DO SCHOOLS PROMPT TERRORIST ATTACKS? EVIDENCE FROM PERU USING SPATIAL ECONOMETRICS

Francisco Romero-Haaker y Carmen Salazar-Sosaya

Lima, mayo de 2013

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

Programa Académico de Economía

UNIVERSIDAD DE PIURA

FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES
PROGRAMA ACADÉMICO DE ECONOMÍA

DO SCHOOLS PROMPT TERRORIST ATTACKS? EVIDENCE FROM PERU USING SPATIAL ECONOMETRICS

Tesis que presentan los Bachilleres en Economía, Señor Francisco Javier Romero Haaker y Señorita Carmen Andrea Salazar Sosaya para optar el Título de Economistas.

Lima, mayo 2013

César Calvo Cervantes

Asesor

ABSTRACT

In this paper we pin down the mechanism whereby the violence of the Peruvian civil conflict spread across districts. Using district-level data and spatial econometric techniques, we find that schools were the key element that propitiated the contagion. This finding is unique in the literature and emphasizes not only the importance of neighborhood effects but also the need to go beyond simple nearness as the driver of these impacts. Spatial nearness requires a channel to deliver its potential contagion effects. In our case, our findings on the role of schools are supported by the qualitative research of the Truth and Reconciliation Commission.

GENERAL INDEX

I. INTRODUCTION 5

II. RELATED LITERATURE 11

III. EMPIRICAL STRATEGY 14

i. Basic Setup 15

ii. Multiple Interconnectivity Measures 17

iii. Robustness Checks 18

IV. DATA 21

i. Sampling 22

ii. Summary Statistics 23

V. RESULTS 25

i. A First Group of Regressions 25

ii. Unmasking the contagion effect 26

iii. Robustness checks 28

VI. CONCLUSIONS 32

VII. REFERENCES 34

VIII. APPENDIX 37

I. INTRODUCTION

The abysmal consequences of Civil Wars[1] have become the subject of an increasing number of economic studies in the last decade. Several authors have analyzed the effects of violence on both nation-level outcomes, such as growth rates (Collier 1999, Rodrik 1999), and micro-level variables mostly related with health, poverty, and education (Camacho 2008, Justino 2007).

On the other hand, maybe because of the important effects civil wars have on society, a flourishing number of social scientists have tried to unmask their puzzling causes. For instance, Fearon and Latin (2003) documented that conflicts are usually positively correlated with poverty, political instability, terrain suitability, and large populations, while Collier and Hoeffler (1998, 2004) linked warfare duration and likeliness to social grievances, ethnic fractionalization, and natural resources. Although these core studies triggered a considerable amount of research on this area, some shortages remain. In particular, much of the analysis has been conducted on a cross country basis where the aggregate country-level figures are probably poor measures of the actual circumstances in the violent zone (Buhaug and Luhala 2005). Thus, the literature seems to have an incomplete understanding of what is actually happening in the warzone,

[1] See Blattman and Miguel (2010) for a comprehensive survey on the literature about civil wars.

being of special interest to come up with more microeconomic evidence (Blattman and Miguel 2010).

We try to address this particular claim by studying the Peruvian civil war (terrorism) at a district level —the smallest geopolitical division in Peru[2]. According to the Truth and Reconciliation Commission (TRC), the official entity in Peru in charge of studying the conflict, the violence of this war has no precedent in the republican history of Peru. In fact, in terms of human deaths, it was worse than the independence war and the war against Chile, the two main conflicts in which Peru have participated[3].

At first sight, the Peruvian civil war seems to be the regular political conflict in which a rebel group claims the flag of social justice and fights against the government. Therefore, the common factors mentioned in the economic literature may play an important role in explaining the causes and conduct of the war, but could a more careful look reveal some particular patterns not previously mentioned in the literature?

Let us consider the following example: the district Santiago de Lucanamarca was pretty similar to that of Coronel Castañeda in 1980, the year in which the conflict started. Both districts were located in Ayacucho, in the highlands of Peru, 700 to 800 Km far from Lima, the country's capital city. In each of them, less than 2% had access to drinking water, and no household had access to electricity or sewage. Moreover, in both districts the average age of the labor

---

[2] The average district in our simple is 538.93 km^2, and almost 90% of them are below the 1000 km^2 threshold. (Buhaug and Rod (2006) try to address the specific characteristics of the war zones in Africa by dividing the continent into 100x100 grids).

[3] Final Report. Truth and Reconciliation Commission. P 53.

force was around 22 years old, and about half of them were farmers. The total population was also similar, around 2000 people each. Although their initial conditions looked alike, 4 years after the conflict had started Santiago de Lucanamarca was still in peace while Coronel Castañeda already registered more than 180 human rights violations[4].

If after controlling for the typical factors considered in the literature, substantial and systematic differences on the intensity of the violence across districts remain; what are the factors that are behind the violence incidence? The answer seems to be twofold. On the one hand, some econometric studies have already started to account for explicit spatial interactions across the units of observation, suggesting that space (or interconnectivity) do matter, and that observations are not independent. For instance, some studies show that the likeliness of war on a nation may be affected by both the presence of war and some characteristics of the neighbor nations, such as the level of democracy, commercial interaction, and ethnic diversity (Gleditsch 2007).

On the other hand, the TRC sheds some light on the conduct of the Peruvian civil war highlighting the importance of schools[5] for the rebel strategy. According to the Final Report of the TRC, the educational system played a prominent role during the conflict as it provided the ideal conditions for the rebel ideology to spread out. Since school teachers are attributed important roles of leadership in the Peruvian rural villages and also are very close to the youth, the rebel group considered them key elements in their task of enrolling new and young people

[4] Human rights violations were collected by the TRC and include arbitrary detention, death, forced disappearance, internal displacement, kidnapping, lesion, sexual violence, and torture.

[5] Hereafter, schools makes reference to secondary or higher educational institutions.

in the terrorist forces.

Considering both points of view, we try to answer why very similar districts in initial conditions end up presenting extremely different patterns of violence. To this end, we ask how important are spatial interactions after having controlled for a number factors usually related with civil wars[6]. Specifically, if schools were such an important asset for the rebels and really played a particular role in the recruiting process of the rebel forces; were they actually the mechanism whereby violence was transmitted across districts?

Unfortunately, the standard econometric approach may not be useful in addressing these interrogatives since its units of observation are assumed to be independent —spillovers and contagion effects are difficulty modeled. Therefore, we try to cope with these potential shortages by using spatial econometric techniques in analyzing the Peruvian conflict for the period 1980-1984 at a micro-level basis. In particular, we address the conduct of the war by allowing spatial interactions across districts and test for the importance of schools as determinants of violence contagion.

Thus, our empirical strategy will be based in the estimation of multi-parametric Spatial Autorregressive Models (SAR), contributing to the literature by truly understanding the conflict and thus taking into account not only geographic notions of connectivity but also political-economic connectivity measures (Beck, Gleditsch and Beardsley 2006).

In particular, we will first show the existence and importance of a geographic spatial effect in the transmission of the conflict after controlling for a set of

---

[6] For a sensitivity analysis on typically used controls in the literature see Hegre and Sambanis (2006).

widely used explanatory variables, polynomial terms, and fixed effects. Then, we will demonstrate that, in fact, the spatial effect of the Peruvian conflict was not merely geographic, but that it was driven by a specific channel of transmission: schools. Our results are robust to a number of specifications and placebo tests, and show that the spatial lag generated by the transmission of the conflict through schools accounts in average for the 65.74% of the actual number of human rights violations in those districts affected by the conflict.

Therefore, this study contributes to the literature in a number of ways:

First, it provides evidence from the conduct of an internal conflict at a micro-level basis. This disaggregation level is quite uncommon in both the civil war literature and the spatial econometrics literature related to political science, and allows us to describe the warzones in a better fashion than the usual cross-country analysis. More importantly, it also allows us to understand the exact mechanism whereby violence spread out in the most important conflict in the republican history of Peru.

Second, in contrast with most studies that apply spatial econometrics techniques, we introduce notions of political-economic interconnectivity and show that this specification is the one that actually matters. We emphasize the importance of a qualitative understanding of the phenomenon under study. Although cross-country analysis and others studies have been useful in many ways, we suggest to go beyond the geographic spatial considerations of interconnectivity measures and use crucial notions of space depending on each phenomenon.

Third, our results have important policy implications and calls for further micro-level research on the conduct of war. It is clear that having a better understanding of the rebel strategy and the dispersion of violence is crucial in avoiding both current and future conflicts, as well as invaluable economic losses and human lives. Unfortunately, there is a dearth of studies depicting the exact mechanisms whereby rebel forces spread out violence; more case studies in this direction are needed

II. RELATED LITERATURE

A good number of empirical papers has been written trying to identify the causes of civil wars. Probably one of studies that prompted the flourishing of this strand of the literature was Collier and Hoeffler (1998). The importance of this particular paper relies on the formalization of a framework in which the decision to fight is explained by economic rationales. Moreover, they show empirical evidence on key factors that determine the likeliness and duration of civil wars (initial income, the amount of natural resources, the size of the initial population, and ethno-linguistic fractionalization).

Likewise, the study conducted by Fearon and Latin (2003) has largely influenced this literature. Interestingly, although they studied a similar sample of countries than Collier and Hoeffler (1998), they found some differences. For instance, they concluded that ethnic and religious conditions were not as important as poverty, political instability, large populations, and rough terrains. Moreover, their econometric techniques were adopted in a number of further studies (Blattman and Miguel 2010).

On this basis, much research followed the cross-country level scheme, finding a wide range of factors related to civil wars. In fact, Hegre and Sambanies (2006) conducted a sensitivity analysis of 88 variables usually included in this type of studies and found that many of them depended on the some particular characteristics of the empirical design —sample period, the set of control variables included, etc.

Although these cross-country analysis can be useful in addressing some particular questions, other researches have started to focus on natural experiments in an attempt to find a straight relation between civil wars and its causes. In this vain, probably on the the most explored areas is the positive relation between negative income shocks and war. Some remarkable studies are Miguel, Satyanath and Sergenti (2004) and Ciccone (2008). Both of them use rainfall as an exogenous source of income variation when studying the break outs of civil wars in Sub-Saharan African countries.

However, despite the remarkable progress on this field, many questions remain unanswered. In particular, more evidence is needed at a micro-level of analysis, among others topics, on the conduct of war (Blattman and Miguel 2010).

To this end, an emerging literature is addressing a number of topics from a different viewpoint, mainly considering spatial interactions and going beyond country-specific factors. For instance, contributing to the literature on the causes of conflicts, Gleditsch (2007) find the existence of a spatial effect between the status (civil war or peace) of a given country and the status of its neighbors, showing that some conditions such as greater cultural diversity, higher levels of democracy, and weaker economic integration, positively affect the probability of war in neighbor countries.

However, these studies are usually affected by the level of aggregation of the data. In fact, Buhaug and Luhala (2005) demonstrate by comparing country-level data and warzone-level data that figures at a nation scale are poor approximations of the characteristics of the real conflict zone. In this vain, Buhaug and Rfd (2006) analyze African civil wars between 1970 and 2001. To do so, they disaggregate the African continent into grids of 100x100Km and allow for the existence of spatial effects between grids. Thanks to this setting, they are able to show that different types of conflicts ("territorial conflicts" and "conflicts over state governance") respond to different types of factors.

Moreover, this strand of the literature includes work not only on the causes of war but also on its consequences. For instance, Murdoch and Sandler (2004) find that civil wars do have an effect on the economic growth of both the country that experienced the conflict and its neighbors.

Nevertheless, to the best of our knowledge little has been done to exploit the interconnectivity measures behind neighboring effects in order to uncover the channels through which conflicts spread out. It seems that some specific spatial econometrics techniques are needed. Beck, Gleditsch and Beardsley (2006) illustrates this on a closely related topic. They explore the importance of political-economic measures of nearness (trade) among countries in comparison to a standard geographic measure. When testing both patterns of spatial dependence, they find that in many cases it is preferable to consider a non-geographic approach.

Contributing to this literature, we will study the Peruvian civil conflict at a fairly low level of aggregation while using multi-parametric Spatial Autoregressive Models. Our particular setting and technique will allow us to depict a specific feature of the Peruvian civil war behaviour. In particular we will pin down the mechanism whereby violence spread out across districts.

III. EMPIRICAL STRATEGY

Recall that we are interested in depicting the mechanism whereby the violence was transmitted across districts. Thus, our strategy must test (i) if there actually was any spatial effect across districts, and (ii) if it was caused by the presence of political-economic factors. That being said, it is important to realize that standard regression models do not consider the spatial structure that the data generating process may have. Therefore, spatial econometrics techniques must be used (Anselin 1999).

Our empirical strategy will be based on the following steps. First, we will show the existence and importance of a standard geographic spatial lag in the severity of the Peruvian conflict. Second, we will demonstrate that this spatial lag is better specified when it is defined as not merely geographical but political-economic —following the TRC Final Report, we will test if the presence of schools played a particular role in the contagion of the conflict. Finally, a number of robustness checks will be presented.

i. Basic Setup

The basic framework is presented by the Spatial Autoregressive Model (SAR) in equation 1:

$$y_i = \alpha + \rho_1 WY + X_i \beta + u_i \qquad (1)$$

Where the dependent variable y_i is the total number of human rights violations in district *i* through 1980 and 1984, per 10 000 citizens. Since *n* is the number of districts, *Y* is a *n*×*1* vector that contains all the y_i of our sample and is pre-multiplied by a *n*×*n* matrix, *W*, which contains the interconnectivity measures (spatial weights) between districts. Each term w_{ij} of matrix *W* measures the nearness between districts *i* and *j*, being different from zero if some interconnectivity exists between districts *i* and *j*, and zero if it does not. It can be easily noticed that the *i-th* row can be considered the neighborhood of district *i*. The spatial weights w_{ij} can be constructed following a number of criteria, but we will use the reciprocal of the distance between districts, a widely used measure (Ward and Gleditsch 2008, LeSage and Pace 2009). Thus, *W* will be defined as:

$$W = \begin{pmatrix} 0 & 1/d_{12} & 1/d_{13} & \dots & 1/d_{1n} \\ 1/d_{21} & 0 & 1/d_{23} & \dots & 1/d_{2n} \\ 1/d_{31} & 1/d_{32} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1/d_{n1} & 1/d_{n2} & 1/d_{n3} & \dots & 0 \end{pmatrix}$$

Where d_{ij} is the euclidean distance between the main squares *("plaza de armas")* of districts *i* and *j*. No district is allowed to be its own neighbour.

Moreover, the interconnectivity among districts can be restricted with maximum and minimum possible distances. For this study we use a maximum distance of 30Km, which is the least distance at which every district has at least one neighbour[7]. In addition, it is important to realize that *W* is symmetric, which makes the estimation procedure considerably easier. Also, for convenience in interpreting the estimated coefficients, we will row-standardize *W*. As a consequence, each row will sum up to unity and no problems with the units of interconnectivity will be generated. Thus, the scalar coefficient ρ_1 will measure the strength and significance of the geographical spatial lag *WY*.

Finally, α is a constant term and X_i is a matrix with observable variables that include*: institutional* controls, such as dummies for the presence of schools and medical facilities, the number of public workers per citizen, and access to drinking water, sewage, and electricity, measured as percentages of households; *socio-economic* controls including population density per squared kilometer, percentage of females, average age, unemployment, percentage of farmers and closely related jobs, and the percentage of people with primary or no education; *geographic* controls such as dummies for whether the district was the capital of the department, the distance in kilometers from the capital of the department, and the area of the district in squared kilometers. In addition, sometimes we will also include *polynomial terms* as well as *fixed-effects* for geographic regions (Chala, Cordillera, Puna, Quechua, Selva Alta, Selva Baja, Suni, and Yunga).

[7] This is an arbitrary decision. However, higher upper bounds could increase the measurement error of our notion of distance. In the following section we will briefly discuss the sensitivity of our results to changes in this band.

ii. Multiple Interconnectivity Measures

Then, we will turn to test the significance and strength of a spatial lag generated by the transmission of the conflict across districts through schools. This hypothesis could be preliminarily tested through a multi-parametric Spatial Autoregressive Model as described in equation 2:

$$y_i = \alpha + \rho_1 WY + \rho_2 W^s Y + X_i \beta + u_i \qquad (2)$$

Where ρ_2 is the coefficient of the spatial lag $W^s Y$. The key term here is the *nxn* matrix W^s, which differs from *W* as described below:

$$W^S = \begin{pmatrix} 0 & 1/d_{12} & 0 & \dots & 0 \\ 1/d_{21} & 0 & 1/d_{23} & \dots & 1/d_{2n} \\ 0 & 1/d_{32} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1/d_{n2} & 0 & \dots & 0 \end{pmatrix}$$

The spatial weight w_{ij}^s is still equal to the reciprocal of the euclidean distance $1/d_{ij}$, but now only those districts with schools *built before 1980* are allowed to transmit the conflict to other districts with schools —the conflict started in 1980, and it was not until 1982 that the violence was substantial and generalized[8]. With this in mind, this specification will test for an additional clustering among districts with schools in addition to that generated by the simple geographic nearness (i.e., *WY*). In other words, the significance of ρ_2 in equation 2 will give preliminary yet important evidence on the hypothesis about schools as the

[8] Final Report. Truth and Reconciliation Commission.

mechanism whereby the violence was transmitted.

It is important to realize that our setting in equation 2 requires a very strong effect of the spatial lag W^sY since *WY* already appears in the same equation. We also report regressions only considering the spatial lag generated by schools, but we only see them as illustrative and preliminary since they could be driven by the effect of the geographical clustering of violence instead of pinning down the effect of geographic contagion through schools in particular. On the other hand, considering both spatial lags in the same equation may substantially increase our standard deviations. Thus, rejecting the null hypothesis in our individual significance tests will be much more difficult.

Finally, as well as in our geographic measure of interconnectivity, W^s is symmetric and row-standardized. All other terms in equation 2 are those already present in equation 1.

iii. Robustness Checks

Equation 2 may be useful in demonstrating the significance of the hypothesis that schools played an important role in the transmission of the conflict across districts, and that this effect is free from pure geographic clusterings. However, some concerns about the exogenous nature of the spatial lag W^sY may still remain. Therefore, further tests are described below.

First, although our regressions consider a large number of controls, polynomial forms, and fixed-effects, we may want to go even beyond and clarify whether schools were really transmitting the conflict across districts. Alternatively, some other institutions may be typically located in the same district as schools and actually drive the result. Equation 3 shows a test for this specific concern:

$$y_i = \alpha + \rho_1 WY + \rho_2 W^s Y + \rho_3 W^P Y + X_i \beta + u_i \quad (3)$$

Where $W^P Y$ is a spatial lag generated by the placebo matrix W^P, which will be constructed similarly to W^S, but allowing some other institution to take the place of schools and determine whether w_{ij} is different from zero.

All other terms in equation 3 are those already contained in equation 1.

Second, we will try to validate our results with a more exogenous sample of schools. Thus far, the schools considered in our sample were created before 1980, the year in which the conflict started. Thus, our sample of schools is free from those that were created after 1980 with the idea of not considering schools that may have been created by any possible reason related to the ongoing conflict.

On this basis, we will perform similar regressions to that of equation 2 but considering a sample of older schools. Although the conflict actually started in 1980, many non-violent actions could have taken place in certain districts even before this year. Thus, maybe the effect of schools created during these non-violent actions could be biased in some way. For instance, *(i)* more schools could have been created as a response to grievances in districts where the conflict was more likely to occur and it finally did (upward bias); *(ii)* more schools could have been created as a response to grievances in districts where the conflict was more likely to occur and it finally did not (downward bias); *(iii)* simply more schools could have been created in districts where the conflict was less likely to occur (possible downward bias).

Therefore, we will restrict our sample to those schools that were created before 1970, the year in which “Shining Path”, the main rebel group in the conflict, was founded[9].

[9] There were mainly two characters in the conflict: the Shining Path, which is the rebel group, and the Government of Peru. A second rebel group was also involved since 1984, but it was almost insignificant in comparison to the Shining Path.

IV. DATA

We use data from four main sources: the National Census of 1981, the Registry of Victims of the conflict (RUV for its Spanish name), and two complementary databases, one with information about schools and universities and another with information about medical facilities.

The Registry of Victims was elaborated by a public institution in charge of creating a database of all the Peruvians affected by the conflict: the Truth and Reconciliation Commission. The data have victim level information with details about the type of human rights abuse (murder, kidnapping, etc.), the district and date, and some demographic characteristics of the victim. Thus, we have 80 546 human rights violations that we have collapsed to a district level in order to create our dependent variable. Importantly, this database excludes human rights violations against the rebel group. Our measure of violence is thus limited to human rights violations against civilians and the military forces of the government.

Some of our control variables at the district level are obtained from the National Census of 1981. We construct these controls following a number of papers[10]. Therefore, our main controls include but are not limited to education conditions,

[10] Mainly, Collier and Hoeffler 1998, Humphreys and Weinstein 2008, Sambanis 2001, and Hengre and Sambanis 2006.

water, sewage and electricity access, female proportion, age, unemployment, the largest economic sector within the labor force, among others.

Our complementary dataset with information of schools and universities was constructed from the National Census of Schools of 2005 and the List of Universities of the National Assembly of University Presidents. Both datasets include the date of creation of each institution. This allowed us to construct a dummy variable for whether there was at least one university or secondary school in the district before 1980, the year when the conflict started. We also use information regarding the type of institution and its administration (private or public).

Finally, our other complementary dataset with information about medical facilities was constructed from the National Registry of Medical Facilities and Services Support (RENAES for its Spanish name) available at the Ministry of Health's website. This database has the date of creation of all medical facilities in Peru, allowing us to construct another dummy variable for whether there was at least one health facility in the district before 1980. We mainly use this information to construct placebo tests for our empirical specification.

i. Sampling

Firstly, we need to remark that our study focuses on the central-southern mountains and forests of Peru. We have excluded the coast, which is the richest zone and was the least affected region of the country, as well as the northern areas that remained mainly in peace until the mid 80s. In practice, we selected non-coastal districts located below latitude -11°, and thus our sample contains 750 districts[11]. In terms of violence from 1980 to 2000, these districts account for 76.79% of the total number of human rights violations.

Secondly, our study focuses on the period 1980-1984 for a number of reasons. First, according to the TRC, the emergence and full display of the violence occurred mostly during those years[12]. Second, in 1985 a new government was established and many economic and political changes took place —Peru experienced rapid economic growth during the first two years of this government and the counter-terrorist strategy also changed. In addition, around 1984 a new rebel group, although much smaller, joined the conflict. Therefore, confining our sample to 1980-1984 seems to be ideal in terms of both the severity and the expansion of the conflict, as well as due to new macroeconomic conditions and counter-terrorist strategy. On this basis, our sample of 750 districts accounts for 93.99% of the total number of human rights violations for the period 1980-1984.

ii. Summary Statistics

Due to the low intensity of the conflict during 1980 and 1981, we consider that our independent variables are sufficiently exogenous to the warfare. Figure 1 shows that these two years only represented 1.55% of the total number of human rights violations during the conflict. Thus, our control variables are hardly contaminated by the effects of violence, and indeed can be seen as initial conditions.

Table 1 shows summary statistics and mean tests of some of our controls. For ease of reading, districts are grouped by peace and war zones. According to the

---

[11] Our results are robust to slightly changes in this boundary.

[12] In fact, the TRC also considers the years 1985 and 1986 as exhibiting great expansion. However, they also mention a number of political and social changes after 1984. Our main results are robust to changes in the sample period.

reported mean tests, those districts that end up exhibiting some level of violence had on average larger and younger populations, larger proportions of females and farmers, and more employment. Moreover, more districts in the warzone had at least one school, as compared with districts in peaceful zones, and a larger number of them were capital cities of departments. On the other hand, districts in conflict zones had less access to water and electricity.

V. RESULTS

i. A First Group of Regressions

Table 2 shows the maximum pseudolikelihood estimation of equation 1, which is a first test for the existence of a spatial lag in the severity of the Peruvian conflict. For all columns, the dependent variable is the total number of human rights violations for the period 1980-1984, per 10,000 citizens, and robust standard errors are used.

The first three columns test the significance of a standard geographical spatial lag. Our estimated coefficient of interest is $\hat{\rho}_1$. Column 1 reports a simple regression in which the spatial lag and one control variable, namely the dummy for the presence of schools in the district, are included. In column 2 a whole set of control variables is included while in column 3 polynomial forms of our controls and geographic fixed-effects are incorporated. Our results show a positive and strongly significant $\hat{\rho}_1$ in all specifications. These findings confirm the intuitive idea that violence was spatially clustered. In other words, the incidence of human rights violations does not followed a random process across districts (conditional on observable initial conditions), but contagion effects were present and conformed violent and peaceful zones.

Although this first result is quite intuitive, it is important to keep in mind that standard econometric models do not allow spatial interactions between observations. In fact, the importance of some control variables usually mentioned in the literature may diminish when spatial interactions are considered (Beck, Gleditsch and Beardsley 2006). However, this discussion is not our main purpose.

Columns 4 through 6 provide a quite unexpected result in the literature. In this case, the specification of equation 1 drops the standard geographic interconnectivity matrix, *W*, and considers W^S instead, which only allows districts with schools to spread out and to receive the violence of the conflict. The coefficient of interest is now $\hat{\rho}_2$. As can be seen, the spatial lag is significant at the 1% level even when polynomial and geographic fixed-effects are included (Column 6). Moreover, recall that our sample of schools contains only those that were built before 1980 as an attempt to isolate the effect shown by the spatial lag from the decision of building new schools when the war already started.

We can now start shedding some light on the clustering process of the violence. While we already expected it to be non-random, we know see that the pattern of clustering was apparently driven by the presence of schools. Furthermore, note that this specification is more restrictive in terms of the shape of the cluster, however, the spatial lag seems to be strongly significant and positive. In fact, the fit of the models specified from columns 4 to 6 are always slightly better than those from columns 1 to 3.

ii. Unmasking the contagion effect

Next, we estimate a multi-parametric Spatial Autoregressive Model in which both the standard geographic spatial lag and the spatial lag generated by schools are included, as proposed in equation 2. On the one hand, this estimation procedure will enable us to determine if the spatial effect generated by schools is significant and additional to that generated by the geographic clustering. On the other hand, we must keep in mind that we will require a very strong effect since both interconnectivity matrices are included and, as a consequence, standard errors will substantially increase.

Table 3 shows the maximum pseudolikelihood estimation of equation 2. As before, $\hat{\rho}_1$ is the estimated coefficient of the standard geographic spatial lag, while $\hat{\rho}_2$ is the estimated coefficient of the school-driven spatial lag.

Column 1 reports a first regression in which only the dummy for schools is used as a control variable. The estimated parameter for the contagion effect generated by schools remains positive (0.3492) and significant at a 5%. However, $\hat{\rho}_1$ is statistically not different from zero. This finding attracts our attention since it reveals that the clustering effect was actually conducted by the presence of schools. In other words, in places where the war unfolded, a contagion effect did occur across districts with schools. Thus, the significance of the standard geographic clustering of Table 2 may have been driven by the existence of schools in the neighborhoods, and not vice-versa. However, this last interpretation must be treated with caution. Recall that our standard errors, although robust, are now substantially larger. While this entails that the strength of $\hat{\rho}_2$ is remarkable, we may also be applying a too strong test on $\hat{\rho}_1$.

In columns 2 and 3 we report the same model but including new control variables. Our results remain. In column 2 we add our sets of institutional, socio-economic, and geographic controls, while in column 3 we also add polynomial terms. Both columns show that the spatial lag generated by schools prevails highly significant at a 5% level.

Column 4 reports the model of equation 2 but with geographic fixed-effects added. The estimated parameter for the schools-driven spatial lag is now significant only at a 10% level.

Thus far, we have demonstrated the statistical significance of a geographic spatial lag driven by the presence of schools in particular. Our findings suggest the violence did spread through schools and not only through simple neighborhood effects, allowing us to identify an specific feature of the Peruvian civil war behaviour. This idea is quite new in the literature and let us contribute to it by showing micro-level evidence on the conduct of war (Blattman and Miguel 2010). Moreover, our results are in harmony with some aspects of the qualitative study conducted by the TRC. In particular, we find quantitative evidence on the important role schools played during the years of violence.

Although these results seem quite strong, some additional robustness check may be required in order to address potential shortages of our empirical design. These plausible problems and their solutions are described in the next section.

iii. Robustness checks

<u>Nothing else but schools and violence?</u>

Although the qualitative literature on the Peruvian conflict suggests the importance of schools for the development of the conflict, some concerns on their quantitative validation may arise. In particular, we may cast doubt on whether schools were the actual channel whereby the violence was transmitted across districts. Put it differently, was there something but schools that propitiated the clustering and that could be misleading our conclusions?

Equation 3 of our empirical specification describes a simple placebo test to reject this hypothesis. We will consider an additional interconnectivity matrix, similar to that generated by the presence of schools, but with a different type of institution. In particular, we will only allow those districts with at least one health facility (hereafter: hospitals) to participate in the interconnectivity of the model. If schools were actually the channel of transmission, then the coefficient of their spatial lag should remain unchanged when competing with the effect of the hospital-based interconnectivity.

Column 1 of Table 4 reports the statistical significance of $\hat{\rho}_3$, which is the parameter associated with the hospital-driven spatial lag. Notice that it is different from zero at a 1% level even when all our controls are included. Thus, the result reported in this regression suggests that the spatiality generated by hospitals do matter for the process of violence.

However, column 2 of Table 4 enlightens us further about $\hat{\rho}_3$. In this model we have considered the dummy for schools, and both the school-driven spatial lag and the hospital-driven spatial lag. The coefficient of our placebo turns out to be statistically not different from zero, while the school-driven spatial lag remains significant at a 1% level. Thus, from columns 1 and 2, we can conclude that although the spatial lag generated by hospitals seemed to be important, it turned out to be driven by the typically simultaneous presence of schools and hospitals[13]. It is also important to notice that standard errors are not as large as in Table 3 since these interconnectivity matrices are more restrictive.

Moreover, column 3 of Table 4 reports the estimation of equation 3 of our empirical strategy. The three spatial lags are incorporated, as well as all our

[13] An unreported simple OLS regression of the schools dummy variable on the hospitals dummy shows a statistically significant coefficient at a 1% level.

control variables including polynomial forms and fixed effects. Neither the standard geographic spatial lag nor the hospital-driven spatial lag are statistically different from zero. $\hat{\rho}_2$ remains at a 10%.

Thus, this simple placebo tests gives us more evidence in support of the hypothesis that schools were the key factor in the mechanism of violence contagion.

In addition, column 4 of Table 4 shows the results of another simple placebo test in which the dependent variable was changed. In an attempt to verify that this mechanism of transmission of violence does not work with other nonsense outcomes, we regressed the percentage of women in each district on the school-driven spatial lag and our control variables. The results are intuitive.

<u>Could the decision of building schools be endogenous to the conflict?</u>

A final concern we address is the validity of our sample of schools. As it was mentioned in our empirical strategy section, our sample of schools is composed by institutions built before 1980. The idea behind this cut-off year is to exclude schools built during the conflict —arguably more schools were intentionally built in places where the violence had already set off.

With this in mind, we could also imagine that, before the violence started in 1980, some non-violent actions could have taken place (i.e., ideological propaganda). Therefore, maybe more (less) schools were built in districts considered to be that were more prone to exhibit violence in the future. Thus, we decided to redefine our sample of schools and retain only those that were built before 1970, the year in which Shining Path, the rebel group, was founded.

$\hat{\rho}_2$ in columns 5 and 6 of Table 4 is the coefficient of this redefined spatial lag. As it can be seen, the estimated coefficient is significant at a 1% level even when all our controls are included as well as the placebo spatial lag and the standard geographical spatial lag.

VI. CONCLUSIONS

In this study we use multi-parametric Spatial Autoregressive Models to demonstrate the existence of neighborhood effects in the severity of the conflict that took place in Peru between 1980 and 1984. Our microlevel strategy let us pin down the mechanism whereby violence was transmitted across districts, finding that schools played a key role in the proliferation of the conflict.

To the best of our knowledge, this finding is new in the literature and let us contribute to it by depicting a specific feature of the war. Importantly, these results are in agreement with the qualitative description of the Peruvian civil war made by the Truth and Reconciliation Commission, the entity in charge of analyzing the conflict.

In particular, our strategy is based on the comparison of two alternative measures of nearness. We first find the statistical significance of a spatial lag defined solely on the grounds of nearness, and then test it against a different interconnectivity pattern in which we only allow neighborhood effects between districts with schools. Our final results confirm that our second definition is the one that matters, and this conclusion survives a number of robustness checks. Thus, we highlight the importance of understanding the channels of interconnectivity.

Our work also calls for the development of higher quality spatial data in Peru. For instance, we are constraint here to euclidean distances between districts, which may exhibit measurement errors in irregular surfaces. (Actual travelling times between districts would be a highly desirable replacement for these distances).

More importantly, our results open up questions for further research. Note that we have thus far described interconnectivity patterns by means of symmetric matrices. While symmetric matrices greatly simplify estimation procedures, they do not necessarily reflect reality. To see why, consider a scenario where violence is transmitted from districts with schools to all their neighbors (districts with or without schools). Alternatively, districts with schools might be more prone to become infected with nearby violence, regardless of the presence of schools in their neighboring districts. Policy implications differ for either scenario, for example, as they may focus on either outward or inward elements of school interactions with their communities. Broadly speaking, we believe this is an instance of a wide array of policy recommendations which will emerge from further work on neighborhood effects in the escalation of conflicts.

VII. REFERENCES

- Anselin, Luc. 1999. Spatial Econometrics. Bruton Center, School of Social Sciences. University of Texas. Unpublished.
- Beck, Nathaniel, Kristian Skrede Gleditsch & Kyle Beardsley. 2006. Space is more than Geography: Using Spatial Econometrics in the Study of Political Economy. International Studies Quarterly, 50: 27-44.
- Blattman, Christopher & Edward Miguel. 2010. Civil War. Journal of Economic Literature, 48(1): 3-57.
- Buhaug, Halvard & Pivi Lujala. 2005. Accounting for Scale: Measuring Geography in Quantitative Studies of Civil War. Political Geography, 24: 399-418.
- Buhaug, Halvard & Jan Ketil Rfd. 2006. Local Determinants of African Civil Wars, 1970-2001. Political Geography, 25: 315-335.
- Camacho, Adriana. 2008. Stress and Birth Weight: Evidence from Terrorist Attacks. American Economic Review, 98(2): 511-515.
- Ciccone, Antonio. 2011. Economic Shocks and Civil Conflict: A Comment. American Economic Journal: Applied Econometrics, 3(4): 215-227
- Collier, Paul. 1999. On Economic Consequences of Civil War. Oxford

Economic Papers, 51(1): 168-183.

- Collier, Paul & Anke Hoeffler. 1998. On Economic Causes of Civil War. Oxford Economic Papers, 50: 563573.
- Collier, Paul & Anke Hoeffler. 2004. Greed and Grievance in Civil War. Oxford Economic Papers, 56: 563-595.
- Fearon, James D. & David D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. American Political Science Review, 97(1): 75-90.
- Gleditsch, Kristian S. 2007. Transnational Dimensions of Civil War. Journal of Peace Research, 44(3): 293-309.
- Gleditsch, Kristian S. & Michael D. Ward. 2008. Spatial Regression Models. Sage Publications, USA.
- Hegre, Havard & Nicholas Sambanis. 2006. Sensitivity Analysis of the Empirical Literature on Civil War Onset. Journal of Conflict Resolution, 50(4): 508-553.
- Humphreys, Macartan & Jeremy M. Weinstein. 2008. Who Fights? The Determinants of Participation in Civil War. American Journal of Political Science, 52(2): 436-455.
- Justino, Patricia. 2007. On the Links between Violent Conflict and Household Poverty: How Much Do We Really Know?. Micro Level Analysis of Violent Conflict Working Paper 1.
- LeSage, James & Pace, Kelly. 2009 Introduction to Spatial Econometrics. Statistics: Textbooks and Monographs. USA.
- Miguel, Eduard; Shanker Satyanath & Ernest Sergenti. 2004. Economic Shocks and Civil Conflict: An Instrumental Variables Approach. Journal of Political Economy, 112(4): 725-753.

- Murdoch, James & Todd Sandler. 2002. Economic Growth, Civil Wars and Spatial Spillovers. Journal of Conflict Resolution, 46(1): 91-110.
- Rodrik, Dani. 1999. Where Did All the Growth Go? External Shocks, Social Conflict, and Growth Collapses. Journal of Economic Growth, 4(4): 385-412.
- Sambanis, Nicholas. 2001. Do Ethnic and Non-Ethnic Civil Wars Have the Same Causes? A Theoretical and Empirical Inquiry (part1). Journal of Conflict Resolution, 45(3): 259-282.
- Truth and Reconciliation Commission. 2003. Final Report. Peru.

VIII. APPENDIX I

Figure 1: Human Rights Violations 1980-2000

Human rights violations from 1980 to 2000 are presented for the entire country. The left axis and the bars indicate the number of human right violations for each year. The right axis and the line indicate the accumulated percentage of human rights violations for each year. From 1980 to 1981 and from 1980 to 1984, the cumulative percentages equal 1.55% and 30.69%, respectively.

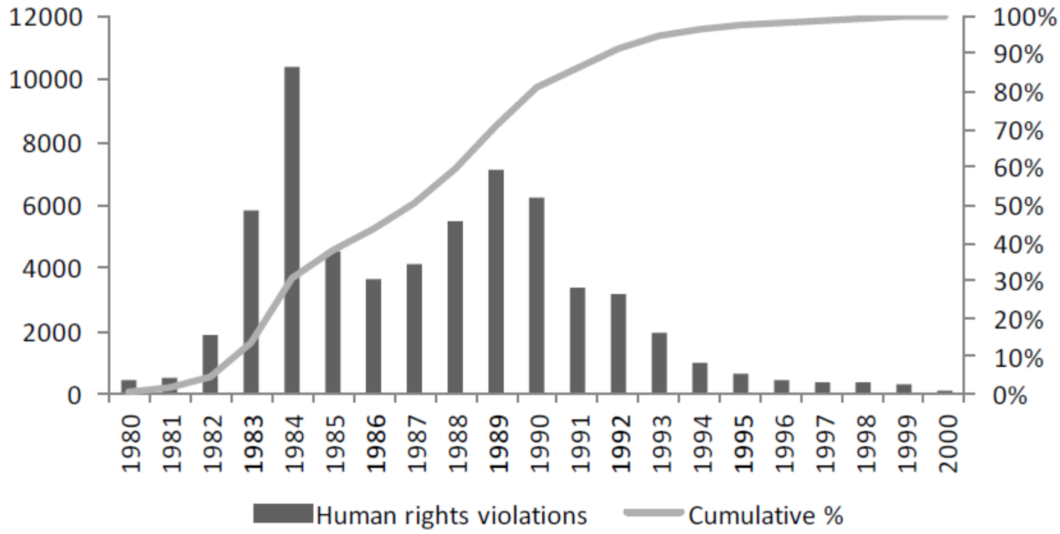

Table 1: Summary Statistics

Summary statistics are presented. We have grouped districts in two: districts in peace are those that presented no human rights violations trough all our sample (1980-1984), while districts in war are those that presented at least one human right violation. The last column presents mean tests. Variables: school is a dummy variable for whether the district had at least one school or not; hospital is a dummy variable for whether the district had at least a medical facility or not, public workers shows the number of public workers per 1000 citizens. Water, sewage and electricity are variables that show the percentage of households with access to drinking water, sewage, and electricity in the district. Population is the number of citizens in the district, population density is the number of citizens per squared kilometer, females shows the percentage of females, average age is the average age of the population, unemployment presents the unemployment rate of the district, farmer is a variable for the percentage of farmers and closely related jobs, and no education is the percentage of people with primary or no education. Finally, capital is a dummy for whether the district was the capital of the department, distance to capital shows the distance in kilometers between the district and the capital of the department, and area is for the area of the district in squared kilometers. All variables presented except school, hospital, capital, distance to capital, and area where based on the National Census of 1981.

Variables	**All** (N= 750)	**Peace** (N= 424)	**War** (N= 326)	**Mean Differences**
School	0.48	0.41	0.58	-0.17***
	(0.50)	(0.49)	(0.49)	
Hospital	0.39	0.38	0.40	-0.03
	(0.49)	(0.48)	(0.49)	
Public workers	23.13	24.28	21.63	2.66
	(24.05)	(27.12)	(19.27)	
Water	16.12	18.16	13.45	4.71***
	(20.14)	(21.37)	(18.11)	
Sewage	4.90	5.44	4.20	1.24
	(12.40)	(13.37)	(11.00)	
Electricity	13.14	16.02	9.39	6.62***
	(22.02)	(24.60)	(17.48)	
Population	5662.75	4601.26	7043.33	-2242.07***
	(9930.36)	(8544.33)	(11357.72)	
Population Density	74.66	88.51	56.65	31.86
	(471.73)	(579.21)	(273.42)	
Females	50.69	50.14	51.41	1.27***
	(3.34)	(3.69)	(2.67)	
Average Age	22.32	22.73	21.78	0.95***
	(2.20)	(2.36)	(1.84)	
Unemployment	4.60	4.89	4.23	0.65**
	(4.26)	(4.68)	(3.62)	
Farmer	42.85	39.53	47.18	-7.65***
	(15.63)	(11.65)	(18.89)	
No education	65.42	65.07	65.89	-0.82
	(9.81)	(10.01)	(9.54)	
Capital	0.011	0.005	0.018	-0.014*
	(0.103)	(0.069)	(0.135)	
Distance to Capital	60.46	62.43	57.90	4.53
	(53.20)	(54.95)	(50.80)	
Area	538.93	580.82	484.45	96.37
	(1535.79)	(1530.02)	(15.43)	

***, **, * are 1%, 5% and 10%, respectively. Means and (Standard Deviations).

Table 2: First-Order Spatial Autoregressive Models

Each column reports a Spatial Autoregressive Model (SAR) with only one connectivity matrix. The dependent variable is the total number of human rights violations in the district through 1980 and 1984, per 10 000 citizens. The estimated coefficient of the spatial lag is labeled $\hat{\rho}_1$ when the geographic connectivity matrix is used (columns 1 through 3), and $\hat{\rho}_2$ when the specified connectivity matrix only allows those districts with at least one school to be interconnected (columns 4 through 6). District-level controls: school is a dummy variable for whether the district had a school or not; institutional controls include a dummy for the presence of hospitals, the number of public workers per citizen, and access to drinking water, sewage, and electricity, measured as percentages of households. Socio-economic controls are population density per squared kilometer, percentage of females, average age, unemployment, percentage of farmers and closely related jobs, and the percentage of people with primary or no education. Geographic controls include a dummy for whether the district was the capital of the department, the distance in kilometers from the capital of the department, and the area of the district in squared kilometers. In addition, polynomial terms for the variables above as well as fixed-effects for geographic regions (Chala, Cordillera, Puna, Quechua, Selva Alta, Selva Baja, Suni, and Yunga) are included.

Model:	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\rho}_1$	0.4696***	0.4351***	0.4054***			
	(0.0541)	(0.0505)	(0.0550)			
$\hat{\rho}_2$				0.4679***	0.4330***	0.3975***
				(0.0535)	(0.0513)	(0.0547)
Controls:						
School	YES	YES	YES	YES	YES	YES
Institutional	NO	YES	YES	NO	YES	YES
Socio-economic	NO	YES	YES	NO	YES	YES
Geographic	NO	YES	YES	NO	YES	YES
Polynomial terms	NO	NO	YES	NO	NO	YES
Geographic fixed-effects	NO	NO	YES	NO	NO	YES
LL	-4745.30	-4726.38	-4683.94	-4743.52	-4724.87	-4683.17
N	750	750	750	750	750	750

***, **, * are 1%, 5% and 10%, respectively. (Standard Deviations).

Table 3: Multiparametric Spatial Autoregressive Models

Each column reports a Spatial Autoregressive Model (SAR) with two connectivity matrices. The dependent variable is the total number of human rights violations in the district through 1980 and 1984, per 10 000 citizens. The estimated coefficient of the spatial lag is labeled $\hat{\rho}_1$ when the geographic connectivity matrix is used, and $\hat{\rho}_2$ when the specified connectivity matrix only allows those districts with at least one school to be interconnected. All controls remain as described in the previous table.

Model:	(1)	(2)	(3)	(4)
$\hat{\rho}_1$	0.1310	0.1226	0.1455	0.1226
	(0.1432)	(0.1326)	(0.1351)	(0.1558)
$\hat{\rho}_2$	0.3492**	0.3230**	0.3121**	0.2583*
	(0.1449)	(0.1367)	(0.1401)	(0.1451)
Controls:				
School	YES	YES	YES	YES
Institutional	NO	YES	YES	YES
Socio-economic	NO	YES	YES	YES
Geographic	NO	YES	YES	YES
Polynomial terms	NO	NO	YES	YES
Geographic fixed-effects	NO	NO	NO	YES
LL	-4743.23	-4724.23	-4699.64	-4682.77
N	750	750	750	750

***, **, * are 1%, 5% and 10%, respectively. (Standard Deviations).

Table 4: Robustness Checks

In this table we present our main placebo tests. The dependent variable is the total number of human rights violations in the district through 1980 and 1984, per 10 000 citizens, except in column 4. The estimated coefficient of the spatial lag is labeled $\hat{\rho}_1$ when the geographic connectivity matrix is used, $\hat{\rho}_2$ when the specified connectivity matrix only allows those districts with at least one school to be interconnected, and $\hat{\rho}_3$ when the interconnectivity restriction is set by the existence of at least one hospital in the district. This latter matrix is used as a placebo for schools in columns 1 through 3. In column 4 we use a different dependent variable (percentage of females) as another placebo for the mechanism of contagion. Finally, in columns 5 and 6 we redefine the matrix that only allows those districts that have a school to be interconnected with a different sample of older schools. All controls are as defined before except in column 4 —the percentage of females is used as the dependent variable and our measure of violence is used as a control.

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Placebo Test:	Hospitals	Hospitals	Hospitals	Different *y*: % Female	Old Schools	Old Schools and Hospitals
$\hat{\rho}_1$			0.0201			-0.1315
			(0.1936)			(0.1741)
$\hat{\rho}_2$		0.3828***	0.2743*	0.0008	0.4270***	0.4339***
		(0.0905)	(0.1460)	(0.0018)	(0.0569)	(0.1342)
$\hat{\rho}_3$	0.3713***	0.1017	0.1272			0.1315
	(0.0542)	(0.0876)	(0.1004)			(0.0977)
Controls:						
School	YES	YES	YES	YES	YES	YES
Institutional	YES	NO	YES	YES	YES	YES
Socio-economic	YES	NO	YES	YES	YES	YES
Geographic	YES	NO	YES	YES	YES	YES
Polynomial forms	YES	NO	YES	YES	YES	YES
Geographic fixed-effects	YES	NO	YES	YES	YES	YES
LL	-4685.73	-4743.08	-4682.41	4941.80	-4681.81	-4681.43
N	750	750	750	750	750	750

***, **, * are 1%, 5% and 10%, respectively. (Standard Deviations).