EC3320

2016-2017

Michael Spagat

Lecture 5

Before diving into the main substance of today's lecture we need to think about the technique known as *logistic regression*.

The standard regression framework is intended for cases where the left-hand-side variable can take on any value, or at least a broad range of values.

In the Berman et al. papers from Lecture 4 the number of violent incidents per capita can take on a lot of possible values. The number of violent incidents cannot be negative but, still, quite a few numbers are possible. So this particular aspect of linear regression seems fairly reasonable in this case.

But some variables have much more of a 0-1, either-or, character.

For example, a person is either employed or unemployed. A country is either at war or it is not at war. Etc.

We can capture such either-or cases numerically by assigning a “0” to one outcome and a “1” to the other outcome.

Also, this reference to 0-1 phenomena should stimulate your fertile minds to think about probability. After all, probabilities are always between 0 and 1.

In our unemployment example it would be nice to have a model that would predict the probability that an individual will be unemployed.

It is unrealistic to hope for a more definite prediction because we will never really know enough to be sure that an individual will be either employed or unemployed. This will depend on many factors, including some random ones.

We can use an old-fashioned linear regression model to make simple probabilistic predictions based just on the independent variable of fatalities. This method is known as a *linear probability model*.

The problem with linear probability models is that they can predict values for the left-hand-side variable that can be negative or greater than 1. Such values make no sense since probabilities must be between 0 and 1.

Logistic regression solves this problem.

The technique is named after the *logistic function* which takes the form $f(x) = \frac{1}{1+e^{-x}}$

Notice that this function will always spit out a number between 0 and 1 no matter what you feed in for x. Logistic regression is, essentially, about estimating an appropriate x to feed into this logistic function. We will return to these technicalities soon.

In 1994 there was a genocide in Rwanda in which hundreds of thousands of people were killed. This occurred between April 6 (when the president was assassinated) and mid July.

For an extremely short overview you can watch this (disturbing) video:

http://www.guardian.co.uk/world/video/2008/dec/18/rwanda-genocide

The video discusses events in Kibuye Prefecture which happens to be the area studied by Philip Verwimp in the paper assigned for this lecture. He focuses on Kibuye because there are good data for this prefecture.

You need to know that there were basically two groups in Rwanda, Hutus and Tutsis, and that genocide in this case refers to Hutus killing Tutsis on a massive scale.

The point of the paper is to study the extent to which the violence was organized, rather than random, in Kibuye Prefecture, which is an area for which Verwimp happens to have good data. He focuses on two aspects which fit well with earlier course material on the Iraq conflict:

1. Weapons choices – this mainly boils down to machetes versus small-arms gunfire.

2. Victim characteristics – these include gender and age as well as profession (on-farm versus off-farm) which is new for this course.

3. Location and timing of atrocities.

Verwimp cites a report of Human Rights Watch that shows a Rwandan Colonel pointing out that the supply of guns is limited so the population should be trained in the use of machetes and other crude, cutting weapons. This already suggests that the killing was fairly well organized. It turns out that the statistics support this suggestion. In fact, this is the main point of the paper.

Here is Verwimp's main hypothesis:

> "The objective of the regime was to kill as many Tutsi as possible under the constraint that firearms and bullets were in short supply."

From this there are secondary hypotheses. Firearms will tend to be used rather than machetes and other crude weapons:

1. Against young adults

2. Against men rather than women

3. Against people killed in large groups

The logic is the same for the above three hypotheses; the people most in a position to defend themselves are killed with firearms rather than machetes (which are harder to kill with).

It is known that the large-scale massacres tended to occur ten days to two weeks into the genocide. Combining this information with hypothesis 3 above suggests:

4. Firearms were relatively more likely to be used a few weeks into the genocide compared to right at the beginning.

Verwimp also notes that people who live in rural areas but have non-farm jobs have relatively high social status in Rwanda and that genocides often prioritize targeting relatively more important people because these are the most able to resist. Combining these two notions gives us the hypothesis:

5. Rural people who work outside agriculture are relatively more likely to be killed with firearms.

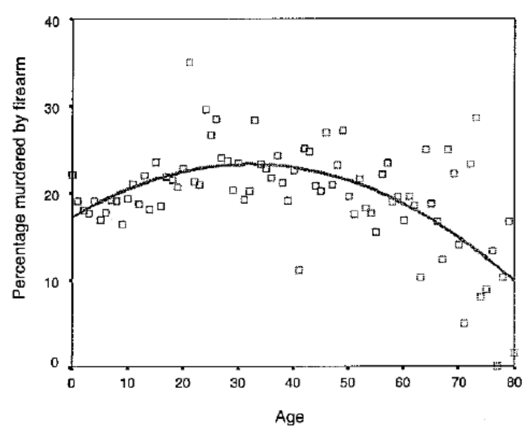
The data come from a survivors group called IBUKA which attempted an exhaustive enumeration using about 200 enumerators. They came up with names of nearly 60,000 victims in Kibuye Prefecture.

Table 4 (next slide) is really horrifying but it gives a good sense of the reality of the situation. Figure 1 (also the next slide) is just a raw, two-way relationship but it is consistent with hypothesis 1 above (slide 8).

Table IV. Killings by Type of Weapon (%)

	Entire file		*Dates of death known*	
Weapon	*Number*	*%*	*Number*	*%*
Machete	31,117	52.8	13,272	51.6
Club	9,779	16.6	4,238	16.5
Gun, rifle	8,706	14.7	4,575	17.8
Grenade	1,058	1.8	609	2.4
Drowned	847	1.4	486	1.9
Hoe, hack	444	0.8	328	1.3
Buried alive	442	0.7	340	1.3
Latrines	437	0.7	150	0.6
Spear	421	0.7	209	0.8
Burnt alive	401	0.7	226	0.9
Pick-axe	337	0.6	192	0.7
Stoned	131	0.2	84	0.3
Hanged	100	0.2	35	0.1
Sword	79	0.1	50	0.2
Starvation	23	0.0	15	0.1
Tractor	12	0.0	7	0.0
Other	636	1.1	197	0.8
Unknown or missing	4,020	6.8	659	2.6
Total	59,050	100	25,719	100

Figure 1. Share of Tutsi Murdered with a Firearm: Realized Probability by Age



N = 25,719

The main analysis is done using logistic regression. In this case firearms (meaning guns, rifles or grenades) will be designated 1 and all other weapons will be designated 0. This is the left-hand-side variable. Right-hand-side variables are:

Age of victim

Age of victim squared

Gender of victim (female is 1 and male is 0)

Gender * age ("*" denotes multiplication)

Occupation of victim (off-farm is 1 and on-farm or student is 0)

Number of days after April 6 the victim was killed

Number of days after April 6 squared

Gender * number of days after April 6

Gender * number of days after April 6 squared

Massacre (1 if victim is known to have been among 100 killed in a specific location within 3 days and 0 otherwise)

Commune dummies (Communes are the administrative layer below Prefectures)

A constant

Table VI. Accounting for Type of Weapon Used

Variable	*Model 1*	*Model 2*
Individual level		
Age	.0329***	.0302***
	(9.86)	(8.31)
Age^2	−.0004***	−.0003***
	(−9.52)	(−7.57)
Gender	−.1958*	.056
	(−1.66)	(0.46)
Gender*age	−.0072***	−.0063***
	(−3.88)	(−3.10)
Off-farm	.5454***	.7065***
	(5.69)	(6.70)
Days after April 6	.1157***	.0522***
	(17.29)	(7.42)
Days after April 6 sq.	−.0014***	−.0005***
	(−12.96)	(−5.15)
Gender*days after	.060***	.018
	(4.64)	(1.49)
Gender*days after sq.	−.0017***	−.0007**
	(−6.32)	(−3.02)
Commune and massacre dummies		
Massacre		2.191***
		(44.23)
Gisovu	1.124***	.6078***
	(6.81)	(3.53)
Gishyata	1.1417***	.1408
	(7.38)	(0.89)
Kivumu	1.5108***	.9590***
	(9.69)	(6.01)
Mabanza	3.152***	2.3064***
	(20.94)	(14.98)
Mwendo	.5965***	
	(3.35)	
Rwamatamu	2.398***	2.6472***
	(16.21)	(17.42)
Constant	−4.851***	−5.060***
	(−28.85)	(−29.35)
N	23,650	21,536
	not weighted	not weighted
	not clustered	not clustered
R^2	0.11	0.20
Log likelihood	−10,577.96	−8,930.64

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.
z-values in parentheses.

I'm going to plug in some numbers to make sure people understand how to interpret the logistic regression table.

First, let's work with model 1. Consider a 25-year-old male who worked off-farm and was killed right on April 6 (just to simplify the calculations by making a lot of the terms into 0's) in Kivumu Commune.

Plugging in these numbers we get: $0.0329x25-0.0004x25^2+0.5454+1.5108-4.851=-2.2223$ The logistic function takes the form $f(x)=\frac{1}{1+e^{-x}}$. In this case we plug in -2.2223 for x which gives almost exactly 0.1 (0.097 to be more exact)

Thus, the model predicts that a 25-year-old male who worked off-farm and was killed right on April 6 in Kivumu Prefecture had a 1 in 10 chance of being killed by a firearm. This is actually a fairly low probability since firearms were the weapon in 18.3% of the cases for which both the date and the weapon are known.

If we switch to model 2 and add that this person was killed in a massacre then the probability changes to 0.35 (You should check this using all the coefficients from model 2.) So switching models has a big effect here.

Remember that Verwimp could have used the common linear regression model rather than logistic regression. If he had done this I suspect the results would have been quite similar to what we see in the paper and the coefficients would have been easier to interpret.

As noted above the main reason for preferring logistic regression is that it forces the predicted probabilities to be between 0 and 1, as probabilities have to be. With linear regression you can end up predicting negative probabilities or probabilities greater than 1.

Note that we ignore some complicated technical issues surrounding the problem of how to deal with missing data. For example, often you don't know dates of deaths. These issues are discussed in the paper.

The biggest effects come from the massacre variable and some of the location variables.

The big effect for Mabanza comes largely from the fact that people there were herded into a football stadium and killed with machine guns.

Verwimp concludes that all five of his hypotheses (slides 8 and 9) are supported, although I would say that hypothesis 2 (that being a man rather than a woman increases the chance of being killed with a firearm) does not get such strong support.

At a more general level, these results support the notion that the genocide was well organized and systematic rather than just being a spontaneous uprising.

Violence Dynamics in the Palestinian-Israeli Conflict

There is a general perception that the Israelis and the Palestinians are locked into a "cycle of violence." A google search of "israeli palestinian conflict cycle" turns up all sorts of stuff:

https://www.google.co.uk/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=israeli%20palestinian%20conflict%20cycle

The general idea is that one side strikes the other side which then hits back, provoking the first side to strike again, etc. In other words, the two sides just go back and forth with each side always thinking of itself as retaliating for the latest atrocity committed by the other side. Such dynamics are sometimes described as "tit-for-tat" or "an eye for and eye and a tooth for a tooth."

The question for today is whether these cycle-of-violence dynamics really fit the reality of the Israeli-Palestinian conflict.

Jaeger and Passerman give two main reasons why you would not necessarily expect to see cycle-of-violence patterns.

1. There could be “incapacitation effects”, i.e., an attack by one side might reduce the capacity for the other side to make attacks, at least temporarily.

2. There could be a “deterrent effects”, i.e, a show of force by one side might intimidate the other side, making it afraid to respond.

The cycle-of-violence notion seems to be based on thinking mostly in terms of what might be called a “revenge effect.”

The Jaeger and Passerman article uses data from B’Tselem, an NGO, to address the cycle-of-violence question.

B’Tselem provides detailed data on conflict fatalities (among many other things). The data go up almost to the present but Jaeger and Passerman did their work several years ago so their analysis covers September 29, 2000 through January 15, 2005. During this period B’tselem recorded 3,244 Palestinian fatalities and 994 Israeli fatalities in the conflict. In addition to media reports, B’Tselem uses their own field work plus a variety of other documents from government sources and NGO’s.

I have uploaded a spreadsheet onto the Moodle page that shows the richness of the B’Tselem data. Jaeger and Passerman use only some of this detail.

VOL. 98 NO. 4 JAEGER AND PASERMAN: THE CYCLE OF VIOLENCE? 1593

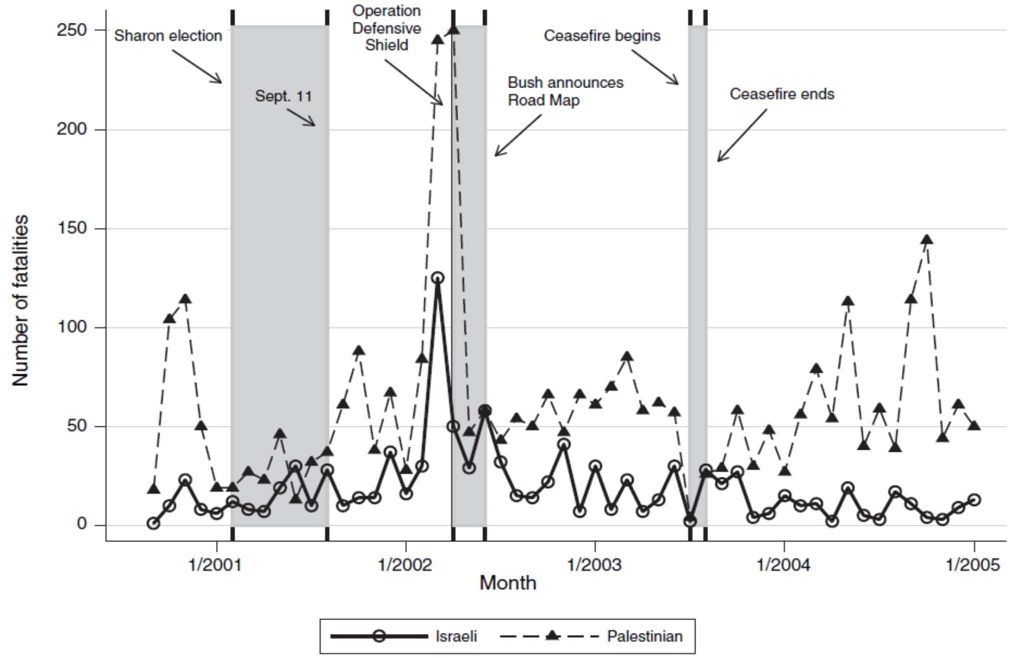


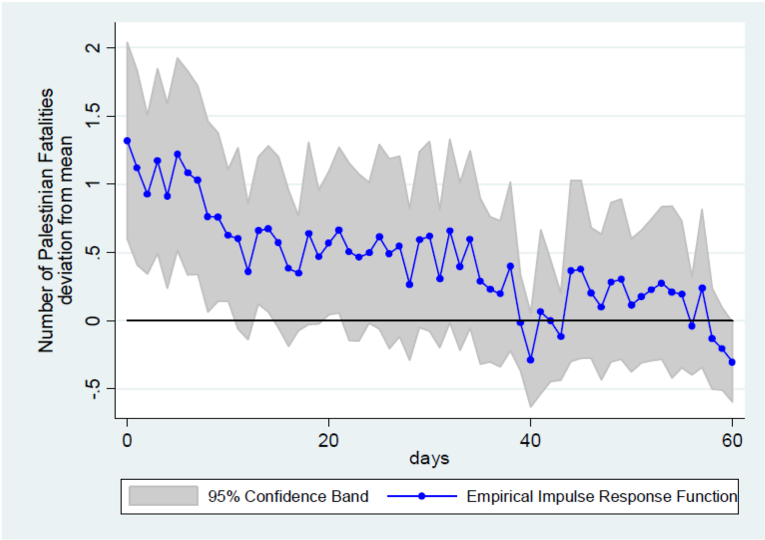
FIGURE 1. MONTHLY NUMBER OF FATALITIES

Source: Authors' calculations from B'Tselem data, from 29 September 2000 to 31 January 2005.

Above you can see the monthly time series for Israelis and Palestinian fatalities with some key dates in the conflict marked.

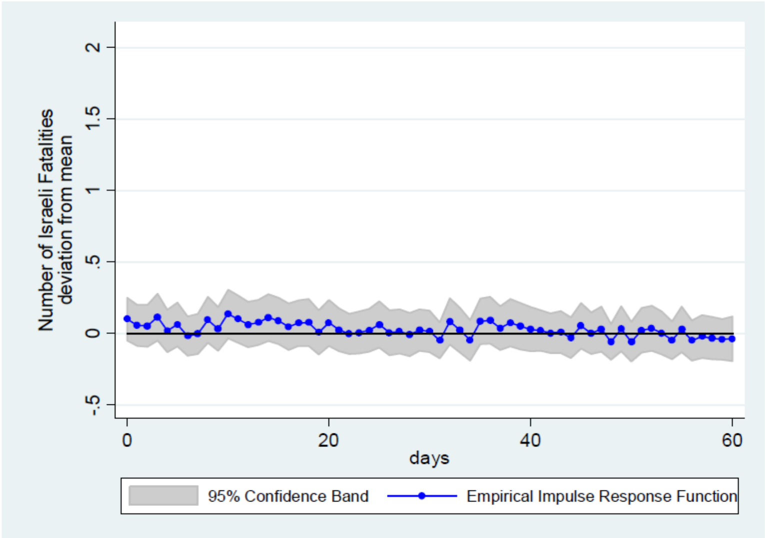
The next two pictures give the main idea and results of the paper.

Figure 2a: Empirical Israeli Response Function



Source: Authors' calculations from B'Tselem data, from 29 September 2000 to 15 January 2005.

Figure 2b: Empirical Palestinian Response Function



Source: Authors' calculations from B'Tselem data, from 29 September 2000 to 15 January 2005.

Let's understand Figure 2a first by considering the following procedures (which we have to spread over two slides, unfortunately).

1. Number all the days, between September 29, 2000 and January 15, 2005 (1, 2, 3, 4, ….). I think the number of days in this range is about 1,570. There are 3,244 Palestinian fatalities during this period so the average number of Palestinian fatalities per day is almost exactly 2.0 during this period.

2. List all the days in which the Palestinians cause at least one Israeli fatality. Suppose these are the days are numbered 3, 7, 9, 14, 18, etc..

3. Now list all the days that are one day after a day in which the Palestinians cause at least one Israeli fatality. Or course, given point 2 just above, these must be days 4, 8, 10, 15, 19, etc.. Compute the average number of daily fatalities of Palestinians on these days. This number turns out to be around 3.4. Subtract off 2.0 (the average daily number of Palestinian fatalities over the entire period) and we have the first data point in Figure 2a which is 1.4.

4. Now list all the days that are two days after a day in which the Palestinians cause at least one Israeli fatality. These will be days 5, 9, 11, 16, 20, etc.. Again, compute the average number of Palestinian fatalities on these days. This number turns out to be around 3.2. Subtract off the average over all days (2.0) and you get 1.2 which is the second data point in figure 2a.

5. Now list all the days that are three days after a day in which the Palestinians cause at least one Israeli fatality. These will be days 6, 10, 12, 17, 21, etc.. And so on and so forth. (Notice that day 10 is both one day after a day with Israeli fatalities and also three days after a day with Israeli fatalities. This is not a problem. As you keep moving through this procedure there will be many days that will be on several lists.)

Following these procedure you get all of the points plotted in Figure 2a. Such a curve is called an “impulse-response function”. In this case killings of Israelis are the impulse and killings of Palestinians are the response.

Notice that for every point on the impulse-response function we always subtract off the average daily number of Palestinians killed over all the days. So each point is interpreted as *excess killing* above and beyond the usual level of killing.

The shaded areas around the points are confidence intervals. For this course we won't worry about how these are computed.

The main take-home point from Figure 2a is that there is some period of time, lasting around ten days or so after days with Israeli fatalities, during which there are an elevated number of Palestinian fatalities on average.

It is easy to understand figure 2b now that you understand figure 2a. Basically, you just go through the algorithm given on the previous slide, interchanging the words “Israeli” and “Palestinian”. Since it is easy for me to cut and paste and change a few words I spell it out in detail.

1. Number all the days, between September 29, 2000 and January 15, 2005 (1, 2, 3, 4, ….). There are 994 Israelis fatalities during this period so the average number of Israeli fatalities per day is about 0.6 during this period.

2. List all the days in which the Israelis cause at least one Palestinian fatality. Suppose that these are the days numbered 3, 7, 9, 14, 18, etc..

3. Now list all the days that are one day after a day in which the Israelis cause at least one Palestinian fatality. Or course, given point 2 just above, these must be days 4, 8, 10, 15, 19, etc.. Compute the average number of Israeli fatalities on these days. This number turns out to be around 0.7. Subtract off 0.6 (the average daily number of Israeli fatalities over the entire period) and we have the first data point in Figure 2b: 0.1.

4. Now list all the days that are two days after a day in which the Israelis cause at least one Palestinian fatality. These will be days 5, 9, 11, 16, 20, etc.. Again, compute the average number of Israeli fatalities on these days. This number turns out to be around 0.6. Subtract off the average over all days (0.6) and you get 0 which is the second data point in figure 2a (It's slightly above 0 actually but right around there.).

5. Now list all the days that are three days after a day in which the Israelis cause at least one Palestinian fatality. These will be days 6, 10, 12, 17, 21, etc.. And so on and so forth.

For figure 2b the take-home message is that following days in which Palestinians are killed by Israelis there are not subsequent periods during which there are elevated rates of Israelis getting killed by Palestinians.

This is it in a nutshell. Jaeger and Passerman argue that there isn't really a *cycle* of violence because the reactions are asymmetric: the Israelis respond just after getting hit but the Palestinians don't respond just after getting hit by the Israelis. Much of the rest of the paper, e.g., the stuff about "Granger causality," is about making this point as rigorously as possible. We won't worry about these details.

I do, however, want to think a little bit about what kind of stories might explain these empirical findings. Here are two ideas:

1. Both sides would like to retaliate promptly when some of their own people are killed but the Israelis are much more powerful militarily than the Palestinians are. Thus, the Israelis are able to respond whereas the Palestinians are not. A slight variation on this theme is that to have effective attacks the Palestinians have to be unpredictable in their attacking so they generally refrain from attacking during the most predictable moments such as just after there have been Palestinian fatalities.

2. The Palestinians actually do respond but these responses aren't captured well by the data that Jaeger and Passerman analyze.