Unemployment and Violence in Northern Ireland: a missing data model for ecological inference.*

James Honaker
Pennsylvania State University †

November 18, 2011

Abstract

Contrary to the body of literature in political violence, and the rhetoric of many parties of the conflict, time-series models of "the troubles" in Northern Ireland by White(1993) and Thompson(1989) have found no evidence that economic conditions effect the intensity, sources or direction of violence. I show that several methodological flaws exist in previous models. They fail to address the discrete, count nature of the data, the contagion present from aggregation over time, pooling issues from different types of violence, and the over dispersal of deaths. However, the key problem, acknowledged even by the authors themselves, is that all measures of unemployment aggregate Protestant and Catholic unemployment rates into one single measure.

Clearly, Republican paramilitary violence should be caused chiefly by the Catholic unemployment rate rather than the aggregate unemployment rate. Possibly also Loyalist violence may be a result of the Protestant unemployment rate. However, little historical data exists on disaggregated rates of unemployment, thus previous studies forced to use aggregate measures have found no connection between economic conditions and the incidences of violent death. Using a model that combines methods of Multiple Imputation to recover missing data (King Honaker Joseph Scheve 2001) and the literature of models for Ecological Inference problems (especially King 1997) I estimate the disaggregated unemployment rates by religion from the available data.

Unemployment is shown to be a leading cause of the violence by Republican factions in Northern Ireland. I also estimate the effectiveness of the various security forces (RUC, UDR, British Army) in lowering the death rate. I show these forces have differential effectiveness depending on both the sectarian source and target of violence using a model of probabilistically distributed lags appropriate for time-series event count data.

*My thanks to Matthew Blackwell for his extensive help in data collection. My thanks are also due for valuable discussions with Jack Buckley, Garrett Glasgow, David Karol, Rafael DeAquino Villar and Christopher Zorn.

†Department of Political Science (207 Pond Laboratory, University Park, PA 16802; tercer@psu.edu)

1 Economics, Politics and Violence

The connection between economic conditions and political violence is central to multiple approaches to the study of conflict. For normative policy makers, economic conditions are often the only policy instruments with the prospect of short term manipulation or improvement. For more positivist observers, measures of aggregate economic conditions are frequently the only variables with consistent explanatory power, other than previous levels of violence. The unfortunate relevance of the topic has brought scrutiny and charged debate to both vantages in this long-standing literature.

At one time the connection seemed concrete. Economic conditions have served as a leading and robust predictor of conflict throughout the quantitative literature on political violence, from its earliest roots (Russett 1964, Feierabend et al 1969, Hibbs 1974). Measures of aggregate economic conditions (Fearon and Laitin 1999, Sambanis 2001), economic growth or instability (Olson, or closely related, Hovland and Sears 1940, Tolnay and Beck 1995) or economic inequality (Gurr 1968, Gurr 1971, Muller 1985, Muller and Seligson 1987) are often the only variable with explanatory power, other than previous levels of violence. Most of these studies are cross-national, sometimes additionally across time.

However, a number of recent counter arguments have formed, partially as the result of increasingly rich and textured data, and partially as the result of shifting units of analysis. Starting with Russell and Miller (1977, 1983) a number of authors have collected detailed intra-national series of violence, support for violence, or the personal histories of instigators of violence (particularly Kruger and Maleckova 2003, Berrebi 2003) and such intra-national level data tends to add no support to a connection to economics. For example, individual instigators of violence generally tend to come from better economic circumstances than average in their country or region. Kruger and Maleckova, from their own work and a synthesis of other chosen studies rather strongly conclude "Any connection between poverty, education, and terrorism is indirect, complicated, and probably quite weak." [1]

In summary, cross-national aggregate studies find a clear connection between political violence and economic conditions, but in sub-national studies there is a growing literature of null results.

[1] In another article, the authors similarly argue that "Instead of viewing terrorism as a direct response to low market opportunities or ignorance, we suggest it is more accurately viewed as a response to political conditions and long-standing feelings of indignity and frustration that have little to do with economics." KM (2003)

1.1 The Case of Northern Ireland

Northern Ireland is probably the most detailed and documented case in the study of political violence, and yet the best empirical analyses yield impossible results. The two most thorough studies, by Thompson (1989) and by White (1993), find little connection between economic conditions and the level of violence, and what little effect they do find is in the wrong direction, with increased unemployment decreasing the level of violence. This exception is important because it stands in opposition to most any theory of political violence (rational actor, modernization, or deprivation theory) and to the preponderance of empirical evidence previously touched on. This has led some outside commentators to speak of Northern Ireland as an innovation in violence, a precursor of instability that will be mirrored in other industrial democracies, or even a "post-materialist insurgency"[2]. However, these exceptional empirical findings are notable not simply because they oppose the current scholarship of violence, but because they stand in opposition to much of the rhetoric of the conflict itself. Republican groups often cast the violence as a fight of Catholics against economic persecution by Protestants.

For the last fifty years Northern Ireland has had a significantly higher rate of unemployment, from 2 to 5 times higher, than the rest of the United Kingdom. Moreover the Catholic unemployment rate in Northern Ireland is significantly higher than the Protestant unemployment rate. One long-standing explanation for the difference of unemployment by religion has been economic persecution by the state and by Protestant dominated business and industry. Other commentators attribute the difference to a natural consequence of demographic differences between the two groups, such as fertility, education and socio-economic structure (Compton 1981, Kovalcheck 1987, Eversley 1989, Howe 1990, Sheehan and Tomlinson 1999). This disparity in unemployment is commonly invoked as one of the chief grievances of the Republican paramilitaries.

As an exceptionally brief overview, simply to define terms, Republican paramilitaries are almost entirely composed of Catholics, and seek a united Irish Republic. In conflict with them are the Loyalist paramilitaries who wish Northern Ireland to remain a part of the United Kingdom, and are composed primarily of Protestants. The third party to the conflict are the Government state forces: the British Army, the Ulster Defense Regiment (a British Army Unit composed of people from

[2]

Northern Ireland almost entirely stationed in Northern Ireland) and the Royal Ulster Constabulary, the police force.

2 Previous Findings

Two key studies have tried to investigate the causal role of unemployment in "the troubles" in Northern Ireland. Thompson (1989) studying the total number of deaths in each year, uses six combinations of variables in three different time-pools of the data for 17 different models[3]. In these seventeen models unemployment is significant only four times. In each of these four coefficients, the effect is also in the wrong direction. Increased unemployment causes the level of violence to decrease. In all these models, the only other variables that explain the level of violence are the lagged levels of violence, and either the level of security forces, or a dummy variable for the period after 1968. The British Army moves into Northern Ireland in 1968 so we might assume these are the same effect.

White (1993) disaggregates the counts of total fatalities by the source (agent) of violence, dividing the total into deaths by Republican, Loyalist, or Government forces. He also disaggregates by time and looks at the monthly level. Again, as with White he finds limited support for unemployment as a cause of violence, and the results he does find are once more in the wrong direction. He finds no support that unemployment causes violence by Republican paramilitaries whereas, contrary to expectation, finds that Loyalist violence decreases with rising unemployment. As White writes, this is contrary to his theory, but may be caused by the aggregation of unemployment measures. "Counter to deprivation theory, percent unemployed leads to fewer deaths by Loyalist paramilitaries. This may reflect the inability to measure unemployment separately for Protestants and Catholics..." (p.582) Similarly, rising unemployment causes decreased numbers of deaths by government forces, although the expectation here is less clear, and White does not discuss this result.

3 Data

For the counts of fatalities in Northern Ireland I use the work of Malcolm Sutton who has collected information about each individual death from a wide range of sources. Indeed the array of sources Sutton synthesizes includes the Irish newspaper sources of Thompson and the book source of White

[3]One set of variables is not appropriate to one of the time-pools, which is why there are not $6 \times 3 = 18$ models.

		Victim			
		Republican	Loyalist	British Forces	Civilian
Agent	Republican	186	45	1088	735
	Loyalist	42	91	14	873
	British Forces	150	14	12	192
	Unknown	17	1	7	56

Table 1: *The numbers of deaths in Northern Ireland by agent and victim, 1969-2001.*

(Irish Information Partnership, 1987). Sutton's work was previously published (Sutton 1994) and has been updated to 2001 by the author, online (CAIN 2003).

Similar to White, I am interested in disaggregating the total death measure into the groups responsible for the individual killings. The source or agent of the death is broken down into four categories, Republican paramilitaries such as the Irish Republican Army, Loyalist paramilitaries such as the Ulster Volunteer Force, and British Forces which include both the British Army and Ulster Defense Regiment as well as government run Northern Irish forces such as the Royal Ulster Constabulary. A fourth category is created when the agent of the death is unknown. However, in addition to this I also want to disaggregate by categorizing the victim. The combination of these two categories gives a "direction of death" count. This is summarized in table 1. The victim categories use the three previous armed groups and a fourth category of civilian.

Several things can be seen in this cross tabulation. The single largest agent-victim death count is the 1088 British forces killed by Republican groups. The number of civilians killed by Loyalist and Republican groups are each of similar scale. Also, significant numbers of killings take place *within* the paramilitary groups. Loyalist members kill 91 other members of loyalist groups, and this in almost entirely due to feuds between different groups and infighting within the same organization. The 186 deaths of Republicans by Republicans are a mixture of similar feuds and infighting, as well as a sizable number of accidental deaths caused by premature bomb explosions.

In addition to the total number of deaths, a sense of the conflict can be seen from looking at the levels and composition of deaths over time. In the top of figure 1 is a rolling average (smoothed mean) of the daily mean number of deaths over the previous year. Clearly the decade of the 1970s saw the most violence, with the early seventies having the peak level of fatalities. At this peak the fatality rate surpasses one death every day. Violence in the late seventies through mid-nineties was at a constant level of about 0.3 deaths per day, or a little more than two deaths every week. In the bottom figure the deaths at any particular point in time is divided compositionally into the percent attributed to the two paramilitaries and the British forces. The only clear pattern is that deaths caused by British forces accounts for the largest percent of recent fatalities very early on in the conflict (before 1972). Loyalist paramilitaries are responsible for their largest relative fraction of deaths early on and late in the conflict, with the majority of deaths in the middle period caused by Republican paramilitaries.

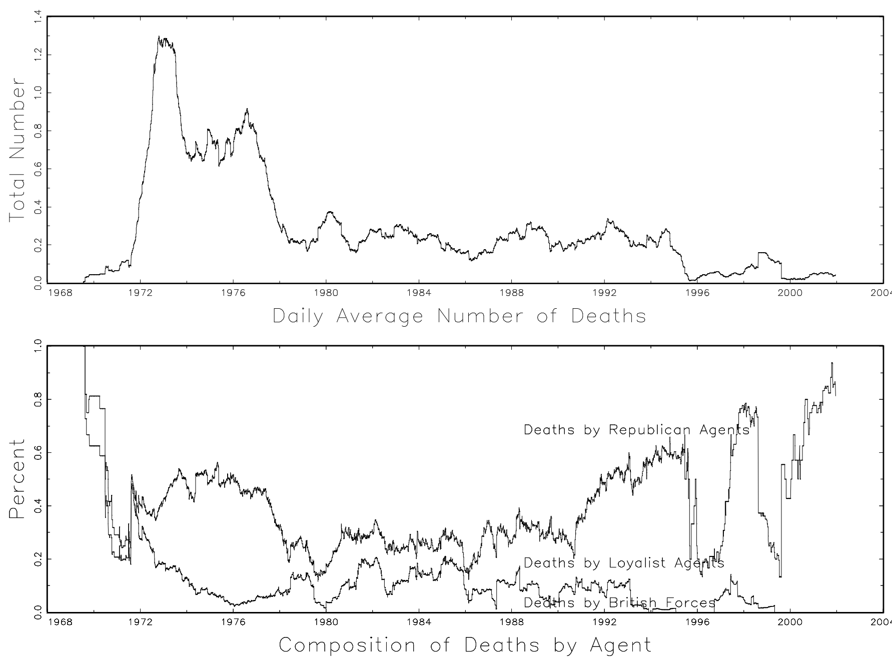

Figure 1: *The top graph shows the mean number of deaths per day, as a rolling average over the last year. The bottom shows the composition of this rolling average, by the agent of violence. At its peak, the violence averaged over one death a day in the early 1970's.*

4 Estimation of Unemployment by Religion

As previously stated, both authors used the total unemployment rate as a variable to explain the death rate. Given the discussion above, however, we should hypothesize that Republican violence is caused by the unemployment level *among Catholics* and Loyalist violence may similarly be caused by unemployment *among Protestants*. White is very aware of this problem and even regrets that "Unfortunately, unemployment figures for Catholics and Protestants separately are not collected systematically in Northern Ireland." and submits that his unexpected results may be an artifact of this problem, as in the quote in section 2. Had they been collected, White would certainly have used these instead of using unemployment aggregated across religions.

Indeed, this data largely does not exist, however, there are several approaches possible to estimate separate unemployment rates by religion. The oldest commonly used approach is Goodman's regression (Goodman 1953). Freedman's neighborhood model (Freedman et. al 1991) is parametrically indistinguishable from Goodman's model but has drastically different interpretations of these parameters, and thus different estimates of these disaggregated quantities of interest. King's model of Ecological Inference (King, 1997) is the most commonly used and accepted model.

Monthly data of the aggregate unemployment rate in different government districts in Northern Ireland is collected by the Central Statistical Office[4]. These districts are "travel-to-work areas" constructed and defined as geographic regions wherein the majority of residents that live in that region also work in region. Therefore the regions can be considered relatively independent self-contained labour markets. Government census data, collected once every decade, also exists describing the composition by religion of these districts. Therefore, in each district we know the total number of Catholics (X_C) and the number unemployed (U), but we don't know the number of Catholics who are unemployed, which is the variable we are interested in. A common perspective on the problem is to lay these terms out in a table and see that we know all the values on the marginals of the table (the Roman Letters) such as the fraction of Catholics, X_c and the unemployment rate, U[5]. However, the elements inside the table (Greek letters) such as the Catholic and Protestant

[4]This sequence of data is published monthly in an economic statistical abstract. The name has changed repeatedly over the years studied, in the following sequence: *Ministry of Labour Gazette*, *Employment and Productivity Gazette*, *Department of Employment Gazette*, *Employment Gazette*, *Labour Market Trends*.

[5]In the increasingly conventional notation, X, is generally the aggregate fraction of some demographic characteristic, and the rate of the dependent variable is labeled T, as the literature is often grounded in $\underline{t}$urnout rates. Here

	Unemployed	Employed	
Catholics	β_C	$1-\beta_C$	X_C
Protestants	β_P	$1-\beta_P$	$X_P(=1-X_C)$
	U	$E(=1-U)$	

Table 2: *The Ecological Inference Problem for Unemployment by Religion: Known data (Roman letters), such as the fraction of the population that is Catholic, or the fraction unemployed, are at the marginals of the table, but the elements inside the table, the unemployment rates by religion (Greek letters), are the quantities of interest.*

unemployment rates, β_C and β_P, need to be estimated. We could construct table 2 for each district, and if we could ascertain the number of unemployed Catholics and Protestants, in each district we then know the Catholic and Protestant unemployment rates in Northern Ireland by the sum of these district values, weighted by the populations of Catholics and Protestants. The three models previously mentioned all propose different methods to estimate these quantities from the known data. In many applications researchers are interested in estimating the elements inside the table for each district, however, we will need only the weighted sum of these individual rates for our final analysis[6]. That is, we only need the national unemployment rates, not the district unemployment rates.

5 Refinements and Complications of King's EI

There is a mounting literature on the applied aspects of King's Ecological Inference model (hereafter, simply EI). Rather crudely, we might divide the literature, like Gaul, into three parts. The first part studies extensions to the model assumptions (see for example the edited volume of King Rosen Tanner 2004, Ferree 2000, Wakefield 2005). The second partition, examines how robust the model is when model assumptions are not met (for example Cho 1998, Ansolabehere and Rivers 2002, Cho and Anselin 2002, Cho and Gaines 2004). The third part has considered how to appropriately use EI estimates when they themselves are not the final quantity of interest, but variables to be included in a latter model (Adolph and King 2004, Adolph et al. 2004, Heron and Shotts 2004, and tangentially Lewis and Linzer 2005).

we use U instead of T as a more relevant mnemonic.

[6]Sometimes referred to in the EI literature as a precinct level analysis.

As previously mentioned, the unemployment rates by religion are generally not known from month to month. However, census data is periodically collected. In April of 1971, and every ten years after, census data is collected in Northern Ireland, from which it is possible to calculate the true value of the unemployment rate by religion. If we had this census data collected every month, we would of course use it, and there would be no need for the ecological model.

Figure 2 shows the available census data. All of the observations sit above the $y = x$ line, thus it can be clearly seen that Catholic unemployment is higher than Protestant unemployment, in every TTWA, in every census. Sometimes this is by as much as a factor of two. Obviously then the assumptions of the Neighborhood model are not met, which premises that these rates are the same in every district, and national differences in unemployment come about because of differences in the densities of religion across different districts. Furthermore, the most narrow interpretation of Goodman's regression is not met as in any year there is clearly a distribution of unemployment across districts. This distribution is centered well inside the unit square, and is approximately doubly symmetric within any year. The center of the distribution shifts substantially across the decades. These observations, and some further assumptions can help us tailor a model for estimating unemployment.

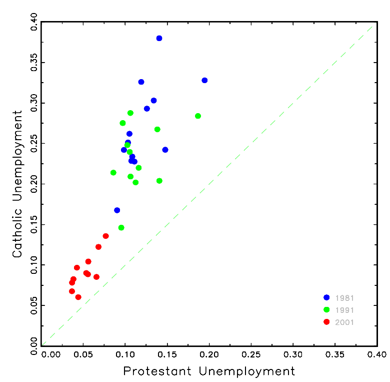

Figure 2: *Relative rates of unemployment by religion in each "travel-to-work area" on the three census dates for which this is available. In every TTWA in every census, the Catholic rate of unemployment is higher than the Protestant rate, by as much as a factor of two.*

If we had the Census collected every month, then we would have exactly the data we wanted. Instead, we have the data we want on three months, and aggregate data throughout the rest of the period. However, these variables are clearly related. We might think of the structure in terms of a missing data problem (Little and Rubin, Schafer 1997, King et al. 2001, Honaker et al. 2002). We possess (very) incompletely observed values of one of our key independent variables. However, if we had good predictors that could forecast, not necessarily causally, predictions of the rate of unemployment in the missing months, then we could proceed. The EI model, to the extent that it is correct, should be a good predictive variable in an imputation model of unemployment by religion. Even if we fall into circumstances where the EI assumptions fail, and the EI estimates are not equal to the true rates of unemployment, the estimates may still be excellent forecasting variables, simply with a coefficient other than one. Moreover, to the extent that EI is mislead in its predictions, we may be able to correct the estimates and find a function of the EI estimates that is a good predictor of the true level of unemployment by religion. Thus even if the first branch of literature applies to our data, we will be fine if the EI estimates are related systematically to the truth, even if they are not themselves correct, or if they can be corrected.

A key complication for general purpose imputation methods is to create an algorithm that can deal with arbitrary sets of missingness patterns. Here the missingness pattern is somewhat simple, as there is only missingness in one variable. Commonly, in such circumstances, the likelihood of the model of interest can be factored into an observed and partially observed contribution and no multiple imputation is required. However, one complication is that the data that we have is at a different level of aggregation (the district level) than the unemployment rates we need (at the national level).

We need to create some number, m, of completed datasets by drawing m simulations of these parameters (as you would do in bootstrapping for standard errors in the model in the fully observed data[7]) and completing one version of the dataset with each one of these sets of parameters. Often m as low as five or ten generally works.

[7] See for example the literature on Bootstrapping, and King, Tomz and Wittenberg.

Specifically here, the model employed is that:

$$(b_{it}^c, b_{it}^p) \sim f_{BivariateNormal}(\mu_{it}^c, \mu_{it}^p, \sigma_c, \sigma_p, \sigma_{cp}) \quad (1)$$

$$\mu_{it}^c = \gamma_0 + \gamma_1 \hat{b_{it}^c} + \gamma_2 G\hat{M}_t^c + \gamma_3 N\hat{E}_{it}$$

$$\mu_{it}^p = \gamma_4 + \gamma_5 \hat{b_{it}^p} + \gamma_6 G\hat{M}_t^p + \gamma_7 N\hat{E}_{it} \quad (2)$$

Where b represents the true unemployment, $\hat{b}$ the estimated unemployment from the EI model, GM the estimated unemployment from Goodman's regression (which does not vary by district, i), and NE the estimated unemployment from the Neighborhood model (which does not vary by religion).

Since many of the diagnostics for the EI method, and the secondary literature on robustness to model assumptions, are concerned with relationships between u_i and x_i, it is tempting to include X_c as a covariate in the imputation model. However, just as figure 2 clearly shows that the unemployment rates varies substantially overtime, then whatever aggregation bias exists between unemployment and X_c would probably not be constant either, and thus not correctly set up by simply including the proportion Catholic as a predictor in the imputation model. Instead, if X_c has a consistent relationship with unemployment, then we would need to interact it in some fashion with the other aggregate data. Including the GM and NE terms allow two specific models of how X may be consistently related to unemployment, while allowing the average level of unemployment, and thus this strength, to vary over time. (In this particular dataset, the Goodman estimates appear to be significant in correcting the EI estimates of Catholic unemployment, and the Neighborhood estimates are significant in correcting the level of Protestant unemployment). The corrected EI estimates produced are graphed against the truth in figure 3, and appear excellent.

Estimated unemployment rates for Catholics and Protestants are plotted in figure 4. Unemployment estimates are color coded by model, and Catholic unemployment rates are above the Protestant rates, that is, every model agrees that Catholic unemployment has been at a higher level throughout this period. It is worth mentioning that Goodman's model, as it is often criticized for doing, gives *negative* unemployment rates for Protestants in the earliest periods. The Neighborhood model gives national levels of unemployment that do not vary much by religion. The corrected EI estimate is plotted in Green. The uncorrected estimate (not shown) more closely resembles the

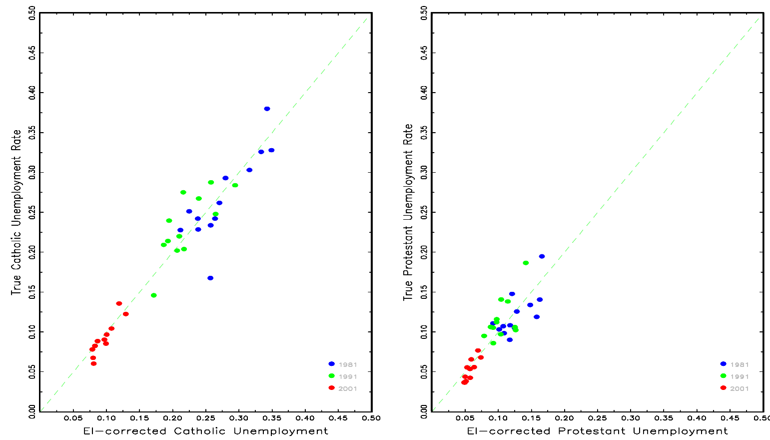

Figure 3: *Plots of corrected EI estimates of unemployment rates against true unemployment, for Catholics (left) and Protestants (right). The plotted points alight close to the $y = x$ dashed line, showing that accurate estimates are produced, across the three census periods.*

Goodman values.[8] [9]

[8]Returning to the literature on EI estimates in second-stage models, the standard concern of heteroskedastic measurement error is not a key concern to this analysis as we are aggregating across districts and the national average should have roughly the same precision across time. However, if we were wanting to use district level estimates, it seems plausible that a correctly specified Multiple Imputation model would be an alternative to the standard Whitewashing algorithms generally proposed, as across datasets, uncertainty between districts should be correctly translated into variance about the mean across imputations.

[9]Similar to this analysis, Quinn (2004) and Lewis (2004) both propose methods to deal with series of EI estimates. Both rely on setting priors so that adjacent observations in time have parameters that move smoothly. As a pragmatic attempt to do this, in each estimation I included the previous and subsequent month's data, heavily downweighted. This had a noticeable dampening effect on the time series of uncorrected EI estimates, and seemed a reasonable practical approach to a similar problem.

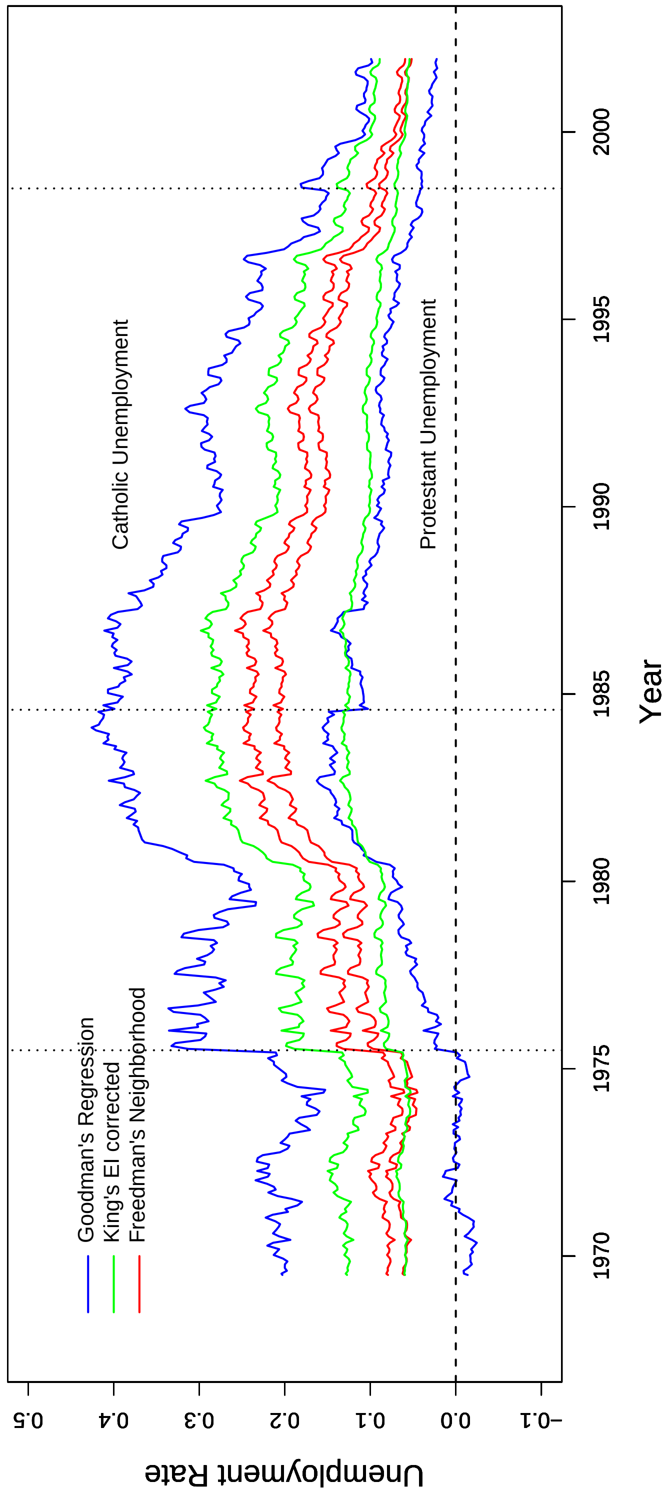

Figure 4: *Estimated unemployment rates by religion by various ecological models. The vertical dotted lines represent revisions of the geographical districts ("travel-to-work areas") and result in small discontinuities.*

6 Additional Model Features

In addition to the central question of estimating unemployment by religion, there are complications presented by the nature of the data and theories of interest. In turn I address the count nature of the dependent violence variables, the contagion present by common aggregation of these counts, and at the end of the next section, address the issue of endogeneity between unemployment and violence.

6.1 Event Counts

The first problem is that linear regression is often not an appropriate model for event count data. This can be particularly true when the number of events is small in each period. The ordinary least-squares model gives predictive probability to negative numbers of events occurring, whereas we know before looking at the data that the numbers of deaths in any period has to be a non-negative integer. When the observed data is not near this boundary at zero, this assumption may pose less of a problem. However, at the monthly level of aggregation that White uses, 19 percent of months have no killings by Republican groups, and a further 13 percent have only one killing. Among the other groups this is even more profound, 36 percent of months have no killings by Loyalist groups and 64 percent have no killings by British forces. Across the entire dataset, 57 percent of observations have zero or one death in that month[10]. A number of models for count data are commonly used. A starting point is Poisson regression, which assumes that the variance is a strict function of (indeed equal to) the mean. However, over dispersal about the mean is common in applied analysis. One common general model that allows for over dispersal is the Negative Binomial.

Two additional models worth investigating are models where dispersal may arise because some of the observations where $y = 0$ may be dissimilar from the rest of the data. A Hurdle model assumes that the probability of an event occurring can be modeled independent of the size of the event, given it's occurrence. Styled to count data, a Hurdle model might be created to predict the probability of an event as Binomially distributed with a Logit link. However, given an event has occurred, we can still employ a truncated the Poisson distribution (Mullahy 1989, Silva and Covas

[10]White does not include deaths where the agent is unknown as observations, so that category is not added to this figure. Since these are the rarest category, this would only increase this figure.

2000).

$$Pr[y_i = 0] = \pi_i \tag{3}$$

$$Pr[y_i = j] = (1 - \pi_i)\frac{e^{-\mu_i}\mu_i^{j-1}}{j-1!}\frac{1}{e^{-\mu_i}} \tag{4}$$

Somewhat similarly, the Zero Inflated Poisson model (or ZIP) assumes the observed variable is a mixture distribution between a Poisson, that of course might ordinarily realize zeros, and a degenerate distribution always equal to zero (Cameron and Trivedi 1998). If the degenerate distribution is ϕ_i of the mixture then:

$$Pr[y_i = 0] = \phi_i + (1 - \phi_i)e^{-\mu_i} \tag{5}$$

$$Pr[y_i = j | y_i > 0] = (1 - \phi_i)\frac{e^{-\mu_i}\mu_i^{j}}{j!} \tag{6}$$

As daily violence data is becoming increasingly common (as for example, King and Lowe 2003), and as there is no clear direction I choose to evaluate these models against each other, and against Ordinary Least Squares. I compared the models by their out-of-sample forecasting ability. The data was partitioned and the same set of covariates (here the model predicting Republican violence against Civilians) was run with each distributional assumption, and functional form (including a spline of unemployment to overcome possible differences in the various functional forms) in one portion of the data. Using the estimated values of the parameters, predicted values of the dependent variable are generated in the other portion of the data, which was not used in the analysis. As the models have different assumptions about the variance of the residuals, and since count data clearly would not be expected to have constant variance, I used the Pearson residuals as my measure of fit in the test portion of the dataset.

Since each partition of the dataset is an equally valid test, by repeated randomly partitioning the dataset, the entire forecasting procedure can be duplicated, and a distribution of test results simulated. The results of one hundred random out-of-sample partitions are reported in table 3. Each number represents the fraction of partitions where the column model had lower Pearson residuals than the row model. Of course the upper diagonal of this matrix would be simply one minus the lower diagonal and is omitted. As can be seen, the models fall into three groups. Here the Negative Binomial model performs only very slightly better than the Poisson model, head-to-head,

	Poisson	Neg.Binomial	ZIP
Negative Binomial	.47		
Zero Inflated Poisson	$> .01$	$> .01$	
Hurdle	$> .01$	$> .01$	.58

Table 3: *Estimated p-values of model superiority from iterated out-of-sample forecasts. Values represent the probability that the column model out performs the row model. The ZIP and Hurdle models significantly out perform all but each other.*

which is surprising as the Negative Binomial model is parametrically more flexible. However, overly flexible models are generally punished in out-of-sample forecasting, so possibly the mechanism of over dispersal in the Negative Binomial is not well suited to the current data.

The ZIP and Hurdle models significantly outperform all other models except each other. That performance between these two models is similar has been strongly conjectured and shown in other applied settings (Zorn, 1998). In none of the one hundred partitions of the dataset did the Poisson or Negative Binomial models out perform either the ZIP or Hurdle model. This degree of dominance was unexpected. One suggestion this brings is that in the data, given the available covariates, the probability of an event is much more predictable than the size of the event. If we take the example of bombings, this makes enormous sense. There may be exogenous factors that predict when bombings will take place, but the total number of deaths that occur is a highly stochastic realization. However, the preponderance of deaths are not bombings, so this is not the lone answer. Returning to the concept of the ZIP model, there have been numerous official and unofficial truces between the parties. If the rate of violence is Poisson distributed, except during truces, when the rate of violence is zero, and truce days can not be identified, then the mixture model is explicitly derived. Of these two models, since the Hurdle model is simpler to estimate, and the parameters can be more readily partitioned into those predicting the occurrence of violence and those predicting the magnitude of violence, which are intuitive terms in this analysis, I continue the analysis with the Hurdle model choice.

6.2 Contagion and Cross-sectional Correlation

Evident contagion exists in the monthly aggregated data. It makes sense that if one type of violence provokes another, and we aggregate over a period in which we might expect to see the initial action and the response, then we may see high correlations within time periods across types of violence. To mitigate this straightforwardly, and to help determine which event came chronologically before the other, the data is temporally disaggregated to the daily level. Since the economic and force data only exist at the monthly level, they are interpolated to the daily level. The three largest remaining correlations are British forces killing Republicans with British forces killing Civilians (0.15), Republicans killing British Forces with Republicans killing Civilians (0.10), and Republicans killing British forces with British forces killing Civilians (0.07). This may also help address any serial correlation as well as the cross sectional correlation (see Brandt et. al 2000, Brandt and Williams 2001, for one explicit time-series model for Poisson processes, Davis et. al 2003, Davis et. al 1999, Zeger and Qaqish 1988) as the data at the daily level will be sparse across time. Lagged autocorrelation plots, again of the Pearson residuals show no significant serial correlation (even before adding the lagged dependent variable terms in the next section), and the Poisson Pearson residual test passes (Li 1991, Cameron and Trivedi 1998). However this test may have little power as the data becomes sparser at such a low level of aggregation. Moreover, the original studies found support for an AR1 process in their monthly and yearly aggregated data, so a more thorough investigation of the time dynamics is warranted.

6.3 Probabilistically Distributed Lags

An extremely general model for the lag structure of violence is the distributed lag model. If we briefly ignore all other variables and assume that the dependent variable is a function of up to k lagged values of some independent variable, then in the Poisson model we might have:

$$\hat{y} = e^{\theta} \tag{7}$$

$$\hat{\theta} = \sum_{i=1}^{k} \beta_i x_{t-i} + \alpha \tag{8}$$

This requires estimating k parameters. If we think that violence by one agent directly causes retaliatory violence by another party later in time, with no violence in-between, then k must be larger than this time span. If it takes the British Army a month (or two) to gather intelligence, plan, and respond to a particular incident, then in this model, we might need to estimate 30 (or 60) parameters. This is clearly excessive and unwieldy, and some additional assumptions on the distribution of the β's can make this manageable. Most fundamentally, we might assume that all the β's have the same sign. That is, violence by agent A cannot both increase and decrease the probability of violence by agent B depending on the time after the event, *ceteris paribus.*

The simplest implementation of this assumption is to restrict all k parameters to the same value. If this is done, having k different variables (lags), or simply creating a new variable counting the total number of events that have occurred in the last k periods before the current observation, are exactly equivalent strategies. Lagged dependent variables are commonly seen in social science data, but with counts these are problematic as at low aggregation as there will be sparse events, and the lag will generally be zero even if zero is a poor measure of what has recently occurred. Instead, what is required is either a direct model of the lagged *latent* value of the dependent variable, or generally more feasibly, a rolling average of all the lags over some period. Averaging over the previous k periods, instead of simply the last period is a better strategy in sparse count data.

Second, in addition to the assumption of all β's having the same sign, we might further assume that the distribution of β_i over time is unimodal. That is, there is a most likely time of retaliation by agent B, and the probability of retaliation is decreasing as we move away from that most likely time, *ceteris paribus.* Two common models that make these assumptions are geometric distributed lags and quadratic distributed lags (a special, but common case of polynomial distributed lags).

$$\hat{\theta} = \beta_1 \sum_{i=1}^{k} z_i x_{t-i} + \alpha, \qquad z_i = g(i, \phi) \tag{9}$$

$$z_i = \phi^{(i-1)} \tag{10}$$

$$z_i = \phi_0 + \phi_1 i + \phi_2 i^2 \tag{11}$$

anchored such that $z_{-1} = 0$, $z_{k+1} = 0$

Geometric distributed lags fit skewed effects well, but assume the modal response is in the next period. There is no reason to believe that retaliation is most likely the very next day after an event. The quadratic form allows the mode to vary, but is known to poorly fit skewed effects. Also, when k is not known, the quadratic form requires testing between different models with k set to different values[11]. Instead, employed here is a unimodal, discrete event probability distribution as the shape parameter on the lags. The negative binomial works well since it gives us flexibility in the mean and variance.

$$z_i = pdf_{\text{negative binomial}}(i, \mu, \alpha) \tag{12}$$

This model allows us to have a skewed, unimodal distribution of lags with variable length k, since the probability function can approach zero before $i = k$. One additional benefits is that $\sum z_i = 1$ if the mean is far from k, thus the coefficients β (on the lag structure z) are reasonably comparable. The second benefit is that there is a micro-foundational story that is plausible. The distribution of z is the probability distribution of the timing of retaliatory events (given the number of events of retaliation).

7 Results

I estimate five models to predict the five most frequent types of political violence, that is "agent-to-victim" events[12], in Northern Ireland:

- Republican violence against British Forces,
- Republican violence against Civilians,
- Loyalist violence against Civilians,
- British Forces violence against Republican Paramilitaries,
- and British Forces violence against Civilians.

Included in table 7 are first differences of key variables, principally the unemployment variables, but also the effects of changes in the size of British Forces. The hurdle model has two sets of parameters for each variable, one predicting whether violence occurs and another parameter predicting the level of resulting violence, but each variables effects are here combined to give the

[11] This requires the number of model estimations to increase linearly with the possible range of k and exponentially with the number of variables for which we have distributed lags.

[12] Republican paramilitary violence against other Republicans is also as common as some of these events, but as presently coded in the data, this mixes both feuding or infighting, with accidental deaths from premature bomb explosions. The latter are quite common, so the pooling of these two types of events is problematic. The solution is to recode this portion of the data.

	Republican on Civilian Violence	Republican on British Forces Violence	Loyalist on Civilian Violence	British Forces on Republican Violence	British Forces on Civilian Violence
Unemployment Ratio	4.8^*	0.94	1.7^*	0.19	0.21
CI 95	(3.0,7.2)	(-1.1,2.8)	(0.38,3.5)	(-0.22,1.0)	(-1.3,0.53)
British Army	1.4^*	-1.7^*	5.8^*	-0.17	-0.38^*
CI 95	(0.18,3.8)	(-2.5,-0.84)	(3.9,8.6)	(-0.34,0.79)	(-0.54,-0.21)
Ulster Defense Regiment	0.29	9.4^*	-1.1^*	0.89	1.4^*
CI 95	(-0.69,1.4)	(6.2,14)	(-1.8,-0.21)	(0.36,2.6)	(0.51,3.8)
Royal Ulster Constabulary	-0.55	-1.7^*	0.06	-0.08	-0.41^*
CI 95	(-1.0,0.12)	(-2.3,-0.89)	(-0.44,1.0)	(-0.30,1.0)	(-0.54,-0.30)

n=11854

The * denotes zero is not in the 95 percent Confidence Interval

Table 4: *First differences (with 95 percent confidence intervals) of the change in the expected number of daily deaths, for 10 percent changes in unemployment or 1 standard deviation changes in force size. Differences are multiplied by 100, so a value of 1, means there is an increased expectation of 0.01 extra deaths every day, (or more loosely, a 1 percent increase in the daily probability of a death occurring).*

resulting change in the total expected number of deaths. The first differences are presented with their 95 percent confidence intervals, and are evaluated for a one standard deviation change from the mean. Although the principle aim was to measure the separate rates of unemployment by religion, the functional form that best fits the data is the ratio of the two unemployment rates. Obviously, to compute the ratio requires knowing both rates, which is not possible in the aggregate data. The ratio is measured as the national Catholic unemployment rate divided by the Protestant unemployment rate. This ratio ranges from 1.6 to 2.4, with a mean and median around 2.1. The first differences are presented for a change from 2 (that is, Catholic unemployment is exactly twice the Protestant rate) to 2.25, a change close to one standard deviation, which historically is close rapid change actually observed in the series between 1980 and 1982.

7.1 Unemployment

The most important and central result is the large effect of unemployment on violence perpetrated on civilians by paramilitaries. Both Republican and Loyalist violence against civilians increases as the ratio of unemployment increase. If the ratio between unemployment rates changes by around 0.25, the expected number of Republican killings of civilians goes up by .048 civilians per day in expectation, or about one extra killing every three weeks (21 days).

7.2 State Force

The largest single effect is the increased violence of Republican forces against British forces when the Ulster Defense Regiment increases by a standard deviation. The increase of 0.094 deaths per day represents a death every eleven days. This is possibly consistent with the idea that members (and former members) of the UDR make easier targets than the British Army, as they have more social interactions with the community, and spend less time barracked. Thus, when the UDR increases in strength, Republican forces instigate much more violence against British forces. It is interesting to note that when the size of the regular British Army forces increase, the largest effects are of Republican and Loyalist violence against citizens. These are interesting preliminary results, however, they may be influenced by not properly accounting for the time dynamics. Also, authors such as White and Hibbs (1974) argue that such measures of force strength or "regime repressiveness" maybe very non-linear, with moderate amounts of troops provoking increasing backlash, until there are enough forces to begin to effectively control the opponent. This is another potential topic for additional study.

7.3 Timing in Retaliation

The models in table 7 all include lagged levels of all five agent-to-victim counts in each of the five models. In table 7 these are simply the average levels of deaths in the preceding 30 days, as discussed previously. To get a more detailed understanding of the relative timing and recipricol cycle of violence, we estimate the same models with distributed lag terms, and graph the shape of these in figure 6. We can see some patterns in the timing of violence and response. Some results are intuitive and others problematic. Starting in the top left, Republican violence against British forces seems to be predicted by two lagged terms, Republican violence on Civilians and lagged violence against British forces. Both are stories of continuing capacity for violence, although the timings are very different. Republican violence against British forces seems to be predicted by recent violence against citizens, and the long term cumulative violence against British forces. In the graph below this we can see the most immediate provocation of Loyalist violence against civilians is Republican violence against civilians, suggesting and urgent and immediate tit-for-tat dynamic. British violence against Republican paramilitaries is only predicted by Republican violence against British forces, not violence against civilians, which might have been hypothesized to provoke a response from state sanctioned forces.

7.4 Causality and Endogeneity

A possible question of endogeneity remains unexamined so far. While there seems to be a strong relationship between unemployment and violence in Northern Ireland, the possibility exists that the direction of causality is mixed or even reversed. It might be plausibly hypothesized that violence is causally prior in this relationship. Once violence occurs, there then are economic consequences. Violence might cause workers to quit or change jobs when their travel is no longer safe. Industry and commerce might experience inefficiencies in the labor force or marketplace, and skilled workers might migrate away. In the long run foreign investment might leave or never arrive as a result of violence. Thus, plausibly, the relationship found between unemployment and violence might have been causally in the other direction. The existence of violence might cause unemployment to rise, rather than the reverse. Understanding the direction of causality is crucial to understanding how policies lowering unemployment and raising economic opportunity might mitigate political violence. Often these are the only policy instruments available.

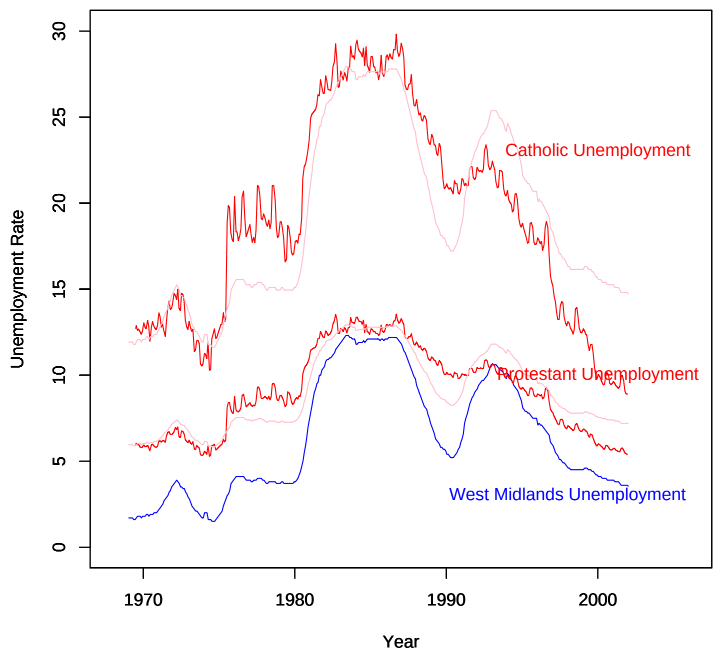

Figure 5: *This figure graphically demonstrates the effectiveness of the instrumental variable. Unemployment and violence in Northern Ireland may by hypothesized to be endogenous, but unemployment in England is unlikely to be influenced by violence in Northern Ireland. However, the economic crisis Northern Ireland experienced was part of a broader crisis, at a lower magnitude, across much of the United Kingdom. Unemployment in the West Midlands predicts unemployment in Northern Ireland and serves as an exogenous instrument. West Midland unemployment is in blue, Northern Ireland unemployment is in red, and the predicted levels of unemployment given the instrument are in pink.*

In this case, however, the direction appears to predominantly be that unemployment causes violence. The economic shocks that Northern Ireland experienced in the 1970's and 1980's were similar to the economic problems experienced across the whole of the United Kingdom at the same time. Using instrumental variables we can measure what fraction of the relationship can not be attributed in the reverse direction. The common problem with such methods is finding an exogenous variable that might cause or predict one variable in the relationship but can not cause the other. Often no such instruments exist, or they have very weak predictive power. Here, however, we can use unemployment rates elsewhere in the United Kingdom as an instrument for unemployment rates in Northern Ireland. We use the unemployment rate in the West Midlands region of England as an exogenous instrumental variable[13]. Reasonably it can be assumed that unemployment in the West Midlands of England is not caused by violence in Northern Ireland. To the extent that the West Midlands predicts unemployment in Northern Ireland, then that predicted unemployment is a part of the larger economic crisis throughout the United Kingdom.

In the first stage I predict the disaggregated unemployment rates by religion solely by the unemployment rate in the West Midlands[14] The fit of this bivariate relationship can been intuitively gauged in figure 5. The blue line shows the unemployment rate in the West Midlands, the red lines the unemployment rates by religion in Northern Ireland. Largely the same sweeping trends in unemployment are seen in the Midlands in England as in Northern Ireland, although at different levels of magnitude. The pink lines are the predicted values of unemployment from the first stage of the instrumental variables model, and appear to match up well to the actual (red) values. These predicted values of the unemployment rates from this (first stage) model are then substituted into the original (now second stage) model of violence.

These results are presented in table 7.4 for the two models in which unemployment can be shown to be related to violence in the original results, that is, Republican and Loyalist violence against civilians. The relationships remain largely unchanged, except that here the Royal Ulster Constabulary appear effective in decreasing violence by paramilitaries as the number of officers

[13]The West Midlands region includes the large urban area of Birmingham, as well as predominantly rural areas along the Welsh border such as Shropshire and Staffordshire. Unemployment rates in the North West region of England are a slightly better fit than the West Midlands for unemployment in Northern Ireland at this period, but the available data is less complete in the earlier 1970's.

[14]First stage models can be made increasingly efficient by adding the regressors of the second stage, but here this runs the risk of introducing other likely endogenous variables.

	Republican on Civilian Violence	Loyalist on Civilian Violence
Predicted Unemployment Ratio	3.6	3.7
CI 95	(1.1,7.2)	(1.2,7.3)
British Army	1.3	6.5
CI 95	(0.07,3.3)	(4.4,8.6)
Ulster Defense Regiment	0.80	−1.2
CI 95	(-0.37,2.6)	(-1.8,-0.24)
Royal Ulster Constabulary	−2.0	−1.2
CI 95	(-2.5,-0.91)	(-1.8,-0.37)

n=11854

The * denotes zero is not in the 95 percent Confidence Interval

Table 5: *First differences (with 95 percent confidence intervals) of the change in the expected number of daily deaths, for 10 percent changes in unemployment or 1 standard deviation changes in force size. Differences are multiplied by 100, so a value of 1, means there is an increased expectation of 0.01 extra deaths every day, (or more loosely, a 1 percent increase in the daily probability of a death occurring).*

increase, which could not be seen in the earlier results. The expected number of deaths from the previously described economic shock slightly decreases for Republican paramilitaries when unemployment is instrumented, while it increases for Loyalist paramilitaries. A 0.25 increase in the ratio of unemployment rates, for example as a result of Catholic unemployment rising while Loyalist unemployment remains constant, leads to 3.6 additional deaths every 100 days, or roughly one additional death every four weeks.

Once again, as in the previous table, and similar to the finding of White, violence by Loyalist paramilitaries *increases* as economic times improve for Protestants relative to Catholics. This result is robust to plentiful different specifications of unemployment and relative unemployment rates. That it increases further when unemployment is measured essentially as unemployment in England, rather than unemployment in Northern Ireland, adds another layer to this puzzle.

8 Conclusion

While previous studies have found no link between unemployment and violence in Northern Ireland, this troubling result occurs because available data aggregates unemployment by religion, and the unemployment rates of each religion have opposite effects. By estimating the separate unemployment rates for Catholics and Protestants, unemployment becomes a significant causal mechanism for the intensity of conflict, as expected both by substantive specialists in this conflict and the body of literature and theory in political violence, and not previously witnessed quantitatively outside of cross-national studies nor seen within *intra*-national analyses of specific conflicts. The sectarian differences in unemployment rates was a leading predictor of violence in Northern Ireland, and furthermore can be demonstrated to be an important causal factor.

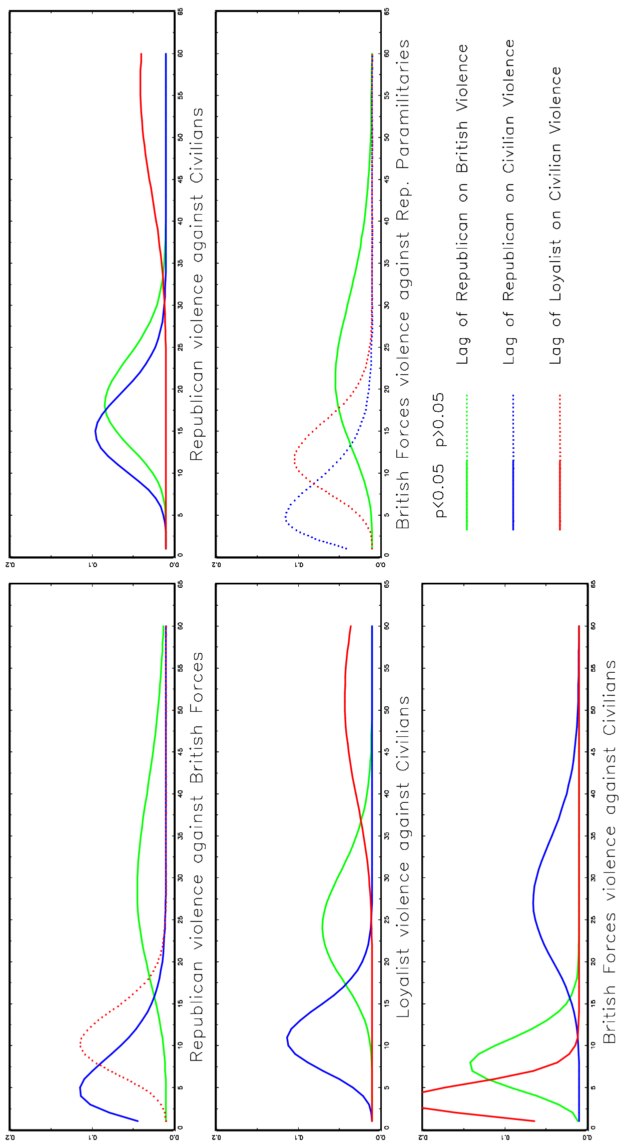

Figure 6: *Distributed Lags.*

References

Adolph, Christopher and Gary King. 2003. "Analyzing Second Stage Ecological Regressions." *Political Analysis* 11(1):65-76.

Adolph, Christopher, Gary King, Michael C. Herron and Kenneth W. Shotts. 2003. "A Consensus on Second Stage Analysises in Ecological Inference Models." *Political Analysis* 11(1):86-94.

Anselin, Luc and Wendy K. Tam Cho. 2002. "Spatial Effects and Ecological Inference." *Political Analysis* 10(3):276-297.

Beck, Nathaniel, Gary King and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review.* 94(1):21-35.

Beck, Nathaniel, and Simon Jackman. 1998. "Beyond Linearity by Default: Generalized Additive Models." *American Journal of Political Science.* 42(2):596-627.

Brandt, Patrick and John. T. Williams. 2001. "A Linear Poisson Autoregressive Model: The Poisson AR(p)" *Political Analysis* 9(2).

Brandt, Patrick, John T. Williams, Benjamin O. Fordham, Brian Pollins. 2000. "Dynamic Modeling for Persistent Event Count Time Series." *American Journal of Political Science* 44(4):823-843

CAIN. Accessed 2003. *Conflict Archive on the Internet.* `http:\\cain.ulst.ac.uk`

Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data* Econometric Society Monographs, Cambridge University Press: Cambridge.

Cho, Wendy K. Tam. 1998. "Iff the assumption fits . . . : A Comment on the King Ecological Inference Solution." *Political Analysis* 7:143-163.

Cho, Wendy K. Tam and Brian J. Gaines. 2004. "The Limits of Ecological Inference: The Case of Split-Ticket Voting." Forthcoming, *American Journal of Political Science.*

Compton, P. A.(ed.) 1981. *The Contemporary Population of Northern Ireland and Population Relatated Issues.* The Queens University: Belfast.

Eversley, David. 1989. *Religion and Employment in Northern Ireland.* Sage Publications: London.

Feierabend, Ivo K., Rosalind L. Feierabend, and Betty A. Nesvold. 1969. "Social Change and Political Violence: Cross-National Patterns." in *H*ugh D. Graham and Ted R. Gurr (eds.), *Violence in America: Historical and Comparative Perspectives* Signet Books: New York.

Ferree, Karen. 2000. "Beyond 2x2 Tables: An R x C Ecological Inference Model." *Working Paper*

Freedman, D.A. S.P. Klein, J. Sacks, C.A, Smyth and Charles G. Everett. 1991. "Ecological Regression and Voting Rights." *Evaluation Review.* 15(6):673-711.

Herron, Michael C. and Kenneth W. Shotts. 2004. "Using Ecological Inference Point Estimates as Dependent Variables in Second Stage Linear Regressions." Forthcoming, *Political Analysis.*

Goodman, Leo. 1953. "Ecological Regressions and the Behavior of Individuals." *American Sociological Review.* 18:663-666.

Hayes, Bernadette C. and Ian McAllister. 2001. "Sowing Dragon's Teeth: Public Support for Political Violence and Paramilitarism in Northern Ireland." *Political Studies* 49: 901-922.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery and Chris T. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science.* 14:382-417.

Honaker, James, Jonathan N. Katz and Gary King. "A Fast, Easy, and Efficient Estimator for Multiparty Data." *Political Analysis* 10:1(Winter):84-100.

Hovland, Carl I. and Robert R. Sears. 1940. "Minor Studies of Aggression: Correlations of Economic Indicies with Lynchings." *Journal of Psychology* 9:301:10.

Howe, Leo. 1990. *Being Unemployed in Northern Ireland: An ethnographic study.* Cambridge University Press: Cambridge.

Hibbs, Douglas A. Jr. 1973. *Mass Political Violence: A Cross- National Analysis.* Wiley-Interscience, New York.

Imai, Kosuke and Gary King. 2004. "Did Illegal Overseas Absentee Ballots Decide the 2000 U.S. Presidential Election?" forthcoming *Perspectives on Politics.*

Irish Information Partnership. 1987. *Irish Information Partnership: Agenda: A database of Facts and Information on the Northern Ireland Conflict and Anglo-Irish Affairs.* 5th ed. London, England: Irish Information Partnership.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual*

Behavior from Aggregate Data, Princeton: Princeton University Press.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1)(March): 49-69.

King, Gary, Michael Tomz, and Jason Wittenbery. 2000. "Making the Most of Statistical Analysis: Improving Interpretation and Presentation." *American Journal of Political Science.* 44:341-55.

Kovalcheck, Kassian A. 1987. "Catholic Greivances in Northern Ireland: appraisal and judgement." *British Journal of Sociology* 38(1):77-87

Lewis, Jeffrey B. 2004. "Extending King's Ecological Inference Model to Multiple Elections Using Markov Chain Monte Carlo." in King, Gary, Ori Rosen and Martin Tanner (eds.) " Ecological Inference: New Methodological Strategies." Cambridge University Press: Cambridge.

Lewis, Jeffrey B. and Drew Linzer 2005. "Estimating Regression Models in Which the Dependent Variable is Based on Estimates." Forthcoming, *Political Analysis.*

Li, W.K. 1991. "Testing Model Adequacy for Some Markov Regression Models for Time Series," *Biometrika* 7:83-89.

Quinn, Kevin M. 2004. "Ecological Inference in the Presence of Temporal Dependence." in King, Gary, Ori Rosen and Martin Tanner (eds.) " Ecological Inference: New Methodological Strategies." Cambridge University Press: Cambridge.

Russett, Bruce M. 1964. *World Handbook of Political and Social Indicators.* Yale University Press: New Haven.

Schafer, Joseph L. 1997. *Analysis of Incomplete Multivariate Data.* New York: Chapman and Hall.

Sheehan, Maura and Mike Tomlinson. 1999. *The Unequal Unemployed: Discrimination, unemployment and state policy in Northern Ireland.* Ashgate Publishing: Aldershot, England.

Silva, J.M.C. Santos, and Francisco Covas. 2000. "A Modified Hurdle Model for Completed Fertility." *Journal of Population Economics* 13:2.

Simpson. 1953. *American Journal of Sociology.*

Sutton, Malcolm. 1994. *An Index of Deaths from the Conflict in Ireland 1969-1993* Beyond the Pale Publications: Belfast.

Thompson, J. L. P. 1989. "Deprivation and Political Violence in Northern Ireland, 1922-1985. A Time-Series Analysis." *Journal of Conflict Resolution* 33(4): 676-699.

Tolnay, Stewart E. and E. M. Beck. 1995. *A Festival of Violence: An Analysis of Southern Lynchings, 1882-1930*. University of Illinois Press: Urbana.

Wakefield, John. 2005. "Ecological Inference in 2 x 2 Tables." *Journal of the Royal Statistical Society*. Series A. 167:1-42

White, Robert W. 1993. "On Measuring Political Violence: Northern Ireland, 1969 to 1980." *American Sociological Review* 58:575-585.

Zorn, Christopher. 1998. "An Analytic and Empirical Examination of Zero-Inflated and Hurdle Poisson Specificiations." *Sociological Methods and Research* 26(3):368-400.