Evaluating UN Peacekeeping with Matching to Improve Causal Inference

Michael J. Gilligan and Ernest J. Sergenti*

Department of Politics

New York University

October 20, 2006

*We thank Kristian Gleditsch, Birger Heldt, Macartan Humphreys, Nicholas Sambanis and the other participants in the Peacekeeping Working Group, which is generously funded by the Folke Bernadotte Academy, Sweden.

Abstract

Previous statistical studies of the effects of UN peacekeeping have generally suggested that UN interventions have a positive effect on building a sustainable peace after civil war. However recent methodological developments have called these findings into question because the cases in which the UN intervened were quite different in a variety of factors than those in which they did not. We argue that this problem arises because of the nonrandom assignment of UN operations to civil conflicts. As a result of this problem estimates of the effects of the UN operations were largely a result of the researchers' *assumptions* about the proper statistical model to use rather than the data. Therefore we cannot be sure if the estimated causal effect of peacekeeping is due to peacekeeping itself or to the functional form of the model that the researchers chose. We correct for this problem by pre-processing our data using matching techniques on a sample of UN interventions in post-Cold-War conflicts and find that the UN interventions are indeed effective in the sample of post-civil conflict interventions, but that UN interventions while civil wars are still ongoing have no effect.

1 Introduction

Does peacekeeping *cause* peace? Establishing causality as opposed to mere correlation is one of the most difficult endeavors an empirical researcher can undertake in the social sciences. Nowhere within the field of international relations is that endeavor more pressing than in evaluating the effectiveness of UN peacekeeping operations. Previous statistical studies of the effects of UN peacekeeping have generally suggested that UN interventions have a positive effect on building a sustainable peace after civil war (Doyle and Sambanis (2000); Fortna (2004)). However recent methodological developments have cast something of a shadow over these results (King and Zeng, 2006*b*). The problem is that the cases in which the UN intervened were quite different from those in which they did not, and therefore estimates of the effects of the UN operations were largely extrapolations from the available data. As a result the findings were primarily driven by the researchers' *assumptions* about the proper statistical model rather than the data, and therefore we cannot be sure if the estimated causal effect of peacekeeping is due to peacekeeping itself or to the functional form of the estimation technique that the researchers chose. As we discuss below this extreme-counterfactual problem arises because of the nonrandom assignment of UN peacekeeping missions to civil conflict. If we fail to control for this problem we cannot be sure if the estimated causal effect of peacekeeping is due to peacekeeping itself or to other factors that are common among cases that received peacekeeping operations and not common among cases that did not.

We correct this problem with a new method (Ho et al., 2005) in which we pre-processing our data using matching techniques on a sample of UN interventions in post-Cold-War conflicts. Our estimates suggest that failing to correct this problem leads to biased causal inferences about the effects of UN intervention in both post-civil-conflict settings and ongoing civil wars. In the former sample our results indicate that failing to correct this problem substantially *underestimates* the effectiveness of

UN intervention. In the latter sample our results indicate that failing to correct the problem leads to a significant *overestimate* of the effectiveness of UN intervention.

When evaluating whether UN peacekeeping accomplishes its goals, an obvious question—indeed a question that must be answered before we can even begin evaluating peacekeeping—is "What are those goals?" Put another way what should be one's criterion for peacekeeping success? Clearly there is more than one reasonable answer to such a question.[1] In this paper we employ a simple answer to that question, namely "Does UN intervention increase the likelihood that a country will transition to a state peace?" We find that the answer to this question is a very clear "yes" in post-conflict situations but that UN interventions have no effect on transitions to peace while conflicts are still ongoing.

In short, we establish the causal efficacy of UN peacekeeping (or lack thereof depending on the samples) by employing techniques that match "treated" observations (those that received a UN operation) with similar "control" cases (those that did not receive a UN operation). In that way we are able to establish that the effect on peace if any was truly due to the presence of a peacekeeping mission and not to other extraneous factors because we only compare cases where those extraneous factors are quite similar, and therefore could not be an explanation for the observed variation.

The paper is organized as follows. In the next section we briefly review some earlier quantitative efforts to evaluate the effectiveness of UN peacekeeping missions. In section 3 we discuss some of the methodological problems present in any attempt to evaluate UN peacekeeping and our method for addressing those problems. In section 4 we turn to the first of our samples—the post conflict sample—and discuss the cases included in the sample, the confounding variables that we controlled for in our analysis and our results. We show that the effect of UN peacekeeping in this sample is even greater than previous studies have indicated once the nonrandom assignment of peacekeeping missions is taken into account. Section 5 proceeds similarly to section 4 except that we discuss our "in-war" sample. In this sample we show that the UN has no effect in reducing or increasing the length of the war. We also show that controlling for nonrandom assignment is very important for this sample because doing so produces estimates that indicate that UN operations do *not* shorten wars, whereas failing to correct for nonrandom assignment would lead incorrectly to the inference that the UN is quite successful at shortening wars.

2 Previous Evaluations of UN Peacekeeping

In this section we review the previous large n quantitative evaluations of peacekeeping that have been undertaken. It should go without saying that our intent is not to be critical of the work we discuss in this section. Each piece reviewed here is an important contribution to our understanding of peacekeeping. We are simply making the point that new techniques can improve on those contributions. For the purposes of discussion we divide these studies into two categories—those that evaluate the UN decision to deploy missions and those that evaluate the performance of peacekeeping missions once deployed.[2]

Regarding the studies in the first category, Gilligan and Stedman (2003) and Fortna (2004) evaluate whether those missions deployed to the appropriate civil wars according to some set of normative criteria, and they find that the record of Security Council deployment decisions is something of a mixed bag. The UN seems to deploy missions (as most would hope that it would) to those civil wars that suffer the most casualties. However there appears to be some regional bias in the way the missions are deployed—Africa and especially Asia appeared to be substantially less likely to receive a peacekeeping mission than similar civil wars in Europe or Latin America. The analyses in this first category have a very important implication for studies that do, namely they show that UN interventions are not randomly assigned. The two articles mentioned above indicate that the cases in which the UN intervenes have clearly identifiable differences from cases where the UN does not intervene. As such we must worry about the effects of nonrandom assignment when making causal inferences about UN peacekeeping—a point to which we return in greater detail below.

Turning to the second category the first study of which we are aware is Doyle and Sambanis (2000). They judge a peacekeeping mission as a success if, following that mission, three criteria are all met: a) the war ends and does not resume for some specified period of time b) residual violence does not resume within the same period of time and c) there is some democratization within that time period. They find that, for the most part, UN peacekeeping is positively correlated with their measures of peace building success.

The Doyle and Sambanis article has come under criticism of late, however. The analysis in the article has been shown to be highly "model dependent." In other words, the estimated causal effect of UN peacekeeping, conditional on covariates, changes dramatically with seemingly minor modifications to the specification (King and Zeng, 2006*a*). King and Zeng show that simply adding an interactive term between conflict duration and the presence of a UN operation altered Doyle and Sambanis' results such that UN peacekeeping was least effective for short wars whereas the original Doyle and Sambanis results suggested that the UN should be most effective in such wars.

Model dependence arises when researchers make faulty counterfactual inferences by comparing samples that are too disjoint. In Doyle and Sambanis' analysis, cases where the UN intervened were poorly matched to counterfactuals in which the UN did not intervene. For example as shown in Figure 1 there are no cases of post-conflict settings in which the UN intervened and in which the previous war had log cumulative battle deaths of less than five whereas there are many cases where the UN failed to intervene in that category. Thus predictions about what would have happened had the UN intervened following a conflict with relatively low casualties are largely extrapolations from the data and not based on the data itself.

Speaking metaphorically, Doyle and Sambanis' analysis compares "apples and oranges" or more precisely it makes inferences about the effect of a given treatment on apples based on observations of the effect of the treatment on oranges. With no data on which to base such inferences their causal claims were based entirely on modeling assumptions rather than real world observations. Our paper explicitly corrects this problem by using matching techniques to reduce model dependence. To continue the metaphor we make counterfactual inferences about the effect of the treatment on apples by examining its effect on other apples.

A second contribution of our research over that of Doyle and Sambanis is our treatment of ongoing wars. Doyle and Sambanis examined seven cases of ongoing conflicts in their analysis however they pooled them with civil wars that had ended. We argue that there are good reasons to suppose that the data generating process of ongoing wars is different than that for wars that have ended. Indeed the literature on civil war implicitly recognizes that pooling is inappropriate by treating "war onset" differently from "war duration."[3] In this paper we allow the data generating process of war onset (i.e. the duration of the post conflict period) to be different from that of war duration.

Another important contribution to the evaluation of the effectiveness of peacekeeping missions is provided by Fortna (2004).[4] A key improvement of her analysis is that it uses a more straightforward measure of UN effectiveness—whether UN peacekeeping missions prolong ceasefires and she finds that they do. Although there are some differences between Fortna's "ceasefires" and our "post conflict periods" our analysis is similar to Fortna's in that it shows that UN interventions increase the duration of post conflict peace. Once again, however, we cannot tell if these results are due to UN intervention or other characteristics of the civil wars into which the UN tends to intervene. Furthermore Fortna does not evaluate the effectiveness of UN peacekeeping operations in civil wars that have not yet ended.

The contributions of our approach beyond Fortna's, then, are twofold: First we also assess the effect of UN interventions on the likelihood of transition to peace in states that are still experiencing civil war, a portion of the sample she misses by only including countries in which there are ceasefires. Second we also correct the bias that arises from the nonrandom assignment of UN interventions and find that failing to correct for the nonrandom assignment underestimates the effectiveness of UN missions in keeping peace.

3 Methodology

In evaluating the effect of peacekeeping missions in prolonging periods of peace or shortening periods of war it is useful to make an analogy to a laboratory experiment. In a laboratory setting the researcher is careful to make sure that the treatment she wishes to study is randomly assigned. In that way she can be sure that any difference in the outcome between the treated and untreated are due to the treatment and not some other differing characteristics of the treated and the control groups. By virtue of randomizing the treatment, any such characteristics will not be correlated with whether or not the object under study received the treatment. By analogy we can think of UN intervention as the treatment, the effects of which we wish to ascertain. The duration of peace or war is the affected outcome. Unfortunately the treatment, in this case a UN intervention, is not randomly assigned. As such the researcher must be careful to make sure that any differences between the treated group and the control group are in fact due to the treatment and not due to some other characteristic that was common to the treated group but not the control group. For example, if, as shown in Figure 1, the post conflict periods in which the UN intervenes are characterized by higher levels of cumulative battle deaths from the previous war and higher levels of cumulative battle deaths are associated with shorter periods of peace, then not controlling for the confounding influence of cumulative battle deaths will bias results.

The prevailing methodology to address these problems is to add a variable to a parametric estimator (ordinary least squares regression, probit or what have you) in order to control for that variable's confounding influence. But, as King and Zeng

(2006*a*,*b*) and Ho et al. (2005) argue, if the distributions of the confounding factors do not overlap—in our case, for example, the distribution of the log of the cumulative battle deaths is farther to the right (higher) for those cases in which the UN intervened than it was for the cases where the UN did not intervene—then the causal inferences are sensitive to the unverifiable specification assumptions of the parametric model.

Notice the relationship between the extreme counterfactual problem discussed by King and Zeng and the nonrandom selection problem—nonrandom selection can *cause* the extreme counterfactual problem. If the UN allocated peacekeeping missions randomly conflicts with low casualty counts would receive missions at roughly the same rate as conflicts with high casualty counts. Counterfactuals would no longer be extreme because the distributions of the treated and the untreated cases would overlap. Randomization of the treatment eliminates extreme counterfactuals. It insures that the factual (treated) cases are similar on average to the counterfactual (untreated cases). However if the UN is choosing cases with relatively high casualties to allocate peacekeeping missions then the distributions will not overlap and counterfactuals will be extreme.

To improve the reliability of one's causal inferences in these circumstances, Ho et al. (2005) recommend pre-processing one's data using matching techniques. By dropping, grouping, and/or repeating observations, matching creates a sample of the data in which the difference between the treated distribution and the non-treated distribution is reduced or eliminated. When matching is perfect, and the distributions match exactly, then a simple difference in means of the treated and the untreated is sufficient to obtain the effect of treatment. But in most cases matching is not perfect. In order to adjust for any remaining imbalance (and in our case to account for right-censoring) one should then use the same parametric model one would have applied to the full data set on the matched sub-sample. This, as Ho et al. (2005) describe, greatly reduces the role of functional form and specification assumptions of the parametric model, leading to more reliable causal inferences. This method is "double robust:" if the matching is not complete but the parametric model is properly specified, or if the model is not properly specified but the matching is adequate, causal estimates will be consistent.

In this paper, we follow the recommendations of Ho et al. (2005). With each of our samples, we first identify the confounding factors. Confounding factors are those variables that (1) influence the dependent variable conditional on treatment, (2) are correlated with the treatment variable, and (3) are causally prior to treatment. Thus, for example, the growth rate of a country after a UN intervention is not a confounding factor and does not need to be included. This is so even though growth does appear to influence periods of peace (see Collier and Hoeffler (1998, 2004) and Miguel, Satyanath and Sergenti (2004)) and is correlated with a UN presence, because it is not causally prior to a UN intervention and therefore the fact that the UN is there may be creating the circumstances for higher growth. Including growth after a UN intervention in our analysis would lead to post-treatment bias, incorrectly controlling for the consequences of treatment, thus biasing the overall causal effect (Rosenbaum, 2002; King and Zeng, 2006*a*,*b*). Therefore, when selecting our confounding factors, we must be careful that they meet all three requirements outlined above.

Once we have developed a list of confounding factors, we use one-to-one nearest neighbor matching with replacement to obtain matches from all of our control (nonintervention) cases for each of our treatment (intervention) cases.[5] One-to-one nearest neighbor matching was preformed with GenMatch (Sekhon, 2006*b*), which uses a genetic search algorithm to determine the weight each confounding factor should receive in order to obtain optimal balance (Sekhon, 2006*b*; Sekhon and Mebane, 1998). Once we obtained matches, we analyzed the subset of a data with a Cox proportional hazards model. The Cox model corrects for the problem of right censoring and any remaining imbalance. We also ran the same specification with different types of parametric model (exponential, weibull, and lognormal) in order to calculate expected duration ratios. We prefer the Cox results as it is semi-parametric technique and is therefore less dependent on any particular parametric assumption. Last, we present results of the Cox regression on the full sample to compare the causal effects from the different procedures.

Before we proceed to our analysis, we should discuss the other commonly used methods for dealing with nonrandom assignment of treatment: instrumental variables and the modified sample-selection model (Przeworski (2005) and Greene (2003) pages 787-89). Neither of these techniques addresses the need for a valid instrument or the point made by King and Zeng (2006*a*,*b*) and Ho et al. (2005) about strong modeling and functional form assumptions. With regard to the need for an instrument, there are good theoretical reasons in our particular case to presume that one does not exist. If the Security Council is being careful about how it allocates its scarce peacekeeping dollar, anything that makes the war last longer (or shorter) should be taken into account by peacekeeping planners. With regard to the model dependence problem, the way in which these techniques deal with the nonrandom assignment/extreme counterfactual problem is completely model driven. Effectively the researcher creates a model of the treatment-assignment (selection) process, uses that model to generate cases (i.e. predictions) of counterfactuals and then compares the factual cases to these predicted counterfactuals. However, if the assumed model is not the correct one, the counterfactual predictions may be invalid and therefore the counterfactual inferences will be.

No method is without its drawbacks however. The drawback of our method is that it cannot control for unobserved variables that might affect both war outcomes and UN intervention the way the above mentioned approaches can (assuming that the researcher has a valid instrument and knows the proper functional form of the data generating process). Our own view is that we have sound theoretical reasons for thinking that a valid instrument does not exist and furthermore that we, as political scientists, have a much better grasp of potential confounding covariates (variables that are correlated with both war outcomes and UN intervention) than we do of the precise functional form of the relationship between civil war, UN intervention and these confounding covariates.

4 The Effect of UN Intervention in Prolonging Peace: The Post-Conflict Sample

4.1 Sample Definition

We start with an analysis of the post-conflict sample. Our dependent variable in this section is the number of months a given country did not experience war as defined by the PRIO/Uppsala Armed Conflicts Dataset, Version 3.0 (Gleditsch et al., 2002). Our data set includes the dates at which wars started and ended through December 2003, so that is the the point in time at which our dependent variable is right censored.[6] However we have covariates only through December 2002 and so we do not include any UN missions that began after that date or other covariates past that date. As we focus on the post-Cold War period, we restrict our attention to periods of peace that existed between January 1988 and December 2003, following the end of a civil war.[7] A period of peace either ends when a civil war starts or is right censored at December 2003. In all, there were 107 such periods of which the UN intervened into 20. Note that we are using a country-level, not a war-level, measure of peace. We do not examine the length of peace after the end of a particular war; instead we examine the length of peace for an entire country. One complication is that some countries may experience several conflicts concurrently. These conflicts are often difficult to disentangle and are almost certainly not independent observations, which is why we have adopted the country-level approach for our dependent variable.[8]

Data for our treatment variable, whether the UN was present in a country during a particular month, come from the United Nations Peacekeeping web site.[9] In this section we examine all UN missions that address internal conflicts that ended in the period 1988 through 2002.[10] Table 1 lists all such UN missions. From these data, we created an indicator variable equal to 1 if the UN was present at any point during a period of peace, and 0 otherwise. In other words a post-conflict period either receives a UN treatment or it does not. We adopt this measure of the UN intervention because we wish to test if a UN intervention led to a lasting peace, that is peace even after UN troops left the country. Using a time-varying approach for UN intervention (that is coding UN intervention 1 only for the periods the UN was present) would only show if the UN intervention reduced the hazard rate of war while it was stationed in the country.[11] Our UN mission variable is a country-level variable so it does not measure the effect of any particular UN mission, e.g. UNMIH present in Haiti from September 1993 to June 1996. Instead we examine the entire effect of a continuous UN presence in a country during a particular period of peace, e.g. the total effect of the UN presence in Haiti under various different acronyms (UNMIH, UNSMIH, UNTMIH and MIPONUH) from September 1993 to March 2000.[12]

4.2 Confounding Factors

In this subsection we outline the confounding factors on which we match our treated and control cases.

Log (Cumulative Battle Deaths) from Last War Our battle death data come from Lacina and Gleditsch (2005) who measure the number of soldiers and civilians killed in a given year of civil conflict. We take the log of the total number of persons killed from the previous conflict. Battle deaths is a confounding factor because it is negatively correlated with the duration of the subsequent peace and positively correlated with the probability that the UN will intervene (Gilligan and Stedman, 2003). Indeed the “before-matching” balance statistics in Table 3 show that the cumulative battle deaths from the previous war of the treated group are significantly greater than those in which it does not intervene, and the Kolmogorov-Smirnov (K-S) test indicates that the two distribution are very different with a p-value of zero.

Duration of Last War. This measure is simply the duration of the previous war in months. The duration of the previous war is positively correlated with the duration of the subsequent peace (Hartzell, Hoddie and Rothchild, 2001; Fortna, 2004) and in previous studies has been shown to be positively correlated with the probability that the UN will intervene (Gilligan and Stedman, 2003). The before-matching balance statistics in Table 3 corroborate this finding—the UN tends to become involved following wars of longer duration. The p-value on the K-S test is quite small indicating that the two distribution are indeed significantly different.

Ethnic Fractionalization. Some authors have argued that ethnic fractionalization affects civil war onset and therefore would be negatively correlated with the duration of the peace period, however this finding is far from robust (Fearon and Laitin, 2003). No previous studies have linked ethnic fractionalization to likelihood of UN intervention and indeed the before-matching balance statistics in Table 3 indicate that there is no difference between the treated and control groups on this dimension. Still we include the variable as a confounding covariate to avoid any omitted variable bias. Our measure comes from the Fearon and Laitin (2003) dataset, which is derived from the Atlas Narodov Mira 1964. It is time invariant allowing us to extrapolate the data forward for the years not in the original Fearon and Laitin dataset.

Log (Population Size) Population size is one of the Collier and Hoeffler (1998, 2004) "robust" findings regarding civil war. In their results countries with large populations were more prone to civil war. As shown in Table **??** populations of countries in which the UN intervened were smaller than the countries in which they did not intervene and both the t-test on the difference of means and the K-S test on the difference of the distributions indicates that this difference were highly significant. Our population data come from the World Bank World Development Indicators, 2005.

Log (Mountainous) Fearon and Laitin (2003) found that countries with more mountainous terrain are more likely to have outbreaks of civil war. Therefore mountainous terrain is negatively correlated with the duration of the peace period. As shown in Table 3 the UN also appears to intervene in more mountainous countries (as would be predicted by the contention that the UN becomes involved in more intractable wars). Although not strictly significantly different at the 0.05 level the p-values on both the t-test and K-S test are low enough for concern and so we match on this variable. Our measure comes from the Fearon and Laitin dataset. It is time invariant and so we extrapolate the data forward for the years not in the original dataset.

Log (Military Personnel) The strongest result in Gilligan and Stedman (2003) was that the size of a country's government army was negatively correlated with the likelihood that the UN would intervene in that country. Furthermore periods of peace should be longer in states with large armies because large government armies should deter would-be rebels *ceteris paribus*.[13] Thus, we include a measure of the log of military personnel at the beginning of the post-conflict period here. Our measure of military personnel comes from the Correlates of War National Material Capabilities dataset, Version 3.02 (David, Bremer and Stuckey, 1972; Singer, 1987)).

Polity before Last War Hartzell, Hoddie and Rothchild (2001) suggest that democracies are less prone to civil war, so democracy score should be positively correlated with duration of peace. Andersson (2000) suggests also that the UN is more likely to intervene in democracies. Gilligan and Stedman (2003) found no evidence for this contention. However as shown in Table 3 our before-matching balance statistics do indicate that the treated and control groups may be different, however they are different in the opposite of Andersson's hypothesis. The Polity scores of our treated group were significantly lower (less democratic) than were those of the control group. Thus we include it as a confounding factor. This variable comes from Polity IV dataset. We use the Polity2 measure from the last month before the previous civil war started.

Regional Controls Gilligan and Stedman (2003) and Fortna (2004) found that the probability of a peacekeeping mission being allocated varied by region. We also found that duration of peace varied by region—Using a log-normal parametrization and picking sub-Saharan Africa as the base category, the expected duration of peace was 9.8 times longer in Latin America, 2.3 times longer in Asia, 3.3 times longer in Eastern Europe and the countries of the former Soviet Union, and 6.0 times longer in North Africa and the Middle East. With respect to a UN presence, according to probit estimates (not shown,)the UN did not intervene in any of our post conflict Asian cases, while it was more likely to intervene in Latin America, Eastern Europe and North Africa/Middle East by 40, 19 and 24 percentage points respectively. For these reasons we include region as a confounding factor.

4.3 Matching, Balance, and Proportional Hazard Results

Having identified the confounding factors, we turn to the creation of our matched sub-sample. As we mentioned in the Methodology section, we employ GenMatch to generate matched observations using one-to-one matching with replacement.[14] Table 3 presents "before" and "after" balance statistics. There are several methods for describing the balance of a sample (see Imai, King and Stuart (2006); Sekhon (2006*a*)). In what follows, we look at three indicators: the difference in means, the *p*-values from a *t*-test on the difference of means, and, where possible, the *p*-values from a

Kolmogorov-Smirnov test of similar distributions.

The after-matching statistics presented in Table 3 indicate that we have achieved very good balance. We have achieved excellent balance on eleven of our 14 confounding factors. On two of the remaining three, duration of the last war and log of military personnel the, p-values on the K-S test are low at 0.152 and 0,038 resepctively, but the p-vales on the t-tests are very good (0.331 and 0.274). On the other remaining covariate, Polity before the War, the p-value on the t-test for is 0.167 (somewhat low) but the p-value on the K-S test for that variable, 0.234, is quite respectable. It is not surprise that we were unable to achieve balance on right-censored observations. because, if UN intervention is working, post-conflict peace periods should be failing less frequently following a UN intervention and therefore should be right censored more frequently.

The traditional method for calculating the causal effect of a treatment is to take a difference in means. However, given that our dependent variable is the duration of a period of peace and, as is evident from Table 3, 75 percent of periods with a UN intervention and 46 percent of periods without were right censored, we employ a duration model instead. Doing so also allows us to control for the right censoring as well as any remaining imbalance with the confounding factors.

Table 4 presents the results of a Cox regression on the unmatched and matched samples. A UN presence results in the reduction of the hazard rate of transition to war by over 80 percent or by a factor of 0.187. The comparable estimate from the unmatched sample is that UN intervetion reduces the hazard rate of transition to war by a factor of 0.32, which indicates that estimates of the effect of UN intervention, when they do not correct for the nonrandom assignment of the UN intervention are biased. They underestimate the effectiveness of UN intervention in prolonging post-conflict peace. We should mention that the coefficients on the remaining covariates in the specifications are not very meaningful because we have matched on all of these variables. The only reason for including them is to control for any remaining imbalance in our data. Furthermore only the coefficients for Sub-Saharan Africa regional variable are presented because estimates for other regional variables failed to converge in the matched sample.[15]

We chose the Cox model for our baseline specification because we did not want to impose a particular parametrization on the duration process. However, the results were substantively the same using the exponential, Weibull, and lognormal models. Those results indicated that a UN intervetion prolongs the expected duration of the peace by 6.2, 6.2, 3.4 with the exponential, Weibull, and lognormal respectively.

Figure 2 illustrates the monthly probability of transition from peace to war when the UN has intervened and when it has not. The top panel of Figure 2 graphs the probability of transitioning from peace to war estimated with a log normal regression. The lower panel of Figure 2 graphs the probability of transitioning from peace to war estimated with a Cox proportional hazard regression. The effects of a UN intervention are quite striking. For example as shown by the graphs from the Cox regression after 25 months of peace the estimated probability of transition to war is over one percent when the UN does not intervene while that probability is less than two-tenths of a percent when the UN has intervened. While these percentages may appear to be small it is important to remember that they are *monthly* probabilities so that over the course of a year or more the difference can be quite substantial.

To summarize, our results from the post-conflict sample clearly indicate UN intervention significantly reduces the proportional hazard of returning to war. The proportional hazard of returning to war in cases where the UN intervened is about 0.187 *ceteris paribus*. Our results also show that it is crucial when evaluating the record of UN missions in post-civil-war settings to take into account the nonrandom nature of how UN missions are allocated. Correcting for the nonrandom assignment of the UN treatment with matching techniques produced an estimated causal effect of UN intervention that was stronger than the effect estimated without correcting for nonrandom assignment. Not correcting for this problem produced an estimate of the proportional hazard of returning to war in cases where the UN intervened of about 0.32 as opposed to the 0.187 when we did correct for it. We will now turn to the other sample of cases that experience UN intervention the “in war” sample-those cases where peacekeepers intervene while the war is still ongoing.

5 The Effect of UN Intervention on Shortening Wars: The In-War Sample

5.1 Sample Definition

The source of our dependent variable and treatment variable are the same as discussed in the previous section. As in the post-conflict sample the data set includes the dates at which wars started and ended through December 2003, so that is the the point in time at which our dependent variable is right censored. However, we also include several variables coded by Fearon (2004). Fearon’s data set ends in 1999 and is missing several wars that are in the PRIO/Uppsala set because Fearon uses a higher battle deaths threshold for a conflict to qualify as a war. Therefore we were forced to exclude wars that began after December 1999.[16] Hence, we examine all conflicts that were identified by both the Armed Conflict Data set and Fearon (2004) and that were either on-going as of January 1988 or were initiated between January 1988 and December 1999. A period of war ends either when the country is no longer characterized by any civil conflict according to Fearon’s data set or is right-censored in December 2003, the point at which the Armed Conflict Data set ends. In all, there were 69 such periods. The UN was present in 16 of them. As in the previous section we could not include UN missions begun after December 2002 because of data limitations of other covariates.

One important feature of the in-war sample, as compared to the post conflict sample, is that the timing of UN interventions is more critical. With the post-conflict cases we were interested in the effect of the UN operation even after the peacekeepers left, which required that we treat the UN operation as time-invariant. It was reasonable to ignore the differences in the timing of the various UN interventions in that sample because most of these interventions occurred in the first month of the peace period. With the in-war case, it is the exception that the UN is present from the start.[17] Hence, to properly examine the effect of the UN, we cannot use a time-invariant UN intervention variable look at the whole period as we did with the post-conflict sample. Doing so would substantially underestimate the effectiveness of the UN mission, because a war could have been going on for many years before the UN intervened, and then, after the UN intervened, stop after a few months. If we were to look at the whole period with the same setup that we used for the post conflict period, the UN would be associated with the whole time of the war, which would seriously underestimate its effect. The relevant metric is instead "how long does the war last once the UN has intervened?"

We are not the first to notice this issue. Regan (2002) also highlights questions of timing with the war duration literature, which is why he argues for the use of a duration model with time-varying covariates, precisely in order to ascertain the effects of the interventions. For the same reason, we also estimate a time-varying covariates model for our in-war sample. In addition to the UN intervention variable, other time-varying variables that we have included are Non-UN Third Party Interventions, Log (Battle Deaths), Log (Military Personnel), and Log (Population).[18]

We have to be concerned about post-treatment bias with time-varying covariates however. For example, after a UN intervention, the number of yearly battle deaths could go down, which might affect the outcome of the war. Controlling for this factor after the UN intervenes will bias down the effect of the UN, but not controlling for it before the UN would fail to address the nonrandom assignment of the UN treatment. Our solution to these issues is to transform our time-varying data set by taking "snapshots" of the data for different durations of the in-war period. With those cases when the UN intervenes, we take a snapshot at exactly the month that the UN intervenes, freezing the time-varying variables at their values in that month. Since the other variables are time-invariant, this amounts to a normal time-invariant duration model. This way we can measure the effect of the UN. However, most of the cases in which the UN did not intervene start with an in-war duration of zero and the length of the war is one of our confounding factors. It would be inappropriate, therefore, to compare the effect of a UN intervention in, say, the fortieth month of a war, with a similar war in, say, its second month. Thus we need to take snapshots of each non-UN case at different durations of the war in order to match the treated UN cases appropriately. Doing so gives us the correct counterfactual: how much longer would the country have remained in a state of war had the UN not intervened? In other words we match on the duration of the war and on the level of the particular time-varying covariates at the time the UN intervened. We then examine the effect of the entire period of in-war from the time after the UN intervened, even if the UN left after a few years as it did in Somalia, Angola, and Cambodia. The difference from the post-conflict sample is that we only start to examine the in-war case once the UN has intervened, not before. Our source for the UN intervention measure is the same as with the post-conflict sample. In-war UN interventions are listed in Table 1.

5.2 Confounding Factors

We now present the confounding variables on which we matched in our in-war sample. Our discussion will be brief because many of the variables are identical to those in the post conflict sample. The data sources for such variables are the same as in the post conflict sample, but the construction of the measure may be different as discussed below. In particular for reasons described above some of the variables may be time varying whereas they were time-invariant in the post conflict sample.

Non-UN Third Party Interventions Regan (2002) shows that international interventions increase the duration of war. Furthermore, our before-matching balance statistics, shown in Table 5, indicate that proportionally more of the treated group (those that received UN operations) experienced a non-UN third party intervention than did the control group. The p-value on the t-test was low enough to cause concern and so we included this variable as a possible confounding covariate. Our data for this variable come from the PRIO/Uppsala Armed Conflict Dataset. The Non-UN Third Party Intervention variable is set to 1 for any month in which any conflict in a country is coded as a Type 4 (Internationalized Internal Conflict).

Log (Battle Deaths) This is a yearly measure of the total number of battle deaths for all wars occurring in a given country. As in the post conflict sample we are concerned that high battle deaths both reduce the hazard rate of peace and increase the likelihood that the UN will intervene. The before-matching statistics in Table 5 indeed indicate that the average battle deaths of the treated group are higher than that of the control group and the p-value of the K-S test indicates that the distribution of this variable is very different across those two groups. Note that this measure is the yearly level of battle deaths for the ongoing war. It is a time-varying measure in contrast to the measure we used in the post-conflict sample, which was cumulative battle deaths from the previous war, which is not time varying. The source, as before, is Lacina and Gleditsch (2005).

Coup/Revolution This is an indicator variable for whether the in-war period was a coup or revolution. Fearon (2004), who is our source for these data, found that such wars are shorter in duration. Our before-matching statistics indicate that it is unlikely that our treated and control groups are different along this dimension but we match on it to avoid omitted variable bias.

Sons of the Soil This is an indicator variable for whether or not the in-war period was characterized by a "sons-of-the-soil" movement. Again this is a Fearon (2004) measure. Fearon found that such wars lasted longer *ceteris paribus* than other types of wars. Our before-matching balance statistics indicate that those conflicts that receive a UN mission are significantly less likely to be "sons-of-the-soil" conflict than those conflicts that do not receive a UN mission.

Rebels Contraband Funded This is an indicator variable for whether or not the in-war period was characterized by a rebel group that was financed by trade in contraband (drugs, diamonds, etc.). Such funding has been shown to increase the duration of the war by Fearon (2004) who is our source for these data. Our before-matching balance statistics do not give great cause for concern but multivariate probit estimates (not shown) indicated the UN was significantly more likely to intervene in conflicts where the rebels were contraband funded and so we include this variable as possible confounding factor.

Ethnic Fractionalization. By some accounts ethnic fractionalization prolongs war. Whereas in the post-conflict sample discussed in the previous section there was virtually no difference between distributions of the treated and control groups there is a strong indication in the in-war sample that those conflicts that received a UN operation had lower levels of ethnic heterogeneity than those conflicts that did not. As shown in the pre-matched statistics in Table 5 the p-value on the K-S test indicates that the distributions of this factor in the treated and the control group are quite different.

Log (Population Size) Similar to the post conflict sample this variable is positively correlated with the duration of war and negatively correlated with the likelihood of UN intervention. The p-values on both the t-test and the K-S test show that the treated cases had significantly smaller populations than the untreated cases. As mentioned above, unlike in the post-conflict sample, however, this variable is time-varying.

Log (Mountainous) Following Fearon and Laitin (2003) we were concerned that wars may be prolonged in countries with more mountainous terrain and the cases in which the UN became involved were less mountainous on average as indicated by the before-matching balance statistics in Table 5. The p-value on K-S test reveals that the distributions of the mountainous variable were very different in the treated and the control groups.

Log (Military Personnel) Our reason for including this variable are the same as in the post conflict sample-UN intervention are negatively and strongly correlated with interventions in countries with large armies and we have theoretical reasons (if not empirical confirmation) for thinking that a large military is correlated with the likelihood that a war will end. The former point is corroborated by the before-matching balance statistics in Table 5. The average number of military personnel in those countries into which the UN intervenes is significantly smaller than the number of military personnel in countries where the UN does not intervene. Our source for these data is the same as in the post conflict sample. However unlike the post conflict sample this variable is time varying in the in-war sample.

Regional Controls Lastly, we control for differences between regions. The before-matching balance statistics in Table indicate that, indeed, treated cases are significantly less likely to be in Asia and North Africa/Middle East than control cases are. Probit analysis indicated that Latin America, Asia and North Africa/Middle East were respectively nine, six and five percentage points less likely and Eastern Europe 41 percentage points more likely to receive a mission than the base category of Sub-Saharan Africa Interestingly treated cases were not any less likely to be in Sub-Saharan Africa than control cases. Furthermore using a Weibull parametrization and picking sub-Saharan Africa as the base category, the expected duration of the war was 0.27 times shorter in Eastern Europe, 0.42 times shorter in North Africa/Middle East and 0.54 times shorter in Latin America.

5.3 Matching, Balance, and Proportional Hazard Results

We employed GenMatch to generate matched observations using one-to-one matching with replacement. As a robustness check we also use two-to-one matching and it did not change the balance statistics or the results appreciably. Table 5 presents before- and after-matching balance statistics. The after-matching statistics indicate that we have achieved good balance overall even with regard to right censoring, where we were unable to achieve balance in the post-conflict sample. The p-value of the t-statistics on the war duration variable is a bit low, but the $p$value of the K-S statistics is excellent. The only balance problem is with log of military personnel, which which was lower in treated cases than control cases even after matching.

In Table 6, we present the results of a Cox regression on the unmatched and the matched sample.[19] The importance of controlling for nonrandom assignment is clear. The results from unmatched sample indicate the the UN is quite good at helping countries transition out of war and into peace. The coefficient on the UN intervention variable suggests that proportional hazard of a country transitioning out of war and into peace is over two and a half times greater in cases where the UN intervened. However within the matched sample a UN presence has no effect. The hazard rate increase by only 1.05 times, and the effect is not statistically significant. In other words failing to correct for nonrandom assignment of UN operations in the in-war sample would lead researchers to overestimate the causal effect of the UN in shortening civil wars. As we mentioned above in our discussion of the post-conflict sample, the remaining coefficients in Table 6 are not meaningful because we have matched on all of these variables. They are included in the analysis only to control for any remaining imbalance. As in the post-conflict estimates we present only those regional-variable coefficients that converged in the matched sample.

Thus while the substantive conclusion of our in-war sample is quite different from the one we reached in the post conflict sample the methodological conclusion is the same—it is crucial that researchers control for the nonrandom assignment when evaluating UN peacekeeping operations. As a sidelight we might also point out that the different causal effects of UN operations in the post-conflict and in-war samples are evidence of the different data generating process occurring in those two samples, which provides further evidence of the inappropriateness of pooling those two samples.

6 Conclusion

Does UN peacekeeping *cause* peace? Our results indicate that an answer to that question is "yes" when the UN intervenes in post-conflict settings and "'no" when the UN intervenes while the war is still ongoing. The innovation of the approach that we have taken in this paper is that we really can claim the causal efficacy (or inefficacy depending on the sample) of UN peacekeeping because of the methodological techniques that we have employed. Other research has claimed to show that UN intervention can prolong post-conflict periods of peace but its methodology has cast something of a shadow over that conclusion because it has failed to take into account into account the fact that those missions are not randomly assigned to conflicts and therefore inferences about what would have happened in conflicts that did not receive UN missions had they received them was based upon making comparison to extreme counterfactuals. As such any claims in that research of causality (as opposed to mere correlation) are brought into question. Was UN intervention truly causing prolonged periods of peace or were the longer peace periods due to other factors that either by chance or design characterized the conflicts into which the UN intervened? Matching techniques have allowed us to get more reliable estimates of the causal effect of UN interventions in civil wars by taking into account the nonrandom assignment/extreme counterfactual problem of UN intervention.

Our results suggest that the UN truly has had an important and independent causal effect in prolonging periods of peace and indeed that causal effect is even larger than would be estimated had we not corrected for the nonrandom assignment of UN missions other scholars have previously indicated. With respect to our in-war sample we could find no causal effect from UN interventions in shortening the war. Controlling for nonrandom assignment was equally important for this result because our estimates when we did not correct for it indicated that UN operations were effective in helping countries transition out of war and into peace.

The implications of our finding for policy and the development of UN peacekeeping doctrine are clear. The UN is quite good at peacekeeping. It is not good at war fighting. Therefore the UN should concentrate it scarce peacekeeping resources on intervening in those cases where there is a peace to keep. While this is a point that has been raised by others, we reach this conclusion based on a more robust analysis of the causal effect (or lack thereof) of UN peacekeeping and therefore we can have much more confidence in the conclusion.

Notes

[1]For a lively discussion on this point see Druckman et al. (1997)

[2]A third type of quantitative study does not fit easily into this dichotomy but has offered important insights into peacekeeping success. Diehl, Druckman and Wall (1998) use factor analysis to provide a taxonomy of UN interventions that provides implications for evaluating those interventions. They point out that peacekeeping operations may be deployed for a variety of reason and that the force requirements for different types of missions are quite different. A force deployed to monitor a cease fire will be insufficient for "nation building" neither type of mission is suitable for a peace enforcement action.

[3]In the former category see Fearon and Laitin (2003), Fortna (2004), Collier and Hoeffler (2004) Miguel, Satyanath and Sergenti (2004) and Hegre and Sambanis (2006) among others. In the latter category see Balch-Lindsay and Enterline (2000), Regan (2002) Fearon (2004) and Hegre (2004) among others. Elbadawi and Sambanis (2002) combine both approaches.

[4]Hartzell, Hoddie and Rothchild (2001)) provide an example of a prior study of the effect

of third party intervention on the prolonging of ceasefires although it is not explicitly a study of UN peacekeeping. They find that peace agreements are more likely to be sustained if there is outside third-party enforcement and in some of their sample the UN plays such a role.

[5] We also used two-to-one matching as well. See below.

[6]Specifically, within the Armed Conflict Dataset, we took all Type 3 (Internal) and Type 4 (Internationalized Internal) armed conflicts. Also note that Version 3.0 of the Armed Conflict Dataset does not provide war end dates. War end dates were kindly supplied by Kristian Gleditsch.

[7]We are aware, of course, that, strictly speaking, the Cold War was still on in 1988, however the sea change in peacekeeping that occurred with the end of the Cold War really began in 1988.

[8]However it does come at a cost. The clearest example of this cost is for Indonesia. During the 1990's, the Armed Conflict Dataset identifies at least two intermittent conflicts, one in East Timor and another in Aceh. Whereas the conflict in East Timor ended in 1998, the one in Aceh started in 1999. Thus, our measure is unable to pick up the post-conflict period of peace in East Timor before it became an independent country in May 2002, because by our country-level measure Indonesia remained in a state of conflict due to the conflict in Aceh. We cannot therefore evaluate the role of the UN Peacekeeping mission in prolonging peace in East Timor. Fortunately, these cases are rare. The UN mission to East Timor is the only UN mission excluded due to this issue.

[9]http://www.un.org/Depts/dpko/dpko/index.asp

[10] Our temporal domain excludes from our sample two on-going missions that were initiated before 1988, UNFICYP in Cyprus (initiated in 1964) and UNIFIL in Lebanon (initiated in 1978), as well as any missions initiated after 2002, such as UNMIL in Liberia, UNMIS in Sudan or UNOB in Burundi.

[11] Our measurement choice is not without its drawbacks however in that it essentially gives the UN intervention "credit" for the months of peace before it intervened in the country. Fortunately this is not a major problem in practice. If the UN intervenes at all it almost always does so in the first month of peace. Of the 19 post-conflict UN interventions, the UN was present within the first month for 15 of them. For the other 4, the UN intervened after 12, 15, 20, and 38 months.

[12] Based on our definitions of our sample and our dependent variable, we exclude two UN peacekeeping presences initiated between 1988 and 2002 from our subsequent analysis, those in the Central African Republic and in East Timor. We exclude the Central African Republic presence because it is not post-conflict and all such peace periods are excluded from our sample by definition. We exclude the East Timor mission first for the reason discussed above—based on our definition of our dependent variable Indonesia does not belong in the post-conflict sample. Second after East Timor's independence the conflict had to be excluded due to missing data.

[13]We were unable to establish this hypothesized correlation between the size of the government army and the outbreak of war. We suspect, however, that this variable is probably endogenous to the threat of civil war. In other words the government must create a larger government army when the threat of a rebel movement is higher. In short countries with large government armies have large government armies because they are more civil-war. Despite the lack of empirical correlation we include the variable because we have theoretical reasons for doing so and to err on the side of caution in taking account of possible confounding factors.

[14]Given the small number of treated cases in our sample, in addition to one-to-one matching, we also tried two-to-one matching. The results of the subsequent analysis were not substantively modified.

[15]This failure to converge is not all that surprising given the small sample size of the the matched sample and the small number of cases in certain categories—there were no treated Asian cases for example.

[16] As a robustness check we also estimated our model excluding the variables that we obtained from Fearon's data set and found that our results are substantively equivalent using the larger sample that resulted from excluding those variables. We are most convinced by the results we present in this paper, however, because they control for confounding factors that Fearon found to be important, notably the availability of contraband to fund the rebel movement.

[17] This happened only in Rwanda and Tajikistan where a UN was already in those countries for other reasons when the civil war broke-out.

[18] Time-varying duration models are still an active body of statistical study (Kalbfleisch and Prentice, 2002; Box-Steffensmeier and Jones, 2004). There are potential problems with endogeneity of some of our covariates. For example, the number of battle deaths per year could be a function of the duration of the war, the number of military personnel could drop as the

war progresses and third party interventions may become more likely as the war progresses. As we mentioned in the previous section UN interventions are also probably correlated with the duration of the war, however we control for any effect this endogeneity has for causal inference with matching techniques as described below.

[19] The size of the unmatched sample in the in-war analysis is much larger than that of the post-conflict sample because of the time-varying approach we adopted in the in-war samples and the time-invariant approach we used for the post-conflict sample at described above.

References

Andersson, Andreas. 2000. "Democracy and UN Peacekeeping Operations." *International Peacekeeping* 7:1–22.

Balch-Lindsay, Dylan and Andrew Enterline. 2000. "Killing Time: The World Politics of Civil War Duration, 18201992." *International Studies Quarterly* 44:615–52.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists.* Cambridge University Press.

Collier, Paul and Anke Hoeffler. 1998. "On the Economic Causes of Civil War." *Oxford Economic Papers* 50:563–73.

Collier, Paul and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers* 56:563–96.

David, Singer J., Stuart Bremer and John Stuckey. 1972. Capability Distribution, Uncertainty, and Major Power War, 1820-1965. In *Peace, War and Numbers*, ed. Bruce Russett. SAGE pp. 19–48.

Diehl, Paul, Daniel Druckman and James Wall. 1998. "International Peacekeeping and Conflict Resolution: A Taxonomic Analysis with Implications." *Journal of Conflict Resolution* 42:33–55.

Doyle, Michael and Nicholas Sambanis. 2000. "International Peacebuilding: A Theoretical and Quantitative Analysis." *American Political Science Review* 94:779–802.

Druckman, Daniel, Paul Stern, Paul Diehl, Robert Johansen, A.B. Fetherston, William Durch and Steven Ratner. 1997. "Evaluating Peacekeeping Missions." *Mershon International Studies Review* 41:151–65.

Elbadawi, Ibrahim and Nicholas Sambanis. 2002. "How Much War Will we see?" *Journal of Conflict Resolution* 46:307–334.

Fearon, James D. 2004. "Why Do Some Civil Wars Last Longer than Others?" *Journal of Peace Research* 41:275–301.

Fearon, James D. and David Laitin. 2003. "Ethnicity, Insurgency and Civil War." *American Political Science Review* 97:75–90.

Fortna, Virginia Page. 2004. "Does Peacekeeping Keep Peace? International Intervention and the Duration of Peace after Civil War." *International Studies Quarterly* 48:269–92.

Gilligan, Michael and Stephen John Stedman. 2003. "Where Do the Peacekeepers Go?" *International Studies Review* 5:37–54.

Gleditsch, Nils Petter, Peter Wallensteen, Mikael Erikssom, Margareta Sollenberg and Håvard Strand. 2002. "Armed Conflcit: 1946-2001." *Journal of Peace Research* 39:615–37.

Greene, William. 2003. *Econometric Analysis.* Prentice Hall.

Hartzell, C., M. Hoddie and D. Rothchild. 2001. "Stabilizing the Peace After Civil War." *International Organization* 55:183208.

Hegre, Håvard. 2004. "The Duration and Termination of Civil War." *Journal of Peace Research* 41:243–52.

Hegre, Håvard and Nicholas Sambanis. 2006. "Sensitivity Analysis of Empirical Results on Civil War Onset." *Journal of Conflict Resolution* 50:508 – 535.

Ho, Daniel, Kosuke Imai, Gary King and Elizabeth A. Stuart. 2005. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametreic Causal Inference." Available at http://gking.harvard.edu/projects/cause.shtml.

Imai, Kosuke, Gary King and Elizabeth A. Stuart. 2006. "The Balance Test Fallacy in Matching Methods for Causal Inference." Available at http://gking.harvard.edu/projects/cause.shtml.

Kalbfleisch, John D. and Ross L. Prentice. 2002. *The Statistical Anlysis of Failure Time Data*. Wiley.

King, Gary and Langche Zeng. 2006*a*. "The Dangers of Extreme Counterfactuals." *Political Analysis* 14:131–59.

King, Gary and Langche Zeng. 2006*b*. "When Can History Be Our Guide? The Pitfalls of Counterfactual Inference." *International Studies Quarterly* .

Lacina, Bethany and Nils Petter Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population* 21:145–66.

Miguel, Edward, Shanker Satyanath and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* 112:725–53.

Przeworski, Adam. 2005. Is a Science of Comparative Politics Posible? In *Oxford Handbook of Comparative Politics*, ed. Carles Boix and Susan Stokes. Oxford University Press.

Regan, Patrick M. 2002. "Third Party Interventions and the Duration of Intrastate Conflicts." *Journal of Conflict Resolution* 46:55–73.

Rosenbaum, Paul R. 2002. *Observational Studies.* Springer.

Sekhon, Jasjeet S. 2006*a*. "Alternative Balance Metrics for Bias Reduction in Matchcing Methods for Causal Inference." Available at http://sekhon.berkeley.edu/matching.

Sekhon, Jasjeet S. 2006*b*. "Matching: Algorithms nad Software for Multivariate and Propensity Score Matching with Balance Optimization via Genetic Search." Available at http://sekhon.berkeley.edu/matching.

Sekhon, Jaskeet S. and Walter R. Mebane. 1998. "Genetic Optimization using Derivatives." *Political Analysis* 7:187–210.

Singer, J. David. 1987. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985." *International Interactions* 14:115–32.

Table 1: UN Missions in Internal Conflicts 1988 - 2002

Country	Mission	Start Date	End Date	Post Conflict Sample	In War Sample
Lebanon	UNIFIL	Mar-78	-	1	1
Angola	UNAVEMI	Dec-88	Jun-91		
	UNAVEMII	Jun-91	Feb-95		
	UNAVEMIII	Feb-95	Jun-97		
	MONUA	Jun-97	Feb-99		1
Namibia	UNTAG	Apr-89	Mar-90	1	
Central America (Nicaragua)	ONUCA	Nov-89	Jan-92	1	
Western Sahara (Morocco)	MINURSO	Apr-91	-	1	
El Salvador	ONUSAL	Jul-91	Apr-95	1	1
Cambodia	UNAMIC	Oct-91	Mar-92		
	UNTAC	Mar-92	Sep-93		1
Croatia	UNPROFOR	Feb-92	Mar-95	1	1
	UNCRO	Mar-95	Jan-96		1
	UNTAES	Jan-96	Dec-97		
	UNPSG	Jan-98	Oct-98	1	
Somalia	UNOSOMI	Apr-92	Mar-93		
	UNOSOMII	Mar-93	Mar-95		1
Mozambique	ONUMOZ	Dec-92	Dec-94	1	
Georgia	UNOMIG	Aug-93	-	1	1
Bosnia/Herzegovina	UNPROFOR	Jun-92	Mar-95		1
	UNMIBH	Dec-95	Dec-02	1	
FYR Macedonia	UNPROFOR	Dec-92	Mar-95		
	UNPREDEP	Mar-95	Feb-99	1	
Liberia	UNOMIL	Sep-93	Sep-97	1	1
Haiti	UNMIH	Sep-93	Jun-96		
	UNSMIH	Jun-96	Jul-97		
	UNTMIH	Aug-97	Nov-97		
	MIPONUH	Dec-97	Mar-00	1	
Rwanda	UNOMUR	Jun-93	Sep-94		
	UNAMIR	Oct-93	Mar-96	1	1
Tajikistan	UNMOT	Dec-94	May-00	2	2
Guatemala	MINUGUA	Jan-97	May-97	1	
Sierra Leone	UNOMSIL	Jul-98	Oct-99		
	UNAMSIL	Oct-99	Dec-05	1	1
Yugoslavia (Kosovo)	UNMIK	Jun-99	-	1	
Indonesia (East Timor)	UNTAET	Oct-99	May-02		1
D.R. of the Congo	MONUC	Dec-99	-	1	1
Total:				**20**	**16**

NOTE: UNFICYP is excluded because the war in Cyprus ended well before our sample period. MINURCA is excluded because there was no war in the Central African Republic prior to or during the UN involvement. UNMISET in East Timor is excluded because of missing data. Tajikistan transitioned out and then back into war during the UN mission according to the PRIO/Uppsala data set which accounts for the two post-conflict and in-war periods recorded in the table.

Table 2: Descriptive Statistics

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**
Post Conflict Sample, n=107				
UN Intervention	0.187	0.392	0	1
Log (Cumulative Battle Deaths)	5.889	3.935	0	13.229
Duration of the Last War	44.150	71.985	0	360
Ethnic Faractionalization	53.201	24.430	0.498	90.163
Log Population	9.240	1.341	6.363	12.139
Log(Mountainous)	2.33	1.417	0	4.421
Log(Military Personnel)	3.668	1.634	0.693	8.006
Log GDP per Capita	6.783	1.101	4.689	9.932
Polity 2	-0.393	5.971	-10	10
Eastern Europe	0.346	0.478	0	1
Latin America	0.112	0.317	0	1
Asia	0.121	0.328	0	1
Sub-Saharan Africa	0.327	7 0.471	0	1
North Africa/Middle East	0.0935	0.292	0	1
In-war Sample, n=3999				
UN INtervetion	0.136	0.343	0	1
Non-UN 3rd Party Intervention	0.126	0.332	0	1
Log(Battle Deaths)	6.749	1.754	0	10.780
Coup/Revolution	0.036	0.187	0	1
"Sons of the Soil"	0.387	0.487	0	1
Rebels Contraband Funded	0.418	0.493	0	1
Ethnic Fractionalization	55.036	26.606	0.498	90.163
Log(Population)	10.001	1.367	6.229	13.864
Log(Mountainous)	2.793	1.243	0	4.407
Log(Military Personnel)	4.384	1.472	0	7.244
Eastern Europe	0.068	0.251	0	1
Latin America	0.125	0.331	0	1
Asia	0..323	0.468	0	1
Sub-Saharan Africa	0.378	0.485	0	1
North Africa/Middle East	0.107	0.309	0	1

Table 3: Balance Statistics, Post-Conflict Sample

Variable	**Mean treated**	**Mean control**	***t*-test *p*-value**	**K-S test *p*-value**
Before Matching				
Log (Cumulative Battle Deaths)	8.026	5.398	0.007	0
Duration of Last War	63.05	39.805	0.161	0
Ethnic Fractionalization	0.493	0.541	0.460	0.336
Log Population Size	8.694	9.366	0.007	0
Log Mountainous	2.769	2.225	0.104	0.042
Log Military Personnel	3.207	3.775	0.088	0.468
Log GDP per Capita	6.651	6.813	0.552	0.846
Polity	-2.15	0.011	0.107	0.122
Eastern Europe	0.4	0.333	0.593	
Latin America	0.2	0.092	0.276	
Asia	0	0.149	0	
Sub-Saharan Africa	0.3	0.333	0.777	
North Africa/Middle East	0.1	0.092	0.916	
Right Censored	0.75	0.46	0.015	
After Matching				
Log (Cumulative Battle Deaths)	8.026	7.954	0.678	0.958
Duration of Last War	63.05	53.9	0.331	0.152
Ethnic Fractionalization	0.493	0.515	0.597	0.958
Log Population Size	8.694	8.730	0.846	0.492
Log Mountainous	2.769	2.423	0.357	0.766
Log Military Personnel	3.207	3.562	0.274	0.038
Log GDP per Capita	6.651	6.970	0.191	0.782
Polity	-2.15	-0.2	0.166	0.234
Eastern Europe	0.4	0.45	0.318	
Latin America	0.2	0.15	0.318	
Asia	0	0	1	
Sub-Saharan Africa	0.3	0.3	1	
North Africa/Middle East	0.1	0.1	1	
Right Censored	0.75	0.5	0.088	

Table 4: Cox Proportional Hazard Estiamtes, Post-Conflict Sample, pre- and post-matching

	Unmatched	**Matched**
Covariate	**Coefficient** ***t*-stat.**	**Coefficient** ***t*-stat.**
UN Intervention	0.320 -2.27	0.187 -2.35
Log Cumulative Battle Deaths, last war	1.045 0.87	1.252 1.68
Duration, last war	0.996 -1.69	0.986 -1.81
Ethnic fractionalization	1.015 2.12	1.000 0.02
Log population	1.230 1.14	1.029 0.05
Log mountainous	1.337 2.2	2.068 2.32
Log Military Personnel	0.942 -0.37	0.945 -0.12
GDP per capita	1.068 0.39	0.876 -0.32
Polity 2	0.963 -1.44	0.982 -0.25
Sub-Saharan Africa	2.528 1.88	2.251 0.83
number of obs.	107	40
Log-likelihood	-211.559	-41.514

Table 5: Balance Statistics, In-War Sample

Variable	**Mean treated**	**Mean control**	t-**test** p-**value**	**K-S test** p-**value**
Before Matching				
Non-UN 3rd-Party Intervention	0.438	0.101	0.019	
Log (Battle Deaths)	6.941	6.613	0.49732	0
Coup/Revolution	0.0625	0.049	0.829	
Sons-of-Soil War	0.188	0.413	0.042	
Rebels Contraband Funded	0.375	0.267	0.402	
War Duration	62.875	95.806	0.062	0
Ethnic Fractionalization	0.488	0.591	0.168	0
Log (Population Size)	8.964	10.013	0.001	0
Log (Mountainous)	2.634	2.7347	0.78069	0
Log (Military Personnel)	3.092	4.386	0.008	0
Eastern Europe	0.375	0.057	0.022	
Latin America	0.0625	0.111	0.450	
Asia	0.125	0.315	0.541	
Sub-Saharan Africa	0.375	0.395	0.878	
North Africa/Middle East	0.625	0.122	0.359	
Right Censored	0.063	0.260	0.007	
After Matching				
Non-UN 3rd-Party Intervention	0.438	0.25	0.255	
Log (Battle Deaths)	6.941	7.108	0.7817	0.628
Coup/Revolution	0.0625	0.0625	1	
Sons-of-Soil War	0.188	0.188	1	
Rebels Contraband Funded	0.375	0.375	1	
War Duration	62.875	55.25	0.137	0.994
Ethnic Fractionalization	0.488	0.420	0.335	0.604
Log (Population Size)	8.964	9.156	0.275	0.18
Log (Mountainous)	2.634	2.4516	0.5144	0.876
Log (Military Personnel)	3.092	3.759	0.160	0.174
Eastern Europe	0.375	0.375	1	
Latin America	0.0625	0.125	0.568	
Asia	0.125	0.188	0.718	
Sub-Saharan Africa	0.375	0.313	0.659	
North Africa/Middle East	0.625	0	0.318	
Right Censored	0.063	0.125	0.318	

Table 6: Cox Proportional Hazard Estimates, In-War Sample, pre- and post-matching

	Unmatched	**Matched**
Covariate	Coefficient t-stat.	Coefficient t-stat.
UN Intervention	2.620 1.85	0.747 -0.51
intervention by a non-UN third party	0..678 -0.81	0.502 -1.04
Log battle deaths, current year	0.808** -2.05	0.662** -2.05
Coup/Revolution	3.510* 2.12	2.260 0.54
"Sons-of-the-Soil" Conflict	0.900 -0.26	0.225 -1.55
Rebels Contraband funded	0.284** -2.77	0.424 -0.63
Ethnic fractionalization	1.006 0.86	1.829 1.04
Log population before war	1.104 0.45	0.810 -0.95
Log mountainous	0.634** -3.23	1.003 0.16
Log military personnel	0.878 -0.72	0.744 -1.35
Eastern Europe	2.358 1.48	2.249 0.54
Asia	0.799 -0.45	0.113 -1.51
number of obs.	3999	32
Log-likelihood	-131.17	-35.63

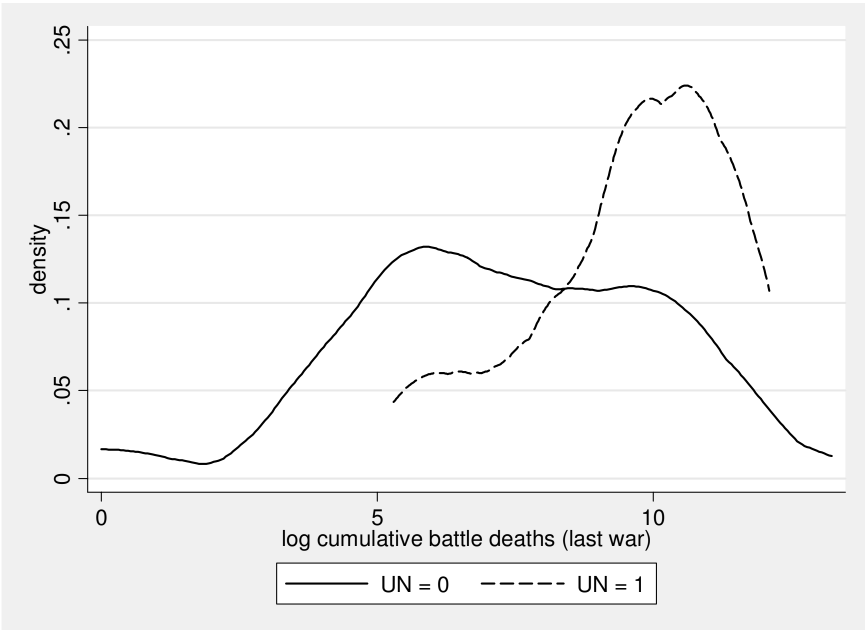

Figure 1: Kernel densities of cumulative battle deaths in cases where the UN intervened and where it did not

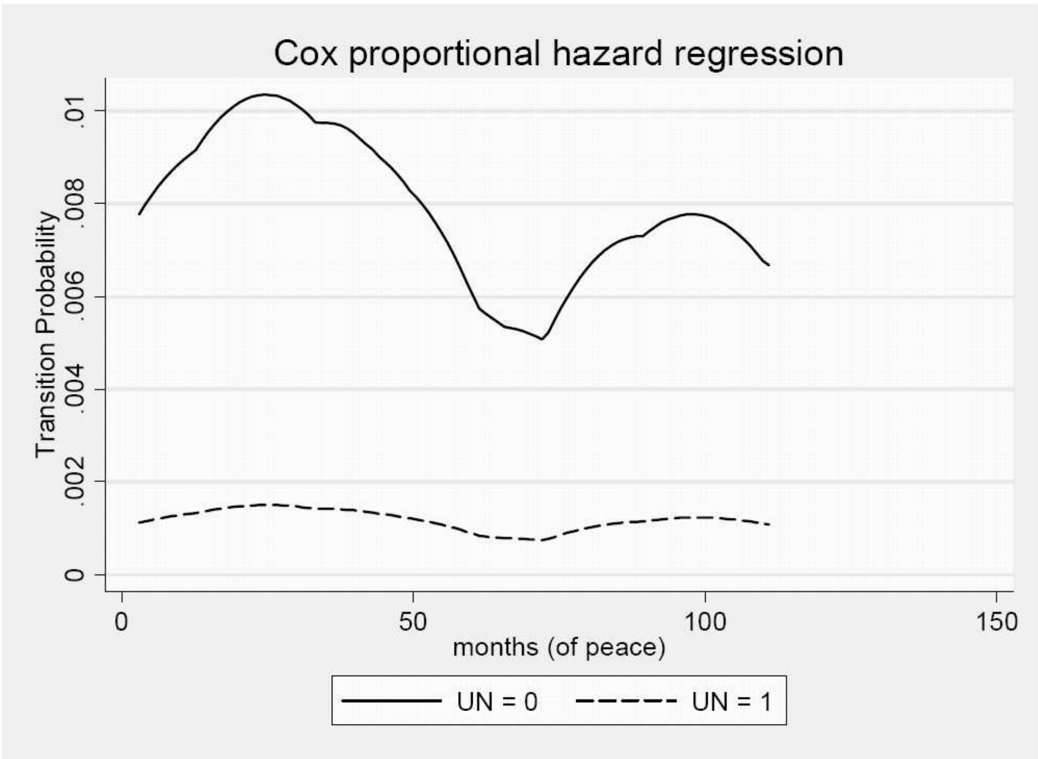

Figure 2: Probability of transitioning to peace, estimated using log-normal regression and Cox proportional hazard regression