Errors do not drive conclusions in our original post: Response to Hsiang and Sekar's response

Quy-Toan Do, Andrei A. Levchenko, and Lin Ma

Hsiang and Sekar respond here to our critique of their study. In this response, we argue that their retort failed to unearth any substantial and/or substantive flaws in our original critique. We take their arguments one by one, in descending order of (perceived) importance. We also publish our replication codes to give anybody the opportunity to see for themselves.

Claim #1: Our sample split selects on the outcome variable. We actually did *not* select on the outcome variable: the outcome variable is *PIKE*, we were splitting the sample by *total number of carcasses found*.

At the most basic level, this requires no further elaboration. An alternative preferred by Hsiang and Sekar is to cut the data by average number of carcasses (calculated pre-2008). Yet another approach would be to split the sites into "small" and "large" using the *initial* number of carcasses first recorded in MIKE. In both cases a single site never switches between being "small" and "large" and is not affected by whatever is believed to have happened in 2008. These two proposed alternatives, however, have the drawback that "large" sites might still have years post-2008 with very low carcass counts, magnifying the very classification problem we are highlighting further below.

The following table reports the results. For illustration purposes, we choose an 8-carcass cutoff so the "large" and "small" subsamples are of roughly equal size. The first column replicates the original finding in the paper. The subsequent columns split the sample according to the current, average, or initial number of carcasses. It is evident that our original finding -- that the significant effect appears only in "small" sites-- holds as starkly as in our original analysis. A look at the magnitudes of the coefficients (columns 2-4 versus columns 5-7) and comparing with column 1 suggests that the results might indeed be driven by "small" sites.

The code is so simple that we are posting it at the end of the blog, together with Hsiang and Sekar's original data. Any reader with access to Stata can run it instantaneously and see for him/herself, and in particular look at alternative cutoffs anywhere between 1 and 20 carcasses for any definition of "large" versus "small". At the conventional statistical significance level (i.e. 5 percent), no matter how we cut the data, we cannot reject that there is no trend break in sites deemed "large". In addition, it is easy to verify that whichever way we split the sample, the combined number of observations in the "large" and "small" bins always equals the total sample size in column (1). (We point this out only because Hsiang and Sekar's post insinuates that we were doing something untoward with the numbers of observations, and using misleading language in our original post.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hsiang and Sekar Original Result	Number of Carcasses in a Site					
		Large Sites			Small Sites		
		Current > 8	Average > 8	Initial > 8	Current <= 8	Average <= 8	Initial <= 8
Effect of Sale	0.115**	0.0826	0.0871	0.0408	0.129	0.135	0.166**
	(0.0508)	(0.0659)	(0.0663)	(0.0668)	(0.0874)	(0.0801)	(0.0723)
Observations	562	246	251	220	316	311	342
R-squared	0.540	0.744	0.664	0.597	0.473	0.450	0.502
Site Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear Time Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Why then the discrepancy between the results above, and the claims/pictures in Hsiang and Sekar's blog post, that apparently show trend breaks? The answer is rather simple: the results in their blog post omit a time trend, which is clearly essential and, incidentally, present in their working paper.

Hsiang and Sekar highlight the following three lines of our code in their post, replicated here:

```
/* ToT > 2 */
replace sample = 1
replace sample = 0 if tot < 3
…
areg pike i.year if sample == 1 , absorb(site) cluster(ccode2)
```

The ellipsis "..." just above the last line of the posted code are actually omitted lines of code that should not have been omitted as they are quite important --- we come back to this just below.

What do these three lines do? Together they estimate a set of year fixed effects. Hsiang and Sekar plot out these estimates, on which they add two lines of time trends that seem to suggest a discontinuity at 2008.

However, the lines suppressed in the ellipsis above are in fact commands estimating the step increase in 2008 (the variable "post_event") while controlling for a linear time trend:

```
areg pike year post_event if sample == 1, absorb(site) cluster(ccode2)
```

This is the same as equation (6) in Hsiang and Sekar's working paper, the key equation underlying the paper's narrative. We simply use the same equation to check if the post-2008 effect exists for sites with more than 2 carcasses, and draw our conclusions from these estimates.

What is the result of running this line? The estimated effect is 0.069, with a standard error of 0.05 and an associated p-value of 0.176. In plain language, this means that the estimated effect of sales for sites with more than 2 carcasses is not really different from zero, and the "discontinuity" in their graph is simply an optical illusion.

Again, all the Stata codes are attached to this post, and we encourage the readers to run them and see for themselves.

Claim #2: The original paper already told everyone that what happened is both an outward shift in demand and an outward shift in supply. Thus, we should not (necessarily) expect the price to move. Hsiang and Sekar's paper estimates neither the demand curve, nor the supply curve, nor the size of the shift in either of those. The claims that both supply and demand shifted out due to the one-off sale in 2008 are a mere *conjecture* that is consistent with increased poaching and no price change. Many others are equally consistent.

Nonetheless, for such conjecture to hold, it has to be the case that not only do these two shifts (demand and supply curves) need to be of just the right magnitudes to offset each other's price impact, but they also have had to kick in exactly at the same time. Ivory is a storable commodity, the demand of which, in the short run, is not just driven by consumer demand but by interest rates and storage levels, so that prices are more sensitive to long-term rather than short-term demand (see Kremer and Morcom 2000 for a theoretical discussion). When plotting the price series of raw ivory (see figure 1), we see that prices increased at roughly constant rates over the periods 1970-1992 and 1993-2014, which is consistent with ivory being stored for speculative purposes. A policy as major as the 1989 CITES trade ban, for example, did not translate into lower prices until 1992! Note also that the 2010 dip observed by some readers is in fact not statistically significant, and certainly due to noise in the data.

Thus, while the effects postulated by Hsiang and Sekar are possible, we do question whether they are plausible both in terms of their magnitudes and timing. This is all the more important since other interpretations are equally (if not more) plausible. This is indeed a generic problem with event studies in non-controlled settings: while one interpretation of the results might be most "desirable," alternative explanations need to be convincingly ruled out before drawing conclusions.

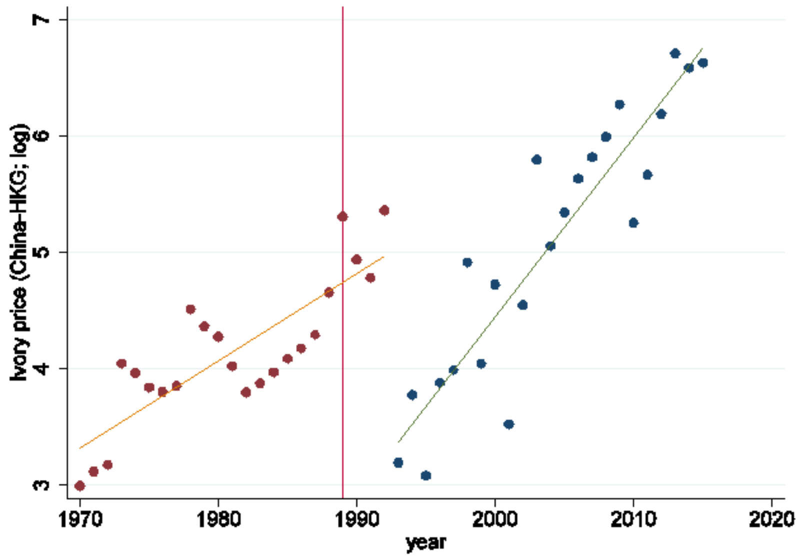

Figure 1: Ivory prices (log of median raw ivory prices in China and Hong Kong SAR, China)

As argued by [Dan](#) [Stiles](#) earlier and in his comment to our earlier blogpost, many other things could have happened in 2008. In particular, the world witnessed the largest economic recession since 1929, and world trade collapsed. Ivory, as a store of value, could well have become in higher demand in times of economic uncertainty; lower trade volumes leading to lower shipping costs could also make smuggling easier. These arguments were actually made *and empirically tested* by Brendan Moyle in a [2014 paper](#).

On the other hand, we have put forward an alternative interpretation of Hsiang and Sekar's findings that has nothing to do with the economics of ivory markets, but with the findings being a statistical artefact.

Claim #3: Our simulation reclassified 7 carcasses out of a possible 10, and claimed this is a small change.

Actually, we are *not* reclassifying carcasses only in 2008 (where the authors calculate there are only 10 illegal carcasses in small sites). We are reclassifying them in any year. We did our simulation re-classifying individual carcasses among all site-year observations with less than three carcasses, which has a total of 203 carcasses in a total of 152 site-year observations. And the absolute number is still 7 out of 203 that need to be reclassified in order for the Hsiang and Sekar's result to lose significance (the codes for the simulation are accessible [here](#)).

Even if we were to reclassify carcasses in 2008 only, while 7 out of 10 is a large number in relative terms, it is still small in absolute terms. No matter the disagreement on the claims made earlier, one fact remains: how less than 10 poached elephant carcasses are classified is what it takes to drive the results documented by Hsiang and Sekar.

Download all the data and codes to replicate our results, including the random reclassification simulations from here.

The STATA code to replicate the tables:

```
clear
use data_est

rename PIKE pike
rename site_id siteid
rename post2008 post_event
rename country ccode2

gen sample = 1

local fname "nber_rep/table_compare"
lab var post_event "Effect of Sale"

/*Comment out the following two lines when running on non-unix
machines*/
!rm `fname'.txt
!rm `fname'.xls

/* ORIGINAL RESULTS */
areg pike year post_event, absorb(site) cluster(ccode2)
outreg2 using `fname'.xls, excel append
addtext("Com.Linear.Trend","Yes") label drop(year) ctitle("NBER
Original") replace

local tot_measure_list totcarc average_tot initial_tot

/* Define average and initial size */

/* Restrict to pre-2008 data*/
gen tot1 = tot

/* Comment out line below if you want to construct definitions over
whole period instead of pre-2008 only */
replace tot1=. if year>=2008
```

```
/* Average size */
bysort site: egen average_tot = mean(tot1)

/* Initial size */
sort site year
by site: gen initial_tot = totcarc[1]

// Loops to run the estimation
// The loops below runs the regressions for large and small sites
// Each loop looks at the 4 definitions of small vs. large
// (current, average, initial)
// For each definition, it loops over every cutoff between 1 and 20

// Large sites
// small_cut is the cutoff for a site to be deemed small

foreach tot_measure in `tot_measure_list'{
	forvalues small_cut = 1/20{
	// tot_measure > small_cut
	replace sample = 0
	replace sample = 1 if `tot_measure' > `small_cut'
	areg pike year post_event if sample == 1, absorb(site)
cluster(ccode2)
	outreg2 using `fname'.xls, excel append
addtext("Com.Linear.Trend","Yes") label drop(year)
ctitle("`tot_measure', > `small_cut'")
	}
}

// Small sites
foreach tot_measure in `tot_measure_list'{
	forvalues small_cut = 1/20{
	// tot_measure <= small_cut
	replace sample = 0
	replace sample = 1 if `tot_measure' <= `small_cut'
	areg pike year post_event if sample == 1, absorb(site)
cluster(ccode2)
	outreg2 using `fname'.xls, excel append
addtext("Com.Linear.Trend","Yes") label drop(year)
ctitle("`tot_measure', <= `small_cut'")
	}
}
```