Problem Set 1

ECONOMICS 172: Issues in African Economic Development

Oscar Chaix

02/20/19

1 - Comparing Demographic and Health Patterns across Regions

Over the past sixty years, the populations of East Asia, South Asia and Sub-Saharan Africa (from now on referred to as Africa for convenience) have significantly increased. In China alone, the population more than doubled since the 1960s from about 650 million to more than 1.4 billion persons, and in India the population similarly increased from less than 500 million to about 1.3 billion persons. While Africa's population is still well below these two countries, it has more than quadrupled in population from only about 250 million persons to more than a billion today. In contrast to the two other regions, Africa therefore has a convex population growth curve, which predicts that it will surpass both India and China's populations by 2035 and reach 2 billion persons by 2045. These predicted changes have numerous implications for the respective roles of these regions in the global economy in the coming decades: Africa will determine a significantly higher share of global demand and will have one of the highest producing population. Moreover, having the largest population in the world, Africa's economic performance will affect an even larger share of humanity and it will become increasingly detrimental that Africa succeeds in developing and lifting its population from poverty.

These population patterns, both in the recent past and in our predictions, are fundamentally linked to changes in fertility rates and life expectancy. In fact, whereas fertility rates have drastically fallen in India and China from around 6 children per woman in the beginning of the 1960s to 2.3 and 1.6 respectively today, they have only fallen from more than 6.5 to just below 5 over the same period in Africa. This explains why the populations of the three regions have increased so dramatically, and why South and East Asia are bound to be surpassed by Africa's population in the near future. Increases in life expectancy and decreases in infant mortality also reinforced these patterns, especially in the African case. For example, the significant decrease in infant mortality in Africa has been a central factor behind its population upsurge, as newborns have increasingly begun to survive diseases while fertility rates have remained high. Meanwhile in East and South Asia, better health care and hygiene levels increased the life expectancy of the population and exacerbated its ageing. The ageing of the population has in fact especially been important in East Asia as life expectancy rose to 74.7 years, compared to 69.3 in South Asia and only 60.7 in Africa. As an illustration of the central role of fertility and mortality rates, in China, the ageing of the population in the near future will largely echo the effects of the One Child Policy and better health care. According to Naughton, "It is estimated the number of Chinese over 65 years old will more than triple from 115 million in 2010 to around 350m in 2050. By 2030, the elderly will make up approximately 20% of China's population."{1}

Therefore, dependency ratios in East Asia and South Asia, which have until now been largely favourable for their economies, will start to deteriorate and increase. For example, China has just reached the end of an extraordinary trough between 2005 and 2015, during which its dependency ratio was more than 10 points below that of the middle-income average. Asian countries like China that largely benefitted from a large working population, very low dependency ratios and high savings rates will see their GDP per capita mechanically decrease with higher dependency ratios and a lower number of producers. In opposition, in Africa, dependency ratios which have until now been largely unfavourable could ameliorate in the near future. According to Bloom and Sachs, Africa's very young population (with a median age of 18 and approximately 0.79 young age person for each working age person) has significantly strained Africa's savings rates and consumption patterns, and as a result has limited both domestic and foreign investment and GDP growth. Similarly, in the Solow model of growth, the steady state capital intensity is expected to increase with more savings in the economy, but to decrease with higher population growth. If this exogenous model of growth holds for Africa, then its high dependency ratios not only limited its economic convergence by constraining capital accumulation, but also its steady state potential in the long run. However, thanks to Africa's huge young population, if fertility rates begin to fall, it could benefit from a golden age of a young and active population with a lower dependency burden and higher ability to spend and invest.

2 - Household Income and Child Health

a)

In any study that seeks to understand the effect of a certain economic policy (or "treatment"), it is essential to randomise who gets assigned to treatment and control groups to make sure that there is no difference between the two aside from the treatment. In other words, it is important to randomise so that we can interpret the results of the randomised controlled trial as underscoring the treatment effect, and not the treatment effect intertwined with other omitted variables (or confounders). Random assignment ensures that the mix of individuals being compared is the same, provided that the sample at hand is large enough for the law of large numbers to work.

In this situation, since we seek to understand the effect of household income on health outcomes, it was important that the cash transfer intervention was randomised across villages so that both those receiving and not receiving the transfer could be properly compared. If the transfer was made according to certain criteria, for example geographic location or village merit, it would have biased the final results as they would have been linked and correlated with other characteristics than the transfer (using our examples, to differences between regions, differences in village wealth or other differences between villages). Distinct persons react differently to cash transfers, and our goal here is to limit differences by having a large pool of randomly assigned villages.

All in all, randomising the cash transfers across villages allows us to deal with the omitted variable bias by giving us confidence, with statistical significance, that the omitted variables are not correlated with the explanatory variable of interest. If we call this indicator variable for cash transfer T and if we call X villager characteristic, we have:

(1) $Yi = a + bTi + dXi + ei$

With:

Y: health outcome (e.g. health measures with regards to nutrition, blood levels, sickness prevalence, etc.)

T: indicator variable (=0 or 1) for receiving the cash transfer

X: villager characteristic (reflecting socioeconomic status)

a, b, d: parameters / "coefficients" to be estimated

e: white noise "error/disturbance" term, $E(e) = 0$

i: denotes person "i" in the population, i from 1, …,N

This first equation models our regression of health outcome on household income, and we focus on the "treatment" effect of the cash transfer (measured by b) and on the effect of possible other villager characteristics (measured by d). If only the cash transfer dummy variable T is observed, and there is no data on household socioeconomic characteristics X, the treatment effect can be estimated as the average of Yi in the population that received the cash transfer minus the average of Yi in the population that did not receive the cash transfer. We have:

$$(2)\quad E(Yi \mid Ti = 1) - E(Yi \mid Ti = 0)$$

$$= [a + b + dE(Xi \mid Ti = 1) + E(ei \mid Ti=1)] - [a + 0 + dE(Xi \mid Ti = 0) + E(ei \mid Ti=0)]$$

$$= b + d[E(Xi \mid Ti = 1) - E(Xi \mid Ti = 0)]$$

In the third line, b corresponds to the true effect of the treatment and the rest of the expression, $d[E(Xi \mid Ti = 1) - E(Xi \mid Ti = 0)]$, corresponds to the omitted variable bias term. This last term indeed measures the difference in Xi between treatment and control group, and the overall effect on the estimated health outcome.

Since randomly assigned treatment and control groups come form the same underlying population, they should be the same in every way, including in their expected *Xi* (villager characteristics). In other words, the conditional expectations $E(Xi \mid Ti = 1)$ and $E(Xi \mid Ti = 0)$, are the same. This means that with randomisation equation *(2)* gives us:

$$
\begin{aligned}
&E(Yi \mid Ti = 1) - E(Yi \mid Ti = 0) \\
&\quad = b + d[0] \\
&\quad = b
\end{aligned}
$$

We have thus shown that in theory, removing differences in populations through random assignment effectively eliminates the possible omitted variable bias. The OVB term in fact is taken off the second equation, and the estimated treatment effect simplifies to *b*. The project's randomised design has helped us address omitted variable bias, and we can measure treatment effect with more precision and confidence.

b)

Average differences between the treatment households (cash=1) and control households (cash=0) for each of the characteristics can be analysed through the regression of these characteristics on the cash dummy variable, which distinguishes households that received the cash transfer and those that did not. If these regressions generate slopes coefficients close to 0, then they are shown to be uncorrelated and we can consider that there are no significant differences between both groups. We have:

(1) Regression of the female variable on the cash transfer

Gender Differences

	Dependent variable:
	female
cash	0.022
	(0.029)
Constant	0.676***
	(0.021)
Observations	1,000
R^2	0.001
Adjusted R^2	-0.0004
Residual Std. Error	0.464 (df = 998)
F Statistic	0.562 (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Our regression of the female variable on the cash variable gives an estimated coefficient of 0.022 with a standard error of 0.029 and t-value of 0.75. The constant coefficient, corresponding to how much female participation there is overall regardless of treatment or control, is equal to 0.676 (or 67.6%) with high statistical significance ($p<0.01$): there is therefore a majority of females in both groups. The slope coefficient is very small, which shows that being in the treatment or control group does not explain why a household respondent is a male or female. In fact, an increase in the cash variable of one unit (meaning that we only consider treated households) is only associated with an average increase in female proportion of 2.2% (since 1 is where all persons are female). In other words, the coefficient does not highlight any substantial difference between treatment and control groups in terms of the proportion of females both have.

Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.029. In fact, a 95% confidence interval around our parameter is (*ß* - 1.96 • SE , *ß* + 1.96 • SE) <=> (0.022 - 1.96*0.029 , 0.022 - 1.96*0.029) <=> (-0.035 , 0.079). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95% level. In other words, we are not 95% sure that the real coefficient *ß* is not 0. The t-statistic gives us the same information: since t = 0.75, we have t < 1.96, and we cannot reject the null with 95% confidence.

Thus, treatment and control groups have similar proportions of women and being in one group or another is not correlated with being a man or a woman.

(2) Regression of the age25 variable on the cash transfer

Age Differences

	Dependent variable:
	age25
cash	0.002
	(0.025)
Constant	0.812***
	(0.017)
Observations	1,000
R^2	0.00001
Adjusted R^2	-0.001
Residual Std. Error	0.390 (df = 998)
F Statistic	0.007 (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Our regression of the age25 variable on the cash variable gives an estimated coefficient of 0.002 with a standard error of 0.025 and t-value of 0.81. The constant coefficient, corresponding to how much age25 participation there is overall regardless of treatment or control, is equal to 0.812 (or 81.2%) with high statistical significance ($p<0.01$): there is at baseline therefore a lot of people older than 25 in both treatment and control groups. The slope coefficient is also very small, which shows that being in the treatment or control group does not explain why a household respondent is older or younger than 25 years old. In fact, an increase in the cash variable of one unit (meaning that we take only treated households) is only associated with an average increase of the proportion of people that are at least 25 of 0.2% (since 1 is where all persons are at least 25). In other words, the coefficient does not highlight any substantial difference between treatment and control groups in terms of the proportion of people at least 25 years old both have.

Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.025. In fact, a 95% confidence interval around our parameter is (-0.047, 0.051). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95% level. In other words, we are not 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since t = 0.081, we have t < 1.96, and we cannot reject the null with 95% confidence.

Thus, treatment and control groups have similar proportions of people that are at least 25 and being in one group or another is not correlated with being at least 25 years old.

(3) Regression of the schooling variable on the cash transfer

Education Differences

	Dependent variable:
	schooling
cash	-0.012
	(0.031)
Constant	0.378^{***}
	(0.022)
Observations	1,000
R^2	0.0002
Adjusted R^2	-0.001
Residual Std. Error	0.484 (df = 998)
F Statistic	0.154 (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

Our regression of the schooling variable on the cash variable gives an estimated coefficient of -0.012 with a standard error of 0.031 and t-value of -0.392. Again the constant is highly significant, and it shows that at baseline there are more people that did not complete a primary education than the contrary for both the treatment and control groups. The slope coefficient is also very small, which shows that being in the treatment or control group does not explain why a household respondent has had primary education or not. In fact, an increase in the cash variable of one unit (meaning that we only consider treated households) is only associated with an average decrease of the proportion of people that went to primary school of 1.2% (since 1 is where all persons have completed primary school). In other words, the coefficient does not highlight any substantial difference between treatment and control groups in terms of the proportion of people that went to primary school.

Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.031. In fact, a 95% confidence interval around our parameter is (-0.073, 0.049). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the

95% level. In other words, we are not 95% sure that the real coefficient *ß* is not 0. The t-statistic gives us the same information: since t = -0.392, we have t < 1.96, and we cannot reject the null with 95% confidence.

Thus, treatment and control groups have similar proportions of people that had primary education and being in one group or another is not correlated with having went to primary school.

<u>Differences in gender and age, with schooling restricted at 1</u>

When we restrict our attention to only those respondents who have completed primary school (schooling=1), we obtain the following regressions of gender and age on the cash transfer:

(4) Regression of the female variable on the cash transfer for primary-educated respondents

Gender differences when schooled

	Dependent variable:
	female
cash	0.073
	(0.051)
Constant	0.545***
	(0.036)
Observations	372
R^2	0.005
Adjusted R^2	0.003
Residual Std. Error	0.493 (df = 370)
F Statistic	2.008 (df = 1; 370)
Note:	*p<0.1; **p<0.05; ***p<0.01

Even though regressing the female variable on the cash variable while keeping schooling restricted at 1 does improve the estimated coefficient for the female parameter and its t-value, both are still insufficient. In fact, we have an estimated coefficient of 0.073 with a standard error of 0.051 and t-value of 1.417. The coefficient is still small, which shows that being in the treatment or control group does not explain why a household respondent is a male or female. In fact, an increase in the cash variable of one unit (meaning that we take only treated households) is only associated with an average increase in female proportion of 7.3% (since 1 is where all persons are female). In other words, the coefficient does not highlight any substantial difference between treatment and control groups in terms of the proportion of females both have. Interesting to note too, the constant slightly decreases when restricting on education, highlighting how fewer females in this population completed primary school for both treatment and control groups.

Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.051. In fact, a 95% confidence interval around our parameter is (-0.027 , 0.173). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95% level. In other words, we are not 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since t = 1.417, we have t < 1.96, and we cannot reject the null with 95% confidence.

Therefore, even though the correlation slightly increases in importance and precision, we can conclude that even when we restrict the population to primary-educated household respondents treatment and control groups also have similar proportions of women.

(5) Regression of the age25 variable on the cash transfer for primary-educated respondents

Age differences when shcooled

	Dependent variable:
	age25
cash	-0.029
	(0.045)
Constant	0.772^{***}
	(0.031)
Observations	372
R^2	0.001
Adjusted R^2	-0.002
Residual Std. Error	0.429 (df = 370)
F Statistic	0.434 (df = 1; 370)
Note:	*p<0.1; **p<0.05; ***p<0.01

Similarly, even though regressing the age variable on the cash transfer while keeping schooling restricted at 1 slightly increases the estimated coefficient for the age parameter, it is still low and statistically insignificant. In fact, we have an estimated coefficient of -0.029 with a standard error of 0.045 and t-value of -0.659. The coefficient is very small, which shows that being in the treatment or control group does not explain why a household respondent is at least 25 years old or not . In fact, an increase in the cash variable of one unit (meaning that we take only treated households) is only associated with an average decrease in this age level of -2.9% (since 1 is where all persons are female). In other words, the coefficient does not highlight any substantial difference between treatment and control groups in terms of the proportion of household respondents at least 25 years old both have. Interesting to note too, the constant again slightly decreases when restricting on education, highlighting how there are more educated people in the population that are below 25 years old and underscoring some progress in completion of primary education in rural Kenya.

Furthermore, the precision of this estimated parameter is statistically insignificant, as the standard error is equal to 0.045. In fact, a 95% confidence interval around our parameter is (-0.117 , 0.059). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95% level. In other words, we are not 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since $t = -0.659$, we have $t < 1.96$, and we cannot reject the null with 95% confidence.

Therefore, we can conclude that even when we restrict the population to primary-educated household respondents, treatment and control groups also have similar proportions of household respondents that are at least 25.

Taken together, the randomisation indeed appears to have succeeded in creating comparable cash transfer treatment and control groups at baseline. In fact, there are no consequent and statistically significant associations between the characteristics and being in the treatment or control group, even when we restrict on one characteristic such as primary education.

c) The average difference between treatment and control households schools in terms of their health status in the endline survey ("health") is as follows:

(6) Regression of the health variable on the cash transfer

Health on Cash Transfer

	Dependent variable:
	health
cash	0.024
	(0.064)
Constant	-0.001
	(0.045)
Observations	1,000
R^2	0.0001
Adjusted R^2	-0.001
Residual Std. Error	1.008 (df = 998)
F Statistic	0.142 (df = 1; 998)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see that the impact of receiving a large cash transfer approximately two years earlier on respondents' health status is small and statistically insignificant. Our regression of the health index on the cash variable in fact gives an estimated coefficient of 0.024 with a standard error of 0.064 and t-value of 0.377. This coefficient is very small, which shows that being in the treatment or control group does not explain why a household respondent is healthier or not after two years, at least according to the index. In fact, an increase in the cash variable of one unit (meaning that we observe only treated households) is only associated with an average health amelioration of 0.024 in the health index. In other words, the coefficient does not highlight any substantial effect of the cash transfer on household respondents' health.

Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.064. In fact, a 95% confidence interval around our parameter is (-0.101 , 0.150). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95%

level. In other words, we are not 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since t = 0.377, we have t < 1.96, and we cannot reject the null with 95% confidence.

Thus, the treatment and control groups have similar health outcomes two years after the cash transfer and being in one group or another is not correlated with being healthier. It seems that there is no impact of receiving a large cash transfer approximately two years earlier on the respondents' health status, as it is not significantly different from zero at 95% confidence.

The average difference between treatment and control households in terms of "health" when restricting attention to only those respondents who had completed primary school is as follows:

(7) Regression of the health variable on the cash transfer for primary-educated persons

Health on Cash Transfer (Education Restricted)

	Dependent variable:
	health
cash	0.049
	(0.100)
Constant	0.096
	(0.070)
Observations	372
R^2	0.001
Adjusted R^2	-0.002
Residual Std. Error	0.963 (df = 370)
F Statistic	0.241 (df = 1; 370)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see that the impact of receiving a large cash transfer approximately two years earlier on respondents' health status is slightly higher but still small and statistically insignificant when restricting schooling=1. Our regression of the health index on the cash variable in fact gives an estimated coefficient of 0.049 with a standard error of 0.100 and t-value of 0.490. This coefficient is very small, which shows that being in the treatment or control group does not explain why a primary-educated household respondent is healthier or not after two years, at least according to the index. In fact, an increase in the cash variable of one unit (meaning that we observe only treated households) is only associated with an average health amelioration of 0.049 in the health index. This is twice as much as in the preceding regression, where we did not focus only on primary-educated respondents, but it is still inconsequential. In other words, the coefficient does not highlight any substantial effect of the cash transfer on household respondents' health, even when looking only at primary-educated respondents.

Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.100. In fact, a 95% confidence interval around our parameter is (-0.147 , 0.245). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95% level. In other words, we are not 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since $t = 0.490$, we have $t < 1.96$, and we cannot reject the null with 95% confidence.

Thus, the treatment and control groups have similar health outcomes two years after the cash transfer and being in one group or another is not correlated with being healthier, even when restricting our attention only to primary-educated respondents. It seems that there is no impact of receiving a large cash transfer approximately two years earlier on the primary-educated respondents' health status, as it is not significantly different from zero at 95% confidence.

We would expect to find meaningful differences between the results of these two regressions, as better educated persons should know more how to better spend the cash transfer to ameliorate their health status. However, even though the estimated coefficient has approximately doubled, showing a stronger effect, we cannot affirm with confidence that there is any meaningful difference between the results of both regressions, since the results are not statistically significant and could be just due to chance (with the actual effects actually equal to zero).

Taken together, these results imply that the cash transfer actually did not ameliorate the health of the recipients within two years of the transfer, and not even the health of educated recipients, at least if we believe the health index to be a good measure of health. The causal impact of higher household income on health outcome cannot be established because the effects are very small and statistically insignificant. These results could reflect various phenomenas: maybe there is a transition period (or lag time) before respondents spend extra income on health care and/or before results manifest; maybe are health outcomes not easily ameliorated with more spending in health; or maybe do people very rarely spend extra-income on health and prefer consuming other goods and services. In any case, these regression results do not support a causal impact of income on health, and even less a positive one.

d) If we re-run the two regressions in part (c) but include "female" and "age25" as additional explanatory variables, we have:

(8) Regression of the health variable on the cash transfer, the female variable and the age25 variable

Health on Cash Transfer, Gender and Age

	Dependent variable:
	health
cash	0.030
	(0.063)
female	-0.215***
	(0.069)
age25	-0.379***
	(0.082)
Constant	0.452***
	(0.098)
Observations	1,000
R^2	0.027
Adjusted R^2	0.024
Residual Std. Error	0.996 (df = 996)
F Statistic	9.175*** (df = 3; 996)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see that the impact of receiving a large cash transfer approximately two years earlier on respondents' health status is still small and statistically insignificant when having the control variables of female and age25. Our regression in fact gives an estimated coefficient of 0.030 with a standard error of 0.063 and t-value of 0.468 for the cash variable. This coefficient is very small, which shows that being in the treatment or control group does not explain why a household respondent is healthier or not after two years, at least according to the index, even when having the gender and age controls. In fact, an increase in the cash variable of one unit (meaning that we observe only treated households) is only associated with an average health amelioration of 0.030 in

the health index. This is even less than in the preceding regression, where we did not add control variables. In other words, the coefficient does not highlight any substantial effect of the cash transfer on household respondents' health, even when adding the female and age25 controls.
Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.063. In fact, a 95% confidence interval around our parameter is (-0.093 , 0.153). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95% level. In other words, we are not 95% sure that the real coefficient $ß$ is not 0. The t-statistic gives us the same information: since t = 0.468, we have t < 1.96, and we cannot reject the null with 95% confidence.

Thus, the treatment and control groups have similar health outcomes two years after the cash transfer and being in one group or another is not correlated with being healthier, even when we add the controls of age and gender. It again seems that there is no impact of receiving a large cash transfer approximately two years earlier on the respondents' health status, as it is not significantly different from zero at 95% confidence even with control variables.

(9) Regression of the health variable on the cash transfer, the female variable and the age25 variable for primary-educated persons

Health on Cash Transfer, Gender and Age (Education Restricted)

	Dependent variable:
	health
cash	0.058
	(0.099)
female	-0.221^{**}
	(0.104)
age25	-0.249^{**}
	(0.120)
Constant	0.409^{***}
	(0.139)
Observations	372
R^2	0.019
Adjusted R^2	0.011
Residual Std. Error	0.957 (df = 368)
F Statistic	2.407^{*} (df = 3; 368)
Note:	*p<0.1; **p<0.05; ***p<0.01

Finally, the impact of receiving a large cash transfer approximately two years earlier on respondents' health status is again small and statistically insignificant when we have the control variables of female and age25 and when we focus our attention only on those respondents who had completed primary school. Our regression in fact gives an estimated coefficient of 0.058 with a standard error of 0.099 and t-value of 0.468 for the cash variable. This coefficient is the highest we found yet on the cash transfer's effect on the health outcome, but it is still small, which shows that being in the treatment or control group does not explain why a household respondent is healthier or not after two years, at least according to the index, even when having the gender and age controls and restricting schooling=1. In fact, an increase in the cash variable of one unit (meaning that we observe only treated households) is only associated with an average health amelioration of 0.058 in the health index. This is a bit more than in the preceding regression, but it does not highlight any substantial effect of the cash transfer on household respondents' health.

Furthermore, this estimated parameter is statistically insignificant, as the standard error is equal to 0.099. In fact, a 95% confidence interval around our parameter is (-0.136 , 0.252). Since the confidence interval includes zero, we cannot reject the null hypothesis of no correlation at the 95% level. In other words, we are not 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since $t = 0.468$, we have $t < 1.96$, and we cannot reject the null with 95% confidence.

Thus, the treatment and control groups have similar health outcomes two years after the cash transfer and being in one group or another is not correlated with being healthier, even when we add the controls of age and gender and restrict our attention to only those who had completed a primary education. It again seems that there is no impact of receiving a large cash transfer approximately two years earlier on the respondents' health status, as it is not significantly different from zero at 95% confidence even with control variables and the restriction.

The inclusion of these two covariates therefore do not change the conclusions in (c), which is not surprising since these control variables have already been regressed against the indicator variable "cash" in (b) and there has not been any statistically significant correlation noted. Therefore our coefficient estimates for the cash variable were expected to remain unchanged with the addition of these controls, and the effect of the cash transfer on health outcome to remain inconsequential and insignificant.

However, the two other variables of gender and age both have high and statistically significant estimated coefficients. In both regressions, their estimated coefficients are higher than 0.2 with statistical significance of 95% or 99%. These statistical significances are denoted by the stars that show p-values of less than 0.05 for two stars and less than 0.01 for three stars. These results are what we would expect since, intuitively, different genders and age levels should have different effects on one's health, especially in the context of a developing country. Moreover, the four coefficients are negative. This intuitively also makes sense for the age variable, since persons are usually less healthy as they get older, or at least they feel like they are less healthy. However, I would not have expected a negative coefficient for the female variable, because intuitively there are no reasons why a woman should be less healthy than a man. Probably, in the context of rural western Kenya, this could be because women generally make less income and/or are treated with less care by their families, which could provoke a negative effect on their health.

A final note is that the coefficients for the constant variables are high and have high statistical significances in the last two regressions. We can interpret these coefficients as underscoring a generally positive outcome of the health index survey for both treatment and control groups, before the negative female and age25 variables, as well as the statistically insignificant cash variable, are taken in account. All respondents on average feel positively healthy at a level of about 0.4.

Appendix

{1} “Chapter 8: Population.” *The Chinese Economy: Adaptation and Growth*, by Barry Naughton, MIT Press, 2018, pp. 203.