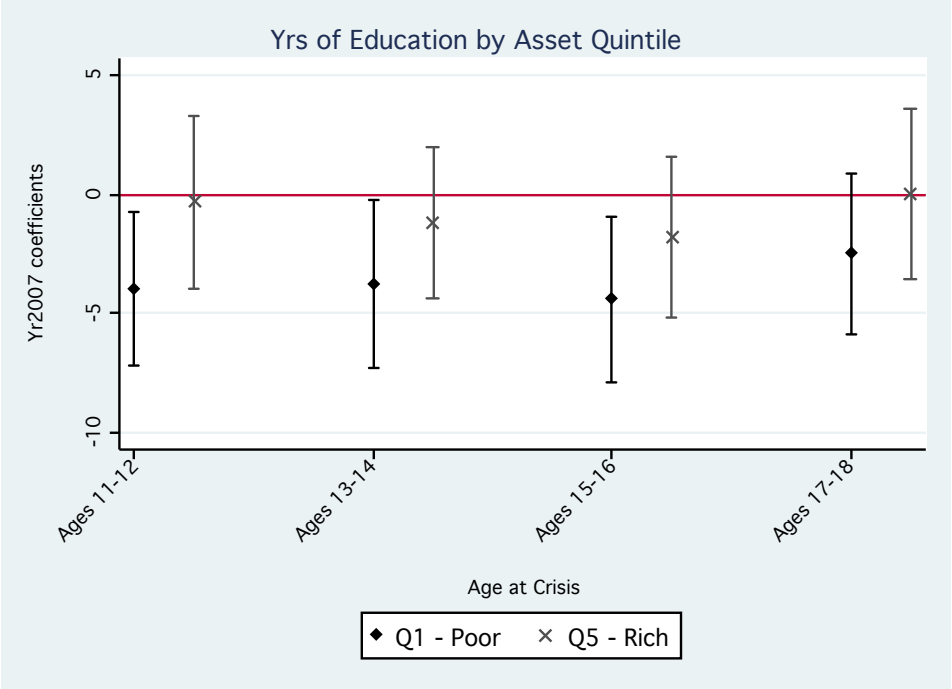
Note: Depicts the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is the log real wage of the individual and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all individuals from ages 19 to 30, divided between female and male. Left panel does not control for education and experience (gross effect), while left panel does (pure crisis effect). Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.11: Long-term Effects of the Crisis on Attendance by Urban/Rural



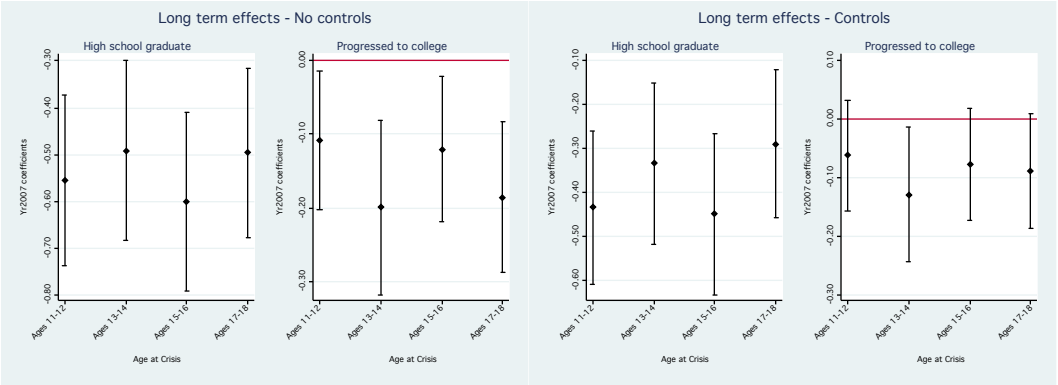
Note: Depicts the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is the log real wage of the individual and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all individuals from ages 19 to 30, divided between urban and rural children. Left panel does not control for education and experience (gross effect), while left panel does (pure crisis effect). Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.12: Long-term Effects of the Crisis on Attendance by Asset Quintile



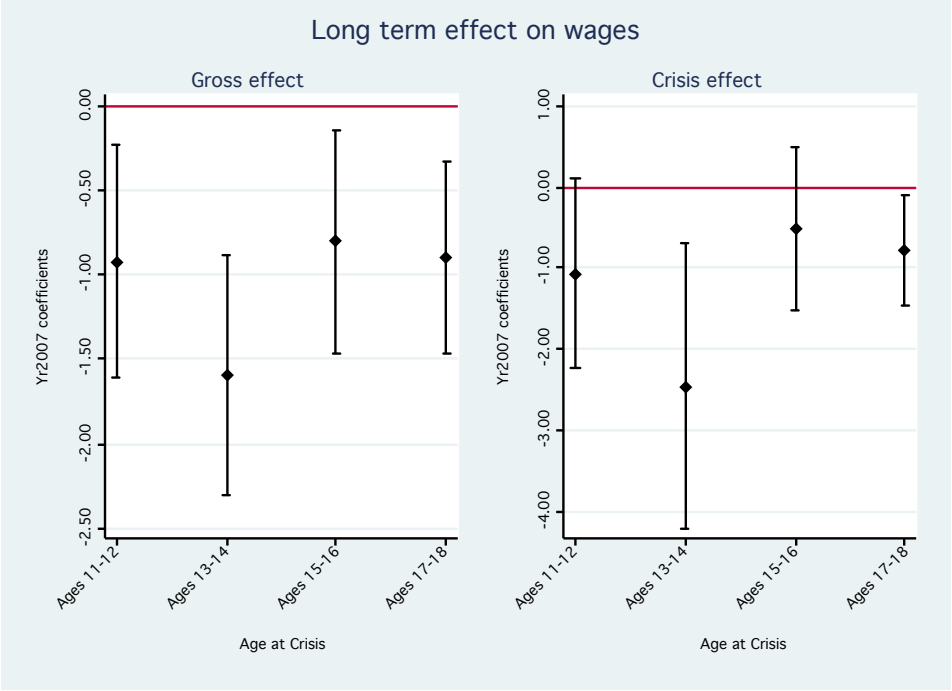
Note: Depicts the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is the log real wage of the individual and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all individuals from ages 19 to 30, divided by quintile 1 (poor) and quintile 5 (rich). Left panel does not control for education and experience (gross effect), while left panel does (pure crisis effect). Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.13: Long-term Effects of the Crisis: Other Indicators



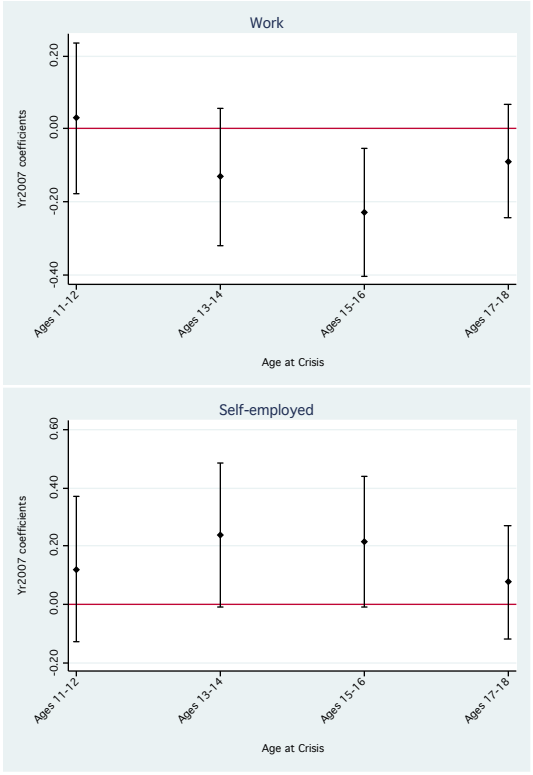
Note: Depicts the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). The left panels use as independent variable an indicator of whether the individual graduated high school while the right panels use as independent variable an indicator of whether the individual progressed to college (university). Similarly, the upper panel doesn't include any covariates while the lower panel includes: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Uses sample of all individuals from ages 19 to 30. Robust standard errors.

Figure 1.14: Long-term Effects of the Crisis on Log Wages



Note: Depicts the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is the log real wage of the individual and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all individuals from ages 19 to 30. Left panel does not control for education and experience (gross effect), while left panel does (pure crisis effect). Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.15: Long-term Effects: Work and Self-Employment



Note: Depicts the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). The upper panels use as independent variable an indicator of whether the individual is working or not, while the lower panel uses an indicator variable of whether the individual is self-employed. In both cases, the following covariates are included: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Uses sample of all individuals from ages 19 to 30. Robust standard errors.

Table 1.1: Recent Severe Macro-economic Recessions

Country	Year	Change in GDP per capita
Argentina	1981	-7.00%
	1985	-8.30%
	2002	-12.00%
Brazil	1981	-6.50%
Colombia	1999	-6.50%
India	1979	-7.00%
Indonesia	1982	-5.74%
	1998	-14.80%
Malaysia	1998	-9.50%
Philippines	1998	-2.70%
Singapore	1998	-4.80%
	2001	-5.20%
Taiwan	2001	-2.10%
Turkey	2001	-9.10%
Venezuela	2002	-10.40%

Source: Barro and Ursua (2008)

Note: Year and magnitude of recent (since 1970) severe macroeconomic crisis in countries characterized by macroeconomic instability.

Table 1.2: Summary Statistics of Education Variables

Year	School attendance	Years of education	I(work)	Hours worked	Hours usually worked	Grade differential	Repeated a grade
1993	0.87	3.94	0.30	28.02	31.25	0.10	0.22
1997	0.82	5.59	0.31	29.06	31.46	0.12	0.21
2000	0.80	5.83	0.32	26.99	30.04	0.10	0.18
2007	0.87	5.56	0.23	29.95	32.82	0.46	0.14
Total	0.83	5.38	0.29	28.29	31.12	0.21	0.18

Note: Children ages 7 to 18.

Year	High school graduate	Progessed to College	Monthly real wage*
1993	0.21	0.05	681,806.90
1997	0.31	0.07	704,426.10
2000	0.32	0.08	589,538.10
2007	0.39	0.11	746,221.90
Total	0.32	0.08	680,411.90

Note: Individuals ages 18 to 30.

* All wages are in 2007 Indonesian Rupiah. Exchange rate was around 1USD = 9,000IR in 2007.

Note: This table present the summary statistics of the main education (and wage) variables that are used as dependent variable throughout the paper. Data comes from the four different waves of the IFLS (1993, 1997, 2000, 2007). Upper panel considers individuals ages 7 to 18 and the lower panel 18 to 30.

Table 1.3: Incidence of the Economic Crisis

Year		Change in log (HH expenditure per capita)		
		Rural	Urban	Difference significant at 95%?
1997	Mean	8.73%	7.46%	No
	N	1,625	1,352	
2000	Mean	-7.96%	-13.14%	Yes
	N	2,070	1,695	
2007	Mean	0.08%	-1.46%	No
	N	1,333	1,243	

Year		Change in log (HH expenditure per capita)				
		Asset quintile				
		1 (poor)	2	3	4	5 (rich)
1997	Mean	3.94%	6.66%	11.30%	7.58%	11.65%
	N	584	624	599	626	542
2000	Mean	-16.60%	-14.03%	-5.88%	-10.86%	-5.22%
	N	637	755	800	824	748
2007	Mean	-14.05%	-2.89%	4.24%	1.35%	6.09%
	N	426	515	539	566	524

Note: Change in the log per capita household real expenditure by different categories: urban versus rural, less affected (household head working in the tradable sector - food production) versus more affected (household head working in the non-tradable sector), by asset quintile (1 being the poorest and 5 the richest). Observations are the household level and include all households whose individuals are used in the empirical exercises.

Table 1.4: Summary of Main Mid-term Effects of the Crisis

	(1)	(2)	(3)	(4)	(5)	(6)
	School Attendance		School Attendance		Years of Education	
Age at Crisis	OLS	OLS	Probit	Probit	OLS	OLS
Ages 5 - 6	0.0187	0.0252	0.0468***	0.0490***	-0.2489	-0.1959
	[0.015]	[0.015]	[0.017]	[0.016]	[0.154]	[0.149]
Ages 7 - 8	-0.0106	-0.0097	-0.0379	-0.0333	-0.2638*	-0.239
	[0.015]	[0.015]	[0.027]	[0.027]	[0.154]	[0.150]
Ages 9 - 10	-0.0121	-0.0032	-0.0228	-0.0136	-0.3492**	-0.2225
	[0.016]	[0.016]	[0.020]	[0.020]	[0.155]	[0.151]
Ages 11 - 12	-0.0300*	-0.0229	-0.0337**	-0.0241	-0.3007*	-0.1419
	[0.018]	[0.018]	[0.017]	[0.016]	[0.158]	[0.152]
Ages 13 - 14	-0.0532**	-0.0566***	-0.0422**	-0.0453***	-0.4872***	-0.4317***
	[0.021]	[0.021]	[0.016]	[0.017]	[0.163]	[0.157]
Ages 15 - 16	-0.0769***	-0.0737***	-0.0542***	-0.0517***	-0.4830***	-0.4586***
	[0.022]	[0.021]	[0.017]	[0.017]	[0.172]	[0.165]
Constant	0.4078***	0.4226***			5.1220***	2.6607***
	[0.014]	[0.020]			[0.154]	[0.176]
Controls	No	Yes	No	Yes	No	Yes
Observations	35,550	34,846	35,550	34,846	33,405	32,718
R-squared	0.280	0.309			0.506	0.545
Pseudo-R			0.261	0.299		

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

Note: The first two columns show the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). The second two show the marginal coefficients of a probit model using the same dependent and independent variables while the third two show the coefficients of a linear regression where the dependent variable is years of completed education. All regressions use the sample of all children from ages 7 to 18. When controls are used, they include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Table 1.5: Mid-term Effects of the Crisis: Work Indicators

	(1)	(2)	(3)
		OLS	
Age at Crisis	I(work)	Hours worked	Hours usually worked
Ages 13 - 14	0.0235	2.7685**	1.8448
	[0.016]	[1.171]	[1.125]
Ages 15 - 16	0.0214	2.0453*	2.2377**
	[0.017]	[1.146]	[1.071]
Observations	36,329	20,863	20,493
R-squared	0.260	0.042	0.040

	(4)	(5)	(6)
	Probit	Tobit	
Age at Crisis	I(work)	Hours worked	Hours usually worked
Ages 13 - 14	0.0136*	3.4129***	2.2410*
	[0.019]	[1.309]	[1.204]
Ages 15 - 16	0.0113	2.2690*	2.5115**
	[0.018]	[1.289]	[1.149]
Observations	36,714	20,863	20,493
R-squared	0.0853	0.005	0.004

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

Note: The first three columns depict the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression. The fourth column uses a probit model and the last two a tobit model (censored at 0). In all specifications the independent variables of interest include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). First and fourth column show the effect on working or looking for a job, the second and fifth on the hours worked last week and the third and sixth on the hours usually worked per week. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Table 1.6: Mid-term Effects of the Crisis: Other Indicators

Age at crisis	(1) OLS	(2) OLS	(3) Probit
	Grade differential	Repeated a grade	
Ages 5 - 6	-0.1400***	0.0258*	0.0476**
	[0.046]	[0.013]	[0.023]
Ages 7 - 8	-0.1083**	0.0199	0.0317
	[0.048]	[0.015]	[0.020]
Ages 9 - 10	-0.0470	0.0162	0.0291
	[0.050]	[0.016]	[0.019]
Ages 11 - 12	0.0136	0.0097	0.0239
	[0.053]	[0.016]	[0.019]
Ages 13 - 14	-0.0598**	0.0174	0.0302
	[0.029]	[0.017]	[0.020]
Ages 15 - 16	-0.1454**	-0.0119	0.0024
	[0.068]	[0.017]	[0.019]
Constant	0.0171	0.1541***	
	[0.051]	[0.015]	
Observations	22,478	33,559	33,559
R-squared	0.101	0.041	0.0494

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

Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). The first column shows the effect on grade differential (current grade - ideal grade), the second and third column show the effect on an indicator variable of whether the child repeated a year using correspondingly a linear probability model and a probit model. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Table 1.7: Effects of the Crisis on EBTANAS Test Scores

	(1)	(2)	(3)	(4)
	Ebtanas scores (primary school)		Ebtanas scores (middle school)	
Post	-8.1851**	-7.3479**	8.1262**	8.7875**
	[3.357]	[3.350]	[3.880]	[3.904]
Time trend	0.0157	0.0642***	-0.0510	-0.0144
	[0.017]	[0.017]	[0.031]	[0.031]
Time trend*post	0.4195***	0.3852**	-0.3814**	-0.4107**
	[0.162]	[0.161]	[0.188]	[0.189]
Post effect	-1.854	-1.534	1.857	2.036
Joint test (post and ttrend*post)				
F-stat	6.849	8.290	2.807	3.344
p-value	0.001	0.000	0.061	0.035
Controls	No	Yes	No	Yes
Observations	9,654	9,558	5,570	5,511
R-squared	0.003	0.090	0.003	0.055

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

	(5)	(6)
	Ebtanas scores (high school)	
Post	-4.7433	-5.3916
	[6.129]	[6.229]
Time trend	0.1182	0.1668
	[0.113]	[0.114]
Time trend*post	0.1604	0.1904
	[0.300]	[0.305]
Post effect	-1.909	-2.028
Joint test (post and ttrend*post)		
F-stat	3.391	3.656
p-value	0.034	0.026
Controls	No	Yes
Observations	2,378	2,354
R-squared	0.002	0.038

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

Note: Individual's test scores are regressed on a post crisis variable ($>$ 1998), a linear time trend, the interaction between these two and socioeconomic controls. The sample consists of individuals ages 7 to 30 who took the EBTANAS test between 1985 and 2001. The first two columns use the primary test score, the second two columns use the middle school test score and the final two the high school test score. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Table 1.8: Selection Analysis of the EBTANAS Test Scores Results

PANEL A. Heckman correction

	Ebtanas scores (primary school)	Ebtanas scores (middle school)
Post	0.2974	8.7914***
	[2.536]	[2.749]
Time trend	0.0819***	-0.0118
	[0.030]	[0.031]
Time trend*post	-0.0458	-0.4112***
	[0.122]	[0.132]
Lambda (correction term)	-4.658	6.490
	[0.413]***	[5.066]**
Observations	8,121	12,141

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

PANEL A. Predicted score (mean) by attendance

Not attending	32.39	33.34
Attendending	32.48	33.44
Difference significant	Yes	Yes

Note: Individual's test scores are regressed on a post crisis variable (>1998), a linear time trend, the interaction between these two, socioeconomic controls and the inverse Mills ratio of the attendance regression. The sample consists of individuals ages 7 to 30 who took the EBTANAS test between 1985 and 2001. The first two columns use the primary test score, the second two columns use the middle school test score and the final two the high school test score. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Table 1.9: Long-Term Effects of the Crisis: Wage

PANEL A. Summary of Long Term Effects

	(1)	(2)	(3)	(4)
Age at Crisis	Years of Education		Log real monthly wage	
Ages 11-12	-1.0495	-1.0972	-0.9232**	-1.0613*
	[0.806]	[0.750]	[0.385]	[0.590]
Ages 13-14	-1.0196	-0.6273	-1.6397***	-2.4483***
	[0.832]	[0.762]	[0.420]	[0.892]
Ages 15-16	-1.8929**	-1.6512**	-0.8789**	-0.5201
	[0.834]	[0.765]	[0.379]	[0.513]
Ages 17-18	-2.1321***	-1.4270*	-1.0504***	-0.7692**
	[0.823]	[0.743]	[0.318]	[0.346]
Constant	5.7592***	2.0106***	13.0490***	11.7263***
	[0.099]	[0.121]	[0.072]	[0.095]
Controls	No	Yes	No	Yes
Observations	54,433	53,302	8,014	4,627
R-squared	0.409	0.491	0.035	0.265

PANEL B. Heckman Correction for Wage Effects

	(5)	(6)
Age at Crisis	Log real monthly wage	
Ages 11-12	-1.9032**	-1.0126
	[0.839]	[0.845]
Ages 13-14	-3.1686***	-2.4251***
	[0.654]	[0.672]
Ages 15-16	-0.9153*	-0.5367
	[0.472]	[0.460]
Ages 17-18	-1.2596***	-0.7766**
	[0.345]	[0.352]
Constant	13.6307***	12.4986***
	[0.368]	[0.416]
Lambda (correction term)	-1.4356***	-1.5643***
	[0.050]	[0.052]
Controls	No	Yes
Observations	26,108	25,650
R-squared		

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

Note: In Panel A, the dependent variable is years of education for the first two columns and log real monthly wage for the second two. The table shows the Yr2007 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies and Yr2007 dummy interacted with the age group specific dummies (variables of interest). In Panel B, the Heckman Selection method is used to control for selection bias. When controls are used, they include: gender, urban, household asset quintile, adjusted household size, % males, average age, # of siblings and female household head. Uses sample of all individuals from ages 18 to 30. Robust standard errors.

Table 1.10: Distribution: IFLS vs. Census

Census Edu. Category	Years of education (IFLS)	(1) *IFLS* Freq.	(1) *IFLS* Percent	(2) *Census* Freq.	(2) *Census* Percent
1	2.76	16,434	49.20%	789,229	53.37%
2	6.46	9,476	28.37%	380,492	25.73%
3	9.43	5,595	16.75%	225,325	15.24%
4	12.04	1,748	5.23%	78,765	5.33%
5	14.44	108	0.32%	4,857	0.33%
6	16.00	44	0.13%	104	1.00%
Total		33,405	100%	1,478,772	100%

Note: This table shows the distribution of the education variable available in the Census data in both the IFLS and the Census data (years 1990/1995/2000/2005/2010). The Census education category is a categorical variable that takes the following values: 1 for less than primary, 2 for primary completed, 3 for junior high completed, 4 for senior high completed, 5 for technical education completed and 6 for university completed. Using the IFLS data, I computed the mean years of completed education for each category.

Table 1.11: IFLS vs. Census data

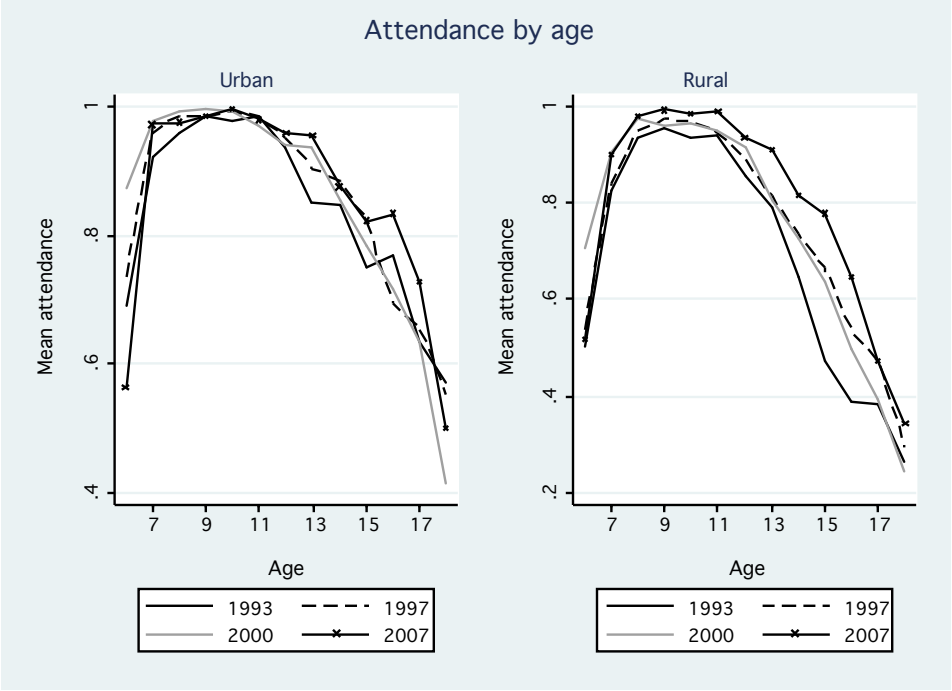
	(1)	(2)	(3)	(4)
Age at Crisis	*IFLS*		*Census*	
Ages 7 - 8	-0.1134***	-0.0957**	-0.0847***	-0.0867***
	[0.040]	[0.039]	[0.002]	[0.002]
Ages 9 - 10	-0.0836**	-0.0751*	-0.0854***	-0.0878***
	[0.041]	[0.040]	[0.002]	[0.002]
Ages 11 - 12	-0.0354	-0.0014	-0.0240***	-0.0264***
	[0.043]	[0.042]	[0.003]	[0.003]
Ages 13 - 14	-0.0949**	-0.064	-0.0297***	-0.0311***
	[0.043]	[0.042]	[0.004]	[0.004]
Ages 15 - 16	-0.0796*	-0.0737*	-0.0392***	-0.0400***
	[0.046]	[0.044]	[0.004]	[0.004]
Ages 17 - 18	-0.0757	-0.0775*	-0.0518***	-0.0511***
	[0.048]	[0.047]	[0.005]	[0.005]
Constant	2.0166***	1.7946***	1.9163***	1.8741***
	[0.039]	[0.041]	[0.002]	[0.003]
Controls	No	Yes	No	Yes
Observations	33,405	33,083	3,031,062	2,969,442
R-squared	0.493	0.519	0.341	0.343

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

Note: The dependent variable is a categorical variable that takes the following values: 1 for less than primary, 2 for primary completed, 3 for junior high completed, 4 for senior high completed, 5 for technical education completed and 6 for university completed. The table depicts linear regressions using two different samples: the IFLS, panel data (columns 1 and 2) and the Indonesian Census, 1990/1995/2000/2005/2010 (columns 3 and 4). For the second regression the controls used are: Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. For the fourth regression, the controls include gender and household size. Robust standard errors.

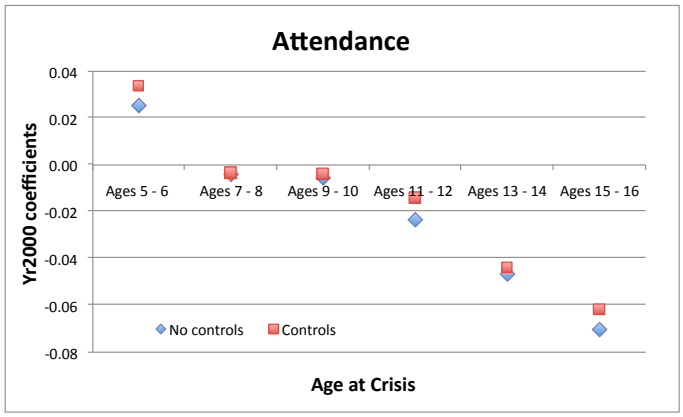
Appendix A

Figure 1.16: Attendance by Age: Urban Versus Rural Areas



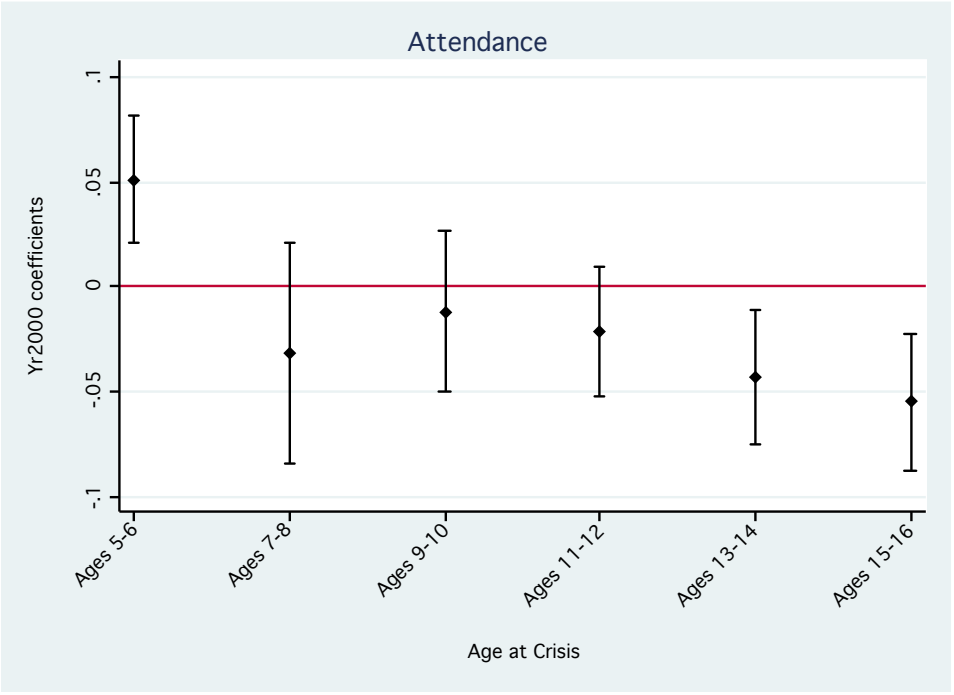
Note: Data from IFLS survey, all waves (1993, 1997, 2000, 2007). Mean attendance for every age in every year using sample of all children from ages 7 to 18. Attendance is an indicator variable that equals one if child is attending to school. Left panel shows urban sample while right panel shows rural sample.

Figure 1.17: Mid-term Effects of the Crisis on Attendance: Regressions by Age Group



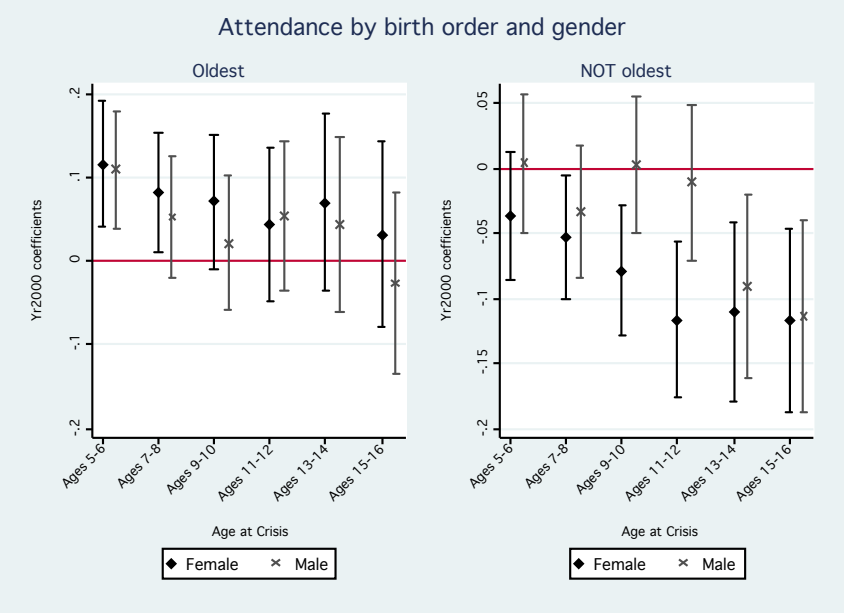
Note: Each point represents the Yr2000 coefficient of separate linear regressions by age group, where an indicator variable for attendance at the individual level is regressed on a Yr2000 dummy. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.18: Probit Model: Mid-term Effects of the Crisis on Attendance



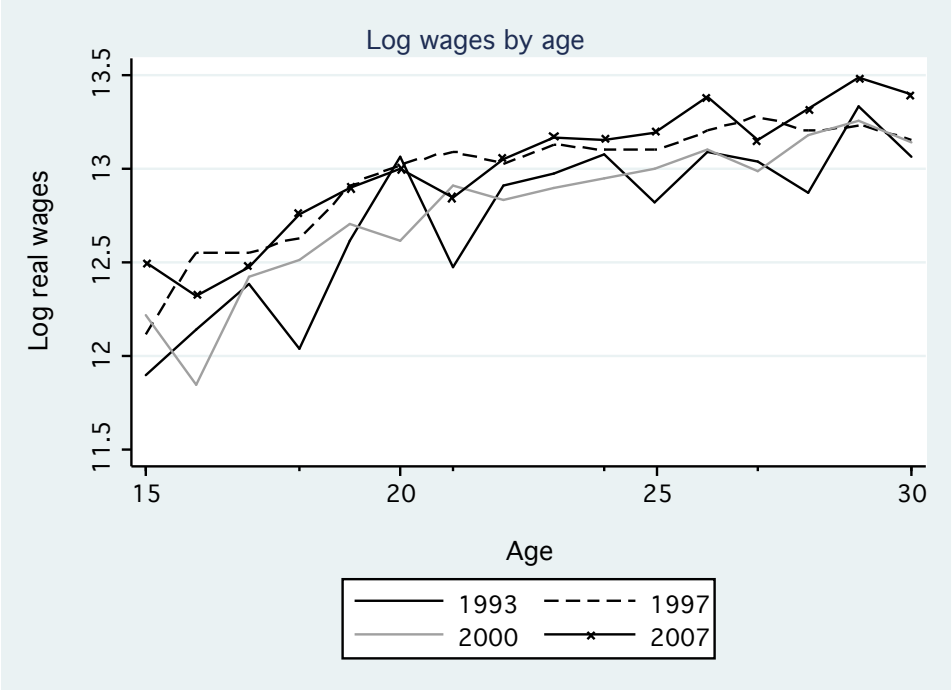
Note: Depicts the Yr2000 age group specific marginal effects (and confidence interval at the 95%) of a probit model in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18, divided by oldest and rest of the children. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.19: Mid-term Effects of the Crisis on Attendance by Birth Order and Gender



Note: Depicts the Yr2000 age group specific coefficients (and confidence interval at the 95%) of the linear regression in which the dependent variable is an indicator variable for attendance at the individual level and the independent variables include: age group specific dummies, linear time trend, the interaction of these last two, Yr2000 dummy interacted with the age group specific dummies (variables of interest). Uses sample of all children from ages 7 to 18, divided by oldest and rest of the children. Controls include: gender, urban, household asset quintile, adjusted household size, % males, % children under 6, average age, # of siblings and female household head. Robust standard errors.

Figure 1.20: Log Real Wage by Age



Note: Data from IFLS survey, all waves (1993, 1997, 2000, 2007). Mean log real wage (2007 Indonesian Rupiahs) for every age in every year using sample of all individuals from ages 19 to 30. Wage includes salary from primary and secondary jobs (when it applies).

Table 1.12: Number of Observations at Every Age

	Number of individuals				
Age	1993	1997	2000	2007	Total
7	499	711	655	916	2,781
8	516	710	683	820	2,729
9	574	685	669	692	2,620
10	631	703	726	645	2,705
11	569	725	709	682	2,685
12	607	786	669	673	2,735
13	603	830	708	613	2,754
14	520	891	727	642	2,780
15	98	759	658	541	2,056
16	88	730	727	566	2,111
17	112	719	737	555	2,123
18	104	675	648	534	1,961
19	95	545	577	473	1,690
20	144	541	554	466	1,705
21	155	448	551	479	1,633
22	154	457	518	493	1,622
23	208	392	492	582	1,674
24	187	405	485	539	1,616
25	373	519	516	595	2,003
26	261	374	395	572	1,602
27	301	438	469	521	1,729
28	358	387	463	536	1,744
29	293	511	423	487	1,714
30	570	543	484	459	2,056
Total	10,471	17,523	17,113	17,176	62,283

Note: Number of observations of every age used in the econometrical exercises, in every wave. Note that for the IFLS 1 (1993), less individuals aged 15-20 are included in the sample.

Table 1.13: Summary Statistics of Education Variables by Age

Age	School attendance	Years of education	I(work)	Hours worked	Hours usually worked	Grade differential	Repeated a grade	High school graduate	Progessed to College	Monthly real wage*
7	0.91	1.18				0.40	0.06			
8	0.97	2.10				0.37	0.10			
9	0.98	2.94				0.30	0.13			
10	0.97	3.78				0.15	0.19			
11	0.97	4.71				0.13	0.20			
12	0.92	5.42				0.08	0.23			
13	0.86	6.23				0.11	0.25			
14	0.80	7.01				0.19	0.23			
15	0.74	7.74	0.19	26.37	28.59	0.17	0.22			
16	0.65	8.41	0.26	25.53	28.79	0.25	0.21			
17	0.56	8.82	0.32	29.94	32.24	0.18	0.22			
18	0.40	9.27	0.40	29.92	33.31	-0.03	0.18			
19	0.21	9.45	0.55	34.29	36.76			0.40	0.06	491,520
20	0.14	9.22	0.60	33.80	36.77			0.41	0.10	526,387
21	0.10	9.54	0.60	33.69	37.76			0.46	0.12	536,657
22	0.09	9.48	0.59	35.09	38.38			0.44	0.12	592,862
23	0.06	9.59	0.66	35.54	38.27			0.45	0.12	657,218
24	0.04	9.50	0.66	36.08	38.65			0.43	0.11	648,613
25	0.03	9.06	0.69	36.50	39.20			0.41	0.10	659,266
26	0.02	9.44	0.67	35.57	37.84			0.43	0.11	736,006
27	0.03	9.29	0.71	35.71	38.31			0.43	0.12	698,844
28	0.01	8.99	0.71	37.41	40.83			0.41	0.10	724,670
29	0.02	9.15	0.70	36.49	39.36			0.42	0.12	820,489
30	0.01	8.32	0.73	35.22	38.42			0.36	0.12	772,764
31	0.01	8.96	0.73	36.70	39.65			0.42	0.13	836,573
32	0.01	8.54	0.72	35.84	38.91			0.37	0.11	828,007
33	0.02	8.22	0.75	36.16	39.75			0.35	0.10	788,514
34	0.01	8.23	0.75	36.55	39.16			0.34	0.10	859,802
35	0.01	7.44	0.76	35.79	39.00			0.29	0.08	787,974
Total	0.48	6.83	0.40	34.96	37.92	0.20	0.18	0.32	0.08	680,412

* All wages are in 2007 Indonesian Rupiah. The exchange rate was around 1USD = 9,000IR.

Note: This table present the summary statistics, by age, of the main education (and wage) variables that are used as dependent variable throughout the paper. Data comes from the four different waves of the IFLS (1993, 1997, 2000, 2007).

Table 1.14: Incidence of the Crisis on PC Household Expenditure Change 1997-2000

	(1)	(2)
	Change in log (pc_exp) 1997-2000	% change in pc_exp 1997-2000
HH age	0.0010	0.0008
	[0.001]	[0.001]
HH female	0.0707	0.1099*
	[0.051]	[0.062]
HH education	-0.0031	-0.0039
	[0.003]	[0.003]
Adjusted size	-0.0273***	-0.0231***
	[0.006]	[0.007]
% under 6 yrs	-0.2405**	-0.2506**
	[0.096]	[0.103]
Urban	-0.0543**	-0.0753***
	[0.023]	[0.025]
Owner	0.0540*	0.0432
	[0.032]	[0.033]
More Affected	-0.0531**	-0.0342
	[0.023]	[0.025]
Quintile 1	-0.1106***	-0.1248***
	[0.040]	[0.045]
Quintile 2	-0.1046***	-0.1204***
	[0.036]	[0.039]
Quintile 3	-0.0624*	-0.0840**
	[0.034]	[0.037]
Quintile 4	-0.0915***	-0.0918**
	[0.033]	[0.036]
Constant	0.1029	0.2436**
	[0.090]	[0.097]
Observations	3,186	3,186
R-squared	0.021	0.018

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

Note: This table shows two linear regressions. The dependent variable in the first column is the change in the log per capita real household expenditure while in the second is the percentage change in the per capita real household expenditure, both at the household level. Sample includes all households in IFLS3 (year 2000 wave).

Table 1.15: EBTANAS Scores, Before and After the Crisis

	All				Primary School			
Variable	Pre	Post	Difference	Significant (95%)?	Pre	Post	Difference	Significant (95%)?
Gender	0.477	0.505	0.028	Yes	0.489	0.452	-0.037	Yes
Urban Household	0.482	0.509	0.027	Yes	0.531	0.550	0.019	No
Asset Quintile	3.060	2.915	-0.145	Yes	3.269	3.136	-0.133	Yes
Adjusted age gender HH size	4.750	5.343	0.594	Yes	5.138	5.329	0.191	Yes
% of males in HH	0.491	0.483	-0.008	No	0.493	0.488	-0.004	No
% HH members under six	0.121	0.118	-0.003	Yes	0.080	0.131	0.052	Yes
Average age	24.297	25.264	0.966	Yes	25.005	26.049	1.044	Yes
Number of siblings	1.950	1.469	-0.481	Yes	2.335	1.422	-0.913	Yes
Female HH head	0.113	0.114	0.000	No	0.110	0.120	0.010	No

	Middle School				High School			
Variable	Pre	Post	Difference	Significant (95%)?	Pre	Post	Difference	Significant (95%)?
Gender	0.484	0.467	-0.018	No	0.500	0.486	-0.014	No
Urban Household	0.623	0.623	0.000	No	0.737	0.679	-0.058	Yes
Asset Quintile	3.511	3.305	-0.206	Yes	3.753	3.413	-0.340	Yes
Adjusted age gender HH size	5.308	5.183	-0.125	Yes	5.565	5.186	-0.380	Yes
% of males in HH	0.490	0.493	0.002	No	0.491	0.497	0.006	No
% HH members under six	0.066	0.141	0.075	Yes	0.050	0.141	0.091	Yes
Average age	25.929	26.165	0.236	No	27.332	26.690	-0.643	No
Number of siblings	2.410	1.346	-1.064	Yes	2.525	1.408	-1.117	Yes
Female HH head	0.115	0.106	-0.009	No	0.135	0.101	-0.034	No

	Pre	Post	Difference	Significant (95%)?
Ebtanas (primary school)	31.695	32.080	0.385	Yes
Ebtanas (middle school)	33.236	33.819	0.583	Yes
Ebtanas (high school)	30.766	31.696	0.930	Yes

Note: This table depicts the mean values for the controls and for the EBTANAS test scores, before and after the economic crisis, and for each of the three samples according to which EBTANAS test the child presented.

1.9 References

Barro, Robert J. and Jose F. Ursua, 2008. "Macroeconomic Crises since 1870," Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, vol. 39(1 (Spring), pages 255-350.

Beaudry, Paul, and John DiNardo. 1991. "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro data." Journal of Political Economy 99(4) August: 665-688.

Betts, Julian R., and Laurel McFarland. 1995. "Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments." Journal of Human Resources 30(4): 741-65.

Black, Sandra E., and Kenneth L. Sokoloff. 2006. "Long-Term Trends in Schooling: The Rise and Decline of Public Education in the United States." Handbook of the Education of Economics, Volume 1, Eric Hanushek and Fins Welch, eds., Elsevier.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2007. "From the cradle to the labor market? The effect of birth weight on adult outcomes." Quarterly Journal of Economics, 122(1):409–439, February 2007.

Chetty, Raj John Friedman, Nathaniel Hilger, Emmanueal Saez, Diane Schanzenbach and Danny Yagan. 2010. "How Does your Kindergarten Classroom Affect your Earnings?" Evidence from Project STAR. NBER Working Paper # 16381.

Cunha, Flavio and Heckman, James. 2007. "The Technology of Skill Formation". AEA Papers and Proceedings on Economics of Human Development, Vol. 97 No. 2.

Currie, Janeth. 1997. "The Effect of Welfare on Child Outcomes: What We Know and What We Need to Know". Working Paper.

Duryea, Suzanne, and Mary Arends-Kuenning. 2003. "School Attendance, Child Labor and Local Labor Market Fluctuations in Urban Brazil." World Development 31(7): 1165-78.

Ferreira, Francisco and Norbert Schady (2008). “Aggregate Economic Shocks, Child Schooling and Child Health”. World Bank Research Working Paper.

Frankenberg, Elizabeth, Duncan Thomas, and Kathleen Beegle. 1999. "The real costs of Indonesia's economic crisis: Preliminary findings from the Indonesia Family Life Surveys". Working paper no. DRU-2064-NIA/NICHD. Santa Monica, CA: RAND.

Frankenberg, Elizabeth, James P. Smith, and Duncan Thomas. 2004. "Economic shocks, wealth and welfare". Journal of Human Resources 38 (2): 280–321.

Friedman, Jed and Duncan Thomas. 2007. "Psychological Health Before, During and After the Economic Crisis: Results from Indonesia, 1993 - 2000." The World Bank Policy Research Paper.

Friedman, Jed and James Levinsohn. 2001. "The Distributional Impacts of Indonesia's Financial Crisis on Household Welfare: A "Rapid Response" Methodology. Discussion paper No. 482.

Funkhouser, Edward. 1999. “Cyclical Economic Conditions and School Attendance in Costa Rica.” Economics of Education Review 18(1): 31-50.

Gershoff, E., J. L. Aber, C. Raver, and M. C. Lennon. 2007. “Income Is Not Enough: Incorporating Material Hardship into Models of Income Associations with Parenting and Child Development.” Child Development 78 (1): 70–95.

Goldin, Claudia. 1999. “Egalitarianism and the Returns to Education during the Great Transformation of American Education.” Journal of Political Economy 107(6): S65-S94.

Hartono, Djoko and David Ehrmann. 2001. "The Indonesian Economic Crisis and its Impact on Educational Enrollment and Quality". Document prepared for the Institute of Southeast Asian Studies seminar.

Hill, Hal and Takashi Shiraishi. 2007. “Indonesia After the Asian Crisis”. Asian Economic Policy Review 2, 123- 141.

Heckman, James. 2007. "The Economics, Technology, and Neuroscience of Human Capital Formation". PNAS, Vol, 104, No. 33.

Jensen, Robert. 2000. "Agricultural Volatility and Investments in Children." American Economic Review, Papers and Proceedings, 90(2): 399-404.

Jones, Gavin and Peter Hagul. 2001. "Schooling in Indonesia: Crisis-related and Longer-term Issues". Bulletin of Indonesian Economic Studies.

Kahn, Lisa B. 2010. "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy." Labour Economics. April, 2010. Vol. 17, No. 2.

Levinsohn, James, Steven Berry and Jed Friedman. 1999. "Impacts of the Indonesian economic crisis: Price changes and the poor". In Managing currency crises in emerging markets, ed. Michael Dooley and Jeffrey Frankel, 393–424. Chicago: University of Chicago Press.

Lundberg, Mattias and Alice Wuermli (eds). 2012. Children and Youth in Crisis: Protecting and Promoting Human Development in Times of Economic Shocks. Washington DC: The World Bank.

Maluccio, John A., John Hoddinott, Jere R. Behrman, Agnes Quisumbing, Reynaldo Martorell and Aryeh D. Stein, 2008. "The Impact of Nutrition During Early Childhood on Education among Guatemalan Adults," Philadelphia-Washington-Atlanta: University of Pennsylvania, IFPRI, Emory.

McKenzie, David J. 2003. "How do Households Cope with Aggregate Shocks? Evidence from the Mexican Peso Crisis." World Development 31(7): 1179-99.

Oreopoulos, Phil, Till von Wachter, and Andrew Heisz. 2012. "The Short- and Long-Term Career Effects of Graduating in a Recession." American Economic Journal - Applied Economics. January 2012, 4(1): 1–29.

Oyer, Paul. 2006. "Initial Labor Market Conditions and Long-Term Outcomes for Economists." Journal of Economic Perspectives 20(3) Summer: 143-160.

Oyer, Paul. 2007. "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income." Stanford GSB mimeo.

Peacock, James. 1973. Indonesia: An Anthropological Perspective. Goodyear Publishing Company, Inc. Pacific Palisades, California.

Schady, Norbert. 2004. "Do Macroeconomic Crises Always Slow Human Capital Accumulation?" World Bank Economic Review 18(2): 131-54.

Strauss, John, Kathleen Beegle, Agus Dwiyanto, Yulia Herawati, Daan Patti- nasaramy, Elan Satriawan, Bondan Sikoki, Sukamdi, and Firman Witolear. 2004. "Indonesian Living Standards Before and After the Financial Crisis". Singapore: Institute of Southeast Asian Studies.

Thomas, Duncan and Elizabeth Frankenberg. 2006. "Household responses to the financial crisis in Indonesia: Longitudinal evidence on poverty, resources and well-being". California Center for Population Research, UCLA.

Thomas, Duncan and Elizabeth Frankenberg. 2007. "Labor market transitions of men and women during an economic crisis: Evidence from Indonesia". In Women in the labour market in changing economies: Demographic is- sues, ed. B. Garcia, R. Anker, and A. Pinnelli, 37–58. Oxford: Oxford University Press.

Thomas, Duncan, and Elizabeth Frankenberg. 2006. "Household response to the financial crisis in Indonesia: Longitudinal evidence on poverty, resources and well-being". Journal of Development Economics 74:53–85.

Thomas, Duncan, Firman Witoelar, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, Cecep Sumantri, Wayan Suriastini. 2012. "Cutting the costs of attrition: Results from the Indonesia Family Life Survey". Journal of Development Economics. Vol. 98 (2012) 108–123.

Thomas, Duncan, Kathleen Beegle, Elizabeth Frankenberg, John Strauss, Bondan Sikoki, and Graciela Teruel. 2004. "Education in a crisis". Journal of Development Economics 74:53–85.

World Bank. 2007. Malawi Poverty and Vulnerability Assessment: Investing in Our Future. Washington, D.C.

Chapter 2

2 Peer Feedback and Teaching Performance: A Randomized Controlled Trial[39]

2.1 Introduction

Recently, there has been a significant effort within the academic literature to understand what types of interventions improve teaching practices, yet very little is known in the context of higher education, and even less with respect to teaching assistants, who play an increasingly important role in instruction in large American universities.[40] In part, this is explained by the fact that most of the quantitative studies on the subject have focused on performance-based incentives at the elementary and secondary school levels, in which teachers are rewarded according to their students' results in terms of standardized tests or grade improvement (Umansky, 2005). The studies that have explored interventions to improve teaching practices at the higher education level, primarily from the education literature, have done so from a qualitative perspective. These are, of course, informative but can not establish causality between a given program or intervention and teaching outcomes, and can not be used to rank different alternatives in terms of their effectiveness (or cost-effectiveness). As Caroll (1980) points out, interventions such as training sessions, the assessment of teacher performance by an education expert or a school administrator, peer observation by colleagues,

[39]This paper is co-authored with Gabriela Rubio, UCLA.

[40]Although there has been limited attention given to teaching assistants in the literature, they do appear to have an effect on students' performance (Hanushek and 2007, Koedel and Betts 2007, Borjas 2000, Watts and Lynch, 1989).

and to a much lesser extent, self-reviews from the individuals themselves, have been analyzed within this strand of the academic literature.

One particularly attractive alternative that has not been fully explored quantitatively is *peer observation*, defined by Bell and Mladenovic (2008) as a "collaborative and developmental activity in which professionals offer mutual support by observing each other teach; explaining and discussing what was observed; sharing ideas about teaching...". Peer observation provides competent asessment from colleagues who perform the same activity (and thus possess comparable academic qualifications), and who are familiar with the context in which the teaching is taking place. Peer observation ideally builds a sense of collegiality between peers, and can improve teaching abilities not only through feedback but through reciprocal observation and learning. On the downside, peer observation may be percieved as intrusive or unconfortable by participants, and could be influenced by the subjective opinion of the observer. This intervention is more likely to be a positive and useful experience when the feeback is non-judgemental and constructive.

Previous qualitative studies have investigated the effect of peer observation on the behavior and performance of university's teaching assistants (TA), and suggest that it is both useful and valuable (Sparks, 1986 and Bell and Mladenovic 2008). Nonetheless, to the best of our knowledge, no quantitative studies have attempted to understand the effect of peer feedback in the performance of teaching assistants. This study represents a first step to overcome this shortcoming of the existing literature. Using a randomized controlled trial (RCT), it aims to establish a causal relationship between trained peer feedback and teaching performance of TAs, and between trained peer feedback and student performance or grade (deviation from course mean). We randomly assigned all the TAs in the Department of Economics of a large public university to either a control group, which was left untouched for the quarter, or a treatment group, which participated in trained peer feedback and receieved a cash reward as compensation. We refer to our intervention as trained peer feedback, because at the beginning of the intervention, the TAs in the treatment group participated in

an interactive workshop in which they learned how to impart feedback in a constructive and positive manner.

The intervention targeted 55 available TAs in the Economics Department of a large public university during the Fall Quarter of 2012 (ten weeks long). Of these TAs, 32 were assigned into the treatment group and 23 into the control group. Only 25 of those assigned to the treatment group agreed to participate in the intervention, which translates into a take up rate of 78%.[41] The treatment consisted of two elements: first, at the beginning of the intervention, the TAs participated in a two hour workshop that covered the most important teaching skills and gave guidance on how to provide constructive criticism to others. Second, all the TAs in the treatment group were observed while teaching and received detailed written feedback from other TAs in the treatment group; over the course of the quarter, all TAs belonging to the treatment group were evaluated on two occasions by two fellow TAs, and, in turn, twice provided feedback to two fellow TAs. Both the observation and the feedback activities were performed using predetermined formats provided by the research group in order to standardize the type of feedback given and to guide the dimensions over which the TAs were evaluated. To promote participation and justify the hours of extra work, the TAs in the treatment group were compensated with a cash reward of a $100.

The analysis of the effect of the RCT on the TAs' teaching skills was performed using data from the students' evaluation for the Fall 2012 and Winter 2013 quarters. The results for the Fall Quarter suggest that the intervention had a positive but non-significant effect on the student evaluations for that quarter. Since the TAs had little time to incorporate the peer feedback into their teaching before the student evaluations took place, this result was not unexpected. We find no evidence of an effect on the students' performance during that same quarter, proxied by deviation of the section mean grade from that of the course mean (most courses, particularly the introductory and lower level courses, have multiple sections and multiple TAs). As mentioned before, it might be possible that the TAs did not fully

[41] For this reason we estimate both intent to treatment effects (ITT) and effects of treatment on the treated (ToT).

adjust their teaching during that same quarter, and thus those students did not receive the full benefit from the intervention. An alternative explanation is a relatively low weight of TA in the students' performance production function.[42]

Nonethless, the intervention had a significant effect of substantial magnitude on the students' teaching evaluations of the Winter Quarter. We find an increase of at least one half of a standard deviation in the TAs' performance both for the intent to treat (ITT) and treatment groups (ToT). Interestingly, the intervention had a positive effect over many dimensions of what is considered good teaching: concern about the students' learning, organization of the class, interaction with students (making them feel welcome), and communication skills.

Regarding the design of the intervention: the post-intervention qualitative survey suggests that while the TAs were aware that they were being observed, and generally felt the program was valuable. They mentioned that the contents of the feedback itself were more a reinforcement of the positive aspects of their teaching and a reminder of problems they already knew they had to work on, instead of specific actions they could take to improve their teaching. The qualitative input also suggests that TAs learned not only through feedback they received, but also by observing their peers teach. In future studies, we should be more emphatic of the importance of having specific suggestions as part of the feeback given to TAs and we should try to distinguish between the effect of feedback and that of learning through observation, in order to understand better the mechanisms through which the intervention is operating.

The rest of the paper is organized as follows. Section 2.2 provides a brief literature review regarding the importance of teaching quality in higher education, and interventions aimed at enhancing teaching quality, with a focus on papers that have addressed the topic of peer observation. The details of the sample and the experimental design are presented in Section 2.3, while the descriptive statistics and the results of the effects of the intervention

[42]As future research, we propose to analyze if there was any effect of the intervention on the students' grades of the Winter Quarter, to better understand the role of TAs in students' performance; the data has not been released to us at the time of writing.

on teaching skills and students' performance are included in Section 2.4. Section 2.5 contains some complementary analyses on the qualitative survey that the treated TAs answered after the the intervention. Section 2.6 concludes.

2.2 Literature Review

The literature on the economics of education has based its analysis on a simple production model in which different inputs interact –school resources, teacher quality, family attributes, among others– to determine student achievement (Hanushek, 2007). Extensive research has been conducted on the effect of each input on student outcomes – e.g. grades, labor market wages, choice of major – with special interest on those inputs that can be affected by public policies, such as institutional aspects of the school system, differing school resource funding models and the teachers' payscale (Hanushek, 2003).

Most of the research on the effect of teacher quality on students' performance has focused on primary and secondary education. However, recent studies have started investigating the effect of professors' quality on students' achievements at the college level (Ehrenberg and Ziang, 2005; Bettinger and Long, 2011; Hoffman and Oreopolus, 2006; Carrell and West, 2008), finding sizable effects on different outputs such as likelihood of dropping a class, number of same-subject courses taken in second and third year, and overall GPA.

While the impact of professors' quality has become a subject of great interest, researchers have paid little attention to the impact of graduate Teaching Assistants (TAs) despite their widespread use in the American higher education system (Park, 2004). Tuckman (1975) is one of the first studies to explore the impact of TAs on students' outcomes. He is concerned, however, with a different question. The main aim of his study is to compare the performance of TAs as instructors to the performance of more experienced faculty, finding that TAs are as effective as experienced faculty. Watts and Lynch (1989) consider several factors affecting students' achievement; among those factors, they stress the effect of non-native English

speaker TAs. Their findings suggest a negative impact on the output of students.

However, their results contrast with Norris (1991) who finds that non-native English-speaker TAs outperformed natives after controlling for "teaching experience". Finally, Borjas (2000) conducts a similar study in which he analyzes the impact of non-native English-speaker TAs on students' grades. He finds that lack of English-language proficiency among foreign-born TAs adversely affects students' understanding of the material, resulting in a lower grade for the course. Nonetheless, he also finds that the results go away for better prepared foreign-born TAs. The results of these papers suggest two key things. First, TAs' "quality" seems to matter for students' performance in a given course. Second, increased effort, better preparation, or increased teaching skills seem to be as important as other characteristics of the TAs; better teaching skills or higher effort compensate for a lower proficiency in English.

Once the importance of teaching quality was established, the literature on the economics of education became greatly interested in understanding how teaching and/or teachers' quality can be improved. There are a wide range of studies that try to understand the effects of monetary and non-monetary incentives. The studies on monetary incentives have found that the effects are concentrated in those areas at which the incentives were targeted, but these effects appear to be short-termed, and non-existent in related but untargeted areas of knowledge (Umansky, 2005, Glewwe et al. 2010). Similarly, this type of incentive can lead teachers either to "teach to the test" or, even more disturbing, to cheat (Levitt and Jacob, 2003).

As mentioned above, the education literature has explored, primarly from a qualitative perspective, the effectiveness of alternative types of incentives and interventions to improve teaching practices, such as training sessions, the assessment of teacher performance by an education expert or a school administrator, peer observation by colleagues in the same discipline, and to a much lesser extent, self-reviews from the individuals themselves (Carroll, 1980). Peer observation seems to be an appealing alternative, as pointed out by Bell and

Mladenovic (2008), given that the feedback provided by peers is competent and relevant, and because the assesment is coming from someone who is familiar not only with the context of the teaching (in terms of the institution, and the expectation of the students) but also with the content of the material.

Supporters of this option emphasize the fact that it is based on constructive feedback and monitoring among colleagues that results in improvements in teaching practices and the enhancement of teaching confidence (Bell, 2005), development of collegiality and an increased respect for colleagues (Quinlan and Akerlind, 2000), and integration of tutors into the department (Allen, 2002). The common methodology of these studies is to engage small/medium (30 people) groups of instructors or lecturers in peer review exercises (of short duration, one or two observations) and assess the success of the intervention based on qualitative surveys (or interviews) of the participants. These surveys usually inquire about how satisfied are they with the peer feedback exercise and how helpful they think it was.

Despite the overall positive appraisal of peer feedback, some studies have found negative aspects to it. It may be considered intrusive and uncomfortable by the teachers who are being observed and it has a subjective component that is sometimes difficult to assess (David and Macayan, 2010). Similarly, as pointed out by Bell (2005), it maybe challenging to engage in critical reflection and providing and accepting feedback. Finally, the participants sometimes believe that the peer evaluation may not reflect their true ability, particularly when it is used as an input for institutional decisions, such as promotions (Allen 2002).

In this context, this study contributes to the existing literature in several ways. This paper provides new insight into the effect of peer observation interventions by conducting (to the best of our knowledge) the first quantitative assement that can establish a causal link between intervention and outcome. Secondly, this paper contributes to the understanding of the impact and importance of teaching assistants by reporting (to the best of our knowledge) on the first randomized control trial that focuses on teaching assistant instructional performance.

2.3 Experimental Design[43]

The intervention took place during the Fall Quarter of 2012 in the Economics Department of a large public university. The class enrollment requirements allowed for a total of 55 TAs to be eligible for the intervention.[44] Every graduate student with a Teaching Assistant Fellowship is responsible for teaching two discussion sessions of a given course per week throughout the academic quarter (the only exception being the three TAs of the core graduate courses, all of which were included in the control group since the beginning, they only teach one section per quarter). The eligible TAs were randomly selected into one of two groups: the control or the treatment group, and those in the treatment group could decline to participate in the intervention. This section provides specific details of the experimental design and the recruitment process of the intervention.

2.3.1 Experimental Design

Most peer review programs are designed in such a way that within a school or department, teachers and/or professors evaluate each other. Along those lines, the TAs assigned to the treatment group acted both as observers and observed subjects. Within the treatment group, each TA was observed and evaluated by two other TAs of the same group while teaching discussion section, twice during the quarter. The evaluation took place around week four and week six of the ten week quarter, but this varied according to the dates in which midterms took place (most discussion sections immediately following the midterm are cancelled or simply review the exam answers) and to individual TA availability. The observation date

[43]Let this be another opportunity to express our deepest gratitute to the Economics Department for the financial, administrative and overall support. Special thanks to Joe Ostroy, Kathleen McGarry and Roger Farmer, who made this project possible.

[44]It must be noted that there were a total of 57 TAs in the Department, but for two of them, the format of the discussion section was significantly different than the rest, so they were discarded from the beginning.

was only announced to the observer but not to the observed TA, in an attempt to prevent any special preparation; however, TAs could have more or less inferred the timing based on their own observation dates.

All observations/assessments followed a detailed format (shown in Figure 2.1) that emphasizes the factors related to teaching effectiveness that are under the TAs control. Observers were instructed not to interfere with the discussion session in any way and to submit written feedback to the project managers within the next couple of days. The feedback format was similar to the observation format but it included suggestions of specific actions in each of the fields related to teaching efficiency that the TA could adopt to improve their discussion sections (see Figure 2.2). This way, the feedback provided to the TA was meant to have a constructive and useful tone, instead of being mere criticisms. Note that because all the observers belong to the treatment group, the intervention also involved attending the TA sections of two different TAs. In that sense, they could have learned or noticed a teaching practice that could have been useful to them in their own teaching. This to underline that feedback was not the only component of the treatment - members of the treatment group were also exposed to teaching practices of their peers, an experience which may also have had an impact on their teaching.

One of our initial concerns was that most graduate students have no training or previous knowledge in assessing teaching performance, which might reduce the efficacy of the feedback. In order to mitigate this problem, we approached the Office of Instructional Development, who provided assistance to design and implement a training workshop before the observations took place. The workshop took place at the beginning of the intervention and was conducted by an education professional who is an expert in evaluating teaching skills. It is important to note that both the contents of the training and of the observation and feedback formats were closely related to the basic components of good teaching skills according to the framework developed by Marsh (1983) and traditionally cited in the education literature. According to this framework, there are some factors commonly related to teacher

effectiveness: i) organization/clarity, ii) group interaction, iii) instructor enthusiasm, iv) learning/value, v) breadth of coverage, vi) examinations/grading, vii) assignments/readings, and viii) workload/difficulty. In the context of this study, only those factors i) through v) are relevant, since the rest are not under the TA's influence.

After each observation round, the project manager emailed the two anonymous feedback formats to each TA, who had to acknowledge their receipt. Even thought there is no way to ensure that all TAs read the feedback formats, the qualitative surveys suggests that they indeed read it shortly after the observation took place. Another intital concern was the cash compensation may not be enough incentive to provide thoughtful and careful feedback. Therefore, to promote better quality and more useful feedback each treated TA was assigned two different observers who would observe simultaneously and provide feedback on the same sessions - this increases the likelihood that a TA would receive useful feedback from at least one person, as well as providing some peer pressure for the observing TAs to attend and take the process seriosuly. This is an essential component of the experimental design to mitigate low-quality feedback as a major issue for the intervention. As stated before, the TAs in the treatment group received a cash reward of $100 as compensation for approximately 5 hours of work throughout the quarter (an effective rate slightly lower than the $25 hourly rate TAs generally receive for teaching).

2.3.2 Recruitment Process

With the help of the Department's Graduate Advisor, we contacted all the TAs that were assigned to the treatment group via email and let them know they had been selected to participate in a "teaching training program", for which they would be compensated if they agreed to participate. Of the initial 32 TAs assigned to the treatment group, only 25 agreed to participate in the program, which translates into a take up rate of 78%. The TAs who agreed to participate attended a two hour workshop in which we explained the activities expected

from them and the compensation scheme: they would receive a $100 cash compensation at the end of the quarter if they agreed to: i) observe and provide feedback to two fellow TAs twice during the quarter, using the formats previously described, and, ii) be observed and receive feedback from two fellow TAs twice throughout the quarter. We emphasized that even though this project was supported by the Department, there was no penalty for not participating and no additional reward for doing so. After the explanation, an experienced professional on teaching evaluation gave a participative workshop on the key elements of good teaching and on how to provide constructive criticism to peers.

The last section of the workshop emphasized how to provide constructive feedback: both negative and positive aspects should be brought up and any criticism should be accompanied with a suggestion on how to improve. After the workshop, all the attendees signed a consent form in which they agreed to be a part of the program and in which it was clear that failing any of requirements of the program would result in receiving no compensation at all.

2.4 Descriptive Statistics and Results

This section provides some descriptive statistics of the sample of TAs that were eligible for the intervention, as well as the results of the intervention on the main outcomes of interest: the students' evaluation of the TA, for both the Fall (when the intervention took place) and Winter (one quarter later) quarters, which includes an overall assessment of the TA as well as of some particular aspects of his teaching; and the impact on students' course performance (measured as the deviation of the TAs section grade average from the course average) for the Fall Quarter. Note that for the last outcome of interest, the students' grades, the effect can only be identified in courses with many sections, mainly the introductory and lower division courses (see Table 2.2).

As discussed above, the observations were planned to take place on weeks four and six of the ten week quarter, but the dates were adjusted according to the midterm calendar of

each particular course (to avoid the abstenteeism observed in discussion sections right after the midterm) and the observer's individual availability. For the most part, all first round observation took place in weeks four or five (98%), but only 86% second round observations took place in weeks six or seven, the reminder of which took place in week eight. Usually, the TA evaluations are distributed by TAs in week nine or week ten (last week) of the quarter, so it may be the case that for those TAs who were observed later in the quarter there was not enough time to incorporate the second round of feedback before the evaluations of that same quarter took place.

We present the effects of the intervention on both the Intent to Treat group (ITT), all the TAs that were assigned to treatment and were offered the chance to participate in the program, and on the Treatment group (ToT), those TAs who agreed to participate in the program and actually received treatment. The ITT group is selected at random, and is not subject to the concern that those choosing to participate in treatment might be those who believe they will get a particularly strong benefit from the treatment and might therefore differ in unobserved ways from the (small number of) TAs who chose not to participate. An additional rationale for examing the ITT effects is that all TAs who were offered the chance to participate in the program recieved a signal that the Department was interested in improving the teaching skills of the TAs, because it was explicitly acknowledged that the Department was strongly supporting the intervention; this signal was not given to the control group. Even though we explained that the intervention was not mandatory and that there was no punishment for not participating, the signal may have reminded them of how important teaching is to the Department, incentivizing them to exert more effort when teaching.

2.4.1 Descriptive Statistics and Comparison of Means

As in many graduate economics programs, TAs came from different countries and backgrounds. Figures 2.3 and 2.4 show the country and undergraduate major of the TAs in the sample. As illustrated in Figure 2.3, a large portion of the TAs, roughly 80%, come from outside the US, mainly from China, Korea and Latin America. Not surprisingly, most of them (55%) majored in Economics for their undergraduate degrees, or Economics and Math (15%).[45] Figure 2.5 shows that among this group of PhDs, the most popular field is Macroeconomics (40%), followed by Theory (24%), Labor (14%), Econometrics (11%) and Industrial Organization (11%). Finally, as can be seen in Table 2.1, the average age of the TAs is 27, approximately three fourths are male and, consistent with the information of country of origin, only 22% are native English speakers. In terms of teaching experience, they have taught an average of six quarters in the university, and roughly half of them had taught that same course in the past.

The assignment to the treatment and control groups was done randomly (except for the three TAs of graduate courses), but stratified by course, as shown in Table 2.2. The table shows the distribution both for the ITT group (where TAs were assigned to treatment but did not necessarily agree to participate) and the treatment group across the courses offered. The purpose was to minimize the effect of course specific traits, such as difficulty, teaching skills of the main lecturer, individual student interest on the subject, etc, on the TA evaluations. After randomization, we verifed that the randomization created balance among treatment groups in terms of the observable characteristics of the participants. Table 2.3 shows that this was the case both across the control and ITT groups, and across control and treatment groups. In both cases, the difference in means between groups is not significantly different from zero (Table 2.3) for any of the eight variables. This is a key aspect of the experimental design, given that characteristics such as age and previous teaching experience may affect

[45]TAs are PhD students that are in their second year or above that are making satisfactory progress in the program and are not hired as research assistants or obtain funding from other fellowhships.

the teaching skills of the TAs.

Even though covariates are balanced between the treatment and the control groups, we should analyze whether there was selection into treatment. That is, whether the complier TAs (the ones assigned to the treatment group who agreed to participate in the intervention) are inherently different from the non-complier TAs (those who were assigned to the treatment group but chose not to participate). Table 2.4 illustrates that there was no observed selection into treatment in terms of the covariates, except perhaps for the PhD year. TAs who are more advanced in the program were less willing to participate in the intervention, likely due in part to the time constraints and stress created by the job market process as students near the end of their program. However, it does not appear that selection into treatment is a big issue in this study.[46]

As mentioned above, the main outcomes of interest for the intervention are: (i) the student evaluations of the TA performance that students fill at the end of the quarter (both for the Fall 2012 and Winter 2013 quarters); and (ii) the students' grades, more specifically, the deviation of the TA's section grade average from the average for the whole course (which in most cases comprises many TA sections) for the Fall Quarter. The purpose of taking deviations from the course mean was to reduce the noise caused by differences between courses and focus on differences betweentreated and non-treated TAs. Regarding the student evaluations, we were particularly interested in two questions: i). What is your overall rating of the teaching assistant? (TA evaluation) and ii). The overall value of the sections justified your time and effort (section evaluation). Both questions were answered on a scale from 1 (Very Low) to 9 (Very High).[47] It must be noted that for the empirical analysis we used only the evaluations of sections with more than 10 responses to the evaluations, in order to obtain a valid measure of the TA's performance and teaching skills.

For the Fall Quarter, the average overall TA evaluation by section is 7.8, while average

[46]The comparison of the outcome variables (both the average overall TA evaluation, the average section evaluation and the raw grade for students) between compliers and non-compliers also show that there is no statistical difference between the groups.

[47]Table 2.24 of the appendix shows a copy of the actual evaluations that the students fill out.

section evaluation is slightly lower, 7.6 (in a 1 to 9 scale). Tables 2.5 and 2.6 show the section average and median for these two questions, as well as the deviation of the TA section average of students' grades from the course average of students' grades, for the control, intent to treat and treatment groups.[48] As suggested in Table 2.5, peer feedback seems to have a positive ITT effect on the average evaluation of TAs of around 0.14, which is almost 1/5 of the standard deviation; nonetheless, it is not statistically significant from zero. There does not seem to be any ITT effect on the evaluation of the section or on grades.

The results for Treatment on Treated (ToT) are quite similar, the effect over the average TA evaluation is around 0.13 but still not statistically significant, and no effect on the section evaluation or final grades. These results (Fall 2012 Quarter) are not surprising since the peer review program took place between weeks four and nine, leaving only a few sections for the TAs to internalize the feedback and adjust their teaching practices.

In order to fully estimate the impact of the program, we conducted a follow up of the Teaching Assistants during the Winter Quarter of 2013. The allocation of TAships is made on a quarterly basis depending on the needs of the deparment and the availability of the graduate students. For the Winter Quarter, only four of the graduate students involved in the intervention were not followed. Three of them belonged to the original control group; the fourth was a non-complier from the original treatment group. Tables 2.7 and 2.8 show the descriptive statistics and how covariates are balanced for the subsample of TAs who taught during the Winter Quarter of 2013. The samples are very similar between the Fall and Winter quarters, except that in the Winter a greater number of TAs (25%) reported having been asked to meet with the TA coordinator at some point in their career - a corrective step taken following poor student evaluations. Nonetheless, the covariantes remain balanced for both the IIT and the ToT groups, which reduces any potential concerns regarding sample bias for the Winter results.

For the Winter Quarter, the average overall TA evaluation by section is also 7.8, when

[48]Note that the treatment is at the TA level and each TA has two sections. This will be relevant for the clustering of errors at the TA level.

compared to the Fall Quarter, while the average section evaluation is 8.0, an increase from the previous quarter. Table 2.9 shows the differences in means of these main outcomes between the ITT and the control groups. The results are larger in magnitude than those for the Fall Quarter. They show an effect of 0.37 points, which represent almost one half of a standard deviation from the mean. Table 2.10 shows comparable results for the ToT group. The magnitude is slightly higher, 0.39, which is also close to one half of a standard deviation. Importantly, in both cases the effect is statistically significant, suggesting that the intervention was succesful. Once the TAs had enough time to incorporate the suggestions made by their peers and to adopt the lessons from their own observations, they improved their performance considerably. It would be interesting to evaluate if the results are long lasting, but at least they suggest that a peer review program might help to boost the performance of TAs.

2.4.2 Regression Analysis

The effects of the peer review intervention can be assessed by comparing outcomes across groups in a simple Ordinary Least Squares regression model. For each TA-section outcome we estimate the following specification:

$$y_{i,a} = \alpha + \theta Treat_a + \beta X_a + \delta_i + \varepsilon_{i,a} \tag{2.1}$$

where, $y_{i,a}$ is the outcome of interest for section i of TA a, $Treat$ is the intent to treat (ITT) or the treatment (ToT) indicator at the TA level, X_a are a set of controls at the TA level[49], and δ_i are course-specific fixed effects used in some specifications, and $\varepsilon_{i,a}$ are robust errors, clustered at the TA level. The coefficient of interest is θ, which should be an unbiased indicator of the causal effect of the intervention, because the unobservable characteristics of

[49]The covariates include: age, male, English native speaker, PhD year, Master degree, number of quarters that they have taught, whether they have taught the course before and whether they have met with the TA coordinator.

the TAs should be distributed randomly across the groups due to the experimental design.

Table 2.11 shows the effects of the ITT on the average TA evaluation, while Table 2.12 shows the effect of ToT both for the Fall Quarter. The estimated effect is positive but statistically insignificant. As discussed before, these results (Fall 2012 Quarter) are not surprising since the peer feedback program took place between weeks four and nine, leaving only a few sections for the TAs to internalize the feedback and adjust their teaching practices.

Tables 2.13 and 2.14 repeat this exercise for the students' grades for the Fall Quarter, more specifically the deviation of the TA average grade from the course average grade. The empirical analysis suggests that there was no effect of peer feedback on the students' performance; both ITT and ToT coefficients are very close to zero, and again not statistically significant. It must be noted that the effect of any treatment at the TA level will depend not only on how early they were able to adjust their teaching behavior in response to feedback, but also on how important the TA is for the students' performance (what is the weight of this in the production function), relative to other factors such as the students' effort or ability at the section level.[50]

Another difficulty in the interpretation of these results is that students may switch TA sessions within a same course, depending on the time of the discussion section or if they don't feel the TA met their expectations. There is no way to account for this problem in the sense that the evaluations are anonymous and there's no way to track the switching. This issue could be an alternative explanation of why we do not observe any statistical effect of the treatment on the students' performance.

These results for the Fall Quarter suggest that the peer feedback did not produce large effects within the quarter, but (as described above) it is possible that the timing of the intervention did not allow the TAs to fully incorporate feedback early enough to observe an impact in TA evaluations or student grades in the same quarter. If the intervention did in

[50]Similarly, the teaching skills of the TA are only relevant for those students who attend to TA section, which is known to be way below the number of students enrolled (all of which are used in the calculation of the average grade by section).

fact provide TAs with valuable and actionable feedback, we would expect to see a larger impact on outcomes in the following quarter, when the treated TAs could incorporate the Fall Quarter feedback to inform their teaching throughout the whole Winter Quarter.

Table 2.15 contains the regression analysis of the ITT effect using the Winter TA evaluations. Column 1 presents the mean difference between samples clustering errors at the TA level (most TAs taught two sections): the effect size is 0.366, equivalent to one half of a standard deviation, and significant at a 10% level. Column 2 shows that the size of the ITT effect does not change after introducing the first set of covariates, as expected if the covariates are balanced (as shown before). The next column adds the lagged TA evaluation reducing slightly the size of the coefficient to 0.337, but it remains significant[51]. The results in column 3 also show that there is consistency in the TA evaluations, those who obtained higher grades in the Fall Quarter also obtained higher grades in the Winter Quarter.

Column 4 controls for the fields of the TAs, the coeffient is significant at a 11% level, and the magnitude does not change significantly showing robustness to the inclusion of additional controls. Column 5 includes course dummies in the specification, which restricts the identification of the effect to those course with various sections. According to this result, the effect of peer feedback was almost one standard deviation and statistically significant at the 1% level. The increase in the effect of the intervention may be explained by the fact that most of the courses that have many sections are introductory courses, for which good communication skills are particularly relevant, an aspect of teaching which was largely affected by the intervenion, as will be discussed in detail in the next section.

The last column adds dummies for nationality, with the coefficient on the ITT increasing to 0.46. While we would not normally expect the results of an RCT to change with the addition of covariates, some nationalities are represented in only one of the groups. The slight increase in the coefficient could indicate that the intervention is particularly effective

[51]Note that this specification may control away some of the treatment effect - if the insignificant positive effect seen in the Fall represents some small improvement thanks to the intervention, by controlling for the Fall evaluation scores we restrict ourselves to examining the incremental improvement in outcomes between the Fall and Winter quarters, rather than the full impact of the intervention.

for one or more of the larger nationality groups, e.g. China (35%) or the US (20%), who make up a larger proportion of the effective sample in this specification.

Table 2.16 shows the results for the ToT. The coefficients of the six columns mimic the results of the previous table. The effect under the first three specifications remains unchanged around 0.36-0.39 and is statistically different from zero. Overall, the results from both tables suggests that the peer feedback was sucessful in terms of improving the TAs teaching skills and the randomization was successfull. In addition, they show that even after controlling for those variables for which it was not possible to randomize, the results are large and, in most cases, statistically significant. Finally, it shows that it was correct to expect an effect on the ITT group, given that those TAs who did not accepted treatment still perceived the signal that the Department was highly interested in the teaching skills of the graduate students.

Tables 2.17 and 2.18 show the results for the log of the TA evaluations. The results can be directly interpreted as percentage changes when the covariates are dummies. Both the ITT and the ToT would suggest an increase of between 4.69% and 5.36% in the evaluations caused by the peer review program (columns 1 to 4 in both tables). If we control for the course differences by adding course fixed effects, the change would increase to almost 10%. While if we control by the differences in nationality (which may mask differences in teaching styles), the change would be around 6%.

2.4.3 Decomposition of TA evaluations

The previous section discussed the results on the overall TA's evaluations and students' grades. However, the evaluation formats have six areas that are assessed by the students and that refer to more concrete skills: (i) The first category refers to the knowledge of the TA in the course taught; (ii) The second one evaluates the concern of the TA regarding the students understanding of the material; (iv) The third category focuses on the preparation and organization of the course; (iv) The fourth refers to the scope of the TA session relative

to the course, more specifically whether the TA helped the students to improve their understanding on the material and expand on the topics covered in class; (v) The fifth area looks at the interaction between the TAs and the students outside the classroom; (vi) Finally, the sixth component evaluates the communication skills of the TA referring to the ability to transmit ideas. As before, all questions are framed in a scale from 1 (Very Low) to 9 (Very High).

Given that the overall TA evaluation is an assessment of all these categories, we believed that the intervention should have different impact across categories. In particular, we expected improvement in pedagogical areas which correspond to categories two, three, five and six, related to concern, organization, interaction and communication of the TA. It is unlikely that the intervention can modify the knowledge of the TA (although it could be correlated with organization and preparation of the course) or the scope of the sessions (because most TAs follow instructions from professors about what topics to cover).

We analyze if there was improvement in the areas in which we originally expected using the Winter 2013 TA evaluations, because it was for this quarter that students perceived a change for the better in the TAs.[52] Table 2.19 shows the summary statistics for each of the six categories (knowledge, concern, organization, scope, interaction and communication). Overall, students seem to consider that the TAs of the Economics department are knowledgable of the topics teaching (8.05 average grade) and concerned about the students' learning (7.84 average grade). However, the TAs seem to lag behind precisely in the areas at which the intervention is aiming at -organization, interaction and communication-, ranging from 7.37 to 7.75 average grades.

Tables 2.20 and 2.21 explore the differences in means between groups for the ITT and the ToT, respectively, relative to the control group. The results on both tables are almost identical in magnitude and statistical significance. As expected, knowledge does not respond to the intervention, which is a skill that would require a different type of intervention to

[52]We also analyzed the individual components of the TA evaluations for the Fall Quarter, but as with the overall TA evaluations, we found no significant effect of the intervention in any individual category.

affect. It might seem surprising that there is an effect on scope of around 3/7 of a standard deviation if the professors dictate the pace of activities. However, the TAs might have learnt from their peers how to approach the content indicated by the professors in a better way.

Concern, organization, and interaction increase by 0.3 points each which represent also a 3/7 of a standard deviation of each component. The largest effect is seen on communication, which increased by 0.68, which is slightly more than 2/3 of a standard deviation of the mean. This result suggests that after the intervention TAs were more concerned with how they expressed themselves and how to convey the material clearly.

Tables 2.22 and 2.22 show the results of the regression analysis for each sub-score, following the exact same specification as before, for the ITT and ToT groups respectively. For the former, 2.22 shows that with the exception of knowledge and organization, the magnitude of the coefficients barely change once the controls are added. More encouragingly, we find that for concern, scope, interaction and communication, the results are robust to the inclusion of covariates. The effect of the treatment is signifcant at the 11% level for the concern and interaction sub-score regressions, significant at the 10% level for the scope sub-score, and at the 5% level for communication. As can be seen in Table 2.22, the results for the ToT group are very similar except for the fact that the effects on concern, organization and interaction are significant at 11%.

Among the covariates, the dummy for native english speakers is not surprisingly positive and highly significant for communication skills. Also for that category, having met with the TA coordinatot has a negative and important effect on the average grade. Overall, the intervention seems to have an important effect on the communication skills of the TAs, a smaller effect on organization, scope, concern and interaction, and no effect on knowledge.

2.5 Complementary Analyses

This section presents two complementary analyses relevant for scaling up the intervention. The first part presents the main findings of the qualitative surveys that the treated TAs completed shortly after the intervention. The responses provide valuable information on which aspects of the intervention worked reasonably well and which aspects could be improved in the future. The second part discusses alternative experimental designs that could be implemented in order to distinguish the mechanisms at work (incorporation of feedback, incidental learning from observation of others, mitigation of potential moral hazard from TAs, among others) or to perform comparisons across different types of interventions.

2.5.1 Qualitative Survey

We conducted a qualitative survey at the end of the peer feedback project in order to assess some of the key aspects of the intervention. In particular, we had two sets of questions: (i) The first set tried to elicit information regarding the experience of the TAs while they were being observed; (ii) The second set of questions tried to gather information on the TAs as observers and evaluators.

For the first part, our objective was to qualitatively assess whether the TAs were aware (self-conscious) of the presence of observers and whether they modified their behavior while teaching or preparing for class, and finally, whether they found the feedback received useful. Overall, the project ran smoothly, all TAs claim to have received their feedback shortly after being observed and all TAs claim to have read it carefully. Most of the TAs agreed that the feedback was useful; however, they believe that it mostly contained positive reinforcement or that it pointed out problems of which the TAs were already aware, instead of pointing specific actions they could take in order to improve their teaching. Regarding the observation, the responses are mixed, some TAs did not notice the observers, while others felt somewhat uncomfortable while they were being observed.

Despite self-awareness, there are mixed responses regarding their attitude towards the preparation of the class: not all TAs modified their behavior knowing that a fellow TA could be present. Finally, we also included a question regarding what type of observer would the TAs rather have (peers or experts) - our concern was that since all observers were other graduate students of the same department, the TAs would feel more nervous or uncomfortable compared to having a stranger observe them. Even though the responses to this question are also mixed, most of them still prefer having somebody within the department evaluating their classes.

The objective of the second part of the survey was to evaluate the perception of the TAs regarding their qualifications as evaluators. The first question, which referred to the initial training workshop, showed that most of the TAs do no believe it helped them to improve their abilities as evaluators. This is an important point to consider for future interventions: the training workshop should be carefuly tailored to the needs of each department and the contents should be revised. The next question tried to assess whether they felt capable of performing the task. Most of them agreed that they could evaluate the teaching skills of their colleagues and that the feedback formats helped them to transmit their thoughts and comments. Regarding the perceived change in teaching "skills" between the two observations, most TAs did not feel that there was any improvement, which is consistent with the results for the Fall Quarter (and unsurprising given that the two observatinos were only 2-3 weeks apart). We also wanted to know if they would feel more comfortable evaluating a stranger from a different department, but most of the answers expressed a preference for observing TAs in the same Department, possibly because they know the content of the courses better.

The last question elicited their perception of the project overall. In particular, we were interested in knowing whether they believed that it was useful and had potential for a large scale implementation. 80% of them answered that they liked the project and that they believed that it had potential. It seems that the TAs took their role seriously during this experiment, that they considered themselves fit to observe and assess the teaching skills of

their peers, and that they preferred both to observe and to be observed by peers of the same department.

2.5.2 Discussion of Potential Future Interventions

Considering the non-trivial effect of the peer feedback intervention, it is interesting to ponder slightly different experimental designs that can both disentangle the mechanisms of this past intervention and compare this to other options to improve teaching skills. In terms of distentangling the different mechanims of the peer feedback observation, it is worth distinguishing the effects of feedback (and learning by observation) from a mere Hawthorne effect in which TAs modify their behavior because there is someone different from their regular students observing him teach the class. This is plausible, particularly because in the qualitative survey the TAs mentioned that they did not purposively modify the way they were preparing to teach their sections.

Similarly, at various points through out this intervention, we thought about the option of an alternative intervention: having a professional in teaching to observe the TAs and provide them feedback. This option solves some of the concerns of having peers observing each other, mainly the quality of the feedback and the possibility of having subjective opinions interfering with the provision or reception of the feedback. Nonetheless, when compared to the peer feedback intervention, it would mean giving up the opportunity of TAs learning teaching practices by observing their peers, it could make observers more uncomfortable since the observer would be an outsider and, lastly, the observer is likely to be less familiar with the contents of the courses being taught.

Along these lines and as part of future research, it would be interesting to design a larger intervention with three different treatments, each of them involving a number of TAs similar to the the number of treated TAs in this study. The first treatment would involve pure observation (no feedback at all), in order to capture any possible Hawthorne effects for

the observed TAs, and any learning though observation effects for the observer TAs. The second treatment would replicate the peer feedback intervention of this study, TAs would observe each other and provide feedback based on some guidelines. The third treatment would consist of having an external expert to observe the TAs and provide them with some feedback. The purpose of the latter would be to compare which alternative works better, peer review or an external observer. The main idea would be to carry out a cost-effectiveness comparison of treatments two and three.

2.6 Conclusions

Peer feedback is an attractive alternative to improve teaching practices, but even though some qualitative studies of the educational literature have assessed its effectiveness, to the best of our knowledge, there is no study with a solid quantitative approach on the subject. Our study is a first step to fill this gap in the literature by using a randomized intervention in the Department of Economics of a large public university to establish a causal relationship between peer feedback and teaching skills of TAs.

The results from the study suggest that peer feedback at the TA level has a positive but not significant effect on the overall TA student evaluations during the quarter that the experiment took place. The RCT, however, had a non-trivial and significant effect in the following quarter (Winter 2013): it increased the TAs evaluations by one half of a standard deviation. The results are robust to the addition of covariates, showing that the randomization was sucessful in terms of balancing the observable characteristics between the control and both the ITT and the ToT groups. In terms of the specific areas of improvement, the results show that the intervention had an important effect on the communication skills, and a smaller, less significant effect on organization, scope, concern and interaction with the students. As expected, the intervention had no effect on how knowledgeable about the material covered in the section was the TA.

The analysis of the students' performance indicates that the intervention had no effect on the students' grade in the quarter of the intervention. This could be partially explained by the fact that during that quarter, the TAs did not have enough time to implement the suggestions made by their observers, and thus students could not benefit from the intervention. Also, it is a common practice among students to switch TA sections throughout the quarter and there is no way to track whether students changed sections, which introduces some measurement error.

The qualitative survey provided valuable information on the components of the intervention that worked well and those which must be improved for future interventions. Regarding the first aspect, it seems that having TAs from the same department to observe each other was an effective choice, not only because they are familiar with the content of the classes, but also because the TAs expressed a preference for both observing and being observed by a peer from their Department. Also, it appears that the observation and feedback formats were useful tools to guide observers, and that the TAs took the exercise seriously and found it valuable. Nonetheless, there are aspects of the program that require some adjusment. In particular, TAs pointed out that the training workshop was not very helpful for their tasks later on in the intervention. Similarly, they mentioned that a drawback from the feedback they received was that it did not contain enough specific actions that they could take to improve their teaching. As emphasized before, in order for peer feedback to be a positive activity, constructive criticism is crucial. One potential solution to this issue is to include in the feedback format a brief reminder of how important constructive cricitism and a list of examples of proactive actions to help improve the different aspects of teaching.

2.7 Main Figures and Tables

Figure 2.1: Observation Format

TA being observed: _____

Time and place:_____

Observer:_____

OBSERVATION FORMAT

INSTRUCTIONS:

Read the format before attending the session, so that you know what to look for. Make sure to know what topics and concepts are going to be covered during the session beforehand.

		Not aplicable	Strongly disagree		Disagree		Neither agree nor disagree		Agree		Strongly agree
	Organization/Clarity										
1	The aims, objectives and structure of the session were clear.	N/A	1	2	3	4	5	6	7	8	9
2	The topic and concepts covered were prepared beforehand.	N/A	1	2	3	4	5	6	7	8	9
3	The TAs speech was easy to understand.	N/A	1	2	3	4	5	6	7	8	9
4	The board or other teaching aids were used appropriately.	N/A	1	2	3	4	5	6	7	8	9
5	The TA managed properly the time of the session	N/A	1	2	3	4	5	6	7	8	9
	Specific comments on this factor:										
	Group Interaction										
6	The TA effectively managed the group interaction.	N/A	1	2	3	4	5	6	7	8	9
7	The TA encouraged students to actively participate in the session.	N/A	1	2	3	4	5	6	7	8	9
8	Students were engaged in the explanation and discussion of the section.	N/A	1	2	3	4	5	6	7	8	9
	Specific comments on this factor:										
	Instructor Enthusiasm										
9	The TA was enthusiastic about and interested in the topic.	N/A	1	2	3	4	5	6	7	8	9
10	The TA developed good rapport with the students and responded to their needs.	N/A	1	2	3	4	5	6	7	8	9
	Specific comments on this factor:										

	Learning/Value										
11	The TA explained things well and the examples used helped the students to understand the topic.	N/A	1	2	3	4	5	6	7	8	9
12	Ideas were transmitted clearly and in a way students would understand them.	N/A	1	2	3	4	5	6	7	8	9
13	The TA's feedback/answers to questions helped students to learn.	N/A	1	2	3	4	5	6	7	8	9
	Specific comments on this factor:										
	Breadth of Coverage										
14	The session was well integrated with the rest of the course (following the syllabus).	N/A	1	2	3	4	5	6	7	8	9
15	The concepts discussed were framed into the broad scope of the course.	N/A	1	2	3	4	5	6	7	8	9
16	The TA linked the topics in a coherent manner.	N/A	1	2	3	4	5	6	7	8	9
	Specific comments on this factor:										
	Comments										
17	Please list the three best things about the TA.										
18	Please list three suggestions for improving the session.										
19	Comments on the lesson plan e.g. activities, structure and timing.										

Note: Based on the Danielson framework (Danielson, 2011) of assessing teaching skills and also, based on the students' evaluations used in the large public university. It is also consistent with the framework developed by Marsh (1983) on what set of factors are important for good teaching.

Figure 2.2: Feedback Format

TA being observed: _____

Time and place:_____

Observer:_____

FEEDBACK FORMAT

INSTRUCTIONS:

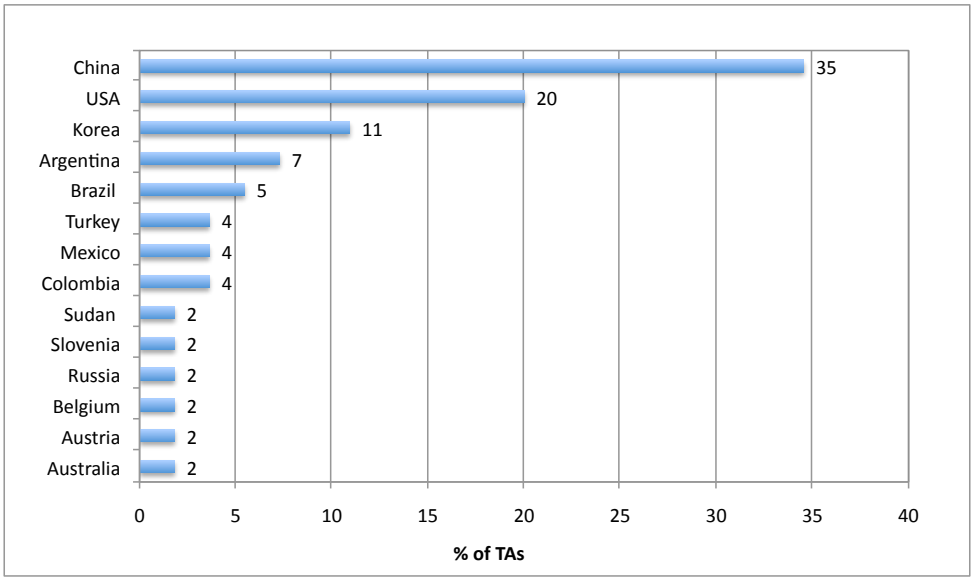
Please record the main comments and feedback points you would like to provide to your peer TA. Please be very specific about the actions she can take in each field to improve her performance.

1	*Organization/Clarity*
	Specific comments on this factor:
	Specific actions towards improvement:
2	*Group Interaction*
	Specific comments on this factor:
	Specific actions towards improvement:
3	*Instructor Enthusiasm*
	Specific comments on this factor:
	Specific actions towards improvement:
4	*Learning/Value*
	Specific comments on this factor:
	Specific actions towards improvement:

5	*Breadth of Coverage*
	Specific comments on this factor:
	Specific actions towards improvement:
6	Comments
	Other comments

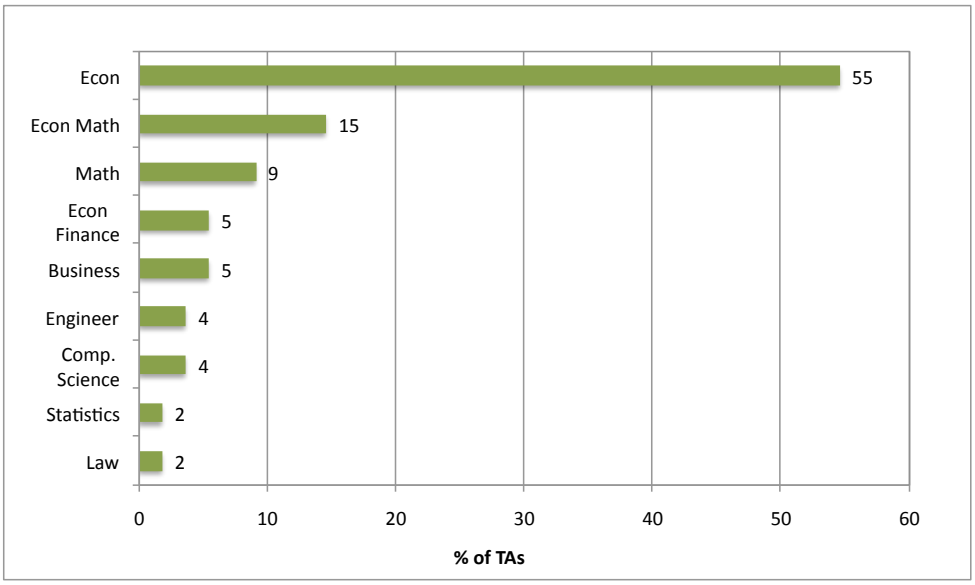
Note: Based on the Danielson framework (Danielson, 2011) of assessing teaching skills and also, based on the students' evaluations used in the large public university. It is also consistent with the framework developed by Marsh (1983) on what set of factors are important for good teaching.

Figure 2.3: TAs Nationality



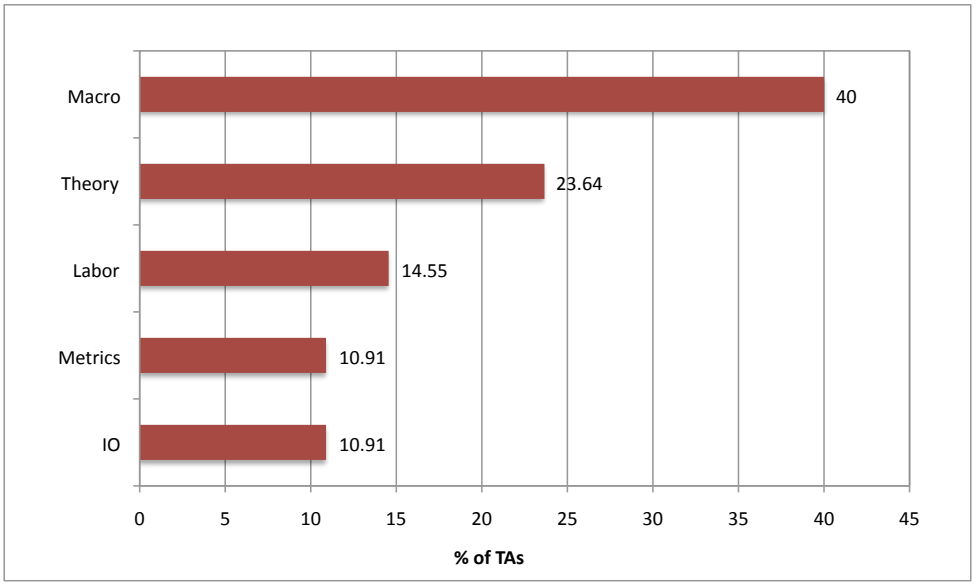
Note: This figure shows the country of origin of all TAs, as reported by them in the post-intervention survey.

Figure 2.4: TAs Undergraduate Major



Note: This figure shows the TAs undergraduate major, as reported by them in the post-intervention survey.

Figure 2.5: TAs PhD Main Field



Note: This figure shows the TAs main Field of specilization in the PhD, as reported by them in the post-intervention survey.

Table 2.1: Descriptive Statistics: Covariates

Variable	Mean	Median	St. Dev.	Min	Max
Age	27.02	27.00	2.32	23	32
I(male)	0.75	1.00	0.44	0	1
I(English native)	0.22	0.00	0.42	0	1
PhD year	3.15	3.00	0.89	2	5
I(MA)	0.40	0.00	0.49	0	1
Quarters taught	6.40	6.00	4.65	0	18
I(taught this course before)	0.45	0.00	0.50	0	1
I(coordinator)	0.16	0.00	0.37	0	1

Note: This table presents the descriptive statistics of the obsevable characteristics of the TAs, as reported by them in the post-intervention survey.

Table 2.2: Randomiztion by Course (Number of TAs)

	Course	Control	ITT	Treatment
Principles of Economics	Econ 1	2	3	2
Principles of Economics	Econ 2	3	3	2
Microeconomic Theory	Econ 11	5	7	5
Statistics for Economists	Econ 41	2	5	5
Microeconomic Theory	Econ 101	2	3	3
Macroeconomic Theory	Econ 102	2	2	2
Introduction to Econometrics	Econ 103L	2	3	3
Economics of Technology and E-commerce	Econ 106TL	1	2	0
Investments	Econ 106VL	1	2	1
Public Economics	Econ 130L	0	1	1
Economic Growth	Econ 164L	0	1	1
Microeconomic Theory (Grad)	Econ 201A	1	0	0
Macroeconomic Theory (Grad)	Econ 202A	1	0	0
Econometrics (Graduate)	Econ203A	1	0	0
	Total	23	32	25

Note: The table displays the courses offered by the Economics Department that has one or more TAs. As noted in the text, most of the introductory courses offered have various TAs, and we were able to stratify the randomization accordingly. Nonethless, this was not the case for the most advanced courses.

Table 2.3: Balancing of Covariates

Variable	Control	ITT	p-value (equal means)	Control	Treatment	p-value (equal means)
Age	27.04	27.00	0.946	27.04	26.84	0.764
I(male)	0.70	0.78	0.481	0.70	0.76	0.625
I(english native)	0.13	0.28	0.188	0.13	0.28	0.211
PhD year	3.13	3.16	0.917	3.13	3.04	0.716
I(MA)	0.52	0.31	0.123	0.52	0.32	0.163
Quarters taught	6.39	6.41	0.991	6.39	6.04	0.784
I(taught this course before)	0.43	0.47	0.807	0.43	0.48	0.760
I(coordinator)	0.17	0.16	0.865	0.17	0.16	0.900
N	23	32		23	25	

Note: This tables depicts the summary statistics of the observable characteristics of the TA participating in the intervention, which include age, indicator variable for male, indicator variable for being an English Native, the PhD year the TA is currently attending to, an indicator variable for obtaining a Masters Degreeb before entering the PhD, number of quarters as a TA in the current university, an indicator variable of whether the TA has taught the course before, and finally and indicator variable of whether the TA has been called by the TA coordinator of the Department due to obtaining very low scores in previous students' evaluations.

Table 2.4: Selection into Treatment: Takers vs. Non-takers

Variable	Takers	Non-takers	p-value (equal means)
Age	26.84	27.57	0.461
I(male)	0.76	0.86	0.597
I(English native)	0.28	0.29	0.977
PhD year	3.04	3.57	0.106
I(MA)	0.32	0.29	0.868
Quarters taught	6.04	7.71	0.344
I(taught this course before)	0.48	0.43	0.817
I(coordinator)	0.16	0.14	0.916
N	25	7	

Note: This table shows the difference of the means of the observable characteristics described above between taker (TAs that decided to participate in the intervention) and non-takers (TAs that refused to participated but were offered to participate in a program intended to improve teaching abilities of the teaching assistants of the Economics Department). The third column shows the p-value of the test of equality of mean between two groups.

Table 2.5: Difference of Means: ITT

Variable	Control	ITT	Difference	p-value (equal means)
Average evaluation of TA	7.76	7.90	0.14	0.338
Median evaluation of TA	8.15	8.21	0.06	0.689
Average evaluation of discussion section	7.63	7.63	0.00	0.994
Median evaluation of discussion section	8.15	8.04	-0.10	0.578
Grade (dev. from course mean)	0.004	0.003	-0.002	0.769
N	41	58		

Note: This table shows the difference of the averages of the main outcome variables of interest between the control group and the ITT group (TAs who were offered to participate in the program). The outcomes of interest are mainly the TAs overall evaluation, the section overall evaluation (value of the section to the students), and the deviation of the section average grade from the course average (reacll that most of the courses had many sections). The last column shows the p-value of the test of equality of mean between two groups.

Table 2.6: Difference of Means: ToT

Variable	Control	Treatment	Difference	p-value (equal means)
Average evaluation of TA	7.76	7.89	0.13	0.422
Median evaluation of TA	8.15	8.18	0.04	0.828
Average evaluation of discussion section	7.63	7.68	0.05	0.786
Median evaluation of discussion section	8.15	8.03	-0.11	0.575
Grade (dev. from course mean)	0.004	0.001	-0.004	0.565
N	41	44		

Note: This table shows the difference of the averages of the main outcome variables of interest between the control group and the treatment group (TAs who actually participated in the program). The outcomes of interest are mainly the TAs overall evaluation, the section overall evaluation (value of the section to the students), and the deviation of the section average grade from the course average (reacll that most of the courses had many sections). The last column shows the p-value of the test of equality of mean between two groups.

Table 2.7: Descriptive Statistics Winter 2013: Covariates

Variable	Mean	Median	St. Dev.	Min	Max
Age	27.00	27.00	2.31	23	32
I(male)	0.76	1.00	0.43	0	1
I(English native)	0.22	0.00	0.42	0	1
PhD year	3.06	3.00	0.83	2	5
I(MA)	0.41	0.00	0.50	0	1
Quarters taught	6.12	6.00	4.34	0	17
I(taught this course before)	0.55	1.00	0.50	0	1
I(coordinator)	0.25	0.00	0.52	0	2

Note: This table presents the descriptive statistics of the obsevable characteristics of the TAs, as reported by them in the post-intervention survey.

Table 2.8: Balancing Covariates Winter 2013

Variable	Control	ITT	p-value (equal means)	Control	Treatment	p-value (equal means)
Age	27	27	1.000	27	26.84	0.817
I(male)	0.75	0.77	0.846	0.75	0.76	0.940
I(english native)	0.15	0.26	0.369	0.15	0.28	0.308
PhD year	2.95	3.13	0.460	2.95	3.04	0.701
I(MA)	0.55	0.32	0.111	0.55	0.32	0.126
Quarters taught	5.40	6.58	0.348	5.4	6.04	0.594
I(taught this course before)	0.45	0.61	0.263	0.45	0.56	0.475
I(coordinator)	0.25	0.26	0.958	0.25	0.28	0.855
N	20	31		20	25	

Note: This tables depicts the summary statistics of the observable characteristics of the TAs participating in the intervention, which include age, indicator variable for male, indicator variable for being an English Native, the PhD year the TA is currently attending to, an indicator variable for obtaining a Masters Degree before entering the PhD, number of quarters as a TA in the current university, an indicator variable of whether the TA has taught the course before, and finally and indicator variable of whether the TA has been called by the TA coordinator of the Department due to obtaining very low scores in previous students' evaluations.

Table 2.9: Difference of Means Winter 2013: ITT

Variable	Control	ITT	Difference	p-value (equal means)
Average evaluation of TA	7.58	7.95	0.37	0.029**
Median evaluation of TA	7.84	8.24	0.39	0.0510*
Average evaluation of discussion section	7.35	7.82	0.47	0.011**
Median evaluation of discussion section	7.61	8.15	0.54	0.014**
N	35	59		

Note: This table shows the difference of the averages of the main outcome variables of interest between the control group and the ITT group (TAs who were offered to participate in the program). The outcomes of interest are mainly the TAs overall evaluation, the section overall evaluation (value of the section to the students), and the deviation of the section average grade from the course average (recall that most of the courses had many sections). The last column shows the p-value of the test of equality of mean between two groups. Also, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.10: Difference of Means Winter 2013: ToT

Variable	Control	Treatment	Difference	p-value (equal means)
Average evaluation of TA	7.58	7.94	0.36	0.042*
Median evaluation of TA	7.84	8.21	0.37	0.096*
Average evaluation of discussion section	7.35	7.83	0.47	0.017**
Median evaluation of discussion section	7.61	8.10	0.49	0.040*
N	35	48		

Note: This table shows the difference of the averages of the main outcome variables of interest between the control group and the treatment group (TAs who actually participated in the program). The outcomes of interest are mainly the TAs overall evaluation, the section overall evaluation (value of the section to the students), and the deviation of the section average grade from the course average (recall that most of the courses had many sections). The last column shows the p-value of the test of equality of mean between two groups. Also, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.11: Regression Analysis ITT: TA Evaluation

	Independent Variable: Average Evaluation of TA				
	(1)	(2)	(3)	(4)	(5)
Intent to Treat	0.1441	0.0458	0.1232	-0.0679	-0.1769
	[0.182]	[0.182]	[0.169]	[0.172]	[0.221]
Age		0.0849*	0.1210**	0.0805	0.1164*
		[0.045]	[0.046]	[0.051]	[0.067]
I(male)		0.1350	0.5783**	0.0616	0.0937
		[0.192]	[0.240]	[0.187]	[0.195]
I(English native)		0.1685	0.0186	0.3338	0.0028
		[0.196]	[0.185]	[0.247]	[0.372]
PhD year		-0.3096	-0.4212**	-0.4776*	-0.4520*
		[0.205]	[0.205]	[0.268]	[0.228]
I(MA)		-0.2858	-0.3233	-0.3396	-0.5779**
		[0.228]	[0.203]	[0.255]	[0.280]
Quarters taught		0.0630	0.0601	0.0935*	0.0523
		[0.045]	[0.047]	[0.047]	[0.053]
I(taught this course before)		0.2064	0.1842	-0.0362	0.3087
		[0.242]	[0.224]	[0.240]	[0.263]
I(coordinator)		-0.2486	-0.1947	-0.2727	-0.0873
		[0.236]	[0.223]	[0.291]	[0.253]
Labor			0.4655*		
			[0.256]		
Macro			-0.5697*		
			[0.322]		
Metrics			-0.2583		
			[0.370]		
Theory			-0.1616		
			[0.300]		
Constant	7.759***	5.999***	5.303***	6.315***	5.485***
	[0.138]	[1.088]	[1.116]	[1.177]	[1.847]
Course dummies	No	No	No	Yes	No
Nationality dummies	No	No	No	No	Yes
Observations	99	99	99	99	99
R-squared	0.009	0.136	0.236	0.324	0.309

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is the TA's average student overall evaluation by section (in general, TAs are responsible for teaching two sections) and the variable of interest is ITT, an indicator variable of the intent to treat. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.12: Regression Analysis ToT: TA Evaluation

	Independent Variable: Average Evaluation of TA				
	(1)	(2)	(3)	(4)	(5)
Treatment	0.1303	-0.0129	-0.0126	-0.2364	-0.2151
	[0.196]	[0.205]	[0.203]	[0.186]	[0.226]
Age		0.0838*	0.1116**	0.0524	0.1278*
		[0.047]	[0.050]	[0.042]	[0.069]
I(male)		0.0693	0.2912	-0.0016	0.0207
		[0.210]	[0.296]	[0.169]	[0.227]
I(English native)		0.2193	0.0728	0.7958***	0.5604
		[0.206]	[0.211]	[0.253]	[0.407]
PhD year		-0.5285	-0.6278	-1.0586***	-0.6547*
		[0.351]	[0.425]	[0.307]	[0.380]
I(MA)		-0.3958	-0.4557*	-0.3673	-0.6909**
		[0.251]	[0.232]	[0.247]	[0.337]
Quarters taught		0.0938	0.0845	0.2231***	0.0952
		[0.069]	[0.081]	[0.061]	[0.077]
I(taught this course before)		0.2234	0.1588	-0.1386	0.2712
		[0.254]	[0.233]	[0.207]	[0.271]
I(coordinator)		-0.2740	-0.2539	-0.3729	-0.0948
		[0.243]	[0.227]	[0.278]	[0.255]
Labor			0.4806		
			[0.345]		
Macro			-0.3174		
			[0.428]		
Metrics			-0.5252		
			[0.480]		
Theory			-0.0371		
			[0.410]		
Constant	7.75***	6.60***	6.24***	7.57***	5.69***
	[0.138]	[1.202]	[1.345]	[1.006]	[1.931]
Course dummies	No	No	No	Yes	No
Nationality dummies	No	No	No	No	Yes
Observations	85	85	85	85	85
R-squared	0.008	0.146	0.235	0.449	0.306

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is the TA's average student overall evaluation by section (in general, TAs are responsible for teaching two sections) and the independent variable of interest is Treatment, an indicator variable of receiving treatment. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.13: Regression Analysis ITT: Grades (Dev. from Course Mean)

	Independent Variable: Grade (dev. from course mean)				
	(1)	(2)	(3)	(4)	(5)
Intent to Treat	-0.0018	-0.0064	-0.0045	-0.0054	-0.0072
	[0.005]	[0.005]	[0.005]	[0.005]	[0.006]
Age		0.0033***	0.0028**	0.0025	0.0030*
		[0.001]	[0.001]	[0.002]	[0.002]
I(male)		0.0021	0.0077	0.0021	0.0014
		[0.005]	[0.006]	[0.005]	[0.005]
I(English native)		-0.0058	-0.0045	-0.0061	0.0114
		[0.005]	[0.005]	[0.006]	[0.015]
PhD year		-0.0006	0.0011	0.0002	-0.0038
		[0.005]	[0.005]	[0.007]	[0.006]
I(MA)		-0.0200***	-0.0165**	-0.0193***	-0.0209***
		[0.007]	[0.007]	[0.007]	[0.007]
Quarters taught		-0.0006	-0.0006	-0.0004	-0.0010
		[0.001]	[0.001]	[0.001]	[0.001]
I(taught this course before)		-0.0035	-0.0006	-0.0054	-0.0013
		[0.006]	[0.007]	[0.007]	[0.006]
I(coordinator)		0.0116	0.0114	0.0171**	0.0162*
		[0.007]	[0.008]	[0.008]	[0.009]
Labor			-0.0054		
			[0.008]		
Macro			-0.0065		
			[0.008]		
Metrics			0.0177**		
			[0.008]		
Theory			-0.0060		
			[0.008]		
Constant	0.0044	-0.0676**	-0.0645**	-0.0502	-0.0591
	[0.004]	[0.027]	[0.031]	[0.035]	[0.044]
Course dummies	No	No	No	Yes	No
Nationality dummies	No	No	No	No	Yes
Observations	92	92	92	92	92
R-squared	0.001	0.088	0.127	0.128	0.197

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is the deviation of the students' grade by TA section from the overall course average, and the independent variable of interest is ITT, an indicator variable of intent to treat. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.14: Regression Analysis ToT: Grades (Dev. from Course Mean)

	Independent Variable: Grade (dev. from course mean)				
	(1)	(2)	(3)	(4)	(5)
Treatment	-0.0037	-0.0077	-0.0068	-0.0077	-0.0392
	[0.006]	[0.005]	[0.006]	[0.007]	[0.024]
Age		0.0030**	0.0027*	0.0020	-0.0066
		[0.001]	[0.001]	[0.002]	[0.005]
I(male)		0.0019	0.0050	0.0028	-0.0165
		[0.005]	[0.008]	[0.005]	[0.027]
I(English native)		-0.0039	-0.0005	0.0030	0.0807***
		[0.006]	[0.007]	[0.007]	[0.028]
PhD year		0.0035	0.0085	-0.0011	-0.0218
		[0.008]	[0.009]	[0.011]	[0.048]
I(MA)		-0.0186**	-0.0163*	-0.0177**	0.0053
		[0.008]	[0.009]	[0.008]	[0.029]
Quarters taught		-0.0015	-0.0018	0.0003	-0.0022
		[0.001]	[0.002]	[0.002]	[0.009]
I(taught this course before)		-0.0034	-0.0023	-0.0065	0.0656**
		[0.007]	[0.008]	[0.007]	[0.029]
I(coordinator)		0.0135*	0.0138	0.0193**	0.0082
		[0.008]	[0.009]	[0.009]	[0.026]
Labor			0.0013		
			[0.010]		
Macro			0.0062		
			[0.012]		
Metrics			0.0234**		
			[0.011]		
Theory			0.0053		
			[0.011]		
Constant	0.004	-0.069**	-0.083**	-0.047	0.957***
	[0.004]	[0.031]	[0.035]	[0.039]	[0.124]
Course dummies	No	No	No	Yes	No
Nationality dummies	No	No	No	No	Yes
Observations	78	78	78	78	83
R-squared	0.004	0.100	0.121	0.171	0.432

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is the deviation of the students' grade by TA section from the overall course average, and the independent variable of interest is Treatment, an indicator variable of receiving treatment. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.15: Regression Analysis ITT Winter 2013: TA Evaluation

	Independent Variable: Average Evaluation of TA (one quarter later)					
	(1)	(2)	(3)	(4)	(5)	(6)
Intent to Treat	0.3661*	0.3653*	0.3372*	0.3444	0.7209***	0.4632*
	[0.193]	[0.210]	[0.199]	[0.210]	[0.214]	[0.256]
Age		-0.0060	-0.0368	0.0175	-0.0391	0.1043
		[0.062]	[0.063]	[0.068]	[0.042]	[0.101]
I(male)		0.0331	-0.0268	0.1852	0.1978	0.2315
		[0.186]	[0.190]	[0.269]	[0.171]	[0.209]
I(English native)		0.2588	0.1654	0.1776	0.4229**	-0.0388
		[0.243]	[0.224]	[0.252]	[0.174]	[0.382]
PhD year		-0.0810	0.1082	-0.1026	-0.0777	-0.1373
		[0.221]	[0.225]	[0.256]	[0.203]	[0.275]
I(MA)		0.2311	0.2865	0.1870	0.4408**	-0.0566
		[0.272]	[0.265]	[0.276]	[0.212]	[0.322]
Quarters taught		0.0367	-0.0048	0.0236	0.0605	0.0524
		[0.043]	[0.044]	[0.052]	[0.043]	[0.052]
I(taught this course before)		0.0690	0.0796	0.1384	-0.1634	0.3549
		[0.285]	[0.273]	[0.298]	[0.275]	[0.319]
I(coordinator)		-0.2097	-0.0896	-0.1599	-0.0074	-0.3444**
		[0.145]	[0.140]	[0.155]	[0.135]	[0.161]
Lagged TA eval			0.3238*			
			[0.178]			
Labor				0.3641		
				[0.327]		
Macro				-0.1760		
				[0.425]		
Metrics				-0.2024		
				[0.507]		
Theory				0.1275		
				[0.314]		
Constant	7.5809***	7.5691***	5.6157***	7.0116***	8.2518***	4.1761
	[0.142]	[1.507]	[1.816]	[1.710]	[1.069]	[2.896]
Course dummies	No	No	No	No	Yes	No
Nationality dummies	No	No	No	No	No	Yes
Observations	94	94	94	94	94	94
R-squared	0.051	0.114	0.166	0.151	0.467	0.239

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is the TA's average student overall evaluation by section (in general, TAs are responsible for teaching two sections) and the variable of interest is ITT, an indicator variable of the intent to treat. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.16: Regression Analysis ToT Winter 2013: TA Evaluation

	Independent Variable: Average Evaluation of TA (one quarter later)					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.3631*	0.3937*	0.3681*	0.3322	0.6935***	0.4252
	[0.207]	[0.217]	[0.209]	[0.226]	[0.220]	[0.262]
Age		-0.0279	-0.0534	-0.0107	-0.0341	0.1249
		[0.065]	[0.066]	[0.072]	[0.051]	[0.095]
I(male)		0.0623	0.0166	0.1020	0.1681	0.2821
		[0.208]	[0.212]	[0.276]	[0.169]	[0.242]
I(English native)		0.2347	0.1405	0.1432	0.5079**	0.5733
		[0.276]	[0.259]	[0.348]	[0.196]	[0.450]
PhD year		-0.0327	0.1429	-0.0030	-0.0023	-0.0857
		[0.248]	[0.265]	[0.316]	[0.203]	[0.285]
I(MA)		0.2959	0.3707	0.2447	0.4744**	-0.0875
		[0.289]	[0.288]	[0.293]	[0.222]	[0.341]
Quarters taught		0.0428	0.0037	0.0146	0.0613	0.0633
		[0.046]	[0.050]	[0.058]	[0.044]	[0.056]
I(taught this course before)		-0.0570	-0.0176	0.0385	-0.1881	0.2474
		[0.290]	[0.281]	[0.314]	[0.274]	[0.321]
I(coordinator)		-0.2027	-0.0901	-0.1666	-0.0051	-0.3949**
		[0.161]	[0.150]	[0.174]	[0.149]	[0.159]
Lagged TA eval			0.2815			
			[0.180]			
Labor				0.3534		
				[0.492]		
Macro				-0.0532		
				[0.658]		
Metrics				-0.3736		
				[0.639]		
Theory				0.1530		
				[0.479]		
Constant	7.5809***	7.9806***	6.1997***	7.5931***	7.8272***	3.4248
	[0.143]	[1.620]	[1.917]	[1.849]	[1.353]	[2.787]
Course dummies	No	No	No	No	Yes	No
Nationality dummies	No	No	No	No	No	Yes
Observations	83	83	83	83	83	83
R-squared	0.050	0.119	0.156	0.163	0.477	0.299

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is the TA's average student overall evaluation by section (in general, TAs are responsible for teaching two sections) and the independent variable of interest is Treatment, an indicator variable of receiving treatment. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.17: Regression Analysis ITT Winter 2013: Log TA Evaluation

	Independent Variable: Log Average Eval. of TA (one quarter later)					
	(1)	(2)	(3)	(4)	(5)	(6)
Intent to Treat	0.0490*	0.0501*	0.0469*	0.0478	0.0971***	0.0635*
	[0.027]	[0.029]	[0.028]	[0.029]	[0.030]	[0.035]
Age		-0.0018	-0.0058	0.0014	-0.0063	0.0131
		[0.009]	[0.009]	[0.009]	[0.006]	[0.014]
I(male)		0.0038	-0.0041	0.0251	0.0259	0.0300
		[0.025]	[0.026]	[0.037]	[0.024]	[0.029]
I(English native)		0.0349	0.0226	0.0232	0.0571**	-0.0066
		[0.033]	[0.030]	[0.034]	[0.023]	[0.052]
PhD year		-0.0092	0.0149	-0.0136	-0.0085	-0.0169
		[0.030]	[0.031]	[0.035]	[0.028]	[0.037]
I(MA)		0.0383	0.0458	0.0326	0.0674**	-0.0023
		[0.039]	[0.038]	[0.039]	[0.031]	[0.044]
Quarters taught		0.0048	-0.0004	0.0032	0.0079	0.0070
		[0.006]	[0.006]	[0.007]	[0.006]	[0.007]
I(taught this course before)		0.0111	0.0117	0.0204	-0.0186	0.0488
		[0.039]	[0.038]	[0.040]	[0.038]	[0.043]
I(coordinator)		-0.0251	-0.0097	-0.0180	0.0009	-0.0438**
		[0.020]	[0.019]	[0.021]	[0.019]	[0.021]
Lagged TA eval			0.3061*			
			[0.180]			
Labor				0.0441		
				[0.044]		
Macro				-0.0288		
				[0.057]		
Metrics				-0.0313		
				[0.069]		
Theory				0.0131		
				[0.042]		
Constant	2.0192***	2.0347***	1.4812***	1.9659***	2.1260***	1.5833***
	[0.020]	[0.208]	[0.381]	[0.235]	[0.153]	[0.390]
Course dummies	No	No	No	No	Yes	No
Nationality dummies	No	No	No	No	No	Yes
Observations	94	94	94	94	94	94
R-squared	0.048	0.111	0.154	0.147	0.446	0.234

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is a logistic transformation of the TA's average student overall evaluation by section (in general, TAs are responsible for teaching two sections) and the variable of interest is ITT, an indicator variable of the intent to treat. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.18: Regression Analysis ToT Winter 2013: Log TA Evaluation

	Independent Variable: Log Average Eval. of TA (one quarter later)					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0482*	0.0536*	0.0507*	0.0463	0.0932***	0.0579
	[0.028]	[0.030]	[0.029]	[0.031]	[0.031]	[0.035]
Age		-0.0048	-0.0081	-0.0023	-0.0057	0.0161
		[0.009]	[0.009]	[0.010]	[0.007]	[0.013]
I(male)		0.0083	0.0019	0.0155	0.0226	0.0368
		[0.028]	[0.029]	[0.038]	[0.023]	[0.033]
I(English native)		0.0317	0.0193	0.0177	0.0690**	0.0720
		[0.037]	[0.035]	[0.047]	[0.027]	[0.061]
PhD year		-0.0024	0.0202	-0.0007	0.0019	-0.0096
		[0.034]	[0.036]	[0.043]	[0.027]	[0.038]
I(MA)		0.0477	0.0576	0.0410	0.0730**	-0.0067
		[0.041]	[0.041]	[0.042]	[0.032]	[0.046]
Quarters taught		0.0054	0.0005	0.0020	0.0079	0.0084
		[0.006]	[0.007]	[0.008]	[0.006]	[0.007]
I(taught this course before)		-0.0052	-0.0010	0.0075	-0.0216	0.0345
		[0.040]	[0.039]	[0.042]	[0.038]	[0.044]
I(coordinator)		-0.0239	-0.0092	-0.0184	0.0016	-0.0503**
		[0.022]	[0.020]	[0.023]	[0.020]	[0.021]
Lagged TA eval			0.2682			
			[0.182]			
Labor				0.0426		
				[0.067]		
Macro				-0.0150		
				[0.089]		
Metrics				-0.0544		
				[0.087]		
Theory				0.0154		
				[0.065]		
Constant	2.0192***	2.0897***	1.5940***	2.0434***	2.0700***	1.4763***
	[0.020]	[0.224]	[0.393]	[0.254]	[0.192]	[0.375]
Course dummies	No	No	No	No	Yes	No
Nationality dummies	No	No	No	No	No	Yes
Observations	83	83	83	83	83	83
R-squared	0.046	0.115	0.146	0.157	0.453	0.293

Robust standard errors in brackets and clustered by TA.
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variable is a logarithmic transformation of the TA's average student overall evaluation by section (in general, TAs are responsible for teaching two sections) and the independent variable of interest is Treatment, an indicator variable of receiving treatment. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.19: Summary Statistics Winter 2013: Other Outcomes

Average evaluation of TA's:	Mean	Median	St. Dev.	Min	Max
Knowledge	8.05	8.22	0.62	5.83	8.95
Concern	7.84	7.91	0.74	5.50	9.00
Organization	7.75	7.90	0.78	4.80	8.90
Scope	7.65	7.77	0.76	5.20	8.88
Interaction	7.76	7.84	0.76	5.67	8.95
Communication	7.37	7.50	1.04	4.00	8.89

Note: The table presents the summary statistics of the more specific questions of the students' evaluation of the TA regarding how knowlegdable the TA is, how concern is the TA about the student learning, the organization and preparation of the section, the scope of the section, how welcome students felt (interaction), and the TAs communication skills.

Table 2.20: Difference of Means (Other Outcomes) Winter 2013: ITT

Average evaluation of TA's:	Control	ITT	Difference	p-value (equal means)
Knowledge	8.00	8.08	0.07	0.58
Concern	7.66	7.95	0.29	0.06*
Organization	7.56	7.87	0.31	0.05*
Scope	7.44	7.77	0.33	0.04**
Interaction	7.57	7.88	0.31	0.050*
Communication	6.95	7.62	0.68	0.002***
N	35	59		

Note: The table presents the difference in means of the more specific questions of the students' evaluation of the TA, specified above, between the control and the ITT groups. The fourth column presents the p-value of the equality test of means between the two groups. Also, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.21: Difference of Means (Other Outcomes) Winter 2013: ToT

Average evaluation of TA's:	**Control**	**Treatment**	**Difference**	**p-value (equal means)**
Knowledge	8.00	8.06	0.06	0.67
Concern	7.66	7.94	0.28	0.17
Organization	7.56	7.90	0.34	0.05*
Scope	7.44	7.76	0.32	0.06*
Interaction	7.57	7.88	0.31	0.07*
Communication	6.95	7.59	0.64	0.006***
N	35	48		

Note: The table presents the difference in means of the more specific questions of the students' evaluation of the TA, specified above, between the control and the ToT groups. The fourth column presents the p-value of the equality test of means between the two groups. Also, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.22: Regression Analysis ITT Winter 2013: Other Outcomes

	Independent Variable: Average Evaluation of TA (one quarter later)					
	Knowledge		**Concern**		**Organization**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Intent to Treat	0.0739	0.0344	0.2898*	0.2889	0.3131*	0.2467
	[0.140]	[0.144]	[0.165]	[0.174]	[0.187]	[0.201]
Age		0.0074		0.0468		0.0346
		[0.050]		[0.050]		[0.055]
I(male)		0.0271		-0.0006		-0.1195
		[0.140]		[0.173]		[0.170]
I(English native)		0.0906		-0.0536		0.1673
		[0.193]		[0.242]		[0.263]
PhD year		-0.1384		-0.1701		-0.4470
		[0.177]		[0.172]		[0.278]
I(MA)		0.0114		0.0121		-0.0061
		[0.184]		[0.203]		[0.216]
Quarters taught		0.0331		0.0221		0.0900
		[0.033]		[0.033]		[0.057]
I(taught this course before)		0.1433		0.1231		0.1460
		[0.222]		[0.269]		[0.278]
I(coordinator)		-0.2274**		-0.1571		-0.0445
		[0.111]		[0.104]		[0.129]
Constant	8.0046***	7.9496***	7.6629***	6.7417***	7.5574***	7.3776***
	[0.076]	[1.207]	[0.110]	[1.265]	[0.130]	[1.393]
Observations	94	94	94	94	94	94
R-squared	0.003	0.070	0.037	0.077	0.038	0.107
	Scope		**Interaction**		**Communication**	
	(7)	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Intent to Treat	0.3268*	0.3348*	0.3128*	0.3076	0.6760**	0.6757**
	[0.171]	[0.167]	[0.179]	[0.187]	[0.280]	[0.281]
Age		0.0100		0.0094		-0.0289
		[0.054]		[0.052]		[0.080]
I(male)		0.1006		-0.0033		0.1557
		[0.148]		[0.169]		[0.224]
I(english native)		0.1963		0.1551		0.7114**
		[0.234]		[0.241]		[0.271]
PhD year		-0.1665		-0.1241		-0.0012
		[0.180]		[0.193]		[0.311]
I(MA)		0.1613		0.1665		0.5380
		[0.193]		[0.230]		[0.386]
I(taugh before)		0.0253		0.0120		0.0287
		[0.034]		[0.037]		[0.063]
I(taugh this before)		0.1161		0.2480		0.2218
		[0.261]		[0.272]		[0.353]
I(coordinator)		-0.1840		-0.1122		-0.2976*
		[0.126]		[0.142]		[0.171]
Constant	7.4440***	7.2946***	7.5674***	7.4001***	6.9474***	6.9807***
	[0.119]	[1.358]	[0.129]	[1.300]	[0.234]	[1.822]
Observations	94	94	94	94	94	94
R-squared	0.043	0.085	0.040	0.076	0.100	0.248

Robust standard errors in brackets and clustered by TA
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variables are the average scores by TA of the specific questions of the students' evaluation and the independent variable of interest is ITT, an indicator variable of intent to treat. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Table 2.23: Regression Analysis ToT Winter 2013: Other Outcomes

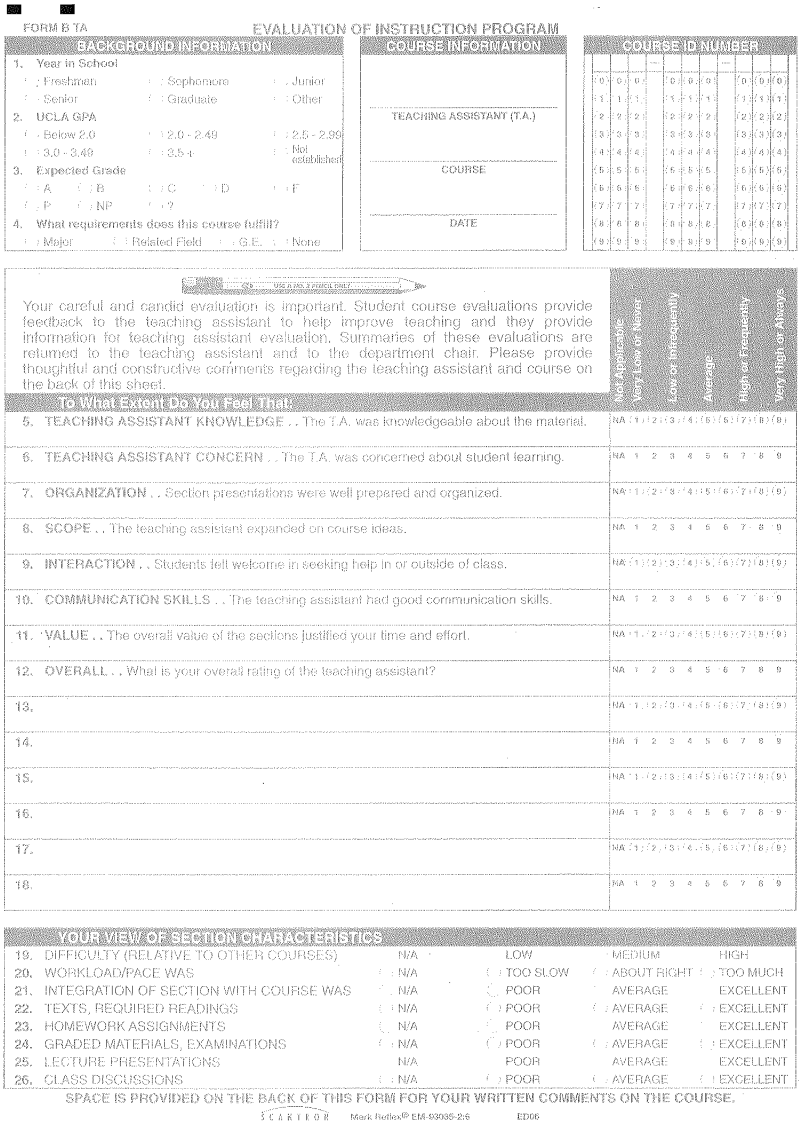
	Independent Variable: Average Evaluation of TA (one quarter later)					
	Knowledge		**Concern**		**Organization**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Treatment	0.0596	0.0354	0.2758	0.2876	0.3444	0.2978
	[0.157]	[0.157]	[0.180]	[0.185]	[0.205]	[0.214]
Age		-0.0032		0.0389		0.0253
		[0.054]		[0.054]		[0.061]
I(male)		-0.0217		-0.0012		-0.1080
		[0.148]		[0.194]		[0.190]
I(English native)		0.0696		-0.0214		0.1359
		[0.214]		[0.266]		[0.292]
PhD year		-0.1516		-0.1592		-0.4457
		[0.201]		[0.187]		[0.297]
I(MA)		-0.0250		0.0004		-0.0026
		[0.201]		[0.220]		[0.238]
Quarters taught		0.0442		0.0290		0.0986
		[0.035]		[0.035]		[0.061]
I(taught this course before)		0.0372		0.0247		0.0856
		[0.235]		[0.285]		[0.304]
I(coordinator)		-0.2394**		-0.1615		-0.0651
		[0.119]		[0.116]		[0.138]
Constant	8.0046***	8.3157***	7.6629***	6.9261***	7.5574***	7.5992***
	[0.077]	[1.351]	[0.110]	[1.409]	[0.130]	[1.548]
Observations	83	83	83	83	83	83
R-squared	0.002	0.070	0.033	0.064	0.044	0.107
	Scope		**Interaction**		**Communication**	
	(7)	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
Treatment	0.3168*	0.3474*	0.3127	0.3173	0.6403**	0.6627**
	[0.185]	[0.177]	[0.193]	[0.200]	[0.296]	[0.293]
Age		-0.0081		0.0004		-0.0447
		[0.057]		[0.057]		[0.086]
I(male)		0.0797		-0.0155		0.1871
		[0.168]		[0.189]		[0.248]
I(english native)		0.1504		0.1650		0.7473**
		[0.262]		[0.269]		[0.309]
PhD year		-0.1432		-0.1062		0.0633
		[0.206]		[0.214]		[0.331]
I(MA)		0.1908		0.1676		0.6056
		[0.205]		[0.250]		[0.415]
I(taugh before)		0.0337		0.0231		0.0328
		[0.037]		[0.040]		[0.067]
I(taugh this before)		0.0025		0.1375		0.0937
		[0.274]		[0.284]		[0.364]
I(coordinator)		-0.1800		-0.1226		-0.2841
		[0.134]		[0.156]		[0.194]
Constant	7.4440***	7.7179***	7.5674***	7.5792***	6.9474***	7.1779***
	[0.119]	[1.515]	[0.129]	[1.440]	[0.234]	[2.002]
Observations	83	83	83	83	83	83
R-squared	0.040	0.077	0.039	0.068	0.088	0.243

Robust standard errors in brackets and clustered by TA
*** p<0.01, ** p<0.05, * p<0.1

Note: Results from OLS regressions in which the dependent variables are the average scores by TA of the specific questions of the students' evaluation and the independent variable of interest is Treatment, an indicator variable of receiving treatment. The controls include age, male, English native speaker, masters degree before the PhD, number of quarters taught, a variable indicating if the TA has taught the same course before, an indicator variable for having met the TA coordinator and field of specilization (Industrial Organization is the excluded category). Robust errors cluster by TA.

Appendix B

Table 2.24: Copy of the TAs Evaluation



FORM B TA

EVALUATION OF INSTRUCTION PROGRAM

BACKGROUND INFORMATION

1. Year in School
   - Freshman
   - Sophomore
   - Junior
   - Senior
   - Graduate
   - Other
2. UCLA GPA
   - Below 2.0
   - 2.0 - 2.49
   - 2.5 - 2.99
   - 3.0 - 3.49
   - 3.5 +
   - Not established
3. Expected Grade
   - A
   - B
   - C
   - D
   - F
   - P
   - NP
   - ?
4. What requirements does this course fulfill?
   - Major
   - Related Field
   - G.E.
   - None

COURSE INFORMATION

TEACHING ASSISTANT (T.A.)

COURSE

DATE

COURSE ID NUMBER

USE A NO. 2 PENCIL ONLY

Your careful and candid evaluation is important. Student course evaluations provide feedback to the teaching assistant to help improve teaching and they provide information for teaching assistant evaluation. Summaries of these evaluations are returned to the teaching assistant and to the department chair. Please provide thoughtful and constructive comments regarding the teaching assistant and course on the back of this sheet.

To What Extent Do You Feel That:	Not Applicable / Very Low or None / Low or Infrequently / Average / High or Frequently / Very High or Always
5. TEACHING ASSISTANT KNOWLEDGE . . The T.A. was knowledgeable about the material.	NA 1 2 3 4 5 6 7 8 9
6. TEACHING ASSISTANT CONCERN . . The T.A. was concerned about student learning.	NA 1 2 3 4 5 6 7 8 9
7. ORGANIZATION . . Section presentations were well prepared and organized.	NA 1 2 3 4 5 6 7 8 9
8. SCOPE . . The teaching assistant expanded on course ideas.	NA 1 2 3 4 5 6 7 8 9
9. INTERACTION . . Students felt welcome in seeking help in or outside of class.	NA 1 2 3 4 5 6 7 8 9
10. COMMUNICATION SKILLS . . The teaching assistant had good communication skills.	NA 1 2 3 4 5 6 7 8 9
11. VALUE . . The overall value of the sections justified your time and effort.	NA 1 2 3 4 5 6 7 8 9
12. OVERALL . . What is your overall rating of the teaching assistant?	NA 1 2 3 4 5 6 7 8 9
13.	NA 1 2 3 4 5 6 7 8 9
14.	NA 1 2 3 4 5 6 7 8 9
15.	NA 1 2 3 4 5 6 7 8 9
16.	NA 1 2 3 4 5 6 7 8 9
17.	NA 1 2 3 4 5 6 7 8 9
18.	NA 1 2 3 4 5 6 7 8 9

YOUR VIEW OF SECTION CHARACTERISTICS

19. DIFFICULTY (RELATIVE TO OTHER COURSES)	N/A	LOW	MEDIUM	HIGH
20. WORKLOAD/PACE WAS	N/A	TOO SLOW	ABOUT RIGHT	TOO MUCH
21. INTEGRATION OF SECTION WITH COURSE WAS	N/A	POOR	AVERAGE	EXCELLENT
22. TEXTS, REQUIRED READINGS	N/A	POOR	AVERAGE	EXCELLENT
23. HOMEWORK ASSIGNMENTS	N/A	POOR	AVERAGE	EXCELLENT
24. GRADED MATERIALS, EXAMINATIONS	N/A	POOR	AVERAGE	EXCELLENT
25. LECTURE PRESENTATIONS	N/A	POOR	AVERAGE	EXCELLENT
26. CLASS DISCUSSIONS	N/A	POOR	AVERAGE	EXCELLENT

SPACE IS PROVIDED ON THE BACK OF THIS FORM FOR YOUR WRITTEN COMMENTS ON THE COURSE.

SCANTRON Mark Reflex® EM-93065-2/6 ED06

Note: This is a copy of the actual evaluation that students fill out at the end of the quarter (usaully in the last two weeks of the quarter). The normal procedure is that the TA hands the evaluations to the students and leaves the room after instructing ond of the students to take them in a sealed envelope to the Department's Office. This way, the TA has no chance to change or modify the scores of the students.

2.8 References

Allen, Liz (2002). "Consenting Adults in Private - Union and Management Perspectives on Peer Observation of Teaching". Working Paper of the Higher Education Academy.

Bell, Amani and Mladenovic, Rosina. (2008). "The Benefits of Peer Observation of Teaching for Tutor Development". High Education.

Bell, Amani. (2005). "Peer Observation and Teaching in Australia". Report of the Higher Education Academy.

Bettinger, Eric and Bridget Long. (2011) "Do College Instructors Matter? The Effects of Adjuncts on Students' Interests and Success." Review of Economics and Statistics.

Borjas, George J. (2000). "Foreign-Born Teaching Assistants and the Academic Performance of Undergraduates". The American Economic Review.

Card, David, and Alan Krueger. (1992) "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States". Journal of Political Economy.

Card, David, and Alan Krueger. (1996) "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina". Journal of Economic Perspectives.

Carrell, Scott E. and James E. West. (2010) "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors". Journal of Political Economy.

Carroll, Gregory. (1980). "Effects of Training Programs for University Teaching Assistants: A Review of Empirical Research". The Journal of Higher Education.

Danielson, Charlotte. (2011). "The Framework for Teaching Instrument" The Danielson Group.

David, Adonis and Macayan, Jonathan. (2010). "Assessment of Teacher Performance". The Assessment Handbook Vol. 3.

Dye, Ronald. (1986). "Optimal Monitoring Policies in Agencies". The RAND Journal of Economics, RAND Corporation.

Ehrenberg, Ronald G. and Zhang, Liang. (2005) "Do Tenure and Tenure-Track Faculty Matter?". Journal of Human Resources.

Glewwe, Paul W. & Eric A. Hanushek & Sarah D. Humpage & Renato Ravina. (2011). "School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010," NBER Working Papers.

Hanushek, Eric (2003). The failure of input-based schooling policies. The Economic Journal.

Hanushek, Eric. (2007). "Education Production Functions". Palgrave Encyclopedia.

Hoffman, Florian and Philip Oreopoulos. (2009) "Professor Qualities and Student Achievement". Review of Economics and Statistics.

Holmstrom, Bengt. (1979). "Moral Hazard and Observability". The Bell Journal of Economics, Rand Corporation.

Koedel, C. and Beets, J. (2007) "Re-Examining the Role of Teacher Quality In the Educational Production Function". University of Missouri Working Paper.

Koedel, Cory and Betts, Julian (2008). "Test-Score Ceiling Effects and Value-Added Measures of School Quality". JSM Proceedings, Social Statistics Section.

Laat, Joost de. (2005). “Moral Hazard and Costly Monitoring: The Case of Split Migrants in Kenya”. Mimeo.

Levitt, Steven and Jacob, Brian. (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." Quarterly Journal of Economics.

List, John and Imran, Rasul. (2011). “Field Experiments in Labor Economics” Handbook of Labor Economics V. 4a. Chapter 2.

Marsh, H. W. (1983). “Multidimensional ratings of teaching effectiveness by students from different academic settings and their relation to student/course/instructor characteristics”. Journal of Educational Psychology.

Norris, Timothy. (1991). “Nonnative English-Speaking Teaching Assistants and Students Performance”. Research in Higher Education.

Park, Chris. (2004). “The graduate teaching assistants (GTAs): lessons from North America”. Teaching in Higher Education.

Perdendergast, Canice. (1999). “The Provision of Incentives in Firms”. Journal of Economic Literature.

Quinlan, K. and G. Akerlind. (2000). “Factors Affecting Departmental Peer Collaboration for Faculty Development: Two Cases in Context”. Higher Education Vol. 40.

Sparks, G.M. (1986). “The effectiveness of alternative training activities in changing teaching practices”. American Educational Research Journal.

Tuckman, Howard P. (1975). “Teacher Effectiveness and Student Performance”. The Journal of Economic Education.

Umansky, Ileana and Vargas, Emiliana. (2005). “Incentives to Improve Teaching: Lessons from Latin America”. World Bank Publications.

Watts, Michael and Gerald J. Lynch. (1989). "The principles courses revised". The American Economic Review, Papers and Proceedings.

Chapter 3

3 Risk Preferences and Risk Sharing:

Evidence from Mexico[53]

3.1 Introduction

Understanding the extent to which individuals or households are able to share risk is essential from a public policy perspective. If individuals are not able to insure themselves against idiosyncratic risk using the available instruments (public or private), then there is room for the government to intervene and improve overall welfare. This intervention may take place through the improvement or regulation of private insurance markets, through the adjustment of existing social welfare programs or through the introduction of new ones. Also, the question of whether risk sharing exists provides a better understanding of the economy overall.

The topic is of particular relevance in developing countries, where access to financial markets is limited to few sectors of society, the coverage of social welfare is relatively low compared to developed countries, and economic fluctuations both at the aggregate and idiosyncratic levels are more pronounced. Moreover, the governments of developing countries tend to have more constraints for rapidly responding to undesirable social circumstances or

[53]I would like to thank Maurizio Mazzocco and Maria Casanova for their insightful comments on this paper. Also, I am thankful to all the participants in the UCLA Applied Microeconomic Proseminar for useful insights.

for changing institutions accordingly. Thus, the consequences of a low degree of risk sharing in developing countries can be more severe and long term.

In a world of complete markets with contingent securities, depicted in the Arrow-Debreu setup, an individual's consumption should respond to aggregate shocks, such as the business cycle, but not to idiosyncratic shocks, such as unemployment or temporary illness. As noted by Cochrane (1991), this implication provoked initial discontent, as the real world seems to deny the existence of formal markets in which consumers can ensure against any possible state of nature.

Nonetheless, in absence of these types of contingent assets, insurance may be provided by other formal and informal alternatives. Among the formal alternatives there are financial assets, such as stocks, bonds and securities and borrowing from financial institutions. Similarly, government welfare programs, such as unemployment and disability insurance, provide other formal options. Insurance can also be provided through informal mechanisms, such as loans and transfers from friends, family and other types of social networks. In this sense, financial markets need not be complete in terms of contingent-securities for the model's implications to hold.

Under this framework, the test of full insurance or efficient risk sharing can be performed by analyzing whether changes in household or individual consumption are determined by changes in aggregate consumption, rather than by changes in individual income or employment status. Overall, the empirical literature that has tested efficient risk sharing in both developed and developing countries has rejected its existence, as will be reviewed in detail in the next section. As Mazzocco and Saini (2009) established, a plausible explanation for this phenomenon is that with very few exceptions, this literature has neglected heterogeneity in risk preferences, which may result in the rejection of efficient risk sharing even when households do share risk efficiently.

A crucial implication of this result is that heterogeneous risk preferences should be somehow taken into account when trying to test the traditional risk-sharing hypothesis that says

changes in consumption should respond to changes in aggregate income but not to changes in own income (or some measure of idiosyncratic shock); and that is derived from the first order conditions of the maximization problem of the central planner.[54] Along these lines, in this paper I introduce a new and simple test that accounts for heterogeneous risk preferences in the traditional efficient risk sharing test when a measure of risk preference is observed in the data. Assuming a finite set of risk aversion parameters, it is possible to group households who share similar attitutes towards risk. In the proposed test, changes in household consumption should depend on the mean consumption change of each risk group and not on the household's own income. There are some data requirements in order to implement the test: in addition to a measure of risk aversion, needed to classify households into risk groups, a panel data set with a considerable number of waves containing income and expenditure information is needed to implement the test. To my knowledge, no dataset fulfills all these requirements at the moment.

Since the traditional risk sharing test only works under the assumption that risk preferences are homogenous, I then develop an alternative way to incorporate heterogeneity in the test: implement it only within households that share the same risk preferences. In other words, test whether households that have similar risk preferences engage in efficient risk sharing. This alternative has a disadvantage with respect to the first test I propose: it can not test risk sharing among households of different risk groups.

The data I use to implement the test within risk groups is the MXFLS (Mexican Family Life Survey, which contains detailed information on household expenditure and income variables, as well as a battery of lottery questions aimed to assess risk attitudes of the houshold members. The MXFLS is a panel data set, but at the moment only two waves are available, which prevents me from performing the test that doesn't need to separate the sample into risk groups because that test requires a long panel for identification purposes. It is important

[54]Traditional, in the spirit of Mace (1991), Cochrane (1991) and Townsend (1994). This differs from the test introduced by Mazzocco and Saini (2009), which is based on the expenditure functions of the households and is robust to risk preference heterogeneity.

to note that implementing the traditional test within each risk group is a valid exercise, because the assumption of homogeneous risk preferences holds and thus, efficient risk sharing should not be rejected if it actually occurs. A drawback of this alternative is that it only tests for risk sharing within the groups, and not across groups.

The results show that within-groups efficient risk sharing is rejected in almost 60% of the cases, mainly when the total household income is considered as the relevant income variable (as opposed to non-labor income).[55] In other words, the results favor the rejection of the ERS hypothesis within risk groups. This way, the results are not very different from implementing the test using the whole sample (and ignoring risk preference heterogeneity), which rejects efficient risk sharing under both Constant Absolute Risk Aversion (CARA) and Constant Relative Risk Aversion (CRRA) preferences.

A potential explanation for the rejection of risk sharing within risk groups is that these groups are not the relevant risk sharing unit, or alternatively, that there's still some preference heterogeneity within the groups formed using the measure available. In an attempt to address the later concern, I employ other risk related questions to further refine the risk groups, which implies obtaning smaller risk sub-groups. The results of implementing the test in these subgroup are highly imprecise, possibly because the small number of observations cause the test to be underpowered.

This paper contributes to the risk sharing literature by introducing a simple way of incorporating risk preference heterogeneity in the traditional efficient risk sharing test when a measure of risk preference is actually observed. In previous papers, heterogeneity of risk preferences was either incorporated theoretically, by using a test based on expenditure functions, such as in Mazzocco and Saini (2009), or by treating households' preferences as nuisance parameters that must be eliminated from the equation, as in Schulhofer-Wohl (2010), but never observed. The new test that may be less demanding in terms of data requirements and theoretical assumptions if there's a measure of risk preferences in the data. This way it

[55]Non-labor income is considered to be more exogenous to idisyncratic shocks but at the same time it is more problematic to measure.

constitutes progress in the way economists test the efficient risk sharing hypothesis.

The paper is organized as follows. Section 3.2 presents a brief literature review on risk sharing and risk preferences. The risk sharing model is derived in Section 3.3, both under risk preference heterogeneity and homogeneity (traditional way), where the first case is more general. Section 3.4 introduces the new test based on the implications of the model and outlines the empirical strategy. Section 3.5 provides a description of the MXFLS data, and presents some relevant statistics. The results for the pooled and within-groups tests are found in Section 3.6, and the last section concludes.

3.2 Risk Sharing Literature

One of the first papers to test the implications of full insurance using the traditional test is Mace (1991), who uses the US Consumer Expenditure Survey data for 1980 - 1983. Overall, the results are mixed. More precisely, they depend on the specification of the utility function, the use of exponential utility (CARA preferences) leads her to accept efficient risk sharing while power or isoelastic utility (CRRA preferences) leads her to reject it. Similarly, Cochrane (1991) tests whether consumption growth is independent of idiosyncratic variables that are exogenous to the individuals (such as involuntary unemployment and illness), as predicted by the full risk sharing result. Again, the results are mixed, full insurance is rejected for some of the shocks while not rejected for others. Moreover, it's not completely clear that these variables are exogenous with respect to consumption.

Hayashi, Altonji and Kotlikoff (1996a) use the PSID data to reject inter and intra-family risk sharing, consistent with their previous results in Hayashi, Altonji and Kotlikoff (1996b) while allowing for non-separable utility in consumption and leisure. Using synthetic panel data for consumption, labor supply and wages for the US, Attanasio and Davis (1997) find a failure of efficient risk sharing between cohort and education groups throughout the 1980s.

None of the above studies considers risk preference heterogeneity.[56]

Townsend (1994), in an important contribution to the literature, tests the risk sharing hypothesis for Indian villages using the ICRISAT data and partially addresses the issue of heterogeneous risk preferences. When identical risk preferences are assumed, he rejects efficient risk sharing, while when the assumption is dropped, the results are imprecise, given that the length of the panel data is too short to obtain good power properties of the test. Despite the rejection, he concludes that household consumption is not much influenced by contemporaneous own income, sickness, unemployment or other idiosyncratic shocks.

For Mexico, Attanasio and Szekely (2004) explore the co-movement of wages and consumption during the 1990s as an alternative test of risk sharing, using the ENIGH data set. Using information from a different data set for the same educational-cohort groups as instrumental variables for the wages, they reject the hypothesis that Mexican households are able to insure idiosyncratic risk in the case of non-durable consumption. Also, they examine how different categories of consumption goods (education, health, durable and non-durable goods) respond to fluctuations in wages and find that long-term investment in human capital is relatively more responsive than other categories. Even though they are able to cover a long period (1989-2000, biannual) and take advantage of the variation of groups over time, they lack panel data and do not incorporate risk preference heterogeneity in the analysis.

More recently, Dubois (2000) incorporated risk preference heterogeneity by using instrumental variables to estimate the household preference parameters and test the full risk-sharing hypothesis in rural areas of Pakistan. Interestingly, his results point out the importance of risk preference heterogeneity and reject efficient risk sharing after assuming that, conditional on observables, risk preferences are homogeneous. Nonetheless, he finds that share-cropping is indeed an important factor in risk sharing when risk markets are incomplete, such as in Pakistan. Also, Schulhofer-Wohl (2010) derives econometric techniques to eliminate the bias created by the fact that income variation is usually correlated with risk

[56]Other studies on risk sharing are Grimard (1997) for ethnic groups in Cote d'Ivore, and Jalan and Ravallion (1999) for wealth groups in China.

preferences, and concludes that welfare losses from the lack of insurance are small. Nonetheless, no direct measure of risk preference is used in either of these studies.

As mentioned above, Mazzocco and Saini (2009) prove that when risk preference homogeneity is assumed, the traditional test rejects the full insurance hypothesis even when households are efficiently sharing risk, leading to mistaken conclusions. The intuition is that in bad times, more risk averse households would have a higher level of consumption than less risk averse households, while in good times the opposite is true. Thus, for the more risk averse, consumption varies less than aggregate consumption, while for the less risk averse it varies more, which will lead to a rejection of the traditional risk sharing hypothesis.

To address this issue, the authors introduce a test based on the expenditure functions of the households to determine whether or not they have homogeneous preferences, and then they present two tests for full risk sharing that are valid in the presence of heterogeneity of preferences. Using data from ICRISAT, the authors find that there is preference heterogeneity across households and that the efficient risk sharing hypothesis is rejected at the village level but cannot be rejected at the caste level.

This paper contributes to the risk sharing literature by accounting for risk preference heterogeneity in a simple way, in the specific case where a measure of risk preference is observed. In a sense, it overcomes the problem of a large set of theoretical assumptions, requirement of previous studies that allowed risk preference heterogeneity. Nonetheless, as will be discussed in detail, the measure of risk aversion of the households must be accurate, otherwise, the proposed test may not be valid.

3.3 Risk Sharing Model

Following the tradition in the literature, the model can be framed as the maximization problem of a social planner in an Arrow-Debreu setup with uncertainty. Consider an endowment economy consisting of N households, which have time separable and leisure-consumption

separable expected utility over the consumption of a basket of non-durable goods.[57] This last assumption allows me to ignore the investment considerations that take place when a household is purchasing durable goods, and that are beyond the scope of this paper. States of the economy are summarized by s_t, which belong to a finite set S.

The Pareto efficient allocations are obtained when the social planner maximizes the weighted sum of the households' utilities subject to a resource constraint. That is:

$$\max \sum_{j=1}^{N} \mu^j \sum_{t=1}^{T} \sum_{s_t} \beta^t \pi(s_t) u^j[c^j(s_t), \delta^j(s_t)] \tag{3.1}$$

subject to the resource constraint

$$\sum_{j=1}^{N} c^j(s_t) \leq \sum_{j=1}^{N} y^j(s_t) \qquad \forall s_t \in S \tag{3.2}$$

where j indexes the households, μ^j is the corresponding Pareto weight for household j, β is the discounting rate for households (assumed to be same for all households), $\pi(s_t)$ is the probability of the occurrence of state s_t, $c^j(s_t)$ is the consumption of household j, and $\delta^j(s_t)$ are taste shifters. One could think of taste shifters as demographic characteristics of the household that affect the utility function such as the average age or the gender composition of the household members. Also, $y^j(s_t)$ is household's j endowment at a time t, and so the resource constraint implies that the sum of all the household's consumptions must be equal or less than the aggregate resource of the economy.

The first order condition for household j at time t is:

$$\mu^j \pi(s_t) \beta^t u_c^j[c^j(s_t), \delta^j(s_t)] = \lambda_t \tag{3.3}$$

where λ_t is the Lagrange multiplier on the resource constraint, and it's constant across all households at time t, so that the household's endowment *does not* determine the efficient

[57] Including savings in the model does not alter the implications of the model.

allocation given aggregate consumption and the Pareto weights. Alternatively:

$$\mu^j u_c^j[c^j(s_t), \delta^j(s_t)] = \widehat{\lambda}_t \tag{3.4}$$

where

$$\widehat{\lambda}_t = \frac{\lambda_t}{\pi(s_t)\beta^t} \tag{3.5}$$

Now, when CRRA preferences are assumed, the utility function is:

$$u^j[c^j(s_t), \delta^j(s_t)] = \frac{\delta^j(s_t)c^j(s_t)^{1-\rho_j}}{1-\rho_j} \tag{3.6}$$

Where, as usual, ρ_j is the risk aversion coefficient of household j. Note that, for simplicity, the taste shifters are assumed to enter the utility function multiplicatively. Then, we obtain the following first order condition:

$$\mu^j u_c^j[c^j(s_t), \delta^j(s_t)] = \mu^j \delta^j(s_t) c^j(s_t)^{-\rho_j} = \widehat{\lambda}_t \tag{3.7}$$

With CARA preferences, the utility function is specified as:

$$u^j[c^j(s_t), \delta^j(s_t)] = \frac{-\exp\left\{-\rho^j[c^j(s_t) - \delta^j(s_t)]\right\}}{\rho^j} \tag{3.8}$$

Once more, the taste shifters are assumed to enter the function multiplicatively. The first order condition is then:

$$\mu^j * u[c^j(s_t), \delta^j(s_t)] = \mu^j \exp\left\{-\rho^j[c^j(s_t) - \delta^j(s_t)]\right\} = \widehat{\lambda}_t \tag{3.9}$$

3.3.1 Heterogeneous preferences

If heterogeneous preferences are allowed in a discrete fashion, the household's risk aversion coefficient belongs to a finite subset of K elements, i.e. $\rho^j \in \{\rho^1, \rho^2, ..., \rho^K\}$. Under CRRA preferences, after taking the logarithm of the first order condition for household j and summing over all N households, it is possible to obtain the following expression for the consumption of household j:

$$\begin{aligned} \log c_t^j &= \frac{1}{N\rho^j}\sum_{i=1}^{N}\rho^i \log c_t^i + \frac{1}{\rho^j}\left(\log \mu^j - \frac{1}{N}\sum_{i=1}^{N}\log \mu^i\right) \\ &\quad + \frac{1}{\rho^j}\left(\log \delta_t^j - \frac{1}{N}\sum_{i=1}^{N}\log \delta_t^i\right) \end{aligned} \tag{3.10}$$

Where the state dependence of consumption and of the taste shifters are suppressed for convenience. Then, household's j consumption growth between time t and $t+1$ corresponds to:

$$\begin{aligned} \log c_{t+1}^j - \log c_t^j &= \frac{1}{N\rho^j}\left(\rho^1\sum_{i=1}^{N_1}(\log c_{t+1}^1 - \log c_t^1) + ... + \rho^K\sum_{i=1}^{N_K}(\log c_{t+1}^K - \log c_t^K)\right) \\ &\quad + \frac{1}{\rho_j}\left(\log \delta_{t+1}^j - \log \delta_t^j\right) \\ &\quad - \frac{1}{N\rho_j}\left(\sum_{i=1}^{N}\left(\log \delta_{t+1}^i - \log \delta_t^i\right)\right) \end{aligned} \tag{3.11}$$

where, the first difference has eliminated the time invariant terms and where the superscript of consumption represents the risk group to which the houshold belongs, i.e. $k \in \{1, 2, 3..., K\}$.

Under the less general assumption of CARA preferences we obtain that the consumption of household j may be expressed as:

$$c_t^j = \frac{1}{N\rho^j}\sum_{i=1}^{N}\rho^i c_t^i + \frac{1}{\rho^j}\left(\log\mu^j - \frac{1}{N}\sum_{i=1}^{N}\log\mu^i\right) + \left(\delta_t^j - \frac{1}{N\rho^j}\sum_{i=1}^{N}\rho^i\delta_t^i\right) \tag{3.12}$$

and thus, the change from period t to period $t+1$:

$$\begin{aligned} c_{t+1}^j - c_t^j &= \frac{1}{N\rho^j}\left(\rho^1\sum_{i=1}^{N_1}(c_{t+1}^1 - c_t^1) + \rho^2\sum_{i=1}^{N_2}(c_{t+1}^2 - c_t^2) + ... + \rho^K\sum_{i=1}^{N_K}(c_{t+1}^K - c_t^K)\right) \\ &\quad + \left(\delta_{t+1}^j - \delta_t^j\right) - \frac{1}{N\rho^j}\left(\sum_{i=1}^{N}\rho^i\left(\delta_{t+1}^i - \delta_t^i\right)\right) \end{aligned} \tag{3.13}$$

As before, all time invariant terms are eliminated by the first difference.

In this fashion, under heterogeneous risk preferences, the efficient risk model implies that household j's consumption is positively correlated with the average consumption of each of the k groups of households that share the same risk aversion coefficient. Under CRRA, this is true for the consumption growth, while for CARA, this corresponds to changes in the level of consumption. Note that in both cases, the consumption change does not depend on the household's own income but rather on the weighted average of the consumption changes of the different risk groups, where the weights are proportional to the ratio of the group's risk aversion coefficient (ρ^k) to that of household j (ρ^j). Put differently, under efficient risk sharing, individuals are perfectly insured against idiosyncratic risk, but not against aggregate risk.

3.3.2 Specific case: Homogeneous risk preferences

Under the stricter assumption that all households have the same risk preferences (as most previous papers have assumed), we have that $\rho^j = \rho \ \ \forall j \in N$. For CRRA preferences, and following the same procedure as before, the logarithm of household's j consumption can be expressed as:

$$\begin{aligned}\log c_t^j &= \frac{1}{N}\sum_{i=1}^{N}\log c_t^i + \frac{1}{\rho}\left(\log \mu^j - \frac{1}{N}\sum_{i=1}^{N}\log \mu^i\right) \\ &\quad + \frac{1}{\rho}\left(\log \delta_t^j - \frac{1}{N}\sum_{i=1}^{N}\log \delta_t^i\right) \end{aligned} \tag{3.14}$$

Thus, the growth of household's j consumption between time t and $t+1$, which eliminates as before the time invariant terms, corresponds to:

$$\begin{aligned}\log c_{t+1}^j - \log c_t^j &= \frac{1}{N}\left(\sum_{i=1}^{N}\log c_{t+1}^i - \sum_{i=1}^{N}\log c_t^i\right) \\ &\quad + \frac{1}{\rho}\left(\log \delta_{t+1}^j - \log \delta_t^j\right) \\ &\quad - \frac{1}{N\rho}\left(\sum_{i=1}^{N}(\log \delta_{t+1}^i - \log \delta_t^i)\right) \end{aligned} \tag{3.15}$$

Likewise, under the assumption of CARA preferences, the consumption of household j corresponds to :

$$c_t^j = \frac{1}{N}\sum_{i=1}^{N}c_t^i + \frac{1}{\rho}\left(\log \mu^j - \frac{1}{N}\sum_{i=1}^{N}\log \mu^i\right) + \left(\delta_t^j - \frac{1}{N}\sum_{i=1}^{N}\delta_t^i\right) \tag{3.16}$$

Now, the change in the level of household j consumption is then:

$$\begin{aligned}c_{t+1}^j - c_t^j &= \frac{1}{N}\left(\sum_{i=1}^{N}(c_{t+1}^i - c_t^i)\right) + \left(\delta_{t+1}^j - \delta_t^j\right) \\ &\quad - \frac{1}{N}\left(\sum_{i=1}^{N}(\delta_{t+1}^i - \delta_t^i)\right) \end{aligned} \tag{3.17}$$

In this case, the change in consumption for household j between period t and $t+1$ depends on the change in the average aggregate consumption, controlling for taste shifters. Note that

these results hold *only* when households have identical risk preferences and also, that this is a specific case ($\rho^j = \rho$) of the first model, which allows for risk preference heterogeneity.

The latter is the traditional derivation used in most of the risk sharing literature, in which the change in the consumption of household j depends solely on the change in average aggregate consumption. Nonetheless, some studies have tried to incoporate non-separability of leisure and consumption into the model, such as Hayashi, Altonji and Kotlikoff (1996a and 1996b), an important extension of the model that is beyond the scope of this paper at present.

3.4 Empirical Strategy: Implications of Efficient Risk Sharing

3.4.1 Risk Sharing Test Allowing for Heterogeneous Preferences

The model that allows for heterogeneous risk preferences, derived in Section 3.1, provides a testable implication of efficient risk sharing that can be taken to the data if the researcher observes an accurate measure of risk aversion. In short: changes in household consumption should depend exclusively on the average change in household consumption of each of the groups of households that share similar risk preferences, controlling for demographic characteristics of the household that represent the taste shifters.[58] This is a new version of the traditional efficient risk sharing test, which improves the old version by incorporating heterogeneous risk preferences.

Under CRRA preferences, the test consists of regressing the growth of the household consumption, as approximated by the change in the logarithm ($\triangle \log c^j$), on the average consumption growth of each of the risk groups ($\triangle C^1, ..., \triangle C^K$), on the growth of the household's income (representing idiosyncratic risk) $\triangle log\, y^j$, and on the change in the demographic variables of the household ($\triangle X^j$). Notice that the error term ε includes possible measurement

[58] Under CRRA preferences, it is the growth of household consumption.

errors in consumption, expenditure and the other covariates. In summation, the specficiation is the following:

$$\triangle \log c_t^j = \phi_1 \triangle C_t^1 + \phi_2 \triangle C_t^2 + ... + \phi_k \triangle C_t^k + \zeta_1 \triangle log\, y_t^j + \zeta_2 \triangle X_t^j + \varepsilon_t \tag{3.18}$$

where

$$C_t^k = \frac{1}{N_k} \sum_{i=1}^{N_k} \log c_t^i \quad for \;\; k = 1...., K \tag{3.19}$$

Under the efficient risk sharing hypothesis, the coefficients accompanying the changes in consumption growth for each of the groups $(\phi_1, ..., \phi_K)$ should be different from each other, while the one accompanying the household's income (ζ_1) should not be significantly different from zero. As the theory predicts, no other variable should enter the equation, except for the average leisure (if we were to consider that separability is not a good assumption). The specification of the test is close to that of the traditional test (outlined below), with the difference that consumption growth is aggregated within each of the K risk groups.

For CARA preferences, the specification of the test translates into:

$$\triangle c_t^j = \alpha_1 \triangle C_t^1 + \alpha_2 \triangle C_t^2 + ... + \alpha_k \triangle C_t^k + \gamma_1 \triangle y_t^j + \gamma_2 \triangle X_t^j + \varepsilon_t \tag{3.20}$$

where

$$C_t^k = \frac{1}{N_k} \sum_{i=1}^{N_k} c_t^i \quad for \;\; k = 1...., K$$

and all other variables are defined as above. Under the null hypothesis of risk sharing, the coefficients accompanying the changes in consumption of the risk groups $(\alpha_1, .., \alpha_K)$ should be different from each other when explaining the change in consumption of household j, while the coefficient accompanying the household's income (γ_1) should not be statistically significant from zero. Note that in either case, it is crucial to be able to group household according to their risk preferences. This implies that the measure of risk preferences in the

data should be very accurate.

In both models, the prediction is that the coefficient accompanying the average of the more risk averse group should be larger than that of a less risk averse group. Consider for example the two extreme groups 1 (least risk averse) and K (most risk averse), with the corresponding risk aversion coefficients ρ_1 and ρ_K. Then the coefficient accompanying the average change in consumption of risk group K should be larger than that of risk group 1 since $\frac{\rho_K}{\rho_j} > \frac{\rho_1}{\rho_j}$.

3.4.2 Traditional Risk Sharing Test - Homogeneous Preferences

If risk preferences are assumed to be homogeneous among all households, efficient risk sharing (or the Pareto allocations) implies that changes in household consumption are determined only by changes in average aggregate consumption rather than idiosyncratic variables, once more controlling for possible taste shifters.

This is the traditional risk sharing test, which can be implemented through a simple specification, depending on the assumed preferences. For CRRA preferences, the test regresses the change in the logarithm of household consumption ($\triangle \log c^j$) on the change in the logarithm of aggregate consumption ($\triangle C$), the change of the logarithm of income (or other idiosyncratic shock, $\triangle \log y^j$) and the change in demographic variables ($\triangle X^j$), as follows:

$$\triangle \log c_t^j = \delta_1 \triangle C_t + \zeta_1 \triangle \log y_t^j + \zeta_2 \triangle X_t^j + \varepsilon_t \tag{3.21}$$

where

$$C_t = \frac{1}{N} \sum_{j=1}^{N} \log c_t^j \tag{3.22}$$

Under the efficient risk sharing hypothesis, δ_1 should be significant and equal to one, while ζ_1 should not be statistically different from zero. As before, the error term contains possible measurement error in the consumption and income variables. This specification is the one

used by Mace (1991), ignoring the idiosyncratic shocks, by Cochrane (1991), where he tests directly whether idiosyncratic shocks affect changes in consumption, and by Altonji, Hayashi and Kotlikoff (1992).

For CARA preferences, the first difference in household consumption ($\triangle c^j$) is regressed on the first difference in aggregate consumption ($\triangle C$), on changes of the household income (or another variable representing idiosyncratic) $\triangle y^j$, and on the change in demographics variables ($\triangle X^j$). Notice that, as before, the error term ε includes possible measurement errors in consumption and expenditure.

$$\triangle c_t^j = \pi_1 \triangle C_t + \gamma_1 \triangle y_t^j + \gamma_2 \triangle X_t^j + \varepsilon_t \tag{3.23}$$

where,

$$C_t = \frac{1}{N} \sum_{j=1}^{N} c_t^j \tag{3.24}$$

Under the efficient risk sharing hypothesis, γ_1 should not be statistically different from zero, $\triangle y^j$ should not determine changes in consumption, and π_1 should be significant and equal to 1. This test is also used by Mace (1991), although she does not include measures of household income, and Townsend (1994).

Note that the models (CRRA and CARA) under heterogeneous preferences are nested in the corresponding models under homogeneous preferences. More precisely, if it is the case that all household share the same risk aversion coefficient, then in the case of the CRRA model, $\sum_{k=1}^{K} \phi_k = \delta_1$ and in the case of the CARA models $\sum_{k=1}^{K} \alpha_k = \pi_1$.

Empirical Approach and Data Requirements

In order to be able to implement the new efficient risk sharing test, there are two requirements: first, as highlighted before, to have an accurate measure of the household's risk aversion, and second, to have access to a panel dataset with several waves. The former is

needed to classify households in a finite number of household groups that share the same risk preferences, while the latter is needed to identify the $\phi_1, ..., \phi_K$ coefficients.

At the moment, and to the best of my knowledge, no dataset fulfills all these requirements. Given that the traditional risk sharing test only works under the assumption that risk preferences are homogenous, I then develop an alternative way to incorporate risk preference heterogeneity into the analysis: implement the test only within households that share the same risk preferences. That is, test whether there is efficient risk sharing among the households that belong to a same risk group. Despite the fact that it restricts the scope of the test, this exercise will be less likely to reject the efficient risk sharing hypothesis if it is true.

For this purpose I use the Mexican Family Life Survey (MXFLS), which contains detailed information on household expenditure and income variables, as well as a battery of lottery questions aimed to assess risk attitudes of the houshold members. The latter feature allows me to classify households in one out of six groups that have similar risk preferences. It is worth noting that the MXFLS is a panel data set, but at this time only two waves are available. This prevents me from using it for the test that doesn't need to separate the sample into risk groups because it requires a long panel for identification purposes. As a benchmark, the results of implementing the test within risk groups are then compared to the results of the traditional risk sharing test using the pooled sample, that is, assuming that all the households share the exact same preferences.

3.5 Data and Stylized Facts

The data used in this paper come from the MXFLS, a longitudinal survey of over 8,500 households (and around 35,000 individuals) containing information on various social, demographic, and economic features of both individuals and households. The data at the household level include detailed information on the household's economy: expenditures and

non-labor income, such as small businesses or farm related activities. At the individual level, the survey inquires in detail about education, labor and non-labor income, health, reproductive and migration history, time allocation and risk preferences, among other things. At the moment, only the 2002 and 2005 waves are available for public use.

In the context of the MXFLS, the following variables at the household level are used:

- *Consumption*: monthly per capita non-durable consumption of the household. It is constructed using weekly and monthly expenditure on food, goods, services, utilities and gift and payments (both in cash and in specie).

- *Income*: monthly per capita income of the household. Two definitions of income are considered in the analysis. The first one includes non-labor income (subsidies, rents, inheritances, occasional sales, etc.) only, while the second considers both labor and non-labor income (total income). The justification for the first one is that ideally, income should not be related to leisure (as labor income allegedly would be), even though it is not included in the equation of interest given the separabilty assumption of the utility function.[59]

- *Socio-economic characteristics (taste shifters)*: various socio-economic traits of the house are used. Also included are dummy variables for: household owns the dwelling place, female household head and urban household. The household head age and years of education, the household average age and fraction of children (under 6 years of age) are also included as socio-economic characteristics.

- *Risk preferences*: The MXFLS contains a battery of questions in which respondents are asked to choose between two lotteries (one certain and the second one risky), so by choosing the preferred lottery the respondent reveals risk preferences to some extent. For a given household, we consider the risk preferences of the household head.

The method of measuring risk preferences using lottery questions is widely used in experimental and finance economics. Ever since Binswanger (1980) used this approach to measure

[59]In other specifications not included in the document, two more definitions of income were used. The third one included any rural income of the household in addition to labor and non-labor income, while the fourth added net profits from family owned business. The results did not differ when these two definitions were used.

attitude towards risk in India, it has become a popular approach (see for example, Holt and Laury, 2002). Studies like Barsky et al. (1997) show that the degree of risk aversion (or in their paper, risk tolerance) predicted by the individuals' response to the lottery questions is correlated with risky behaviors, such as smoking, drinking and having life and health insurance. Dohmen et al. (2005) also reaches the conclusion that the results from the standard lottery approach are consistent with real life risk attitudes.

More importantly, the 2005 MXFLS was accompanied by another survey, the MXFLS-PP (Preferences Pilot), whose objective was to pilot approaches for eliciting attitudes and preferences and whose sample is a subsample of the former. The MXFLS-PP included incentivized tasks with real money, whose payoffs were usually around the daily wage of the average respondent. Using both surveys, Hamoudi (2006) concludes that the MXFLS risk indicator predicts similar patterns and behaviors to those of real-stakes incentivized tasks.[60] This is of course reassuring of the fact that the ordinal classification that is provided by these questions is close to how individuals are expected to behave in real life.

The layout of the questions, presented in Figure 3.1, makes it possible to classify the individual in one out of six risk groups. Each of the rectangles represents one question of the survey. In each question, the respondent had to choose between two lotteries (the first lottery depicted in the blue square and the second lottery depicted in the red square of the figure), each with two equally likely outcomes (50% - 50% chance) that can be seen within each lottery. For example, in the first question, the individual is asked which lottery he will choose: one with a 50% chance to win $1,000 and a 50% chance to win $1,000, or one with a 50% chance to win $500 and a 50% chance to win $2,000. The denomination is in Mexican Pesos, and at the time of the survey the exchange rate was around 10 Mexican pesos per dollar.

The battery of questions is adaptive, so that the next question depends on the option chosen, as illustrated in Figure 3.1. Depending on the path, each individual is classified in

[60] Unfortunately, the data from this pilot is not available to the public.

risk groups 1 through 6, where 6 is the more risk averse group and 1 is the less risk averse group. Two important remarks: in question five only those choosing the dominant option (1000, 7000), over (1000, 4000), are included in risk group 1, and in question four those choosing option (1000, 1000) are not classified into any groups as they are reversing their preferences by choosing the (1000, 1000) option.

Figure 3.2 shows that the distribution of households in the six risk groups is mainly bimodal. Around 36% of the household are classified into the least risk averse group (risk group 1), while around 40% are classified in the moderate risk averse group (risk group 3). Barely 10% of the households are part of the most risk averse group (risk group 6). Finally, 14% of households belong to the other three groups (risk group 2, 4 and 5)

Table 3.1 presents the summary statistics for the housholds' monthly expenditure and income (denominated in 2005 Mexican Pesos), in levels and in per capita terms.[61] First, it's important to note that the stakes of the lotteries used to identify risk aversion were approximately equal to the average household's monthly expenditure. Thus, it's a considerable amount of money but it might not be enough to change the household's location in the income distribution. It's with this consideration in mind that the model is presented under the assumption of both CRRA and CARA preferences. Second, on the one hand, per capita monthly expenditure had a negligible decrease between 2002 and 2005, mainly because family size increased. On the other hand, total per capita income increased by around 8 percent, mainly explained by an increase in labor income.

3.6 Results and Discussion

As explained above and due to data constraints, it's not possible yet to implement the new test proposed in this paper. Therefore, this section presents the results of an alternative way

[61] These correpond to gender-age adjusted per capita variables, following Townsend (1994). The gender-age adjusted size of a household is obtained by adding the following numbers: for adult males, 1.0; for adult females, 0.9; for males aged 13-18, 0.94; for females aged 13-18, 0.83; for children aged 7-12, 0.67; for children aged 4-6, 0.52; for toddlers 1-3, 0.32; and for infants 0.05.

to incorporate risk preference heterogeneity in the traditional efficient risk sharing (ERS): implement the test within each of the risk groups defined by the lottery questions. As mentioned, the test is valid in the sense that it is more likely that all households will share the same risk preferences within these risk groups. However, before doing this and as a benchmark for the upcoming results, it is informative to implement the ERS test to the pooled sample, which implicitly ignores that there is risk preference heterogeneity across households.

As shown in Table 3.2, both under CRRA and CARA preferences, the test rejects ERS within the households of all the sample[62], regardless of the definition of income used in the equation of interest. For the most part, the rejection is the result of the significance of the household income variable. One possibilty is that the test rejects efficient risk sharing because it does not consider risk preference heterogeneity. As Mazzocco and Saini (2009) pointed out, the traditional test may reject the null even in the presence of risk sharing because consumption for less risk averse households varies more than the aggregate consumption.

The next step is to implement the traditional ERS test within the risk groups. The results under CRRA preferences are shown in Table 3.3, using both non-labor and labor income as definitions for the household's income, while those under CARA preferences are displayed in Table 3.4. It must be noted that considering that the distribution of the household in the risk groups was mainly bimodal, some of the groups have very few observations, and thus, the power of the test is very low.[63]

Under CRRA preferences (Table 3.3), the ERS is rejected in almost 60% of the cases that have a considerable amount of observations (bold and black at the bottom of the table), particularly when total income, instead of non-labor income, is used. A similar story occurs for CARA preferences, the ERS test is rejected in 60% of the "credible" regressions. Once more, the test rejects the null in a higher proportion when the total household income definition is used. Despite the fact that ERS is accepted for some risk groups (mainly risk

[62]The survey is representative at the national level, so this test would reject ERS at a national level.
[63]The summary statistics of the sociodemographic variables can be found in Table 3.7 of the Appendix.

group 3), these results can not be considered strong evidence in favor of the ERS in Mexico, considering that total income may be the more appropiate income definition to be used in the equation of interest.

As can be seen in the Table 3.5 summary, the results favor the rejection of the ERS hypothesis among households of the same risk group. There are at least two possible interpretations of this result. First, it could be the case that the appropriate unit of risk sharing is not the risk group and that the test is rejecting the null because, indeed, it is false. A household may belong to the same risk group as another household but have no connection at all: not geographically, not in terms of relatives and not in terms of friends, etc.[64] One possible way to overcome this concern is using the new test proposed in this paper, which doesn't restrict ERS to just the households within the risk group. Second, it could be the case that the measure of risk preferences captured by the lottery questions is not precise enough and thus, there still might be some risk preference heterogeneity within people in the same risk group.

In order to discard the second explanation, two additional questions related to risky behavior were used to subdivide the most numerous risk groups 1 and 3, each into two subgroups. The first question inquires whether the household head has ever smoked and the second asks what's the probability that the household would invest all its monthly income into an informal ROSCA.[65] The first subgroup is comprised of the households whose household head has never smoked and responded 0% to the ROSCA question, while the second subgroup is comprised of all the rest. Roughly, households were distributed evenly across these two subgroups for both risk group 1 and risk group 3.

Once the subgroups were formed, the traditional ERS test was implemented within these subgroups. Table 3.6 shows the summary results of this procedure (the actual results can be found in Tables 3.8 and 3.9 of the Appendix), in which ERS is rejected in 50% of the

[64] A geographical subdivision (households residing in the same state) of the risk groups was attempted. However, there were not enough observations to implement the test in the subgroups.

[65] According to Gugerty (2007, pg. 1), Roscas are "locally organized groups that meet at regular intervals; at each meeting members contribute funds that are given in turn to one or more of the members".

CRRA specifications and only 20% of the CARA specifications. Nonetheless, the coefficients are estimated rather inefficiently; the standard errors are too large and thus, it is difficult to assess the outcomes of the test. In sum, the results show that the risk categories provided by the lottery questions can be refined but also that the risk groups are not the relevant unit at which households share risk in Mexico.

3.7 Conclusion

This paper introduces a new and simple test that accounts for heteregeneous risk preferences in the traditional efficient risk sharing test, when a measure of these risk preferences is observed. According to this test, changes in consumption should depend on the average consumption change of each group of households that share similar risk preferences, and not on the household's own income. Moreover, the test predicts that the more risk averse the group, the larger the coefficients accompanying the change in average consumption. In order to be able to implement the test, a panel data set with a considerable number of waves and containing information on both household's income and expenditure, and risk preferences, is required. Conditional on having an accurate measure of the household's risk preference, the proposed test may be less demanding in terms of data and in terms of theoretical assumptions than the tests proposed by recent papers in the literature which have incorporated risk preference heterogeneity successfully. At the moment and to the best of my knowledge, no dataset fulfills all these requirements. Given that the traditional risk sharing test only works under the assumption that risk preferences are homogenous, I then develop an alternative way to incorporate risk preference heterogeneity into the analysis: implement the test only within households that share the same risk preferences.

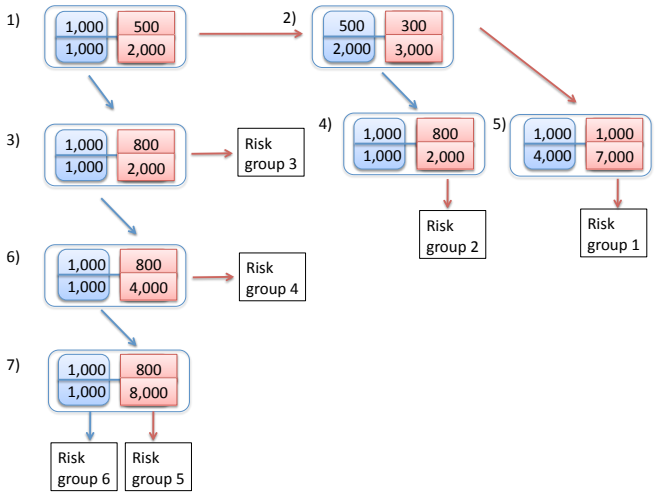
I do this using the MXFLS (Mexican Family Life Survey), which is a panel data set with two waves and contains detailed information on household expenditure and income, as well as a battery of lottery questions that allows me to classify households in groups

that share similar risk preferences. I then implement the traditional test within these risk groups, since the assumption of homogenous risk preferences is likely to hold (minimizing the type I error, or rejecting when the null hypothesis is true). When the test is performed within risk groups, the hypothesis is rejected in most of the specifications with a considerable number of observations, suggesting that there is not perfect risk sharing within these groups, both under CRRA and CARA preferences. This is a similar result to the one I obtain when implementing the traditional ERS test using the whole sample (which ignores risk preference heterogeneity).

The result within risk groups could be explained by the fact that there's still some risk preference heterogeneity within the groups formed using the lottery questions or that in fact, risk groups are not the appropiate risk sharing unit. Further refinement of the risk groups using other risk related questions leads to a lower rejection rate of the ERS hypothesis. Nonetheless, this can not be interpreted as strong evidence in favor of ERS because the low number of observations leads to a very low power of the test. An alternative explanation is that risk groups, or groups of households that share similar risk preferences, are not the correct reference group for risk sharing. It maybe the case that risk sharing is taking place among households with different risk sharing preferences.

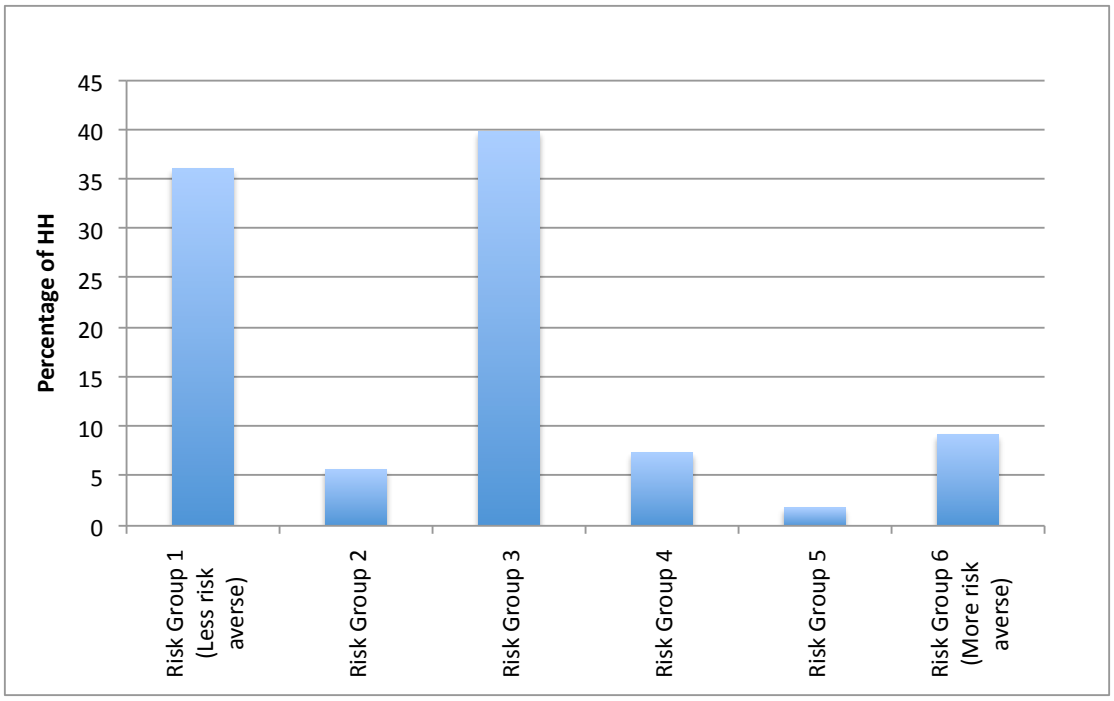
3.8 Main Figures and Tables

Figure 3.1: Battery of Lottery Questions



Note: This figure shows the layout of questions of the Risk module of the MXFLS. The sequence (order) depends on the response to each question. In each of them, the individual is asked to choose between two lotteries (blue and red squares), each with two equally likely outcomes (50-50% chance) that can be seen within each square. This allows me to classify individuals in one out of six risk preference groups.

Figure 3.2: Risk Preferences Groups



Note: Data from MXLFS (2005 wave). This figure shows the distribution of households in the six risk sharing categories, as determined by how the house head answered the lottery questions.

Table 3.1: Households' Expenditure and Income Variables

Year		Monthly expenditure	Per capita monthly exp.	Non-labor income	Per capita non-labor income	Total income	Per capita total income	Household size
2002	Mean	3,399	**1,114**	898	**326**	4,859	**1,507**	3.50
	St. Deviation	2,211	878	1,304	572	6,580	2,251	1.50
	Median	2,927	904	414	126	3,058	917	0.83
	N	6,961	6,961	3,377	3,377	5,847	5,847	6,989
	Mean	3,490	**1,052**	759	**263**	5,884	**1,631**	3.83
	St. Deviation	2,324	856	1,104	530	7,505	2,364	1.63
2005	Median	2,960	835	350	97	4,000	1,053	0.90
	N	6,845	6,845	2,700	2,700	5,661	5,661	6,884

Note: Exchange rate 1U$ = 10 M$. Data from MXLFS (2002 and 2005 waves). The table describes the main income and expenditure variables used in the analysis.

Table 3.2: ERS Test: Pooled

CRRA preferences

	Dependente variable	
	(1)	(2)
	Change in *log* per capita expenditure	
Change in *log* mean per capita expenditure (ϕ_1)	3.756***	2.197***
	[1.205]	[0.706]
Change in *log* per capita non-labor income (ζ_1)	0.0274*	
	[0.0152]	
Change in *log* per capita income (ζ_1)		0.0587***
		[0.00996]
N	**963**	**2,714**
R-squared	0.02	0.02
F-test ($\zeta_1 = 0$)	3.24	34.76
Prob > F	0.07	0.09
F-test ($\phi_1 = 1$)	5.23	2.88
Prob > F	0.02	0.00
Reject Efficient RS (95%)	**YES**	**YES**

Robust standard errors in brackets

Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age.

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

CARA preferences

	Dependente variable	
	(1)	(2)
	Change in per capita expenditure	
Change in mean per capita expenditure (ϕ_1)	2.446***	1.091
	[0.798]	[0.844]
Change in per capita non-labor income (ζ_1)	0.279**	
	[0.110]	
Change in per capita income (ζ_1)		0.0266**
		[0.0130]
N	**963**	**2,714**
R-squared	0.05	0.02
F-test ($\zeta_1 = 0$)	6.48	4.22
Prob > F	0.01	0.91
F-test ($\phi_1 = 1$)	3.28	0.01
Prob > F	0.07	0.04
Reject Efficient RS (95%)	**YES**	**YES**

Robust standard errors in brackets

Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age.

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

Note: The left panel shows the traditional ERS test using the pooled sample and assuming CRRA preferences, while the right panel does the same assuming CARA preferences. The first columns of each panel use non-labor per-capita income as an independent variable, while the second ones use total income per capita. Controls include: change in house ownership, age and schooling of household head, an indicator variable for female household head, share of children under six years old and average household age.

Table 3.3: ERS Test by Risk Group: CRRA

CRRA preferences

	Dependent variable: Change in *log* per capita expenditure					
	(1)	(2)	(3)	(4)	(5)	(6)
	Risk group 1 (less risk averse)	**Risk group 2**	**Risk group 3**	**Risk group 4**	**Risk group 5**	**Risk group 6 (more risk averse)**
Change in *log* mean per capita expenditure (ϕ_1)	7.994	2.560	3.512**	9.435**	-2.666	1.376
	[4.876]	[1.541]	[1.697]	[4.318]	[6.778]	[3.053]
Change in *log* per capita non-labor income (ζ_1)	0.0417	0.151**	0.0223	-0.0609*	-0.0431	-0.0492
	[0.0289]	[0.0640]	[0.0236]	[0.0317]	[0.360]	[0.0617]
N	311	46	324	69	15	82
R-squared	0.03	0.42	0.03	0.25	0.34	0.12
F-test ($\zeta_1 = 0$)	2.09	5.59	0.89	3.70	0.01	0.64
Prob > F	0.15	0.32	0.14	0.06	0.91	0.43
F-test ($\phi_1 = 1$)	2.06	1.03	2.19	3.82	0.29	0.02
Prob > F	0.15	0.02	0.35	0.06	0.61	0.90
Reject Efficient RS (95%)	**NO**	**YES**	**NO**	**NO**	**NO**	**NO**

Robust standard errors in brackets
Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age.
*** p<0.01, ** p<0.05, * p<0.10

CRRA preferences

	Dependent variable: Change in *log* per capita expenditure					
	(1)	(2)	(3)	(4)	(5)	(6)
	Risk group 1 (less risk averse)	**Risk group 2**	**Risk group 3**	**Risk group 4**	**Risk group 5**	**Risk group 6 (more risk averse)**
Change in *log* mean per capita expenditure (ϕ_1)	4.960*	1.627*	0.533	5.592***	-2.106	1.034
	[2.538]	[0.979]	[1.154]	[1.433]	[2.786]	[2.566]
Change in *log* per capita income (ζ_1)	0.0502***	0.159***	0.0610***	0.0478	0.00948	0.00742
	[0.0172]	[0.0500]	[0.0172]	[0.0382]	[0.0678]	[0.0404]
N	886	130	934	174	38	221
R-squared	0.03	0.42	0.03	0.25	0.34	0.12
F-test ($\zeta_1 = 0$)	8.51	10.08	12.59	1.57	0.02	0.03
Prob > F	0.00	0.00	0.69	0.00	0.27	0.99
F-test ($\phi_1 = 1$)	2.43	0.41	0.16	10.26	1.24	0.00
Prob > F	0.12	0.52	0.00	0.21	0.89	0.85
Reject Efficient RS (95%)	**YES**	**YES**	**YES**	**YES**	**NO**	**NO**

Robust standard errors in brackets
Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age.
*** p<0.01, ** p<0.05, * p<0.10

Note: This table presents the ERS test within each of the six risk groups, assuming CRRA preferences (variables are in logs not levels). The key is that the test only works if preferences are homogeneous among households, so it should provide the correct conclusion if implemented within a risk group. The upper panel uses non-labor per capita income as an independent variable, while the lower panel uses total income per capita. Controls include: change in house ownership, age and schooling of household head, an indicator variable for female household head, share of children under six years old and average household age.

Table 3.4: ERS Test by Risk Group: CARA

CARA preferences

	Dependent variable: Change in per capita expenditure					
	(1)	(2)	(3)	(4)	(5)	(6)
	Risk group 1 (less risk averse)	**Risk group 2**	**Risk group 3**	**Risk group 4**	**Risk group 5**	**Risk group 6 (more risk averse**
Change in group mean per capita expenditure (ϕ_1)	5.158	1.413	2.216*	2.856	3.558	1.892
	[3.358]	[1.091]	[1.250]	[3.452]	[5.192]	[1.576]
Change in per capita non-labor income (ζ_1)	0.496**	0.103	0.263*	0.0774	-0.0248	0.297*
	[0.232]	[0.286]	[0.134]	[0.186]	[0.982]	[0.166]
N	311	46	324	69	15	82
R-squared	0.12	0.28	0.06	0.17	0.34	0.32
F-test ($\zeta_1 = 0$)	4.57	0.13	3.83	0.17	0.00	3.20
Prob > F	0.033	0.722	0.051	0.593	0.637	0.078
F-test ($\phi_1 = 1$)	1.53	0.14	0.95	0.29	0.24	0.32
Prob > F	0.217	0.707	0.331	0.679	0.981	0.573
Reject Efficient RS (95%)	**YES**	**NO**	**NO**	**NO**	**NO**	**NO**

Robust standard errors in brackets
Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age. Also,
*** p<0.01, ** p<0.05, * p<0.10

CARA preferences

	Dependent variable: Change in per capita expenditure					
	(1)	(2)	(3)	(4)	(5)	(6)
	Risk group 1 (less risk averse)	**Risk group 2**	**Risk group 3**	**Risk group 4**	**Risk group 5**	**Risk group 6 (more risk averse**
Change in group mean per capita expenditure (ϕ_1)	1.128	0.824	0.329	4.631***	0.103	1.903
	[3.572]	[0.779]	[0.955]	[1.528]	[2.854]	[1.732]
Change in per capita income (ζ_1)	0.0445**	0.0556	0.0307*	0.0867***	-0.0497	-0.0996***
	[0.0203]	[0.0384]	[0.0180]	[0.0199]	[0.0859]	[0.0249]
N	886	130	934	174	38	221
R-squared	0.03	0.10	0.02	0.16	0.15	0.19
F-test ($\zeta_1 = 0$)	4.80	2.10	2.89	19.01	0.34	16.03
Prob > F	0.03	0.15	0.09	0.00	0.76	0.00
F-test ($\phi_1 = 1$)	0.00	0.05	0.49	5.65	0.10	0.27
Prob > F	0.97	0.82	0.48	0.02	0.57	0.60
Reject Efficient RS (95%)	**YES**	**NO**	**NO**	**YES**	**NO**	**YES**

Robust standard errors in brackets
Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age.
*** p<0.01, ** p<0.05, * p<0.10

Note: This table presents the ERS test within each of the six risk groups, assuming CARA preferences (variables are in levels not logs). The key is that the test only works if the preferences are homogeneous among households, so it should provide the correct conclusion if done within a risk group. The upper panel uses non-labor income per capita as an independent variable, while the lower panel uses total income per capita. Controls include: change in house ownership, age and schooling of household head, an indicator variable for female household head, share of children under six years old and average household age.

Table 3.5: Summary: ERS Test by Risk Groups

	Rejects ERS? (95%)							
	Pooled	Risk group 1 (risk lover)	Risk group 2	Risk group 3	Risk group 4	Risk group 5	Risk group 6 (risk averse)	Rejection rate (black only)
CRRA								
Non-labor income	YES	NO	YES	NO	NO	NO	NO	0/2
Income	YES	YES	YES	YES	YES	NO	NO	4/5
CARA								
Non-labor income	YES	YES	NO	NO	NO	NO	NO	1/2
Income	YES	YES	NO	NO	YES	NO	YES	3/5

Note: This table presents the summary of the ERS test within each of the six risk groups. The key is that the test only works if the preferences are homogeneous among households, so it should provide the correct conclusion if done within a risk group. The color red means that particular specification doesn't have a considerable amount of observations.

Table 3.6: Summary: ERS Test by Risk Sub-groups

	Rejects ERS? (95%)				
	Risk group 1		Risk group 3		Rejection rate (black only)
	Sub group 1	Sub group 2	Sub group 1	Sub group 2	
CRRA					
Non-labor income	NO	NO	YES	NO	1/2
Income	NO	YES	NO	YES	2/4
CARA					
Non-labor income	NO	NO	NO	YES	1/3
Income	NO	NO	NO	NO	0/3

Note: This table summarizes the results in Tables 3.8 and 3.9 of the Appendix. Risk groups that had enough observations (groups 1 and 3) were divided into sub-groups according to an additional risk attitude: smoking of the household head. Then, the efficient risk sharing was tested within this sub-groups. The color red means that particular specification doesn't have a considerable amount of observations.

Appendix C

Table 3.7: Summary Statistics of Demographic Variables

Year		Household size	House owner	Age of HH head	Female HH head	Schooling of HH head	Share of under six years	Average age	Urban (=1)
2002	Mean	3.50	0.72	47.82	0.20	6.55	0.46	32.24	0.56
	St. Deviation	1.50	0.45	15.53	0.40	3.96	0.72	16.50	0.50
	Median	0.83	0.00	15.00	0.00	0.00	0.00	4.00	0.00
	N	6,989	6,989	6,971	6,989	4,889	6,989	6,986	6,989
	Mean	3.83	0.78	50.44	0.22	6.81	0.41	34.06	0.59
	St. Deviation	1.63	0.41	15.05	0.42	4.03	0.69	16.61	0.49
2005	Median	0.90	0.00	19.00	0.00	0.00	0.00	7.40	0.00
	N	6,884	6,884	6,633	6,884	4,573	6,884	6,876	6,884

Note: Data from the MXFLS. Summary statistics of demographic variables.

Table 3.8: ERS Test by Risk Sub-groups: CRRA

CRRA preferences

	Dependent variable: Change in *log* per capita expenditure							
	Risk group 1				Risk group 3			
	Sub group 1 (1)	Sub group 2 (2)	Sub group 1 (3)	Sub group 2 (4)	Sub group 1 (5)	Sub group 2 (6)	Sub group 1 (7)	Sub group 2 (8)
Change in *log* mean per capita expenditure (ϕ_1)	1,911*	-29,315	1.413	-14.64	2,442**	-150.0	1.289	2.280
	[1,128]	[34,180]	[0.900]	[15.68]	[936.1]	[3,259]	[0.829]	[2.820]
Change in *log* per capita non-labor income (ζ_1)	91.37	107.2			16.81	33.31		
	[76.54]	[76.56]			[30.95]	[25.33]		
Change in *log* per capita income (ζ_1)			0.0404	0.0567**			0.0363	0.0892***
			[0.0255]	[0.0236]			[0.0256]	[0.0228]
N	163	148	429	457	161	163	444	490
R-squared	0.09	0.05	0.03	0.03	0.04	0.05	0.03	0.03
F-test ($\zeta_1 = 0$)	1.43	1.96	2.52	5.75	0.30	1.73	2.00	15.32
Prob > F	0.23	0.16	0.11	0.02	0.01	0.96	0.16	0.65
F-test ($\phi_1 = 1$)	2.87	0.74	0.21	1.00	6.80	0.00	0.12	0.21
Prob > F	0.09	0.39	0.65	0.32	0.59	0.19	0.73	0.00
Reject Efficient RS (95%)	**NO**	**NO**	**NO**	**YES**	**YES**	**NO**	**NO**	**YES**

Robust standard errors in brackets
Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age.
*** p<0.01, ** p<0.05, * p<0.10

Note: This table presents the ERS test within the risk sub-groups, assuming CRRA preferences (variables are in logs not in levels). The key is that the test only works if preferences are homogeneous among households, so it should provide the correct conclusion if done within a risk sub-group. Columns 1, 2, 5 and 6 use non-labor per capita income as an independent variable, while columns 3, 4, 7 and 8 use total income per capita.

Table 3.9: ERS Test by Risk Sub-groups: CARA

CARA preferences

	Dependent variable: Change in per capita expenditure							
	Risk group 1				Risk group 3			
	Sub group 1 (1)	Sub group 2 (2)	Sub group 1 (3)	Sub group 2 (4)	Sub group 1 (5)	Sub group 2 (6)	Sub group 1 (7)	Sub group 2 (8)
Change in group mean per capita expenditure (ϕ_1)	1.667	-19.61	0.834	5.598	2.145**	0.666	1.392**	3.036
	[1.043]	[30.30]	[0.660]	[26.38]	[0.910]	[2.888]	[0.625]	[2.542]
Change in per capita non-labor income (ζ_1)	0.498	0.615*			0.213	0.331**		
	[0.334]	[0.343]			[0.218]	[0.148]		
Change in per capita income (ζ_1)			0.0490*	0.0399			0.0200	0.0391
			[0.0280]	[0.0290]			[0.0185]	[0.0283]
N	163	148	429	457	161	163	444	490
R-squared	0.14	0.16	0.03	0.04	0.06	0.10	0.03	0.02
F-test ($\zeta_1 = 0$)	2.23	3.21	3.06	1.88	0.95	4.97	1.16	1.91
Prob > F	0.14	0.08	0.08	0.86	0.21	0.03	0.28	0.17
F-test ($\phi_1 = 1$)	0.41	0.46	0.06	0.03	1.58	0.01	0.40	0.64
Prob > F	0.52	0.50	0.80	0.17	0.33	0.91	0.53	0.42
Reject Efficient RS (95%)	**NO**	**NO**	**NO**	**NO**	**NO**	**YES**	**NO**	**NO**

Robust standard errors in brackets
Controls: Change in homeownership, age and schooling of HH head, female HH head, children under 6 yrs, average age.
*** p<0.01, ** p<0.05, * p<0.10

Note: This table presents the ERS test within the risk sub-groups, assuming CARA preferences (variables are in levels not in logs). The key is that the test only works if preferences are homogeneous among households, so it should provide the correct conclusion if done within a risk sub-group. Columns 1, 2, 5 and 6 use non-labor per capita income as an independent variable, while columns 3, 4, 7 and 8 use total income per capita.

3.9 References

Attanasio, Orazio and Miguel Szekely 2003. "Wage shocks and co movement variability in Mexico during the 1990s". *Journal of Development Economics*, Vol 73, pp. 1-25.

Attanasio, Orazio and Steven Davis 1997. "Relative WaSchulhofer-Wohlge Movements and the Distribution of Consumption". *The Journal of Political Economy*, Vol. 104, No. 6, pp. 1277-1262.

Barsky, Robert, et al. 1997. "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *The Quarterly Journal of Economics*, MIT Press, vol. 112(2), pages 537-79, May

Binswanger, Hans 1980. "Attitudes toward risk: experimental measurement in rural India", Washington, World Bank.

Cochrane, John 1991. "A Simple Test of Consumption Insurance". *The Journal of Political Economy*, Vol. 99, No. 5, pp957-976.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jurgen Schupp and Gert Wagner 2005. "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey". IZA Discussion papers.

Dubois, Pierre 2000. "Consumption Insurance with Preference Heterogeneity. May the Sharecropping Institution Help Completing Markets?" Manuscript.

Gugerty, Mary Kay 2007. "You Can't Save Alone: Commitment in Rotating Savings and Credit Associations in Kenya," *Economic Development and Cultural Change* 55 (2007): 251-282.

Grimard, F. 1997. "Household Consumption Smoothing through Ethnnic Ties: Evidence from Cote d'Ivoire", *Journal of Development Economics*, 53:391-422.

Hamoudi, Amar 2006. "Risk Preferences in Household and Families". Job Market Paper, UCLA, Department of Economics.

Hayashi, Fumio, Joseph Altonji and Laurence Kotikloff 1996a. "Risk Sharing between and within families". *Econometrica*, Vol 64, No. 2, pp 261-294.

Hayashi, Fumio, Joseph Altonji and Laurence Kotlikoff 1996b. "The Effects of Income and Wealth on Time and Money Transfers between Parents and Children," NBER Working Papers 5522, NBER.

Holt, Charles A., and Susan K. Laury 2002. "Risk Aversion and Incentive Effects". *American Economic Review*, 92(5): 1644–1655.

Jallan, J. and Martin Ravallion 1999. "Are the Poor Less Well Insured? Evidence on the Vulnerabilty to Income Risk in Rural China". *Journal of Development Economics*, 58, 61-81.

Mace, Barbara 1991. "Full Insurance in the Presence of Aggregate Uncertainty". *The Journal of Political Economy*, Vol. 99, No. 5, pp 928-956.

Mazzocco, Maurizio and Shiv Saini 2009. "Testing Efficiency Risk Sharing with Heterogeneous Preferences". Forthcoming, this version, March 2009.

Schulhofer-Whol, Sam 2010. "Heterogeneity, Risk Sharing and the Welfare Costs of Idiosyncratic Risk". Manuscipt.

Townsend, Robert 1994. "Risk and Insurance in Village India". *Econometrica*. Vol. 62 No. 3, pp 539-591.