Problem Set 3

ECONOMICS 172: Issues in African Economic Development

Oscar Chaix

04/25/19

Drought, Economic Growth, and Armed Conflict

(a) Motivation, data, econometric approach and empirical findings of the MSS (2004) article

The article "Economic Shocks and Civil Conflict: An Instrumental Variables Approach", written by Edward Miguel, Shanker Satyanath and Ernest Sergenti and published in the *Journal of Political Economy* in 2004, principally seeks to estimate the impact of economic conditions on the likelihood of civil conflict. The authors view civil conflict as particularly relevant to study because, in the particular case of sub-Saharan Africa, 29 out of its 43 countries suffered from some form of civil conflict in the 1980s and 1990s, involving hundreds of thousands of persons displaced, very significant disease and disability burdens, and countless of lives lost. However, although a growing body of research has highlighted the strong association between economic conditions and civil conflict (for example, Sambanis 2001), according to the authors they did not adequately address the endogeneity issue in the relationship and hence did "not convincingly establish a causal relationship". Moreover, these studies have often overlooked possible omitted variables like government institutional quality that may have driven both economic outcomes and conflict, producing biased cross-country estimates.

The authors attempt to disentangle the causal link between the two and to address these issues through the use of rainfall variation as an exogenous instrumental variable. Ruling out of other channels through which rainfall may affect civil conflict, they estimate the impact of GDP per capita variations on civil conflict thanks to the 'natural event' and exogenous instrument of rainfall. This instrument is particularly relevant in the study of African development since sub-Saharan Africa is composed of largely unindustrialised agrarian economies without extensive irrigation systems. The authors hence find a strong and statistically significant correlation estimate btw weather shocks and income growth in sub-Saharan Africa (0.055*), proving the relevance of their first stage regression. The intuition behind the IV econometric method is that if rainfall has a large impact on economic growth but does not increase civil conflict risk, then it is unlikely that economic growth has a causal impact on civil war; in contrast, if the IV approach fulfils the three conditions of relevance, exogeneity and exclusion restriction and the instrument has a large impact on war risk, then it is likely that economic growth has a causal impact on war.

The authors use conflict and rainfall data from 41 sub-Sahara African countries in the time period 1981 to 1999. The ‘conflict data’ comes from a comprehensive new database of conflicts developed by the International Peace Research Institute of Oslo, Norway and the University of Uppsala, Sweden (the “PRIO/Uppsala database”). This database is more transparent in its construction than other like the Correlates of War (COW) database and records all conflicts with a threshold of 25 battle deaths per year in addition to classifying conflicts by standard 1000-death threshold, hence including more small conflicts in the analysis. It only captures politically motivated violence, with the limit, among other, of not accounting for the many important types of organized violence in Africa that do not directly involve the state (for example crimes related to drug trade in Lagos, Nigeria). The ‘rainfall data’ comes from the Global Precipitation Climatology Project (GPCP) database of monthly rainfall estimates, uniquely combining actual weather station rainfall gauge measures and satellite information on the density of cold cloud cover in 2.5 latitude and longitude degree intervals. Finally, other country characteristics mostly come from the Fear and Laitin (2003) paper and from World Bank databases, as well as the Soviet ethnographic index *Atlas Marodov Mira* (concerning the ethnolinguistic control variable) and the *CIA Factbook* (concerning the religious fractionalisation control).

The main empirical finding of the paper is that growth is strongly negatively related to civil conflict: a negative growth shock of five percentage points increases the likelihood of conflict by about one-half the following year. In fact, a drop in growth of 5% (a large recession) is associated with an increased civil conflict risk of 12%: since the mean civil conflict in dataset is of 27%, this means that a large economic shock should increase civil conflict to about 40%. This is moreover true across a range of regression specifications (including some country fixed effects), and the impact of growth shocks on conflict is *not* significantly different in richer, more democratic or more ethnically diverse countries.

(b) To investigate the impact of GDP growth on civil conflict, similarly to Miguel, Satyanath and Sergenti's approach, we use the NDVI measure as an instrument for GDP growth and control for year. We also denote time by *t* and country by *i*. Through an Instrumental Variable (IV) approach, we first regress GDP growth on the NDVI (the first stage regression), and then regress civil conflict on GDP growth (the second stage regression). The reduced form regression is hence the regression of civil conflict on the NDVI indicator, and the full causal chain can be simply described as follows:

NDVI (IV) —> GDP Growth (endogenous variable) —> Civil Conflict (outcome variable).

More formally, we have the following equations:

<u>Reduced form regression</u>

$$(\text{Civil Conflict})_{it} = a_0(\text{country})_{it} + b_0(\text{NDVI})_{it} + c_0X_{it} + d_0(\text{year})_{it} + e_{0,it}$$

<u>First stage regression</u>

$$(\text{GDP growth})_{it} = a_1(\text{country})_{it} + b_1(\text{NDVI})_{it} + c_1X_{it} + d_1(\text{year})_{it} + e_{1,it}$$

<u>Second stage regression</u>

$$(\text{Civil Conflict})_{it} = a_2(\text{country})_{it} + b_2(\text{GDP Growth})_{it} + c_2X_{it} + d_2(\text{year})_{it} + e_{2,it}$$

The intuition behind the IV method is that if the NDVI variable (reflecting rainfall) has a large impact on economic growth but does not increase civil conflict risk, then it is unlikely that economic growth has a causal impact on civil war; in contrast, if the IV approach fulfils the three conditions of relevance, exogeneity and exclusion restriction and the instrument has a large impact on civil conflict risk, then it is likely that economic growth has a causal impact on civil conflict. The three IV conditions will be further described and examined in this context in question (g).

<u>We can calculate the estimate of interest ß2 through the Two Stage Least Squares estimator:</u>

2SLS estimator (effect of GDP growth on civil conflict) = IV impact on war / IV impact on growth

= Estimated ß0 / Estimated ß1

<u>In fact, if we plug the first stage regression in the second stage regression we have:</u>

(1) $(\text{Civil Conflict})_{it} = a_2(\text{country})_{it} + b_2(\text{GDP Growth})_{it} + c_2X_{it} + d_2(\text{year})_{it} + e_{2,it}$

(2) $(\text{Civil Conflict})_{it} = a_2(\text{country})_{it} + b_2\{a_1(\text{country})_{it} + b_1(\text{NDVI})_{it} + c_1X_{it} + d_1(\text{year})_{it} + e_{1,it}\} + c_2X_{it} + d_2(\text{year})_{it} + e_{2,it}$

(3) $(\text{Civil Conflict})_{it} = (a_2 + b_2a_1)(\text{country})_{it} + \mathbf{(b_2b_1)(NDVI)_{it}} + (b_2c_1 + c_2)X_{it} + (b_2d_1 + d_2)(\text{year})_{it} + (b_2e_{1,it} + e_{2,it})$

<u>Since this last expression simply corresponds to the reduced form, its coefficient estimate (civil conflict on NDVI, ß0) is:</u>

ß0 = ß2 • ß1

<=> ß2 = ß0 / ß1

This recovers the 2SLS estimator, thereby showing how it should capture the effect of interest ß2 if IV conditions are met.

(c) We first proceed to the second stage regression (also called the second stage model in the MSS paper), corresponding to the regression of the armed conflict outcome (armed_conflict) on GDP growth (gdp_growth). We obtain the following regression results:

Second Stage Regression

	Dependent variable:
	armed_conflict
gdp_growth	-0.597**
	(0.273)
year	0.006
	(0.004)
Constant	-11.049
	(7.228)
Observations	646
R^2	0.010
Adjusted R^2	0.007
Residual Std. Error	0.447 (df = 643)
F Statistic	3.271** (df = 2; 643)
Note:	*p<0.1; **p<0.05; ***p<0.01

In the second stage model, it appears that there is a strong relationship between GDP growth and armed conflict. In fact, the slope coefficient is substantial, at about -0.6, which shows that GDP growth seems to reduce armed conflict, and vis versa that negative GDP growth should increase armed conflict. More specifically, our regression results indicate that a 1% increase in GDP growth, controlling for the year, decreases the predicted probability of armed conflict by around 0.6%. However, we cannot yet infer a causal relationship since we have not yet incorporated our instrumental variable and since we know that the relationship between GDP growth and conflict probably involves significant reverse causality (endogeneity) and omitted variable bias issues.

This estimated parameter is furthermore statistically significant, as the standard error is equal to 0.273. In fact, a 95% confidence interval around our parameter is (β - 1.96 • SE, β + 1.96 • SE) <=> (-0.597 - 1.96*0.273, -0.597 + 1.96*0.273) <=> (−1.132, −0.062). Since the confidence interval does not include zero, we can reject the null hypothesis of no correlation at the 95% level. In other words, we are 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since t = -2.187, we have |t| > 1.96, and we can reject the null with 95% confidence. Finally, the two stars associated to the estimated coefficient indicate that the p-value is lower than 0.05, again showing that we are 95% confident that the real coefficient β is not 0. Thus, in the second stage model, there is a substantial and statistically significant correlation between GDP growth and civil conflict.

The year covariate, being small and statistically insignificant, shows that the armed conflict indicator in our dataset is not significantly different with time. Even though the coefficient on year is slightly positive, its t-statistic is equal to $1.567 < 1.96$ and hence we cannot conclude that it is statistically significant at the 5% level. It is however useful as a control variable to capture the potential effect of time trends in our panel data and avoid omitted variable bias.

(d) We then proceed to the first stage regression (also corresponding to the first stage model in the MSS paper), which is the regression of GDP growth (gdp_growth) on the NDVI (green_index_growth). We obtain the following regression results:

First Stage Regression

	Dependent variable:
	gdp_growth
green_index_growth	0.108^{***}
	(0.027)
year	0.001^{***}
	(0.001)
Constant	-2.868^{***}
	(1.026)
Observations	646
R^2	0.037
Adjusted R^2	0.034
Residual Std. Error	0.064 (df = 643)
F Statistic	12.349^{***} (df = 2; 643)
Note:	*p<0.1; **p<0.05; ***p<0.01

In the first stage regression, it appears that there is a strong positive relationship between GDP growth and the green index growth, which reflects rainfall variation. In fact, the slope coefficient is substantial, at about 0.1, which shows that more rainfall seems to increase GDP growth, and vis versa that less rainfall (and a corresponding lower green index) should decrease GDP growth.

Furthermore, this estimated parameter is highly statistically significant, as the standard error is equal to 0.027. In fact, a 95% confidence interval around our parameter is (0.055, 0.161). Since the confidence interval does not include zero, we can reject the null hypothesis of no correlation at the 95% level. In other words, we are 95% sure that the real coefficient β is not 0. The t-statistic gives us the same information: since t = 4.008, we have $|t| > 1.96$, and we can reject the null with 95% confidence. Finally, the three stars associated to the estimated coefficient indicate that the p-value is in fact lower than 0.01, showing that we are as much as 99% confident that the real coefficient β is not 0. Thus, in the first stage regression, there is a substantial and statistically significant correlation between the green index growth and GDP growth, proving the relevance of our instrument and hence meeting the first IV condition.

The year covariate, being small, shows that the GDP growth in our dataset is not substantially impacted by certain years, and hence that the relationship we witness is largely indifferent to time in our period of study. It is however statistically significant ($t = 2.79 > 1.96$), which makes sense since some years should involve regional-wide declining GDP growth in times of economic shocks.

(e) Third, we proceed to the reduced form regression (also corresponding to the reduced form model in the MSS paper), which is the regression of the armed conflict outcome (armed_conflict) on the NVDI (green_index_growth). We obtain the following regression results:

Reduced Form Regression

	Dependent variable:
	armed_conflict
green_index_growth	-0.332*
	(0.189)
year	0.005
	(0.004)
Constant	-9.705
	(7.195)
Observations	646
R^2	0.007
Adjusted R^2	0.004
Residual Std. Error	0.448 (df = 643)
F Statistic	2.413* (df = 2; 643)
Note:	*p<0.1; **p<0.05; ***p<0.01

In the reduced form regression, it appears that there is a strong relationship between green index growth (rainfall variation) and armed conflict. In fact, the slope coefficient is substantial, at about -0.3, which shows that more rainfall seems to decrease armed conflict, and vis versa that less rainfall (and a corresponding lower green index) should increase armed conflict. In other words, a 1 unit increase in the NDVI, controlling for the year, decreases the probability of armed conflict on average by 0.33%.

Furthermore, this estimated parameter is statistically significant at the 10% level, as the standard error is equal to 0.189. In fact, a 90% confidence interval around our parameter is (β - 1.645 • SE, β + 1.645 • SE) <=> (-0.332 - 1.645*0.189, -0.332 + 1.645*0.189) <=> (−0.643, −0.021). Since the confidence interval does not include zero, we can reject the null hypothesis of no correlation at the 90% level. In other words, we are 90% sure that the real coefficient β is not 0. The t-statistic gives us the same information, however if we test the significance at the 5% level, since t

= - 1.752, we have |t| < 1.96, and we cannot reject the null with 95% confidence. Finally, the star associated to the estimated coefficient indicates that the p-value is lower than 0.1, showing that we are 90% confident that the real coefficient *ß* is not 0. Thus, in the reduced form regression, there is a substantial and statistically significant (at the 10% level) correlation between the green index growth and armed conflict, proving the effect of our instrument on the outcome and thus further legitimising it as a good instrument.

The year covariate, being small and statistically insignificant, again shows that the armed conflict indicator in our dataset is not significantly impacted by certain years, and hence that the relationship we witness is largely indifferent to time in our period of study. In other words, we cannot infer from this regression any substantial differences in armed conflict throughout the years for countries included, something we already observed in (c).

(f) We finally carry out the estimation of the instrumental variables model, in which the NDVI measure serves as an instrument for GDP growth. Regressing armed conflict (armed_conflict) on GDP growth (gdp_growth), using the NDVI as an instrument, we obtain the following results:

IV Regression

	Dependent variable:
	armed_conflict
gdp_growth	-3.067*
	(1.856)
year	0.009**
	(0.005)
Constant	-18.501*
	(9.460)
Observations	646
R^2	-0.116
Adjusted R^2	-0.120
Residual Std. Error	0.475 (df = 643)
Note:	*p<0.1; **p<0.05; ***p<0.01

In this regression form, it appears that there is a very strong causal relationship between armed conflict and GDP growth. In fact, the slope coefficient is substantial, at about -3, which shows that GDP growth is highly negatively correlated to armed conflict. In other words, higher GDP growth should decrease armed conflict, while negative GDP growth should largely increase armed conflict. To be more precise, a one-percentage-point decline in GDP growth increases the likelihood of an armed conflict by more about three percentage points. Thus, a five-percentage-point decline in GDP growth (a large recession) leads to a greater than 15-percentage-point increase in the incidence of armed conflict. This is an increase of more than one half the average likelihood of conflict (27.7% in the dataset, from our summary statistics computation), which would lead to an increase in armed conflict prevalence reaching about 43% of our dataset if it was generalised over all countries for every years included. This is an enormous causal effect that proves to be even higher than the one found in the MSS paper. Based on this IV analysis, the implied causal effect of economic growth on armed conflict is therefore very strong and negative.

Furthermore, this estimated parameter is statistically significant at the 10% level, as the standard error is equal to 1.856. In fact, a 90% confidence interval around our parameter is (−6.120, −0.014). Since the confidence interval does not include zero, we can reject the null hypothesis of no correlation at the 90% level. In other words, we are 90% sure that the real coefficient *ß* is not 0. The t-statistic gives us the same information, however if we test the significance at the 5% level, since t = - 1.652 , we have $|t| < 1.96$, and we cannot reject the null with 95% confidence. Finally, the star associated to the estimated coefficient indicates that the p-value is lower than 0.1, showing that we are 90% confident that the real coefficient *ß* is not 0. Thus, in the IV regression, there is a substantial and statistically significant (at the 10% level) correlation between GDP growth and armed conflict.

The year covariate is still relatively small but statistically significant, highlighting that the armed conflict indicator in our dataset is more significantly impacted by certain years when we run the IV method. However, we cannot infer from this regression any substantial differences in armed conflict throughout the years for countries included, since the difference is only of 0.009.

To check the accuracy of our estimate ß2, we calculate the 2SLS estimator as explained in (b) and we have:

|2SLS estimator| (effect of GDP growth on civil conflict) = IV impact on war / IV impact on growth

= | Estimated ß0 | / | Estimated ß1 |

= 0.332 / 0.108

≈ 3.07

We get around the same estimate as in our regression output, which confirms our above results.

(g) The three IV conditions, as laid out in Lectures 18 and 19 and by Angrist and Pischke, are as follows:

(1) The relevance condition: the instrument needs to have a causal effect on the variable whose effects we are trying to capture; in other words, there needs to be a strong and statistically significant estimate behind our instrumental variable in our first stage regression.

(2) The exogeneity condition: the instrument needs to be randomly assigned or a 'natural event', in the sense that it is not correlated with any omitted variable (an unobserved X factor influencing the instrument). More specifically, there should not be any 'feedback' or endogeneity from the outcome variable to the instrument and no covariation with an omitted variable.

(3) The exclusion restriction condition: there is only a single channel through which the instrument affects the outcome: in our situation, rainfall variation only affects armed conflict acting through GDP growth, our 'endogenous variable'.

Mathematically, the first condition is met when we have *corr*(Zi, Xi =) $\neq$ 0, with Zi corresponding to the instrumental variable and Xi to the endogenous variable. The second and third conditions are met if and only if *corr*(Zi, ε) = 0.

In our analysis, the relevance condition seems to be met, as we saw in (d) a high correlation between the instrument, the NDVI growth measure, and the endogenous variable, the GDP growth measure. Intuitively, since sub-Saharan Africa is composed of largely unindustrialised agrarian economies without extensive irrigation systems, rainfall variation should have important effects on production and GDP levels. We hence observe a correlation of -0.108 that is highly statistically significant ($p < 0.01$), which confirms that our framework meets the relevance condition: *corr*(Zi, Xi =) $\neq$ 0. This assumption might fail in this context if, for example, the NDVI growth measure is a bad indicator of recent rainfall outcomes and/or land greenness has nothing to do with rainfall. In that case, we would either see no significant correlation or a biased correlation due to other omitted factors (for example, irrigation levels).

The exogeneity condition is also likely to hold here, because rainfall can be considered as a considerably random and natural event and it should not be influenced by any unobserved variable. Intuitively, there should also not be any feedback or endogeneity issue because the weather should not be affected by conflicts on the ground and is largely independent of human control. Again, this assumption might fail in this context if the NDVI growth measure is a bad indicator of recent rainfall outcomes or if it is biased in some way by an omitted variable. For example, if land greenness is not only affected by rainfall but is also related to land care or something else that is not random or natural, this may bias the results. Here, we might think about irrigation as a potentially omitted variable that can bias land greenness upward with higher use. Not only it is an unnatural and non-random variable, varying largely with the development level and wealth of specific countries, it is a major agricultural technology widely used in developed economies. Hopefully, as the MSS paper points out, only 1% of cropland is irrigated in the median African country according to the World Development Indicator database, and therefore in our analysis irrigation should not play a significant role and should not bias our results.

However, the exclusion restriction condition does not seem plausible in this framework and context. In fact, variation in precipitation and correspondingly the NDVI growth measure is likely to affect other factors, beyond economic growth, that might impact the risk of violence and armed conflict. In other words, it is likely that other channels exist that link our instrumental variable, land greenness, to our outcome variable, armed conflict. For example, numerous studies have shown that different weather and temperatures have important psychological impacts on persons and increase the level of violence, even in developed countries and cities (see for example Tiihonen et al, 2017). Research on weather's effects on violence explores numerous channels, from the serotonin hormonal channel to higher police brutality. Hsiang, Burke and Miguel's paper "Quantifying the Influence of Climate on Human Conflict", published in 2013 in *Science* in fact shows that an astounding 27 out of 27 studies find that higher temperatures lead to more violence. Another significant channel, especially in the context of sub-Saharan Africa, is the impact of rainfall and weather on road quality and infrastructure. In times of drought or flooding, transportation and commerce are likely to be affected in sub-Saharan countries, maybe leading to different levels of interaction between ethnic groups and possible exacerbation of tensions.

In conclusion, this a relatively valid instrumental variable approach but the effect of economic growth on armed conflict should be interpreted with care, as it is probably overestimated because of other channels linking the NDVI growth measure and armed conflict. In other words, there is probably an upward bias on ß2 since the exclusion restriction condition is likely not to be met and since other psychological and infrastructure channels might also increase armed conflict occurrence in times of extreme rainfall patterns. Nevertheless, our approach does provide an interesting framework to analyse the effect of growth on conflict, and even though there likely is an upward bias, the effect is large enough that we can presume the instrumental variable to have an effect on the outcome also through the endogenous variable. Taken naively and isolated, the estimate we find is not perfect, but it does highlight a substantial and statistically significant effect of GDP growth on armed conflict.

In my opinion, taken for their accuracy, the estimates that are more informative are therefore those from part (e), because they should be less biased than those from (f). In other words, estimates of the reduced form regression of conflict on the NDVI are more informative and less biased than estimates from our IV approach linking conflict to GDP growth, which are probably biased because of the potential failure to meet the exclusion restriction condition. Estimates of the reduced form regression, on the other hand, should not be biased if the NDVI is truly random and reflects the natural variation of rainfall growth, hence delivering the true effect of rainfall variation on armed conflict. They are furthermore interesting and particularly relevant to examine, since temperature and extreme weather events are increasing and expected to further increase in the future with global warming, and since they are particularly going to affect sub-Saharan Africa. Climate change is likely to have serious implications for African economic development and political stability, and thus the link between adverse weather and human violence should be understood so that local policies can be devised to counter its effects.

R-History

```
#PB Set 3
data=read.csv("~/Desktop/Econ172_S19_ProblemSet3_data.csv")
install.packages("stargazer")
install.packages("ivpack")
install.packages("dplyr")
library(stargazer)
library(dplyr)
library(ivpack)
summary(data)
#c
reg1 = lm(armed_conflict ~ gdp_growth + year, data)
reg1
summary(reg1)
stargazer(reg1,out="Table 1.html",type="html",
title="Second Stage Regression")
#d
reg2 = lm(gdp_growth ~ green_index_growth + year, data)
reg2
summary(reg2)
stargazer(reg2,out="Table 2.html",type="html",
title="First Stage Regression")
#e
reg3 = lm(armed_conflict ~ green_index_growth + year, data)
reg3
summary(reg3)
stargazer(reg3,out="Table 3.html",type="html",
title="Reduced Form Regression")
#f
reg4 = ivreg(armed_conflict ~ gdp_growth + year | green_index_growth +
year, x=TRUE, data=data)
reg4
summary(reg4)
stargazer(reg4,out="Table 4.html",type="html",
title="IV Regression")
timestamp(stamp=date())
##------ Thu Apr 25 12:47:10 2019 ------##
savehistory(file="PB_Set_3_Oscar_CHAIX.Rhistory")
```

Comments on code

We first look at the summary statistics of our dataset using the command “>summary(data)”, this will be useful later on for interpreting our coefficient estimates. For the first three regressions (corresponding to question c., d. and e.), we use the function lm(), and always include the year variable as a control. We place the dependent variable before the ~ sign, and the variable whose effect we’re after following that sign. Finally, we use the function ivreg() to run the full IV regression, with armed_conflict: the outcome of interest; gdp_growth: the endogenous regressor; green_index_growth: the instrument; year: another covariate. The “x=TRUE” serves to take into account in the second stage that the first stage produces an estimate, adjusting standard errors to be correct.