EC3320

2016-2017

Michael Spagat

# Lecture 16. Climate and Conflict 2

Couttenier and Soubeyran offer a new approach to climate and conflict by using a drought index as an explanatory variable. The index has two important advantages:

1. The value of the index at any particular moment in time reflects conditions for some time prior to this moment. Thus, a year with little rain that follows several years with plenty of rain will not, properly, get classified as an extremely dry year.

2. The drought index data are available going all the way back to 1945.

Once again the technique is cross-country regression - the dependent variable is civil war incidence.

The authors exclude anti-colonial wars on the grounds that the dynamics of local peoples' driving away colonial occupiers are unlikely to be connected with climate change.

For similar reasons there are also some specifications that exclude what are called "internationalized" wars, i.e., wars that involve some outside actor - again, the argument is that such outside actors may not be driven very much by local climate conditions.

The following table is the main one for the paper – it shows a fairly consistent and positive relationship between drought and internal armed conflict.

## Table 2: Drought and Civil War Incidence: After independence

| Specifications<br>War | (1)<br>Internal & Internationalized | (2)<br>Internationalized | (3)<br>Internal | (4)<br>Internal | (5)<br>Internal | (6)<br>Internal |
|---|---|---|---|---|---|---|
| Drought (PDSI) | 0.719*** | -0.0328 | 0.669*** | 0.352* | 0.460* | 0.371* |
| | (0.264) | (0.138) | (0.246) | (0.188) | (0.254) | (0.223) |
| Lag Incidence | | | | 0.560*** | 0.568*** | |
| | | | | (0.0584) | (0.0216) | |
| Country Fixed Effect: | Yes | Yes | Yes | Yes | Yes | Yes |
| Country Time Trends : | Yes | Yes | Yes | Yes | Yes | - |
| Common Time Trends: | - | - | - | - | - | Yes |
| Observations | 1,597 | 1,597 | 1,597 | 1,594 | 1,594 | 1,597 |
| R-squared | 0.445 | 0.433 | 0.321 | 0.533 | - | 0.278 |

Note: Robust Standard errors in parentheses with ***, ** and * respectively denoting significance at the 1%, 5% and 10% levels. Column (1) to (4) and (6) using OLS. The method of estimation is system-GMM for column (5).

## El Niño/Southern Oscillation (ENSO)

We need to orient ourselves a bit before we can really get going on the Hsiang et al. paper.

First of all we need to know what the El Niño/Southern Oscillation (ENSO) weather phenomenon is – it is a weather pattern that affects the continental tropics.

In this zone there is alternation between periods that are warm and dry, which are known as El Niño periods, and periods that are cold (relatively cold that is: remember this is the tropics) and wet, which are known as La Niña periods.

I can think of at least three reasons to be interested in this weather pattern.

1. There is the general issue of whether global warming causes conflict.

    a. Of course, this is an important question in view of the considerable evidence that global warming is happening.

    b. The ENSO pattern provides us with a quasi experiment that helps us to explore this question. Specifically, we get to observe countries during both years when it is hot and dry (El Niño years) and other years when it is cold and wet (La Niña years). If we find that there tends to be more conflict during the hot, dry years than during the cold, wet years then this constitutes evidence suggesting that global warming may be associated with conflict.

2. Another advantage of looking at ENSO and conflict is that some parts of the world are strongly affected by ENSO whereas other parts of the world are only weakly affected or not affected at all by ENSO.

This regional variation provides us with a second quasi experimental dimension. In particular, we can investigate whether ENSO-affected areas behave differently from non-ENSO-affected areas.

If so, we can have some confidence that we are really observing effects attributable to ENSO.

3. The question of whether El Niño years are associated with increased conflict relative to La Niña ones is an interesting one in its own right.

   Such an association would give us a handle on predicting, and perhaps even preventing, conflict since these shifting weather patterns are predictable to some degree.

There is an issue of how to track the ENSO pattern so that we know when we are in an El Niño period and when we are in a La Niña period.

Hsiang et al. use several measures but the main one they use is called the Niño 3 Index.

This is the average sea surface temperature in the grey area shown in panel a of the next slide which is a part of the Pacific Ocean to the West of South America.

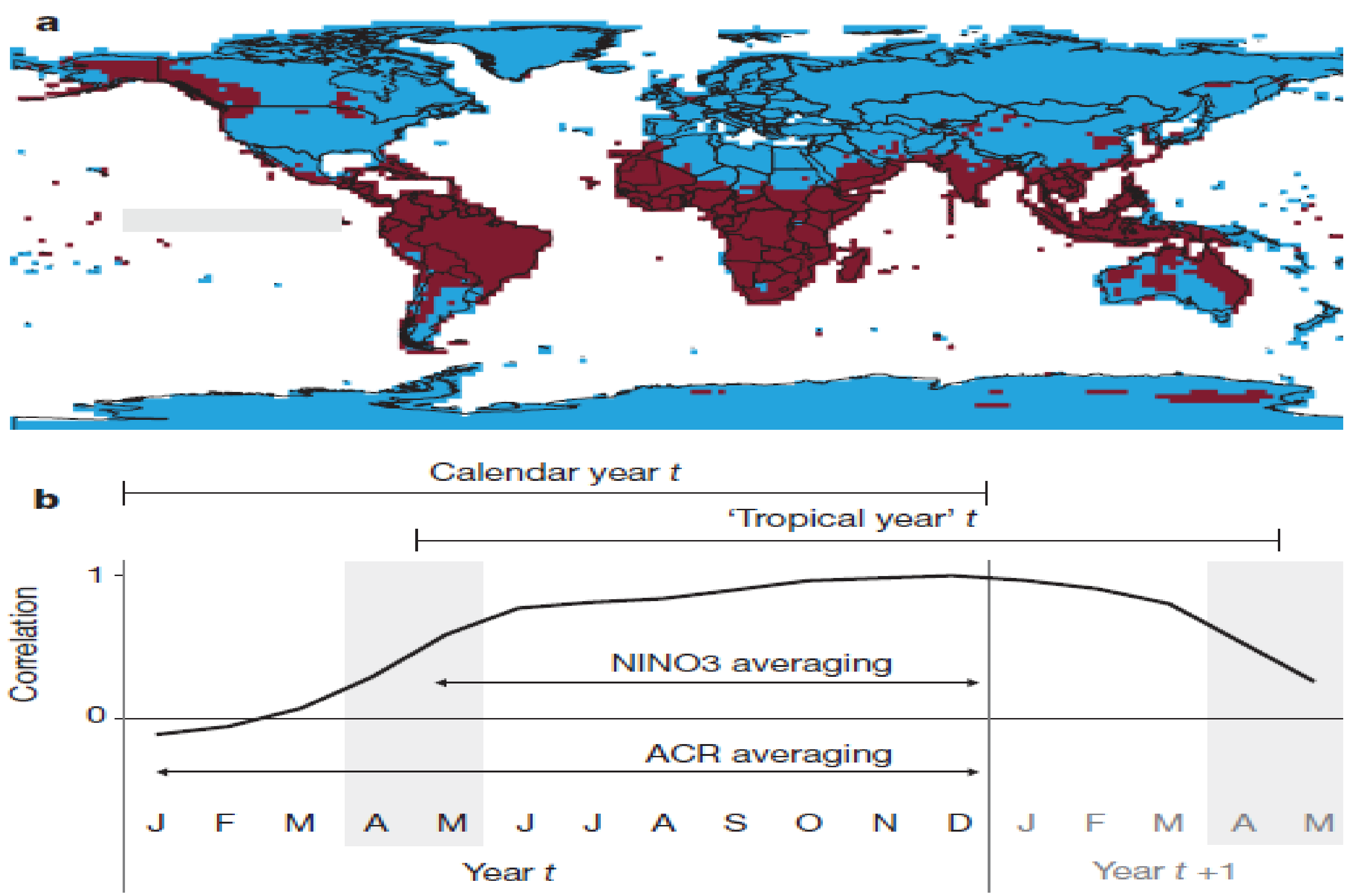



**Figure 1 | ENSO exposure over space and time.** **a,** Red (blue) indicates an ENSO teleconnected (weakly affected) pixel; NINO3 region in grey. **b,** Correlation of monthly NINO3 with NINO3 in December. The natural 'tropical year' begins in May and ends the following April at the 'spring barrier' (grey). To match monthly ENSO data with annual ACR data, an annual ENSO signal is isolated by averaging May–December NINO3.

The red areas in panel a (slide 9) are classified as “ENSO-affected areas” (also known as the “teleconnected zone”) while the blue areas are classified as “weakly affected”.

There will, of course, be some judgment in such a 0-1 type classification but we will just accept these judgments.

The point of panel b is to convince us that we should ignore January-April in measuring temperatures and classifying years into El Niño and La Niña ones – this early part of the year is known as the “Spring Barrier”.

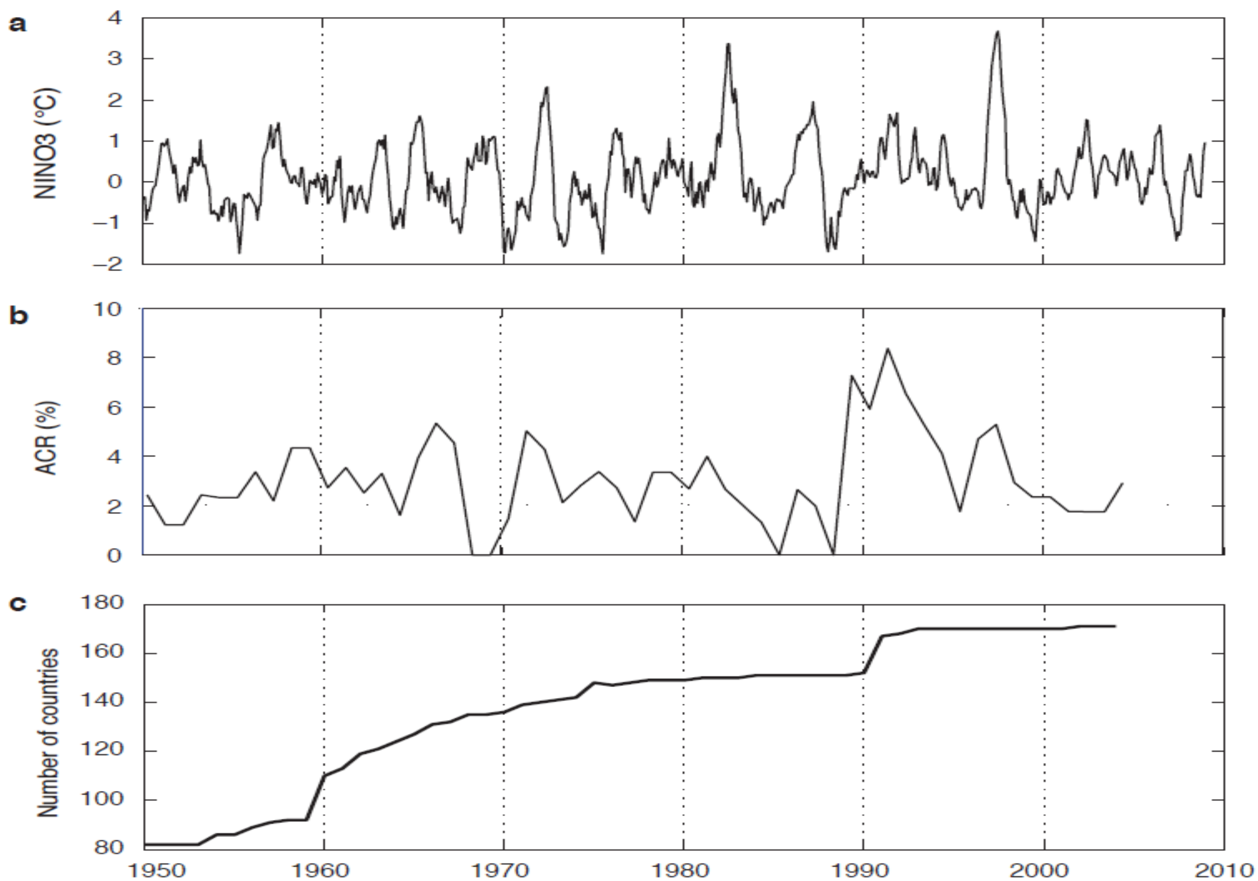
The ENSO phenomenon only really gets going in May and the weather in the early part of the year is only weakly correlated with the weather in the later part of the year.

Finally, one last piece of terminology is what Hsiang et al. refer to as “Annual Conflict Risk”, abbreviated as “ACR”.

Hsiang et al. define the ACR for a group of countries in a particular year as the “probability” that a country in that group will experience the onset of conflict during that year.

Perhaps I am being pedantic but I would prefer to say that the ACR for a group of countries in a given year is the fraction of those countries that experience conflict onset in that year and that this fraction is taken as an estimate of the conflict onset probability for that group.

The following pictures display what the key data streams look like.




Supplementary Figure 1: (a) Monthly sea surface temperature anomaly over the NINO3 region (5°S-5°N, 150°W-90°W). (b) ACR for the whole world. (c) The number of countries in the dataset is growing over time, more than doubling over the period of observation. Because *conflict onset* is coded as a binary variable for each country-year observation, it is necessary to normalize the total number of observed *conflict onsets* by the number of distinct countries being observed. To check that this trend is not driving any of our results, we re-estimate the model while restricting the sample to the period following 1975 (inclusive) when the sample of teleconnected countries was stable; we also estimate a model with the raw panel data using country-specific constants in the regression (country fixed-effects).

Panels a and b show that both the NINO3 index and the ACR vary considerably over time.

Panel c shows that the number of countries in the world has grown sharply over time.

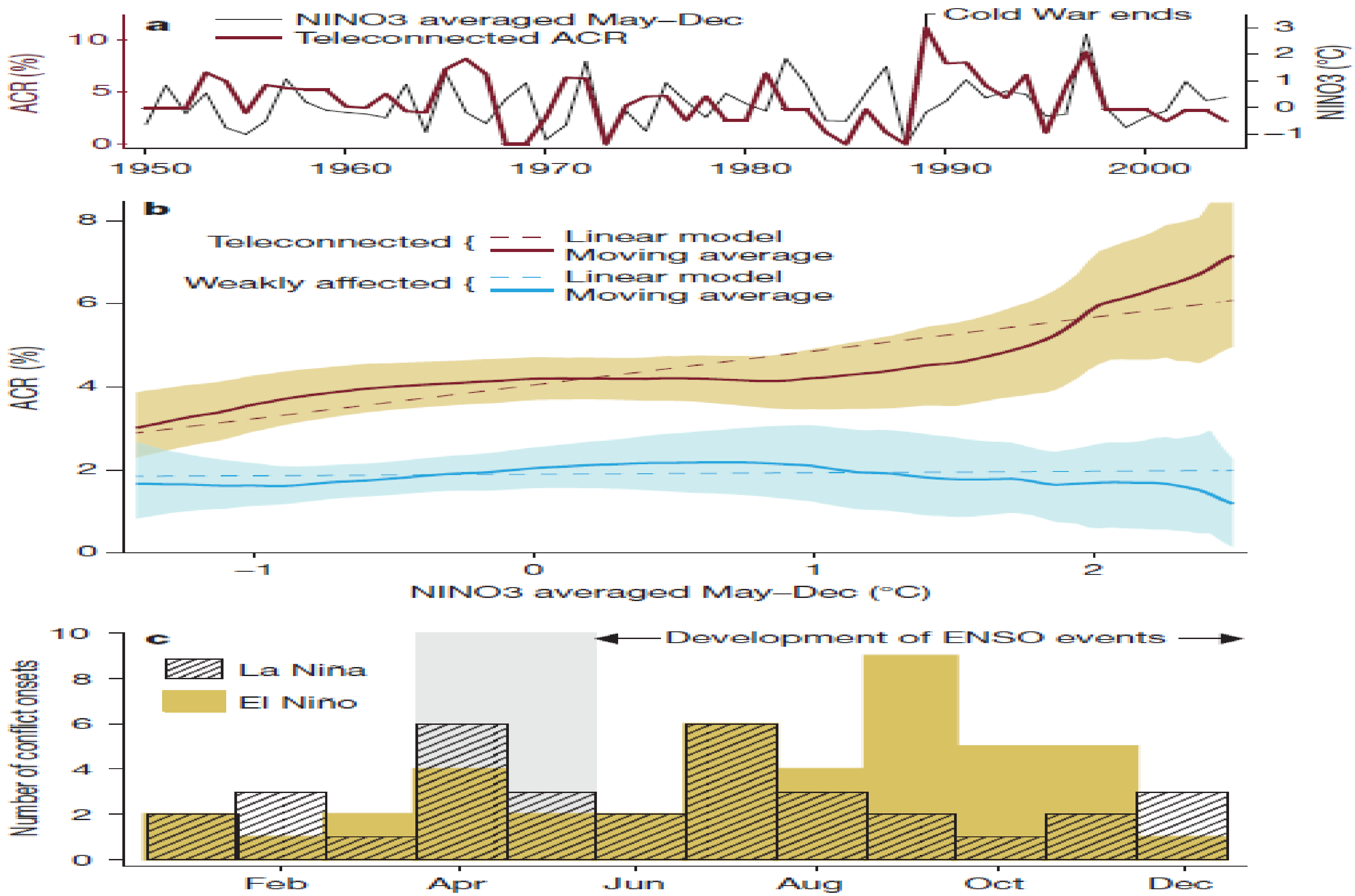



**Figure 2 | Conflict risk associated with ENSO.** a, Time series of NINO3 and ACR for the teleconnected group. b, Linear and non-parametric fit ($n = 54$, weighted moving average, 90% confidence intervals shaded) of ACR against NINO3. Time trends and mean shift after the end of the Cold War are removed. c, Solid (hatched) bars show total monthly conflict onsets in teleconnected countries during one-third of years most El-Niño-like (La-Niña-like). Monthly data are available for only half of the conflicts.

Panel a of slide 14 shows that the Annual Conflict Risk for the area (i.e., the teleconnected zone) tracks the NINO3 Index pretty well.

Panel b shows that for the ENSO-affected group, the ACR ranges between 3% in the (relatively cool and wet) La Niña years to 6% in (the relatively warm and dry) El Niño years.

For the weakly affected countries the ACR always stays around 2% regardless of ENSO.

Hsiang et al. calculate that this excess conflict risk of 6% - 3% = 3% for El Niño compared La Niña periods translates into 48.2 extra conflicts which amounts to about 21% of all conflicts.

Panel c shows that this premium of El Niño conflicts over La Niña ones occurs only toward the later part of the year when the El Niño phenomenon is actually operating.

This is an important reality check gives us some confidence that the findings are not just random accident.

If it had turned out, instead, that the ENSO-affected (teleconnected) zone only behaves differently from the weakly affected zone during the time of year when the ENSO phenomenon is not actually operating (January-April) then Hsiang et al.'s results would have looked suspiciously like some kind of statistical anomaly.

**Table 1 | Regression of ACR on NINO3 averaged May–December 1950–2004**

| Model | Teleconnected (% °C$^{-1}$) | Weakly affected (% °C$^{-1}$) |
|---|---|---|
| (1) Group aggregate | 0.76* | 0.16 |
| | (0.39) | (0.31) |
| | $n = 54$ | $n = 54$ |
| (2) Group aggregate | 0.85** | 0.06 |
| Linear trend | (0.40) | (0.30) |
| | $n = 54$ | $n = 54$ |
| (3) Group aggregate | 0.81** | 0.04 |
| Linear trend | (0.32) | (0.31) |
| Post-1989 constant | $n = 54$ | $n = 54$ |
| (4) Same as (3) | 0.95** | 0.33 |
| 1975–2004 only† | (0.34) | (0.45) |
| | $n = 29$ | $n = 29$ |
| (5) Country-level panel | 0.89** | 0.04 |
| Country-specific trends | (0.38) | (0.29) |
| Country-specific constants | $n = 3,978$ | $n = 3,400$ |
| (6) Same as (5) | 0.84** | −0.01 |
| Non-African countries only | (0.41) | (0.29) |
| | $n = 2,084$ | $n = 3,203$ |

Standard errors in parentheses. 1% °C$^{-1}$ means the probability of conflict (ACR) rises 0.01 for each 1 °C in NINO3. 1989 dropped.
* $P < 0.1$; ** $P < 0.05$.
† After 1974, the set of countries in the teleconnected group stabilized at 87–91.

The table on slide 17 gives a variety of regressions.

They all suggest that there is a positive relationship between temperature and conflict in the ENSO-affected (teleconnected) zone but not in the weakly connected zone.

The restriction to 1975-2004 in row 4 is to ensure that the post-colonial surge of countries isn’t driving the results.

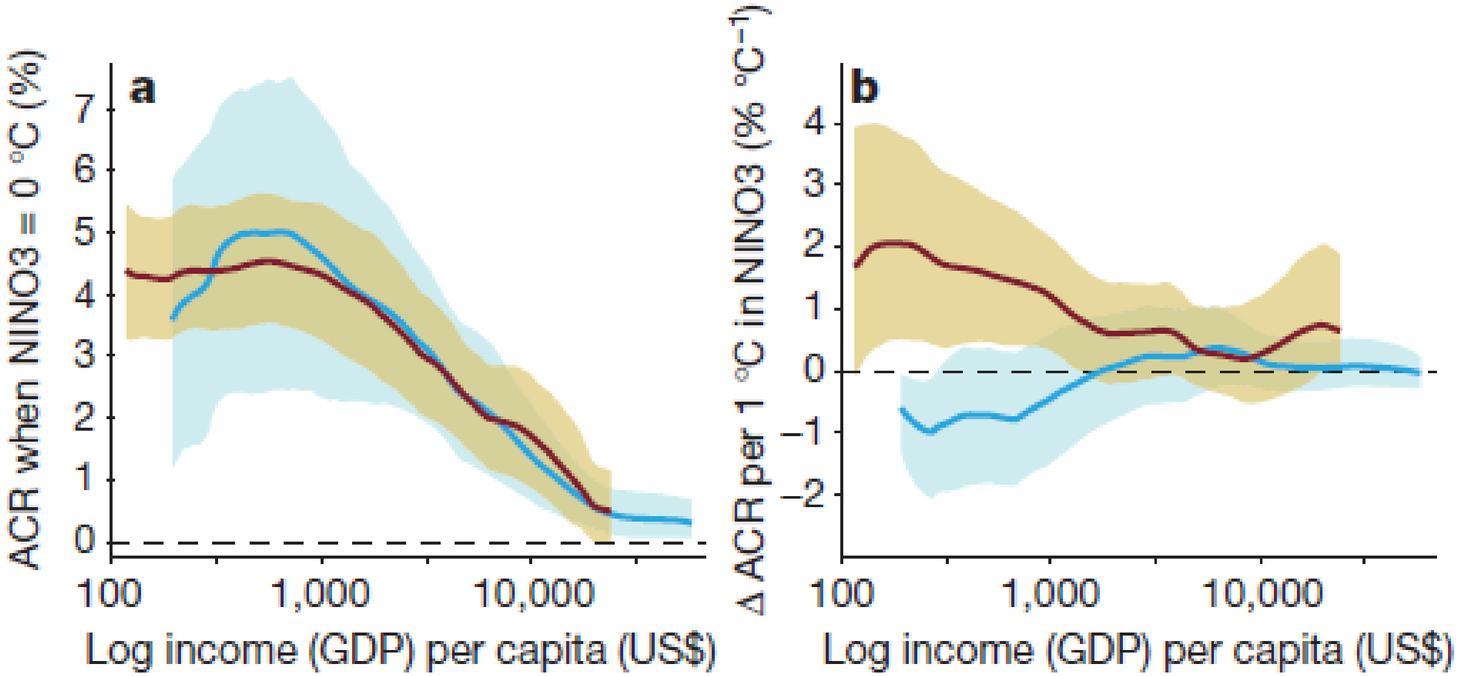



**Figure 3 | ENSO, ACR and income.** For each country $i$, we estimate $ACR_i(t) = \alpha_i + \beta_i NINO3(t)$. a, Baseline ACR ($\alpha_i$) against log income per capita in 2007 (moving average, 90% confidence intervals shaded). Teleconnected (weakly affected) group in red (blue), $n = 85$ ($n = 75$). b, Same, but for the sensitivity of ACR to ENSO ($\beta_i$).

Figure 3 requires a bit of explanation.

First Hsiang et al. run a separate regression for each country - remember that for a single country there will be one observation each year (1950-2004) for temperature and conflict so there are enough data points to do this.

The conflict variable is “1” for a particular country in a particular year if a conflict starts for that country in that year – otherwise the conflict variable is “0”.

For each country you get two estimates – one for a constant and one for a slope.

Also, for each country we have information on GDP per capita.

Panel a on slide 19 graphs the estimated intercept coefficients, denoted by $\alpha_i$ for country i, against GDP per capita.

There is one curve for the ENSO-affected (teleconnected) zone and another for the weakly affected zone.

These curves look pretty much identical.

Panel b on slide 19 graphs the slope coefficients, denoted by $\beta_i$, against GDP per capita.

These slope coefficients behave very differently in the ENSO-affected (teleconnected) zone than they do in the weakly affected zone.

Low income is associated with high responsiveness to ENSO in the ENSO-affected (teleconnected) zone.

However, cause and effect are not clear.

Does high responsiveness of conflict to ENSO cause countries to be poor or does being poor cause high responsiveness of conflict to ENSO?

One could, potentially, tell stories going in either direction.  Somewhat to my surprise Hsiang et al. do not supply such stories and I will refrain from offering my own speculations here.

Ultimately, though, I feel that we need some convincing case studies linking ENSO with armed conflict in specific times and places for the Hsiao et al. work to be fully convincing.

Finally, I offer a little discussion of ***panel data*.**

We have already seen panel data in the course but I will now devote a few slides to this type of data because next week we will cover a paper that heavily stresses its importance.

The defining characteristic of panel data is that you observe a single set of units at *multiple time periods*. Here are some examples.

1. Annual data on study effort and marks for every Royal Holloway student collected three times – once in the first year, once in the second year and once in the third year.

2. Quarterly data on economic output and unemployment rates by county for every county in the UK.

3. Annual data on armed conflict and temperature for every country in Sub-Saharan Africa.

The set of units varies from case to case – students, UK counties, the countries of Sub-Saharan Africa. Each one is observed at least two times.

Having multiple observations on each unit allows us to generate better insights than we could get from a single observation on each unit.

For example, some students at Royal Holloway may study little but still get high marks.  If we only observe each student once then the existence of low-study-high-marks students suggests that studying does not improve your marks, and may even harm them.

But if we observe each student multiple times then we will see that some students always tend to have high marks while other students always tend to have low marks.  This diversity may occur simply because high-mark students are smarter than low-mark students.

Panel data allows us to check whether the high-mark students perform relatively better in their relatively high-study years compared to their performance in their relatively low-study years.  In other words, if high-mark students get *especially* high marks in their high-study years and only somewhat high marks in their low study years then this suggests that studying does actually improve your marks.  Similarly, if low-mark students get relatively decent marks in the years when they study relatively hard then, again, this suggest that studying improve ones' marks.

There are a few regression techniques that researchers use to tease out the useful information generated by the fact of multiple observations on each unit.

One method is to introduce a dummy variable for each unit.

Thus, in the RHUL example the estimated coefficient on the Mary Smith dummy might be a 6.5, indicating that Mary tends to score 6.5 points higher than an average student who studies exactly as much as Mary does. The estimated coefficient on the Joe Blogs dummy (no pun intended) might be -8.0, indicating that Joe tends to score 8.0 points lower than an average student who studies exactly as much as Joe does.

This dummy variable approach is great for developing intuitions and I have used it in this way in earlier lectures. But having such a large number of dummy variables creates large confidence intervals around your estimates so researches usually turn to other methods.

A common technique to exploit panel data without introducing a large number of dummy variables is called a "fixed effects model".

Go through the many tables of regression results presented in this course and you will see more than a few that refer to "fixed effects", sometimes only in the footnotes. (See, for example, the main Burke et al. table given in lecture 15, slide 4.)

The key idea for these fixed-effect models is to transform each data point into a *deviation from the average* for that data point's unit.

In the student example suppose that Mary Smith studies 7 hours per week in her first year, 15 hours per week in her second year and 17 hours per week in her third year for a career average of 13 hours per week.

We then convert her study-time data points into -6 for her first year, +2 for her second year and +4 for her third year.  Similarly, we transform Mary's marks into deviations from her three-year average.  We then perform the same transformation for every student and proceed to our regression analysis.

After these transformations our analysis is now about whether *deviations* from average study levels are associated with *deviations* from average marks.  We could not perform such a deviation-based analysis on a single year of data.

This is the advantage of panel data.