The Determinants of Civil War and Excess Zeros

John Paul Dunne
School of Economics, SALDRU
University of Cape Town
john.dunne@uct.ac.za

Nan Tian[1]
School of Economics
University of Cape Town
tnxnan001@myuct.ac.za

Abstract

This paper considers the determinants of civil conflict, using a zero-inflated Poisson modelling approach that deals with the problem of excess zero observations, which we argue are related to two distinct data generation processes. Despite their continued use in the literature, traditional probit and logit models have limited capacity in dealing with this issue and can lead to misleading results. This is illustrated by estimating the model in Elbadawi and Sambanis (2002) using their data and finding different results and providing substantive information about the data generating process of heterogeneous zeroes in the database. This greed-grievance model is then estimated on a sample of 134 countries, over 54 years and provides some interesting results. While the standard probit and Poisson models tend to emphasise the grievance effects in support of opportunity variables,as found in other studies, the zero-inflated models gives provides more support for grievance effects. In particular, ethnicity and inequality are found to play a signifcant role, in contrast to earlier studies.

Keywords: Civil war; zero-inflation; greed and grievance
JEL code: D74; C3

Preliminary Draft: Comments and suggestions are welcome. Please do not quote without authour's permission.

[1] Corresponding Authour

1 INTRODUCTION

Following Collier and Hoeffler (2004), empirical work on the causes of civil war burgeoned, providing valuable insights into why conflicts start and continue and the role of economic factors (Blattman and Miguel, 2010). In empirical research almost all studies use some form of probit or logit model to estimate a zero-one dependent variable on a panel or cross section of countries. There is, however, a potential in that these models do not deal well in situations where there are a large number of zeros in the dependent variable and this is likely to be the case for conflict as, fortunately many country-year observations are zero. Until recently there has been little recognition of this issue, possibly because the earlier models were estimated on cross section data or five year average panels, but the use of annual data in panels has made it an issue of greater concern. Using a probit or logit model in the presence of excess zeroes can potentially lead to biased estimates results due to the correlation of the error term with the explanatory variables (Bagozzi et al, 2014).

In conflict studies the dependent variable normally takes the value zero when the number of battle-related deaths does not exceed exceed a particular bound, e.g. 1000, but this encompasses a situation when the number of battle related deaths is below the threshold and there is complete peace and one where there are minor conflicts still present. In addition, heterogeneity across countries means that a zero may represent a break between conflicts in a conflict ridden country, or a year of peace in a peaceful coujntry. Consider the difference between a zero for Sweden, Singapore or Australia with a close to zero chance of civil war, and a zero for countries like Ethiopia or Cambodia. Thus, while coded uniformly, the zeros can come from two distinct processes or sources and not taking this into account can lead to statistically biased estimates when evaluating the impact of explanatory variables on the dependent conflict variable.

A second issue is that when the dependent variable has excess zeroes probit and logit models cannot statistically account for the observable and latent factors that generate the high proportions of zeros. The normal probit and logit models generate only one latent equation and are unable to account for or differentiate between the different additional weights put on zero observations, especially if the zeroes relate to different processes. Using probits or logits in hypothesis testing could lead to model misspecification (Harris and Zhao, 2007). A further concern is whether the probit or logit models conform to the process that generated the data. In the case of civil conflict, an event count process that is characterised by having a rare occurrence, a Poisson distribution is likely to be better suited to the data than logistical or normal distributions (Smith and Tasiran, 2012). This indeed has been confirmed by Richardson (1960), Wilkinson (1980) and Benoit (1996).

To get some idea of the likely effect of using logit or probit models, when a split population model might be more suitable, this paper starts by revisiting the Elbadawi and Sambanis (2002) study on the determinants of civil war prevalence and applying a zero-inflated poisson model (ZIP) to their data, considering the impact on the results and its implications [2]. It then uses the zero-inflated model to revisit the greed and grievance debate on an updated dataset of 134 countries for the period 1960 to 2013. The next section presents the zero-inflated probit model, which is then followed by the results of the reestimation of the Elbadawi and Sambanis (2002) study. Section four applies the zero inflated poisson model to a new dataset and the final section presents some conclusions.

2 MODELLING CIVIL CONFLICT

In most analyses of the determinants of conflict, an ordered dependent variable is used in which a given country-year is assigned a zero for peace and a value of one when violence between the state another side reaches a given threshold, classifying it as a civil war. In most studies this will mean that there are a large number of zero observations as peaceful years will dominate conflict years. These zeros can also reflecting rather different states, one where the structural and societal forces ensure a zero probability of civil conflict regardless of greed or grievance incentives and another that relects a break in conflict and a high probability of return to conflict. The first group of zeros will often be advanced economies, such as Norway, Sweden, and Japan and could be labelled "complete-peace". The second group, will often be found in developing regions such as sub-Saharan Africa, Asia or Latin America, from which the zeroes can be labelled as "incomplete-peace". The main difference between the first and second type of zero is that while the probability of transition into war for first type is zero, the observations for "incomplete-peace" group is not. In the case of

[2] Prevalence can be defined as the likelihood of observing a civil war at any point in time $Y(t)$, estimating the probability of that $Y(t) = 1$ is the sum of the probability that war occurs at time t contingent on there being no war at $t-1$ and the probability that war occurs at time t given that war had been ongoing at time $t-1$.

a "incomplete-peace" country conflicts of interest or grievance can induce violent conflict. There is also a third group of zero observation where there is still civil conflict taking place, but the violence has not reached or has gone beyond the threshold that is used to define a conflict, often 1000 battle related deaths. Such cases can be labelled as "incomplete-war". Given the high proportion of heterogeneous zeroes in the analysis, using ordinary probit or logit models may not be an appropriate tool for statistical inference and can potentially give biased estimates Bagozzi et al (2014).

A more statisfactory alternative is the use of split population or two-part models proposed by Harris and Zhao (2007) and Vance and Ritter (2014). This is typically in the form of a zero-inflated poisson (ZIP) model, where estimations follow two stages. The first of the two latent equations, stage one, is a probit selection equation, while the second stage is a poisson equation. This splits the observations into two processes each potentially having different sets of explanatory variables. In the context of civil war incidence, zero observations in process 0 ($w_i = 0$) include inflated zeroes that never experience civil war (e.g. Sweden), while zero observations in process 1 ($w_i = 1$) includes cases for which the probability of transitioning into a civil conflict is not zero and civil war casualties have not reached the lower bound (or limit) of 1 000 battle related deaths. The binary variable w indicates the split between process 0 (with $w_i = 0$ for no war) and regime 1 (with $w_i = 1$ for war). w is related to the latent dependent variable w_i^* so that $w_i = 1$ for $w_i^* > 0$ and $w_i = 0$ for $w_i^* \leq 0$, where w_i^* now represents the propensity for participation in which the response variable Y_i (i.e conflict) has the distribution:

$$\Pr(Y_i = y_i) = \begin{cases} w_i + (1 - w_i)^{e^{(-\lambda_i)}} & , y_i = 0 \\ (1 - w_i)^{e^{(-\lambda_i)}} \frac{\lambda_i^{y_i}}{y!_i} & , y_i > 0 \end{cases} \tag{1}$$

where the parameters λ_i and w_i depend on vectors of covariates x_i and z_i, respectively, which are modelled as:

$$\log(\lambda_i) = x_i^t \beta \tag{2}$$

and

$$\log(\frac{w_i}{1-w_i}) = z_i^t\gamma \tag{3}$$

with mean and variance as $E(Y_i) = (1-w_i)\lambda_i$ and $var(Y_i) = (1-w_i)\lambda_i(1+w_i\lambda_i)$. In this ZIP model, the matrices z and x contain different sets of experimental factor and covariate effects that relate to the probability of the "zero-state" (zero probability of civil war) and the Poisson mean in the "nonzero-state" (probable civil war), respectively. Thus, the γ's have interpretations in terms of the factor level effect on the probability that there is a zero probability of conflict and the β's have the interpretation of the effect on the average risk of civil war when the probability is non-zero. Following Lambert (1992) the ZIP model (equation 1) can be regressed using maximum likelihood with an Expectation-maximum (EM) algorithm[3].

While empirical research on the determinants of civil conflict has generally followed a more standard approach of assuming normality, conflict data is count data produced in a discrete and countable manner and the number of events can also never be negative or a non integer. This does not suggest a normal distribution and the error terms in a regression would not be normally distributed and the observed variables would not be a linear function of the covariates (Benoit, 1996). When civil conflict is a random event, and is observed at the end of each observation period i (common in conflict studies), then the data will conform to a Poisson process which has a rate of occurence λ, where $\lambda > 0$, as long as the zero events occurred at the start of the period and no more than one event occurs at the same time. Both these assumptions are satisfied in civil war research, as no more than one civil war can occur in a given country, by definition, and since civil war is defined as an event that has either more than 25 or 1000 battle deaths in a given year, occurrences in the beginning or previous period are independent events, since conflicts are only recognised if the threshold has been reached by the end of a calender year. The Poisson distribution used in this model seems more suitable and has been shown to fit the distribution of conflicts over time

Using this method allows more accurate estimates to be obtained than using the standard probit model, as the probability of a zero observation is now modelled conditional on the probability of zero from the probit process plus the probability of being in process

[3] For full derivation of the model see Lambert (1992) and Hall (2000).

0 from the splitting equation. It should be noted that the usefulness of the model (i.e. unbiased estimates) declines when the size of the split in the sample population becomes very big or very small and can lead to biased results, with statistal inference increasingly difficult as the proportional of zeroes gets closer to one. Bagozzi et al (2014) suggests that this will become an issue when there are less than 10 percent or greater than 90 percent of zero observations.

3 AN EMPIRICAL INVESTIGATION

Elbadawi and Sambanis (2002) provides an influential contribution to the 'greed-grievance' debate by combining Collier and Hoeffler's (2004) model of civil war onset with Collier, Hoeffler and Soderbom's (2004) model of civil war duration. Their model predicts the prevalence of civil conflict, based on opportunities for rebellion against its constraints. These opportunities are divided into greed versus grievance or rebellions that generate profit versus rebellions triggered by genuine grievance. They code incidents as civil conflicts using five categories, (1) the war caused more than 1 000 battle-related deaths, (2) it challenged the sovereignty of an internationally recognised state, (3) it occurred in the territory of the state, (4) it included the state as a principle combatant, and (5) the rebels were able to mount a organised military opposition to the state. Their sample includes over 150 countries with available data from 1960 to 1999, with a dependent variable having about 81 percent of zero or peace observations. Using a probit model, Elbadawi and Sambanis find that prevalence of civil war is consistent with earlier studies on war onset and duration (Table 1, column 1). It is positively influenced by primary commodity exports as a share of GDP (a proxy for "looting" or economic opportunity), population and previous wars experiences in the past 10 years, while the level of GDP, the growth of per capita GDP and squared term of primary commodity exports as a share of GDP having negative effects. As with most studies in the literature, none of the grievance factors ethnic fractionalisation, ethnic dominance nor democracy (proxy for political rights) were statistically significant.

Given the high proportion of heterogeneous zeroes, which can be thought of as including "complete-peace" or "incomplete peace" as discussed earlier, the use of a simple probit can be questioned and a zero-inflated Poisson model (ZIP) would seem more suitable. The results for both are presented in Table 1. The first set of results (1) provide the standard probit model estimated by Elbadawi and Sambanis (2002) and the next two columns (2) give the ZIP results, the first stage probit estimates (Outcome) and the second stage (Inflation) estimates. The second half of the Table 1, specification's (3) and (4) replicates the estimation method but replaces ethnolinguistic diversity with ethnic dominance. In order to remain consistent, covariates in the outcome equation of the ZIP model must be identical to the normal probit. Moreover, in identifying plausible indicators for conlfict "relevance", variables that are included in the inflation equation should directly influence the probability that a country in any given year always experiences peace (i.e. per capita GDP, ethnic diversity or political freedom).

Comparing the first set of results, the probit and ZIP estimates for the outcome equation (1) and (2) show similar coefficient estimtes for most of the variables, but also a striking difference. The ethnolinguistic diversity variables become significant in the ZIP model. This is also true when ethnic dominance is used (3) and (4), with the polity index also becoming important. The estimates for the zero-inflated inflation equation (2) seem reasonable, with the share of exports in GDP, log population and ethnolinguistic diversity reducing the likelihood that the zero is a 'compete peace' observation, and real GDP increasing it and when ethnic dominance is used (4) real GDP has a positive and significant effect as does the polity index, with log of population negative.

The Vuong (1989) test rejects the hypothesis that the traditional probit or logit models are better, thus favouring the zero-inflated models as a less biased estimator. A robustness check across the different specifications used by Elbadawi and Sambanis reveals the same consistent results. All the coefficients in the ZIP model, except log population and primary commodity exports as a share of GDP have larger coefficients than in the standard probit, suggesting that not allowing for zero-inflation leads to civil war risk being underestimated.

Table 1: Probit, Zero-Inflated Poisson of Civil War Prevalence 1960-1999

	(1)	(2)		(3)	(4)	
	Probit	ZIP		Probit	ZIP	
	Outcome	Outcome	Inflation	Outcome	Outcome	Inflation
Pri Exports/GDP	10.53*	9.488*	-2.276	10.57**	10.14**	
	(4.136)	(4.341)	(5.061)	(3.835)	(3.454)	
Pri Exports/GDP^2	-21.24*	-23.41*		-20.79*	-22.68*	
	(9.325)	(11.31)		(8.646)	(8.966)	
log real GDP	-0.0003**	-0.0004**	0.0018*	-0.0002**	-0.0002*	0.001*
	(0.000)	(0.000)	(0.0009)	(0.0001)	(0.0000)	(0.0005)
RGDPPC Growth	-0.0899**	-0.122**		-0.0723**	-1.054**	
	(0.0286)	(0.034)		(0.0270)	(0.0212)	
Polity Index (1 lag)	-0.0115	0.0135		-0.0105	0.0341*	0.296*
	(0.020)	(0.0184)		(0.0183)	(0.018)	(0.044)
Polity Index^2 (1 lag)	0.0032	0.0035		0.0030	0.0086*	
	(0.0041)	(0.0042)		(0.0039)	(0.0038)	
Ethno Diversity	0.0389	0.0656**	-0.231**			
	(0.0258)	(0.0178)	(0.0759)			
Ethno Diversity^2	-0.0004	-0.0007**	0.0024**			
	(0.0003)	(0.0002)	(0.0008)			
Ethnic Dominance				0.362	0.389*	
				(0.291)	(0.177)	
Log Population	0.599**	0.266*	-2.532**	0.429**	0.142*	-2.412**
	(0.140)	(0.107)	(0.946)	(0.122)	(0.073)	(0.932)
War in Past 10 Years				0.735**	1.442**	
				(0.214)	(0.235)	
Constant	-12.48**	-6.805**		-9.548**	-5.535**	
	(2.569)	(1.962)		(2.223)	(1.508)	
Rho	0.601**	-		0.474**	-	
	(0.086)	-		(0.104)	-	
Observations	783	783		783	783	
Zero Observations	-	692		-	692	
Log likelihood	-189.771	-223.663		-184.481	-209.713	
Wald χ^2	33.11	-		51.82	-	
Vuong test	-	3.41		-	3.68	

Standard errors in parentheses, ** p<0.01, * p<0.05

These results, using the Elbadawi and Sambanis (2002) data, provide a strong case for arguing that the determinants of conflict literature should move from standard probit and logit models to some form of a ZIP model. If not researchers risk both underestimating the risk of civil conflict and making erroneous conclusions regarding the significant determinants. Given the changes in significance of the grievance terms, if such models had been used earlier the trajectory of the greed-grievance debate may well have been different. Having established that the zero-inflated models yields potentially better results for civil conflict, the next section develops the analysis by estimating a more general greed-grievance model based upon the literature and using data for the period 1960-2013.

4 GREED VS GRIEVANCE REVISITED

To provide the opportunity to specify a general greed grievance empirical model a range of variables were collected based upon the debates in the literature. Proxies for greed or opportunity include real GDP, growth in GDP per capita, degree of urbanisation, life expectancy and natural resource dependence. For this study two sets of income variables were collected, from the World Bank and Penn World Tables 8.0. Degree of urbanisation is measured as the proportion of a country's population living in an urban environment, while life expectancy follows the usual measurement.[4] Male secondary school enrolment was not used in the estimations due to poor and incomplete data. Following from the literature, natural resource dependence is measured by the share of primary commodity exports in GDP. The World Bank provides data for the period 1960 to 1999, which was cross reference it with Fearon (2005) for consistency. The remaining 14 years are constructed using export data (primary commodities) provided by the World Trade Organisation (WTO) and GDP from the World Bank.

Taking into consideration the numerous debates on the measure of natural resource dependence and the type of commodities used, three additional measurements are considered. A measure of oil production in metric tons and oil exports greater than 1/3 total exports are used to proxy for oil abundance and dependence respectively.[5] To distinguish between

[4] This data is sourced from the World Bank, the degree of urbanisation can also be thought of as a measurement of geographic dispersion, the greater the urbanisation, the lower the geographic dispersion. All income figures are purchasing power parity (PPP) adjusted.

[5] Oil production in metric tons is provided by Ross (2013), this data goes from 1932 to 2011, the additional

fuel and non-fuel minerals with other primary commodities, a mineral dependence variable was created. A country is considered mineral dependent if its mineral exports constitutes 25% or more of a country's total tangible exports. Mountainous terrain in a given country is included as an indicator of military accessibility or safe havens for rebels

The grievance variables are, for the most part, common to those identified by Collier and Hoeffler (2004) and Fearon and Laitin (2003). This paper considers three measures of grievance: ethnic and religious hatred, political repression or freedom, and income inequality (horizontal inequality). Ethnic fractionalisation is the most commonly chosen indicator to test the linkage between ethnicity and civil conflict.[6] Two measurements of ethnic fractionalisation are borrowed from the two influential contributions to the civil war literature, namely Fearon and Laitin (2003) and Collier and Hoeffler (2004). Another measure of ethnic hatred often used in the literature is ethnic dominance, which is measured as a binary variable taking on the value one if the largest ethnic group in a country consists between 45% - 90% of the population. To measure religious religious haterd, Collier and Hoeffler constructed a fractionalisation index analogous to ethnic fractionalisation.

Other things being equal, poltical democracy or freedom should be associated with less descrimination, repression and civil war. Data from the Polity IV database is used to measure political rights, with the variable *polity* ranging from -10 (high autocracy) to 10 (high democracy). The relationship between political freedom and civil war has often been thought of having a non-linear effect (Hegre et al, 2001), this hypothesis is tested through the inclusion of *polity* squared. Three alternative variables are used as substitutes for ethno-political and economic grievance. In a recent paper by Buhaug, Cederman and Gleditsch (2014), the authours found the new grievance indices of horizontal income inequality and political discrimination to perform much better than conventional indicators. Economic grievance is captured by the relative gap between the mean national income and the income level of the poorest and richest groups (positive and negative horizontal inequality), while ethno-political grievance is measured by demographic size of the largest discriminated ethnic group.[7]

two years were drawn from the same source as the authour, US energy information administration website for international energy statistics: http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm.

[6] Initially used by Easterly and Levine (1997), the fractionalisation index follows in accordance with Herfindahl's formula, and is interpreted as the probability that two randomly selected individual in a population belong to different ethnic groups.

[7] For full description and derivation of the variables see Buhaug, Cederman and Gleditsch(2014)

The control variables included in the model are the standard ones found in the literature. Population, cold war and Africa feature in various specifications with their effects on civil war prevalence, apart from population, subject to much empirical debate. Finally, the dependent variable used in the paper takes on a value of 0 for all peace year observations and a 1 for civil war years with combat deaths ranging between 25-999 and annual battle deaths of above 1000.

Table 2: Descriptive Statistics: Means

	Full Sample	Always 0	Not Always 0	Civil war	No Civil war
Opportunity					
Primary Commodity Exports/GDP	0.156	0.178	0.139	0.109	0.164
GDP per capita (const.US$)	7931	14069	3311	3172	8699
GDP per capita growth	0.018	0.022	0.016	0.010	0.019
Mountainous Terrain %	16.38	14.93	18.11	23.16	15.33
Rate of Urbanisation	46.94	56.00	39.73	40.61	47.92
Life Expectancy	61.61	66.15	57.98	59.41	61.95
Oil Production (Metric Tons in 000's)	17000	13700	19300	19100	16700
Mineral Dependence	0.493	0.415	0.545	0.550	0.484
Oil Exports	0.187	0.155	0.208	0.168	0.189
Grievance					
Ethnic Frac (F&L)	0.489	0.397	0.553	0.576	0.474
Ethnic Dominance	0.470	0.483	0.467	0.549	0.457
Religious Frac (index, 0-100)	36.47	36.07	36.58	0.36	0.37
Polity IV (-10 to 10)	1.13	3.84	-0.73	0.97	1.30
LDG	0.056	0.024	0.081	0.142	0.042
NHI	1.189	1.064	1.278	1.398	1.155
PHI	1.201	1.086	1.287	1.224	1.197

Notes: ldg = largest discriminated ethnic group, phi = positive horizontal inequality (relative gap between mean national income and income level of the richest group), nhi = negative horizontal inequality (relative gap between mean national income and income level of the poorest group)

Table 2, above, presents the descriptive statistics of the above mentioned variables with a breakdown by conflict experience. These results seem support the central thesis that the different zeroes in the sample are formed through completely separate processes. For the always zero group, GDP per capita, per capita growth, rate of urbanisation, life expectancy and political freedom are all higher than the non-always zero group, while "complete peace" countries all experience lower ethnic and religious fractionalisation and income inequality. Estimted correlations suggest some association between income and inequality variables and the likelihood of a country being completely peaceful versus incompletely peaceful. In episodes of civil conflict, GDP per capita, GDP per capita growth, rate of urbanisation, life expectancy and political freedom are all lower compared to times of peace, ethnic divisions, high income inequality, primary commodity dependence and substantial amounts of rough terrain.

Estimating the probability of civil conflict using ordinary probit and Poisson regressions, with the civil conflict dependent variable taking the value of one if deaths total over 25 in a given battle or over 1000 in a given year and zero otherwise, gave the results in Table 3. The standard probit results (1) show GDP and growth in GDP per capita to be highly significant in decreasing the probability of civil conflict, with primary commodity exports as a share of GDP also highly significant and non linear. Primary commodity exports are seen to initially decrease civil war risk, reaching a trough when it constitutes about 33% of GDP,[8] therafter increasing civil war risk. As regards to the grievance terms, the Polity IV index squared is significant and ethnic fractionalisation (Fearon and Laitin (2003)'s definition) and its square, are substantively and statistically significant and non linear, with inceased ethnic diversity first increasing the potential for civil war, peaking at about 68 percent and then decreasing. As for the control variables, population has a positive and significant effect on civil war prevalence while the Cold War dummy is negative and statistically insignificant. The likelihood ratio test of the correlation coefficient (*rho*) suggests panel estimator to be preferred to a pooled estimator.

Starting with the inflation equation, the variables of real GDP, per capita GDP growth, political freedom, ethnic diversity all represent plausible indicators that influence the probability a country always experiences peace. To this end, the inflation equation confirms that higher income and political freedom does indeed lead to greater probability of being

[8] Differentiate the probability of civil war with respect to primary commodity exports (4.197/2(*6.436)) = 0.326

in the "complete peace" group, while ethnic fractionalisation, ethnic dominance and population all have a negative effect. The variation in sign between the outcome and inflation estimates makes sense as one equation calculates the likelihood of countries being in conflict and the other of them being in the in the 'complete peace' group, conditional on them not experience battle related deaths above the threshold.

Moving onto the outcome equation of the zero-inflated Poisson model (2) gives signs that are consistent with the standard probit, but there are diffferences in the significance of the grievance terms. Primary commodity exports as a proportion of GDP shows the same effect as the probit model, albeit at a higher turning point of 35% while income, both its level and growth decrease the likelihood that a country experiences civil conflict conditional on that country being able to experience such events. Proxies for ethno-political grievance are better represented using the zero-inflated models than the normal probits. Political freedom is now significant in explaining civil war prevalence, featuring an inverse u-shape, first increasing civil war risk and then decreasing, while ethnic diversity remains a significant civil war predictor. Another interesting finding is that using the ZIP model leads to lower standard errors. The improvement of this group of variables on civil war prevalence is not as marked as in Table 1, but is still apparent. The proportion of zero observations in the sample is 78.9 and falls within the accepted band of 10 to 90 percent (Bagozzi et al, 2014) and the Vuong test rejects the probit model in favour of the zero-inflated Poisson. Following the suggestion of Cameron and Trivedi (2009), all regressions are estimated using robust standard errors.

Table 3: Probit, Poisson and Zero-Inflated Poisson of Civil War Prevalence 1960-2013

	(1)	(2)	
	Probit	ZIP	
Explanatory Variables	Outcome	Outcome	Inflation
Opportunity Variables			
log real GDP	-0.150*	-0.082*	2.842**
	(0.071)	(0.034)	(0.371)
Real GDPPC Growth	-2.331**	-1.851**	$7.894^{\dagger}$
	(0.504)	(0.540)	(4.513)
Primary Exports/GDP	-4.197**	-5.571**	
	(1.087)	(0.816)	
Primary Exports/GDP^2	6.436**	7.850**	
	(1.522)	(1.323)	
log % Mountainous	-0.050	0.002	
	(0.092)	(0.027)	
Grievance Variables			
Polity Index	0.004	0.022**	-0.649**
	(0.007)	(0.006)	(0.166)
Polity Index^2	-0.011**	-0.009**	0.087**
	(0.001)	(0.001)	(0.022)
Eth Frac (F&L)	6.978**	4.489**	-4.222**
	(2.393)	(0.874)	(0.859)
Eth Frac^2 (F&L)	-5.150*	-3.573**	
	(2.424)	(0.832)	
Ethnic Dominance	0.380	0.110	-8.635**
	(0.299)	(0.085)	(1.347)
Religious Frac	-0.236	0.045	
	(0.729)	(0.200)	
Control Variables			
log Population	0.546**	0.198**	-3.176**
	(0.118)	(0.044)	(0.407)

Cold War	-0.094	0.072	0.017	
	(0.082)	(0.071)	(0.073)	
Rho	0.577**	-	-	
	(0.049)			
Constant	7.783**	-5.026**	-4.480**	-15.841**
	(1.624)	(0.547)	(0.566)	(3.350)
Observations	4286	4286	4286	
Zero Observations	-	-	3382	
Log likelihood	-1311.43	-2050.02	-1985.01	
Wald χ^2	162.45	-	-	
R^2	-	0.113	-	
Vuong test	-	-	6.50	
AIC		4128.04	4014.01	

Notes: Standard errors in parentheses, ** p<0.01, * p<0.05, † p<0.1; AIC = Akaike information criterion

To consider the robustness of the results a number of alternative specifications were considered. Table 4 is one example, with horizontal income inequality and ethnic discrimination added in place of ethnic dominance and religious fractionalisation, which increased the number of observations by over 100. These results remain consistent with the earlier probit and zero-inflated Poisson models, although mountainous terrain now has a substantive positive and significant effect. As for the measure of income inequality, only negative horizontal inequality (relative gap between mean national income and income level of the poorest group) has a significant positive effect on the likelihood of civil war. Higher ethnic discrimination, measured as the proportion of the largest discriminated ethnic group to the group in power, is also estimated to increase the likelihood of civil war. This suggest that discrimination and high income inequality with the poorest group decrease the odd, but as income inequality between the richest group and the country level increases, the likelihood of being a peaceful country also increases. The chances of not being in "complete peace" or likelihood of violence are driven by the lower income, which form the majority of a countries population. As before, the Vuong test and AIC statistics concludes that the zero-inflated Poisson is preferred to the standard probit.

Table 4: Probit, Poisson and Zero-Inflated Poisson of Civil War Prevalence 1960-2013

	(1)	(2)	
	Probit	ZIP	
Explanatory Variables	Outcome	Outcome	Inflation
Opportunity Variables			
log real GDP	-0.186**	-0.086*	1.063**
	(0.072)	(0.034)	(0.234)
Real GDPPC Growth	-2.520**	-2.115**	2.035*
	(0.505)	(0.537)	(1.037)
Primary Exports/GDP	-4.390**	-3.663**	
	(1.095)	(0.834)	
Primary Exports/GDP^2	6.578**	4.524*	
	(1.558)	(1.401)	
log % Mountainous	-0.033	0.056*	
	(0.091)	(0.027)	
Grievance Variables			
Polity Index	0.003	0.040**	-0.098*
	(0.010)	(0.007)	(0.051)
Polity Index^2	-0.010**	-0.004**	0.023**
	(0.001)	(0.001)	(0.006)
Eth Frac (FL)	7.841**	4.528**	-3.322*
	(2.218)	(0.794)	(1.624)
Eth Frac^2 (FL)	-6.151**	-3.723**	
	(2.152)	(0.710)	
ldg	0.795**	1.135**	-18.700**
	(0.218)	(0.168)	(4.932)
phi	-0.275†	-0.084	4.073**
	(0.148)	(0.109)	(1.359)
nhi	0.640**	0.230*	-1.461*
	(0.239)	(0.067)	(0.661)
Control Variables			
Population	0.567**	0.202**	-2.227**
	(0.119)	(0.045)	(0.508)

Cold War	0.014	-0.079	-0.122†	
	(0.083)	(0.073)	(0.074)	
Rho	0.573**	-	-	
	(0.051)			
Constant	7.840**	-5.461**	-3.854*	-5.641**
	(1.566)	(0.542)	(0.590)	(1.126)
Observations	4390	4390	4390	
Zero Observations	-	-	3481	
Log likelihood	-1318.83	-2025.06	-1950.79	
Wald χ^2	172.30	-	-	
R^2	-	0.151	-	
Vuong test	-	-	8.67	
AIC		4080.12	3951.58	

Notes: Standard errors in parentheses, ** p<0.01, * p<0.05, † p<0.1

AIC = Akaike information criterion, ldg = largest discriminated ethnic group, phi = positive horizontal inequality (relative gap between mean national income and income level of the richest group), nhi = negative horizontal inequality (relative gap between mean national income and income level of the poorest group)

Other variants of the zero-inflated Poisson were estimated replacing primary commodity exports with either mineral dependence, oil production and oil export; replacing the polity index with the freedom house measure, democracy and autocracy dummies; substituting income variables with urbanisation rate and life expectancy; and adding an Africa dummy showed the results to be relaively robust., with primary commodity dependence increases civil war risk, democracy, political freedom and higher urbanisation decrease civil war risk, and surprisingly the Africa dummy having special effect.[9]

The main results found in this article are twofold. Firstly, unlike most historical articles, explanations of civil war risk seems to rest with both greed and grievance variables. Secondly, regressions of civil war prevalence seems to perform better using the zero-inflated probit than the ordinary probit. The zero-inflated probit is able to statistically account for observable and latent factors that produce different types of peace observations. In the

[9] See appendix Table A1 and A2 for the additional results.

way the inflation equation was able to provide substantial insight on the different types of peace observations and its data generating process, while the outcome equation showed coefficient estimates that are more accurate on conflict outcome than standard probits.

Although the opportunity variables were always significant in both the standard probit and zero-inflated probits, it is the grievance variables that have provided the greatest insight. In the case of the ordinary probit, model misspecification has biased the estimates giving greater weighting to opportunity variables over grievance variables. This led to most empirical work to find opportunity or income variables as the main determinant of civil conflict, brandishing grievance type variables as having little explanatory power. As one takes a deeper look at what type of country is mostly associated with the always zero or "complete peace" group, the answer is often high income. By not distinguishing the different zeroes, the normal probit gave a likelihood of war calculation that included countries conditioned to not experience such an event. These countries main attribute is high income, and thus income variables were estimated with greater emphasis and significance, crowding out the grievance variables explanatory power.

By using a zero-inflated probit estimation process, greed and grievance variables are given equal emphasis, resulting in a clearer picture that both ethno-political and economic grievance matter, with substantial explanatory power in predicting civil war risk.

5 CONCLUSION

This paper has made a contribution to the burgeoning literature on the determinants of civil conflict by highlighting the possible impact of using the standard probit modes to model a situation when the binary conflict dependent variable is characterised by excess zeros. In such cases the zeros are not homogenous and the standard models do not account for the factors that produce the high proportion of zeroes. In the case of peace there is a big difference between a zero that reflects peace in a peaceful country and one that reflects a lull in conflict, where the number of battle deaths used in the definition of conflict falls below the threshold used to construct the variable. There is a more satifactory approach that has been suggeseted by Bagozzi et al (2014), a zero-inflated model, which treats the excess zeroes as a heterogeneous group of observations, accounting for observable and unobservable factors that produce the different types of zeroes.

Applying this model to Elbadawi and Sambanis (2002) data and using their model specification, showed that, differing results with the grievance terms becaming more significant. An intriguing result, that suggests that if these model can been used earlier the trajectory of the greed-grivance debate might have been different. Doing a similar exercise on an updated dataset of 134 countries for the period 1960 and 2013, and using a general greed-grievance empirical specification, provided further support for the need to recognise the problem of too many zero and deal with it within zero inflated model framework. Again grievance terms had reduced standard errors from the standard probit, with Polity variable showing signifiance and ethnicity and inequality being found to play a signifcant role: a contrast to earlier studies. The implications resulting from this study suggests a need for future research to recognise the heterogeneity of observations that have so far been treated as homogenous peace and assigned values of zero to try to better understand the heterogeneous process that are generating these outcomes.

6 REFERENCES

Bagozzi, B. E., Hill, D. W., Moore, W. H. and Mukherjee, B. (2014) Modelling two types of peace: the zero-inflated ordered probit (ZiOP) model in conflict research. *Journal of Conflict Resolution.*

Available: http://jcr.sagepub.com/content/early/2014/02/03/0022002713520530

Benoit, K. (1996) Democracies really are more pacific (in general): Reexamining regime type and war involvement. *The Journal of Conflict Resolution* **40**(4) 636-657.

Blattman, C and Miguel, E. (2010) Civil war. *Journal of Economic Literature* **48**(1) 3-57.

Buhaug, H., Cederman, L., and Gleditsch, K. S. (2014) Square pegs in round holes: Inequalities, grievances and civil war. *International Studies Quarterly* **58** 418-431.

Cameron, C. and Trivedi, P. K (2009) Microeconometrics using stata. *College Station, TX*: Stata Press.

Collier, P. and Hoeffler, A. (2004) Greed and grievance in civil war. *Oxford Economic Papers* **56**(4) 563-595.

Collier, P., Hoeffler, A., and Soderbom, M. (2004) On the duration of civil war. *Journal of Peace Research* **41**(3) 253-73.

Easterly, W. and Levine, R. (1997) Africa's growth tragedy: policies and ethnic divisions. *Quarterly Journal of Economics* **113** 203-49.

Elbadawi, I. and Sambanis, N. (2002) How much war will we see?: Explaining the prevalence of civil war. *Journal of Conflict Resolution* **46**(3) 307-334.

Fearon, J. D. and Laitin, D. D. (2003) Ethnicity, insurgency, and civil war. *The American Political Science Review* **97**(1) 75-90.

Fearon, J. D. (2005) Primary commodity exports and civil war. *Journal of Conflict Resolution* **49**(4) 483-507.

Hall, D. B. (2000) Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics* **56** 1030-1039.

Harris, M. N and Zhao, X. (2007) A zero-inflated ordered probit model, with an application to modelling tobacco consumption.*Journal of Econometrics* **141** 1073-1099.

Hegre, H., Ellingsen, T., Gates, S., and Gleditsch, N. P. (2001) Towards a democratic civil peace? Democracy, political change, and civil war, 1816-1992. *American Political Science Review* **95**(1) 33-48.

Kalyvas, S. N. and Kocher, M. (2007) How free is free riding in civil wars? Violence, insurgency, and the collective action problem. *World Politics* **59**(2) 177-216.

Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1) 1-14.

Richardson, L. F. (1960) *Statistics of deadly quarrels.* Pittsburgh, PA: Boxwood Press.

Ross, M. L. (2013) Oil and Gas Data, 1932-2011.
Available: http://hdl.handle.net/1902.1/20369 UNF:5:dc22RlDasveOTAJvwIjBTA [Data source].

Vance, C. and Ritter, N. (2014) Is peace a missing value or a zero? On selection models in political science. *Journal of Peace Research* **51**(4) 528-540.

Vuong, Q. H. (1989) Likelihood ratio tests for model selection and non-nested hypothesis. *Econometrica* **57**(2) 307-333.

Wilkinson, D. (1980) *Deadly quarrels: Lewis F. Richardson and the statistical study of war.* Berkeley, CA: The University of California Press.

7 APPENDIX

Table A1: Additional Results of Zero-Inflated Model: Varying in Primary Commodity

	(1)		(2)		(3)	
Explanatory Variables	Outcome	Inflation	Outcome	Inflation	Outcome	Inflation
Opportunity Variables						
Mineral Dependence	0.287**					
	(0.075)					
Oil Export Dependent			0.148			
			(0.108)			
Oil Production					-0.042	
					(0.140)	
Oil Production^2					0.002	
					(0.005)	
log real GDP	-0.092*	0.409**	-0.070	0.416**	-0.306**	0.249*
	(0.042)	(0.064)	(0.042)	(0.061)	(0.069)	(0.114)
Real GDPPC Growth	-2.079**	2.067	-2.342**	1.418	-3.080**	-2.343
	(0.589)	(1.837)	(0.623)	(1.198)	(0.723)	(1.809)
log % Mountains	0.127**		0.101**		0.137**	
	(0.027)		(0.026)		(0.038)	
Grievance Variables						
Polity Index	0.043**	0.078**	0.044**	0.073**	0.080**	0.318**
	(0.008)	(0.021)	(0.008)	(0.019)	(0.009)	(0.062)
Polity Index^2	-0.004**		-0.003**		0.004**	
	(0.001)		(0.001)		(0.002)	
Eth Frac (F&L)	5.265**	1.191**	5.367**	1.187**	5.254**	-1.123
	(0.737)	(0.384)	(0.741)	(0.374)	(1.037)	(0.656)
Eth Frac (F&L)^2	-4.392**		-4.288**		-4.908**	
	(0.672)		(0.672)		(1.007)	
ldg	0.436*	-12.581**	0.374*	-12.392**	1.169**	-20.612**
	(0.191)	(1.704)	(0.192)	(1.621)	(0.288)	(3.585)
phi	0.094	0.596**	0.103	0.589**	0.103	1.246**
	(0.091)	(0.091)	(0.098)	(0.088)	(0.105)	(0.185)
nhi	0.224**	-0.710**	0.195**	-0.774**	0.381**	-0.261
	(0.072)	(0.210)	(0.072)	(0.225)	(0.097)	(0.244)

Control Variables						
log Population	0.126**	-0.966**	0.104*	-0.943**	0.223**	-1.291**
	(0.049)	(0.086)	(0.048)	(0.081)	(0.097)	(0.198)
Cold War	-0.117		-0.095		-0.225*	
	(0.073)		(0.075)		(0.095)	
Africa	0.126		0.172		0.630**	
	(0.096)		(0.096)		(0.143)	
Constant	-2.768**	5.491**	-2.788**	5.128**	0.389	13.058**
	(0.742)	(1.091)	(0.754)	(1.000)	(1.855)	(2.997)
Observations	4460		4460		2478	
Zero Observations	3543		3543		1910	
Log Likelihood	-2335.26		-2341.71		-1308.27	
Vuong test	8.16		8.13		7.13	

Notes: Standard errors in parentheses, ** p<0.01, * p<0.05

ldg = largest discriminated ethnic group, phi = positive horizontal inequality (relative gap between mean national income and income level of the richest group), nhi = negative horizontal inequality (relative gap between mean national income and income level of the poorest group)

Table A2: Additional Results of Zero-Inflated Model Varying in Income and Democracy

	(1)		(2)		(3)	
Explanatory Variables	Outcome	Inflation	Outcome	Inflation	Outcome	Inflation
Opportunity Variables						
log Real GDP					-0.066	0.403**
					(0.037)	(0.056)
Real GDPPC Growth					-2.262**	0.774
					(0.457)	(0.822)
Primary Exports/GDP	-2.212**		-1.838**		-1.605*	
	(0.667)		(0.686)		(0.739)	
Primary Exports/GDP^2	1.958		1.618		1.522	
	(1.129)		(1.147)		(1.253)	
Urbanisation Rate	-0.103*	1.217**				
	(0.049)	(0.242)				
Life Expectancy			-0.887**	1.861**		
			(0.233)	(0.575)		
Δ Life Expectancy			-7.353**	-8.988		
			(2.280)	(6.213)		
log % Mountains	0.048*		0.075**		0.0788**	
	(0.029)		(0.024)		(0.026)	
Grievance Variables						
Polity Index	0.034**	0.277**	0.032**	0.102**		
	(0.006)	(0.058)	(0.007)	(0.028)		
Polity Index^2	-0.006**		-0.008**			
	(0.001)		(0.001)			
Democracy					-0.420**	1.123**
					(0.099)	(0.186)
Eth Frac (F&L)	5.702**	0.253	6.011**	1.116**	5.859**	1.346**
	(0.658)	(0.472)	(0.669)	(0.403)	(0.733)	(0.372)
Eth Frac (F&L)^2	-4.570**		-4.710**		-4.526**	
	(0.593)		(0.603)		(0.674)	
ldg	1.134**	-17.760**	0.767**	-11.626**	0.476**	-12.890**
	(0.139)	(2.687)	(0.187)	(2.006)	(0.181)	(1.645)
phi	0.199*	1.153**	0.186*	0.717**	0.178*	0.575**
	(0.077)	(0.167)	(0.081)	(0.089)	(0.081)	(0.075)
nhi	0.185**	-0.652*	0.207**	-0.439	0.203**	-0.773**
	(0.064)	(0.268)	(0.074)	(0.283)	(0.072)	(0.242)

Control Variables						
log Population	0.093**	-0.703**	0.031	-0.642**	0.072	-0.967**
	(0.294)	(0.085)	(0.032)	(0.089)	(0.046)	(0.077)
Cold War	-0.197**		-0.239**		-0.222**	
	(0.064)		(0.072)		(0.068)	
Africa	0.057		-0.066		0.200*	
	(0.083)		(0.088)		(0.095)	
Constant	-3.802**	4.673**	0.574	1.667	-2.647**	5.228**
	(0.549)	(1.350)	(1.048)	(2.327)	(0.701)	(0.865)
Observations	5083		4998		4446	
Zero Observations	4018		3945		3528	
Log Likelihood	-2775.87		-2730.70		-2342.33	
Vuong test	9.47		7.97		8.46	

Notes: Standard errors in parentheses, ** p<0.01, * p<0.05

ldg = largest discriminated ethnic group, phi = positive horizontal inequality (relative gap between mean national income and income level of the richest group), nhi = negative horizontal inequality (relative gap between mean national income and income level of the poorest group)