US Food Aid and Civil Conflict (Nunn & Qian 2014) – Replication and Extension [1]

Gilad Gaibel (302549977) and Guy Ishai (200485555)

Nunn & Qian (2014) find that wheat aid shipped from the US to developing countries increases conflicts in these countries on average. To deal with endogeneity, they use exogenous variation of US wheat production as an instrument. We find that using better controls for long-run time trends, the effect becomes statistically insignificant. In addition, using Monte Carlo simulations and arbitrary placebo data we find that a supposedly statistically significant effect on conflict patterns is not rare to find, implying that significance levels should be interpreted with caution.

I. Replication

The research question of Nunn & Qian (2014) is: what is the causal effect of US wheat aid on the incidence of armed conflict in recipient countries? Their dataset is a panel of 125 non-OECD countries in the years 1971-2006. OLS regression does not identify the causal effect due to joint determination of aid and conflict and due to reverse causality of conflict on aid. To identify the causal effect correctly, the paper uses the amount of US wheat production with a one-year lag as an IV for the amount of wheat aid sent to recipient countries. Wheat price stabilization policy within the US ensures that in years when production exceeds a threshold, the surplus is stored and some of it is sent as aid to other countries in the next year. This should create a plausibly exogenous variation in US wheat aid, mostly generated by local US weather, which can identify the causal effect on outcome variables of interest. The exact definition of the IV is the interaction of US wheat production with an average measure of the propensity of a country to receive US wheat aid. This limits the compliers population but pronounces the effect. The main result is a positive effect – an average country in the sample suffers from around 4% more conflict incidence if wheat aid is increased by 10%. The effect is statistically significant at the 1% level.

Table 1 is a replication of Table 1 from the original paper, and shows descriptive statistics of the main conflict incidence variable, US wheat aid and US wheat production. Figure 1 is a

[1] Authors' disclaimer: Papers replicating and discussing Nunn & Qian (2014) were published in the recent years. We became aware of this fact along our work process, but we intentionally avoided reading and even looking at any of these papers. We did this to remain focused on the original paper and to make sure that the ideas we propose remain original, at least from our perspective. Just before submitting and after finishing all edits we briefly looked at some of these papers. Christian & Barret (2017) have a detailed criticism with a basic idea similar to ours which looks more generalized. A paper by USAID (2014) criticizes the specific channels of causality that Nunn & Qian suggest. In contrast, Chu & Henderson (2016) support Nunn & Qian's results with additional robustness tests.
To remove any suspicion we declare that this work is ours, and reflects our own thinking process alone.

"reduced form" replication of Figures 1 and 2 in the paper, showing the positive correlation between lagged US wheat production and total US wheat aid sent to recipient countries, which motivates the use of US wheat production as an IV.

Table 1: Summary statistics of the main variables of interest

Variable	Mean	Std. deviation	*N*
Conflict incidence of any type	.217	.412	4089
US wheat aid (1000 tons)	27.6	117	4089
Avg. fraction of years receiving food aid	.374	.312	4089
Lagged US wheat production (1000 tons)	59053	9176	4089

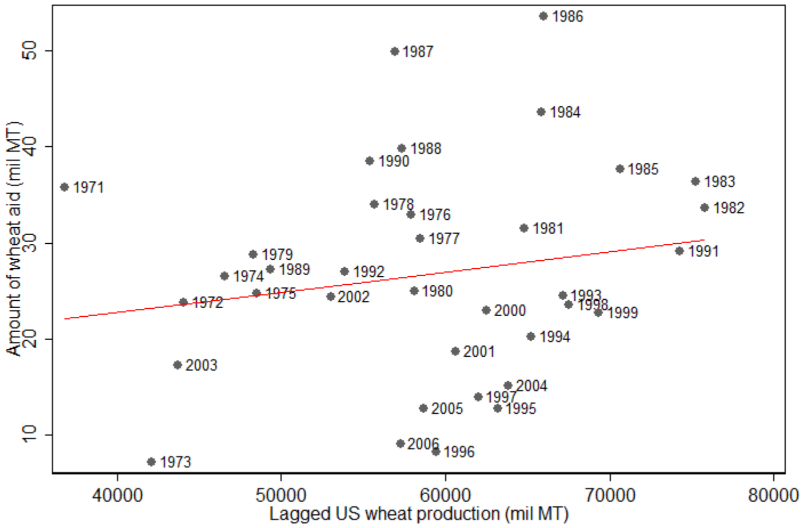

Figure 1: Total US wheat aid as a function of 1-year lagged US wheat production. The observed positive correlation is the motivation to use lagged US wheat production as an insrument for US wheat aid.

Table 2 panel B (replication of Table 2 column 5 in the original paper) shows the baseline regression results. These are the OLS, IV reduced form, IV first stage and IV second stage estimates for the effect of lagged wheat aid on incidence of conflict (of all types). The estimated second stage (or OLS) and first stage equations are respectively

$$C_{irt} = \beta F_{irt} + \mathbf{X}_{irt}\Gamma + \varphi_{rt} + \psi_{ir} + \nu_{irt} \quad (1)$$

$$F_{irt} = \alpha(P_{t-1} \times \overline{D}_{ir}) + \mathbf{X}_{irt}\Gamma + \varphi_{rt} + \psi_{ir} + \varepsilon_{irt}, \quad (2)$$

where i indicates country, which is one of 125 non-OECD countries, r indicates region, which is one of 6 geographic continents, t indicates year between 1971 and 2006, and:

- C_{irt} is conflict incidence dummy (equals 1 if there are at least 25 casualties in conflicts of all types – internal or involving other countries),
- F_{irt} is US wheat aid amount (measured in 1000 tons),
- P_{t-1} is lagged US wheat production (measured in 1000 tons),

- $\overline{D}_{ir}$ is the propensity to receive aid, defined as the fraction of years in which a country receives any amount of aid ($\overline{D}_{ir} \equiv \frac{1}{36}\sum_{t=1971}^{2006} D_{irt}$, where D_{irt} is an indicator that country i received US food aid in year t),
- φ_{rt} is region specific time fixed effect,
- ψ_{ir} is country fixed effect,
- $\mathbf{X}_{irt}$ is a vector of political, economic and weather controls (see full list in the Appendix),
- ν_{irt} and ε_{irt} are error terms.

The results indicate a zero OLS effect and strong effects for both stages of the IV regression, demonstrating both the relevance of the instrument and the positive effect of food aid on conflict. The results are economically significant. The average recipient country receives 27.6 KT wheat per year and suffers 21.7% chance of conflict incidence. The estimated effect implies that sending additional 10% wheat to this country results in an addition of $0.00299 \cdot 27.6 \cdot 0.1 \approx 0.008$ to the probability of conflict incidence. This is an addition of 0.8 percentage points, which is an addition of about 4% relative to the original chance of conflict incidence. US wheat aid can show large volatility and the effect may supposedly be even larger. It should be noted that a more correct approach for this counterfactual calculation is to use average values relevant to the compliers population and not to the entire sample, but this is not done in the original paper and is not our main focus in this work.

Table 2: The effect of US wheat aid on conflict – main regressions replicated

Specification	Explanatory variable	OLS (C_{irt})	First Stage (F_{irt})	Second Stage (C_{irt})	Reduced Form (C_{irt})
A. Alternative	F_{irt}	-0.00000		0.00367*	
		(0.00020)		(0.00209)	
	P_{t-1}		0.00070*		0.00255***
			(0.00038)		(0.00086)
B. Baseline	F_{irt}	-0.00011		0.00299***	
		(0.00017)		(0.00096)	
	$P_{t-1} \times \overline{D}_{ir}$		0.00358***		0.01071***
			(0.00103)		(0.00320)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: These are the main regression results showing the estimated causal effect of US wheat aid (F_{irt}) on conflict incidence (C_{irt}). Both the baseline specification with an instrument of the interaction of lagged US wheat production and the propensity of receiving aid ($P_{t-1} \times \overline{D}_{ir}$), and the alternative specification with an instrument of lagged US wheat production only, show a positive and significant effect. The first stage is weaker in the alternative specification (Kelibergen-Paap F statistic of 3.35 compared to 12.10 in the baseline specification), which is a motivation for using the baseline specification although it is less intuitive. All regressions have 4089 observations of 125 non-OECD countries over the years 1971-2006. Standard errors (in parentheses) are clustered at the country level.

Table 2 panel A (replication of Table 3 column 5 in the original paper) shows the alternative regression results using only lagged wheat aid as the explanatory variable, with lagged wheat production as the IV. The estimated second stage and first stage are respectively

$$C_{irt} = \beta F_{irt} + \mathbf{X}_{irt}\Gamma + \delta_r t + \psi_{ir} + \nu_{irt}, \tag{3}$$

$$F_{irt} = \alpha P_{t-1} + \mathbf{X}_{irt}\Gamma + \delta_r t + \psi_{ir} + \varepsilon_{irt}. \tag{4}$$

The first difference between this and the baseline specification is that the instrument is lagged wheat production only, without interaction with the propensity to receive aid. The second difference is that the region specific time fixed effects are now replaced with region specific linear time trends, $\delta_r t$. Time fixed effects are no longer possible as they are collinear with wheat production in the first stage. This specification is simpler and more intuitive, but although its results are similar quantitatively to the baseline ones, they are slightly less statistically significant and have a weak first stage. Thus, by focusing on a more specific population of compliers, the interacted instrument seems to increase precision without biasing results. The authors claim that this provides a justification for the use of the interaction $P_{t-1} \times \overline{D}_{ir}$ in the baseline regression.

II. Identification Assumptions

To have a correct identification of the local average treatment effect of US wheat aid on conflict using the instrument variables, the four assumptions of exclusion, independence, relevance and monotonicity must hold. In this section we discuss the validity of the assumptions in light of the paper's claims and prepare the ground for our extensions and critique.

1. *Relevance*. This is the testable assumption that the interaction of US wheat production and the propensity to receive aid, $P_{t-1} \times \overline{D}_{ir}$, is highly correlated with US wheat aid, F_{irt}. This is valid since the first stage is strong, as was shown before. A weak first stage using the instrument P_{t-1} in the alternative specification was the main reason to use the interacted version in the baseline specification.
2. *Monotonicity*. This is the assumption that there are no defiers in the sample, i.e. that both under a potential increase in US wheat production and under a potential increase of the average propensity to receive aid, a country should be (weakly) more likely to receive increased aid. It seems reasonable that when the US wheat surplus is larger, policy makers do not decrease the planned amount of aid given to any country. It also seems reasonable that when a country with a larger average propensity to receive aid is considered, its likelihood of receiving more aid in a given year increases. There may be evidence from the decision making process regarding aid shipments that supports the monotonicity assumption, but it is not provided in the paper.

3. *Exclusion*. This is the assumption that the interaction of US wheat production and the propensity to receive aid, $P_{t-1} \times \overline{D}_{ir}$, affects conflict only through US wheat aid F_{irt}, conditional on the controls. The concerns here are:
   a. US wheat production can affect conflicts through an effect on global wheat prices, or through the effect on the price of other cereals. The fact that the US wheat price stabilization policy is designed to detangle US wheat production from prices, and the control for regional time fixed effects which include all regional prices, removes most of the concern here. The authors are extra cautious and also control for imports and exports of cereals in each country, which should proxy for different sensitivities to cereal prices. This is done using interaction of the imports/exports averages with year dummies, to avoid reverse causality of aid on these controls.
   b. There may be influences of US wheat production through country specific channels different from wheat aid. The averages of these influences are controlled for using country fixed effects.
4. *Independence*. This is the assumption that the instrument $P_{t-1} \times \overline{D}_{ir}$ is not correlated with the potential outcomes of wheat aid F_{irt} and conflict C_{irt}, conditional on the controls, where potential outcomes are a function of the instrument $P_{t-1} \times \overline{D}_{ir}$ and the treatment F_{irt}. The concerns here are:
   a. US wheat production can be jointly determined with other variables such as US GDP or US political cycles, which may affect recipient countries and their potential outcomes of conflict as a function of food aid. Their effect can be different for different levels of $\overline{D}_{ir}$, therefore the authors include them and their interactions with $\overline{D}_{ir}$ as controls (see the full list in the appendix).
   b. Weather can affect both US wheat production, conflict potential outcomes and many other variables jointly. Therefore the authors control for temperatures and precipitation both in the US and recipient countries, and also control for their interactions with $\overline{D}_{ir}$.
   c. $\overline{D}_{ir}$ can be jointly determined with the average level of other forms of aid given to country i, affecting the potential outcomes of conflict. Therefore the authors control for the interaction of the average US military aid and economic aid given to each country with year dummies.
   d. The potential outcomes of wheat aid and conflict can be jointly determined with the instrument if global time trends are correlated among all variables. The authors account for this by controlling for region specific time fixed effects. In the alternative specification they include region specific linear time trends, and the strong assumption of linearity is a main motivation for the authors to focus on the baseline specification instead. We consider this a problematic approach. In the baseline specification, regions

are not necessarily groups of countries with similar characteristics for whom common effects can be isolated; it is not straightforward that time fixed effects are meaningful for large groups of different countries. In the alternative specification, the linearity assumption could have been relaxed more easily, for example by including higher polynomial trends.

e. Finally, we note that $\overline{D}_{ir}$ is mechanically related to F_{irt}, potentially introducing the original endogeneity problem in the reduced form equation. Since $\overline{D}_{ir}$ is controlled for in the country fixed effects, the joint determination or reverse causality of propensity to receive aid and *average* conflicts level per country is controlled for. However, there still exists the concern that conflict patterns *net of the average* are jointly determined with the propensity to receive aid, and this is not controlled for. This is directly connected to the issue of time fixed effects stated above, and we revisit it in the next section.

III. Extension 1: Improving the Control for Long-Run Time Trends

The authors control for an impressive set of observables that cover many potential identification issues. However, since the scope of unobservables tends to be much larger and harder to trace, we are concerned with an unsatisfying account of unobservables. Specifically, we are concerned that time de-trending is insufficient. This would mean that the independence assumption, and perhaps exclusion as well, do not hold because of important common time trends that are not accounted for.

First, to illustrate the concern, we focus on the simpler alternative specification with P_{t-1} as the instrument. Figure 2 shows a comparison of the general time trends of average conflict incidence and US wheat production. It can be observed that the trends are quite similar, and

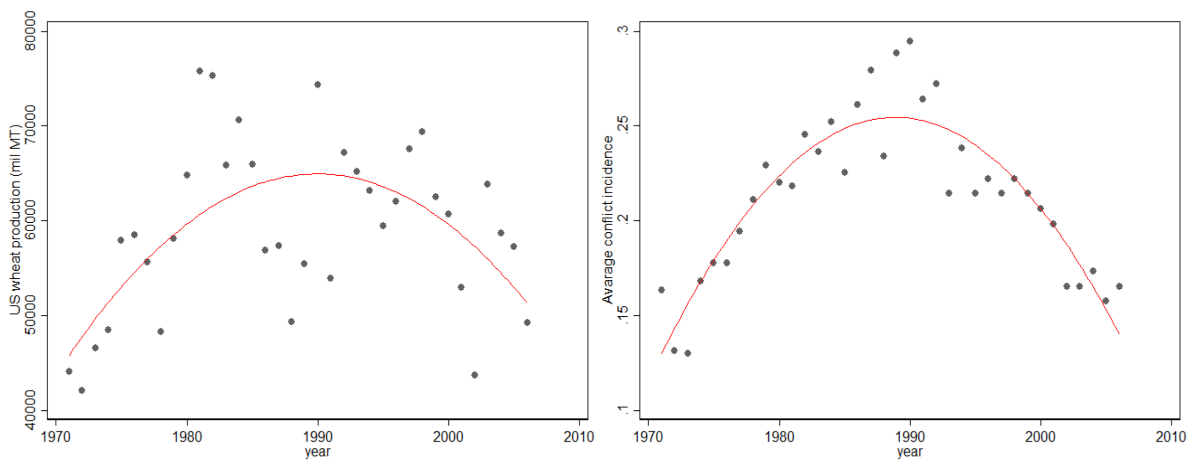

Figure 2: Left panel: US wheat production as a function of time. Right panel: Average conflict incidence across the 125 non-OECD countries in the sample as a function of time. Both quantities show similar time trends, approximated by quadratic polynomials shown by red smooth lines. Our new regression specifications try to account for this long-run trend properly.

this appears to be the main force driving the original results. The conflicts time trend is quite smooth and impressive, and has a good approximation by a quadratic polynomial. The wheat production time trend is noisier, but has a general similar shape. Given the shape of the conflicts trend, it is impossible to de-trend the conflicts using linear polynomials, even if they are fitted at the region level, as aggregation of linear polynomials always results in a linear polynomial. We conclude that in order to account for the trend properly, second order polynomials must be used. To illustrate the importance of this trend, we replace the region specific linear time trends with a single second order polynomial for the entire sample of 125 countries. Thus, the new estimated second and first stage equation are:

$$C_{irt} = \beta F_{irt} + \mathbf{X}_{irt}\Gamma + \delta_1 t + \delta_2 t^2 + \psi_{ir} + \nu_{irt}, \tag{5}$$

$$F_{irt} = \alpha P_{t-1} + \mathbf{X}_{irt}\Gamma + \delta_1 t + \delta_2 t^2 + \psi_{ir} + \varepsilon_{irt}. \tag{6}$$

The regression results are shown in Table 3 panel A. Given the general quadratic trend of conflicts, the remaining predictive power of wheat production in the reduced form is very small and insignificant. The first stage is also no longer significant and the second stage original results vanish as well. The reduced form estimate have confidence intervals that reject the original effect estimated in the paper.

Table 3: The effect of US wheat aid on conflict – improved control for time trends

Specification	Explanatory variable	OLS (C_{irt})	First Stage (F_{irt})	Second Stage (C_{irt})	Reduced Form (C_{irt})
A. Alternative	F_{irt}	-0.00009		-0.00110	
		(0.00016)		(0.00236)	
	P_{t-1}		0.00029*		-0.00031
			(0.00015)		(0.00065)
B. Alternative (no controls)	F_{irt}	-0.00007		-0.00526	
		(0.00016)		(0.01014)	
	P_{t-1}		0.00010		-0.00050
			(0.00014)		(0.00064)
C. Baseline	F_{irt}	-0.00023*		0.00300	
		(0.00013)		(0.01034)	
	$P_{t-1} \times \overline{D}_{ir}$		0.00165		0.00496
			(0.00488)		(0.01155)
D. Baseline (no controls)	F_{irt}	-0.00024*		0.01188	
		(0.00013)		(0.05858)	
	$P_{t-1} \times \overline{D}_{ir}$		0.00076		0.00905
			(0.00367)		(0.01146)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: These regressions are identical to those estimated in Table 2, except for two changes: (1) the alternative specification uses a global quadratic time trend instead of region specific linear time trends, to account correctly for the global long-run trend of conflicts shown in figure 2, and (2) The baseline specification uses an alternative definition of regions, using quintiles of the propensity to receive aid $\overline{D}_{ir}$ instead of geographic continents, in order to account for differential time trends for countries with different levels of $\overline{D}_{ir}$. Each specification is estimated with and without controls to show robustness of results. All regressions have 4089 observations of 125 non-OECD countries over the years 1971-2006. Standard errors (in parentheses) are clustered at the country level.

Recall that the authors are concerned that linear time trends are not sufficient for de-trending the data, which is a main motivation to use the baseline specification with $P_{t-1} \times \overline{D}_{ir}$ as the instrument. In the baseline specification, the general quadratic trend shown before is fully controlled for using the region specific time fixed effects. However, in this specification we are interested in the differential reduced form effect of wheat production on conflicts for countries with different levels of $\overline{D}_{ir}$. Therefore, it makes sense to account for different time trends for countries with different levels of $\overline{D}_{ir}$. A similar approach is already taken by the authors with respect to other controls, as they add an interaction term of each time-varying control with $\overline{D}_{ir}$. In principle, the control for time effects is even more careful, as time fixed effects within groups of countries can provide a more flexible control than an interaction term. However it is questionable if regions, which are defined as continents, are an adequate segmentation of the countries when differential effects are a function of $\overline{D}_{ir}$.

Figure 3 shows the mean time trends of conflict in quintiles of $\overline{D}_{ir}$, minus the region specific mean conflict in each year. This illustrates that accounting for region specific time fixed effects leaves us with time trends that are indeed different for different values of $\overline{D}_{ir}$. These differences are responsible for the significant and positive effect of the baseline specification. Table 3 panel C shows the baseline specification results when the original region segmentation is replaced with segmentation by $\overline{D}_{ir}$ quintiles. The new second and first stage equations can be written the same as eq. (1), (2), with the only difference that the index r now refers to one of the 5 quintiles of $\overline{D}_{ir}$ instead of one of the 6 original geographic regions. Both the reduced form causal effect

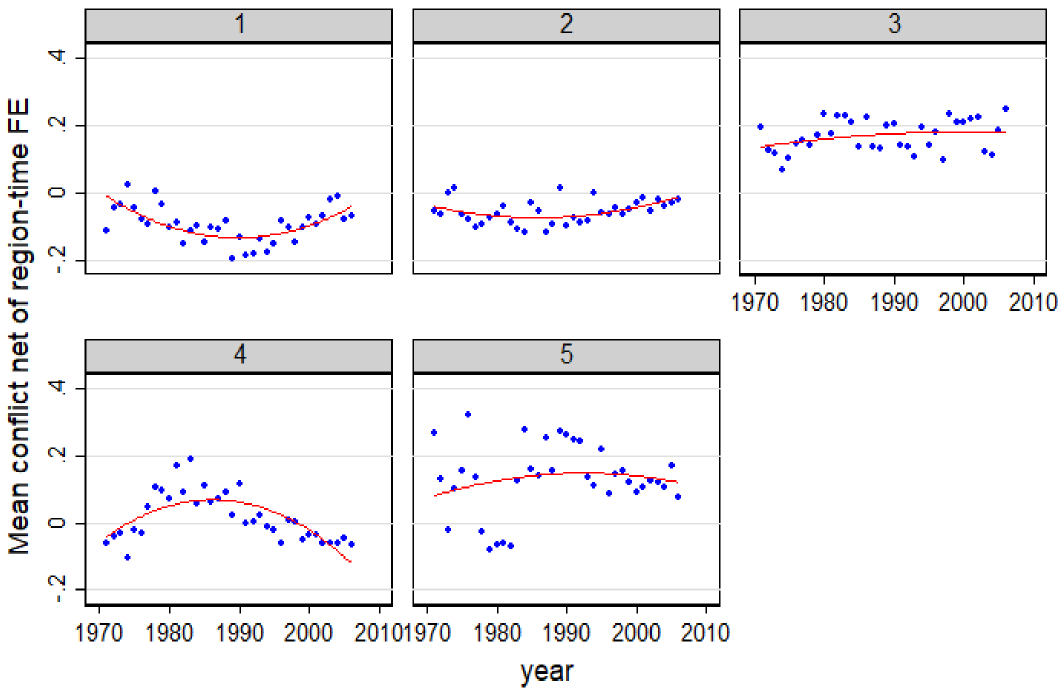

Figure 3: Mean conflict incidence minus the region-specific conflict mean in each year, as a function of time, shown in quintiles of propensity to receive aid, $\overline{D}_{ir}$. Fitted quadratic time trends are shown by red smooth lines. This demonstrates that the segmentation into geographical regions (continents) does not account for important common trends when focusing on differential effects for countries with different levels of $\overline{D}_{ir}$.

and the first stage become statistically insignificant, however here we cannot reject the original reduced form estimates due to large standard errors.

The underlying reason that the new segmentation diminishes the effect of wheat production on conflict is that the original segmentation into continents de-trends together very different countries in the context of wheat aid and conflicts. For example, it seems straightforward that countries that almost never receive aid are very different in trends of conflicts from countries who do receive some amounts of aid. The common conflict time trend for both types may be meaningless. This is a worry because most regions, even Sub-Saharan Africa which is known for its bloody civil conflicts, contain comparable numbers of countries with different levels of $\overline{D}_{ir}$. Table 4 demonstrates that all regions are quite heterogeneous in $\overline{D}_{ir}$ levels. Figure 3 demonstrates that when conflict time trends of aided countries are averaged with those of non-aided, the result is simply an attenuated time trend. Countries who do not get much aid, in the lowest $\overline{D}_{ir}$ quintile, have an artificial attenuated negative conflict trend, which equals their original zero conflict value minus the average trend. Countries who get more aid have the opposite artificial strengthened trend. It is questionable whether or not $\overline{D}_{ir}$ provides the most correct segmentation for de-trending, but since the authors choose to interact all their time-varying controls with $\overline{D}_{ir}$, we find it consistent with their approach to do a similar thing when controlling for time effects.

Table 4: Distribution of $\overline{D}_{ir}$ in the 6 geographic regions (%)

$\overline{D}_{ir}$ quintile	East Asia & Pacific	Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	South Asia	Sub-Saharan Africa
1st	46.1	10.5	20	47.1	0	15.9
2nd	30.8	36.8	12	5.9	14.3	18.2
3rd	7.7	15.8	12	17.7	14.3	34.1
4th	7.7	10.5	24	11.8	28.6	20.5
5th	7.7	26.3	32	17.7	42.9	11.4
N	13	19	25	17	7	44

We find these new specifications the most straightforward for demonstrating the importance of the long-run time trend, however their explanatory power can be further demonstrated using more extreme specifications, which also show the robustness of these results. Table 3 panels B and D show the effect on conflicts using the instruments P_{t-1} and $P_{t-1} \times \overline{D}_{ir}$ respectively, including country fixed effects and not including any controls. For P_{t-1} we only add the global linear and quadratic time trends as controls, and for $P_{t-1} \times \overline{D}_{ir}$ we only add the $\overline{D}_{ir}$-specific time fixed effects. In both specifications the effect on conflicts is insignificant, meaning that time trends have the majority of the explanatory power. This can be compared to the control-less regressions in Table 2 column 1 and Table 3 column 1 in the original paper, in which a positive and significant effect of the instrument was estimated. Robustness of the new baseline specification to different numbers of $\overline{D}_{ir}$ segments can be found in Appendix Table A1.

Our interpretation of the new results is that global time trends of US wheat production and conflict incidence happened to be synchronized in the sample period by an obscure reason. In fact, a possible scenario is that no hidden variable drives both of these quite distinguished phenomena, and instead they only happened to have a long run cycle of a similar length. Without controlling for the long-run trend the identification was incorrect, e.g. because picking a good year in terms of wheat production means picking a good year in terms of all potential outcomes of conflict, thus violating independence.

An alternative interpretation is that our control for time trends is a bad control that confounds the results. This is true if the long-run time trend of conflicts (in countries that are similarly-aided on average) is *caused* by the long-run time trend of US wheat production (interacted with propensity to receive aid). We find this implausible for two reasons. First, and most importantly, we find it hard to believe that a single parameter from the US is such a main driver of conflict patterns around the world; civil and interstate conflicts are complex phenomena and are caused by and related to a great amount of variables, among which US wheat production and food aid are a small portion. We expect them to cause small and local shocks relative to long-run trends of large groups of similar countries, not to cause the trends themselves. Second, if this were the case, it would have been a mistake to account for any time trends or time fixed effects, because any attempt to do so will be a bad control of the same kind. The approach of the original work is to do account for time trends, so we find it consistent with this approach to try and do so in more specifications and with greater detail.

IV. Extension 2: Placebo Tests

The alternative interpretation of our results suggested in the previous section, that conflict long-run time trends are caused by the US wheat production long-run time trend, can seem more likely if the observed time trends are unique and hard to find in time series data. In this section we show that this is not the case by conducting a placebo test, replacing the time series of US wheat production with other exogenous time series. This approach addresses both the uniqueness of the observed long-run conflicts trend and the uniqueness of other patterns of conflicts that we did not characterize, because successfully predicting conflicts using the placebo means reproducing at least one of these two options.

Our placebo tests focus on two different questions:

1. How rare it is to find an exogenous variable that predicts conflicts? This is answered by estimating the reduced form regression using a placebo time series as the explanatory variable.
2. How rare it is that a strong and valid instrument for US food aid generates a significant effect of US food aid on conflicts? This is answered by estimating the second stage regression using a placebo time series instrument that provides a strong first stage.

We begin with the first question. An anecdotal placebo time series is the frequency of appearances of the phrase "Star Wars" in English books indexed by Google Books in the years 1971-2006 (data taken from Google Ngram Viewer, in units of 10^{-6} %). Similarly to the US wheat production series, this series also shows an increasing long-run trend until around 1990 and a decreasing trend afterwards. The reduced form result of the baseline specification using this series instead of US wheat production is an effect of 3.24 (1.40), which is positive and significant similarly to the original results. It is quite common to find series with a similar long-run trend that are clearly unrelated to conflicts and produce such results. More examples from Google Ngram Viewer can be found in Appendix Table A2. This provides a first alert regarding the uniqueness of the conflicts time trend and regarding the causal inference of US wheat production on conflicts.

To answer the first question more thoroughly and also address the second one, a more systematic placebo approach is needed. For this purpose we run a Monte Carlo simulation, in which we generate 1000 standard random walk series with 36 observations each and estimate the baseline model with each of them instead of US wheat production. The data generating equation for the each random walk series is

$$R_t = R_{t-1} + \varepsilon_t, \qquad (7)$$

where $\varepsilon_t \sim N(0,1)$. Table 5 panel A row 1 shows that a significant effect of a random walk series on conflicts is estimated with 5% significance or more for 37.4% of the series, and an effect with 1% significance or more is estimated for 11.6% of the series. The effect can be either positive or negative with roughly equal probabilities. This shows that among random walk series it is quite frequent to find reduced form false positives. Conflict long-run time trends and other patterns are therefore not hard to predict with this test. The implication is that the original estimated standard errors may be wrong, and imply much larger than true significance levels. To the extent that real data can be fully or partly generated by a random walk, this may be a large concern.

We now address the second question of the possibility to find a significant second stage result when all IV assumptions hold. The random walk as an instrument satisfies independence and exclusion by construction. It satisfies monotonicity because the placebo value in a given time clearly doesn't affect potential outcomes of aid, and we are left with the assumed monotonicity with respect to $\overline{D}_{ir}$. We only need to satisfy relevance. 129 out of the 1000 random walk series have a strong first stage with Kleibergen-Paap F statistic larger than 10, and these provide the subsample for this test. Table 5 panel A row 2 shows that in this subsample, a significant effect of US food aid on conflicts is estimated with 5% significance or more for 58.1% of the series, and an effect with 1% significance or more is estimated for 15.5% of the series. This result is interesting because in a way it mimics the steps of building an IV model. First, the researcher searches for an IV among variables that are strongly correlated with the explanatory variable. Only after assuring this and supporting the rest of the assumptions using controls etc., the IV model should be estimated. Then, the probability that the researcher finds a significant result is the one estimated here.

These placebo results suggest that the significance levels reported in the original paper may be misleading. A possible reason is that there is more correlation between observations in the panel than the correlation accounted for. Clustering is done only at the country level but different countries may be also correlated, e.g. countries with similar propensities to receive aid. The fact that only a few global time trends were found sufficient to explain most conflicts variance supports this hypothesis.

Finally, we can answer the two questions again using our new specifications with the improved account for time trends, to see how much of the conflicts data vulnerability to placebo is due to time trends and how much is due to other patterns. Table 5 panel B summarizes the results.

Table 5: Random walk placebo tests

	Placebo series	Reduced form		Second stage		N
		$p \leq 0.05$	$p \leq 0.01$	$p \leq 0.05$	$p \leq 0.01$	
A. Original Paper Baseline Specification	Arbitrary random walk	37.4%	11.6%			1000
	Random walk with a strong first stage			58.1%	15.5%	129
B. New Baseline Specification with improved time controls	Arbitrary random walk	5.2%	1.2%			1000
	Random walk with a strong first stage			Not enough observations		1

Notes: The table shows the percentages of random walk series that produce a significant effect on conflicts when used instead of US wheat production in the regressions. 1000 random walk series are generated randomly according to eq. (7) in a Monte Carlo simulation and are then used to estimate the models. Each panel shows both the percentage of significant estimations out of the entire 1000 series and out of a smaller sample of series which have a strong first stage, with Kelibergen-Paap $F \geq 10$. Significant effects in the original paper's baseline specification appear much more than predicted by the significance levels themselves, but appear in reasonable frequencies when accounting for the long-run time trends. When accounting for time trends only one observation has a strong first stage, therefore results are not available. All regressions estimated in the Monte Carlo simulation have 4089 observations of 125 non-OECD countries over the years 1971-2006 and standard errors clustered at the country level.

Percentages of significant estimates of the reduced form effect now match the significance levels themselves. This supports the assumption that the long-run time trends, which are now controlled for, were responsible for most of the ease of prediction observed using the original baseline specification. Percentage of series with a strong first stage is around 0.01%, so using our simulation sample size we cannot estimate the percentage of significant estimations in this case. However, this implies that it is very hard to find a good instrument for US wheat aid when controlling for the time trends, consistent with our previous results which showed insignificant first stages in the new regressions.

To conclude this section, it should be reminded that the random walk placebo test is an arbitrary benchmark, which is supported only if real data in this context is produced fully or partly by a random walk process. In any case, given the popularity and importance of random walk in Economics, it can at least place a warning sign regarding causal inference in this case.

V. Discussion

Nunn & Qian (2014) estimate a positive causal effect of US wheat aid on conflicts in developing countries. In this work we showed that this is largely a result of common time trends of conflicts incidence and of US wheat production, which is used as an instrument. We believe controlling these long-run time trends of conflicts with sufficient detail in essential, because they are probably determined by the great amount of unobservables which cannot be controlled for. We support this claim using a large set of random placebo variables generated in a Monte Carlo simulation, which show that the long-run trend of conflicts is not unique and is easy to predict. The implication is that the original significance levels reported in the paper may have been misleading.

Using improved controls for the long-run time trends we also find that US wheat production is no longer a strong instrument for US wheat aid, and that generally it may be rare to find a strong instrument which relies on a global time series. Recalling that the US wheat production instrument was weak to begin with, and that only a somewhat controversial interaction of US wheat production with the propensity of a country to receive aid provided a strong instrument, we propose that a better instrument for this research question should be pursued at the country level instead of the global level. In addition, given that a great deal of the time variance in this sample is the result of quite a generic long-run time trend, i.e. that the correlation between countries is strong, a sample with more years might be necessary to make better statistical inference.

VI. Appendix

1. Full List of Controls

The full set of controls in the baseline specification is:

- US real per capita log GDP × $\overline{D}_{ir}$
- Indicator for US democratic president × $\overline{D}_{ir}$
- Oil price × $\overline{D}_{ir}$
- Monthly recipient temperature and precipitation
- Monthly recipient temperature and precipitation × $\overline{D}_{ir}$
- Average recipient US military aid × year dummy
- Average recipient US economic aid × year dummy
- Average recipient cereal import × year dummy
- Average recipient cereal production × year dummy
- Average recipient per capita log GDP × year dummy (*this is not listed in the paper but included in the code provided by the authors*)

The full set of controls in the alternative specification is:

- US real per capita log GDP
- Indicator for US democratic president
- Oil price
- Monthly recipient temperature and precipitation
- Average recipient US military aid × t
- Average recipient US economic aid × t
- Average recipient cereal import × t
- Average recipient cereal production × t
- Average recipient per capita log GDP × t (*this is not listed in the paper but included in the code provided by the authors*)

Note: The paper lists the 5 last controls as being interacted with region-specific time trends rather than global time trends, however their code accounts for global time trends.

2. Robustness of Baseline Specification with Improved Time Controls

Table A1 shows the robustness of the results in Table 3 panel C to an alternative segmentation into groups with similar $\overline{D}_{ir}$ values, using segmentation into 3, 4, 6, and 7 groups in addition to the original segmentation into 5 quintiles. All segmentations but the one into 7 groups maintain the result of not rejecting the null hypothesis of zero effect of US wheat production on conflicts.

Table A1: Robustness of new baseline specification with improved time controls to the number of $\overline{D}_{ir}$ segments

# of $\overline{D}_{ir}$ segments	Explanatory variable	First Stage (F_{irt})	Second Stage (C_{irt})	Reduced Form (C_{irt})
3	F_{irt}		-0.00038 (0.00181)	
	$P_{t-1} \times \overline{D}_{ir}$	0.00400 (0.00245)		-0.00153 (0.00746)
4	F_{irt}		0.01042 (0.04226)	
	$P_{t-1} \times \overline{D}_{ir}$	0.00053 (0.00198)		0.00555 (0.00897)
5	F_{irt}		0.01188 (0.05858)	
	$P_{t-1} \times \overline{D}_{ir}$	0.00076 (0.00367)		0.00905 (0.01146)
6	F_{irt}		-0.00501 (0.04514)	
	$P_{t-1} \times \overline{D}_{ir}$	0.00063 (0.00619)		-0.00318 (0.01244)
7	F_{irt}		0.00383 (0.00253)	
	$P_{t-1} \times \overline{D}_{ir}$	0.00890* (0.00506)		0.03405** (0.01443)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All regressions have 4089 observations of 125 non-OECD countries over the years 1971-2006. Standard errors (in parentheses) are clustered at the country level.

3. *Placebo Tests Full Results*

Table A2 shows the full regression results of the baseline specification with three placebo time series taken from Google Ngram Viewer replacing US wheat production: the frequencies of appearances of the phrases "Star Wars", "Economics" and "Econometrics" in English books in the years 1971-2006.

Table A2: Placebo regressions using arbitrary time series from Google Ngram Viewer instead of US wheat production as the instrument (interacted with $\boldsymbol{\overline{D}_{ir}}$)

Placebo series	Explanatory variable	Reduced Form (C_{irt})
"Star Wars" (frequency units 10^{-7} %)	"Star Wars" $\times \overline{D}_{ir}$	3.24016** (1.40284)
"Economics" (frequency units 10^{-5} %)	"Economics" $\times \overline{D}_{ir}$	7.47579** (3.62692)
"Econometrics" (frequency units 10^{-6} %)	"Econometrics" $\times \overline{D}_{ir}$	17.41331*** (5.53953)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: All regressions have 4089 observations of 125 non-OECD countries over the years 1971-2006. Standard errors (in parentheses) are clustered at the country level.

Graphs showing the time trends of the placebo series can be found in the following links.

"Star Wars":

https://books.google.com/ngrams/graph?content=star+wars&year_start=1970&year_end=2008&corpus=15&smoothing=0&share=&direct_url=t1%3B%2Cstar%20wars%3B%2Cc0#

"Economics":

https://books.google.com/ngrams/graph?content=economics&year_start=1970&year_end=2008&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Ceconomics%3B%2Cc0

"Econometrics":

https://books.google.com/ngrams/graph?content=econometrics&year_start=1970&year_end=2008&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Ceconometrics%3B%2Cc0