

Means to an end: Pro-government militias as a predictive indicator of strategic mass killing¹

Conflict Management and Peace Science

1–24

© The Author(s) 2015

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0738894215600385

cmp.sagepub.com



Ore Koren

University of Minnesota, USA

Abstract

Forecasting models of state-led mass killing are limited in their use of structural indicators, despite a large body of research that emphasizes the importance of agency and security repertoires in conditioning political violence. I seek to overcome these limitations by developing a theoretical and statistical framework that highlights the advantages of using pro-government militias (PGMs) as a predictive indicator in forecasting models of state-led mass killing. I argue that PGMs can lower the potential costs associated with mass killing for a regime faced with an internal threat, and might hence “tip the balance” in its favor. In estimating a series of statistical models and their receiver–operator characteristic curves to evaluate this hypothesis globally for the years 1981–2007, focusing on 270 internal threat episodes, I find robust support for my expectations: including PGM indicators in state-led mass killing models significantly improves their predictive strength. Moreover, these results hold even when coefficient estimates produced by in-sample data are used to predict state-led mass killing in cross-validation and out-of-sample data for the years 2008–2013. This study hence provides an introductory demonstration of the potential advantages of including security repertoires, in addition to structural factors, in forecasting models.

Keywords

Forecasting models, mass killing, political violence, pro-government militias

In this article I argue that pro-government militias (PGMs)² can be used as a predictive indicator to improve the strength of state-led mass killing forecasting models.³ This is because PGMs, as part of a government’s security repertoire, can lower the potential political and economic costs of mass killing for the government. An estimated 13–26 million civilians have died in civil, international and extra-systemic wars between 1945 and 2004, many by directed violence (Valentino et al., 2004: 375). In many of these cases, irregular pro-government groups are used to carry out some of this violence. Although recent research on civilian abuses takes more agent/organizational factors into account, most state-led mass killing

Corresponding author:

Ore Koren, Department of Political Science, University of Minnesota, 267 19TH Ave S, MN 55455, USA.

Email: Koren044@umn.edu

studies focus on the structural (i.e. political, socio-economic, geographic) factors that might give rise to this violence. In this article I complement these approaches by highlighting the effect of security repertoires, in addition to structural factors, on increasing the likelihood of state-led mass killing, as the government seeks to minimize its potential costs. Specifically, I illustrate the improvement in state-led mass killing forecasting models that results from taking potential perpetrators—in this case PGMs—into account.

This argument assumes that some catalyst typically creates an incentive to perpetrate mass killing. I hypothesize that this catalyst is the onset of some form of internal threat (hence I focus only on state-led mass killing perpetrated during episodes of such a threat).⁴ When the opportunity exists, PGMs—as part of the regime’s security repertoire—can “tip the balance” for perpetrating strategic mass killing by lowering anticipated government costs in at least three respects: by providing plausible deniability; by freeing the state’s official military forces to handle other external and internal threats; and by offering lower costs of formation or cooptation (in comparison to the state’s official military force). As my empirical analysis shows, PGMs are not only statistically and substantively associated with an increased likelihood of state-led mass killing, but their inclusion in models of state-led mass killing provides a significant improvement in the predictive strength of these models, even after state capability, conflict type and regional indicators are taken into account.

Prediction has proven useful in the past in improving our understanding of and testing theories about certain forms of conflict (Brandt et al., 2011; Goldstone et al., 2010; Ward et al., 2013); similarly, I argue that militias are useful as a predictive indicator. The composers of these predictive models use various methods and indicators to estimate different levels of political instability and conflict. My aim in this article falls within this framework: I do not claim to provide a model that perfectly predicts the occurrence of state-led mass killing or to replace existing theories of violence against civilians, only to supplement these theories by highlighting the improvement in prediction provided by taking PGMs into account.

Why should PGMs specifically—rather than the entire state security apparatus—become the focus of study? After all, official military forces have proven themselves many times to be willing participants in the mass killing of civilians (Koren, 2014), perhaps most recently in Syria. Recent studies have argued that a relationship exists between PGMs and various types of human rights violations (Mitchell et al., 2014; Nordås and Cohen, 2012). These studies emphasize the importance of the principal–agent framework in allowing the government to shift responsibility for the violence by using troops that are naturally more likely to be violent (Koren, 2014; Mitchell et al., 2014). These conclusions are undoubtedly important, but significant additional insights are gained by focusing on mass killing specifically. The bureaucratic complications involved with identifying members of the targeted civilian group and efficiently deploying forces to kill them suggest that mass killing is unlikely without some degree of centralized government oversight (Goldhagen, 2009; Valentino, 2004). Hence, PGMs are less likely to perpetrate mass killing independently—in contrast to other human rights violations—or to do so against the government’s wishes.

The remainder of this article is divided into four parts. In the first section, I discuss the theoretical framework for my analysis. Beginning with a case study of Rwanda, I build on extant theories related to both political violence and militias to develop my argument that PGMs are an effective indicator of state-led mass killing. In the second section, I discuss my dataset and research design. The third section reports my empirical analysis. I conclude with a brief discussion of some implications of this study.

Theoretical motivations

The theoretical and empirical literature on militias examines the relationship between these groups and conflict (e.g. Ahram, 2011; Carey et al., 2013), genocide (e.g. Ahram, 2014; Alvarez, 2006) or sexual violence (e.g. Nordås and Cohen, 2012). Although some militia-centric analyses examine the connection between militias and violence using comprehensive datasets (Mitchell et al., 2014; Raleigh, 2012), the shortage in large- N studies of PGMs—as well as other military attributes—is still apparent. Moreover, most PGM-related studies usually attempt to ascribe various causal effects to PGMs, focusing on the principal-agent perspective in relation to these groups (Ahram, 2014; Mitchell et al., 2014; Nordås and Cohen, 2012; Raleigh, 2012). Although this is a useful approach, establishing causality in the absence of (quasi-) experimental data is difficult owing to factors such as simultaneity bias, unobserved confounders and the different ways in which political violence is operationalized. However, I argue that, even if the correlative associations between PGMs and state-led mass killing do not necessarily represent a causal relationship, they can nevertheless be used to improve mass killing forecasts. The main contribution of this paper is in showing the validity of using PGMs as predictive indicators in state-led mass killing forecasting models.

To demonstrate the relationship between PGMs and mass killing, and for the purposes of theory generation, it is useful to first briefly discuss a specific case in which a PGM has been directly involved in a mass killing campaign. The role of the *Interahamwe* militia during the Rwandan Genocide illustrates the advantages a PGM might provide to a regime that intends to perpetrate mass killing. In the years following the onset of the conflict with the Rwandan Patriotic Front (RPF), the Rwandan government invested in forming and training militias and civil defense troops (Fujii, 2009: 87; Straus, 2015: 273). Following the assassination of President Juvénal Habyarimana in 1994, the *Interahamwe* militiamen were key actors in promoting and perpetrating atrocities against the civilian population.

There were at least two advantages that made the *Interahamwe* PGM a useful tool for perpetrating mass killing. First, the use of the *Interahamwe* reduced the necessity to rely on the official Rwandan Armed Forces (RAF) for the purpose of perpetrating large-scale killing of civilians. The genocide occurred as part of a military campaign against a well-trained military force, the RPF (Straus, 2006, 2015), which meant that the RAF was primarily occupied with fending off this threat, and hence was unable to allocate a significant number of troops for killing civilians (Mamdani, 2001: 185–233), although members of the military and the police did participate in the killings (Ahram, 2014). Under these constraints, the *Interahamwe* were used to enforce the killing, serving as mediators between the local leaders who orchestrated the violence and a large group of local civilian participants who perpetrated most of the small-scale violence (Mueller, 2000; Straus, 2006: 93). In numerous regions around the country, members of the *Interahamwe* were used as “violence specialists”, dispatched by the regime to spark or boost the genocide (Fletcher, 2007; Fujii, 2009: 87, 129–130). *Interahamwe* troops also perpetrated the largest massacres, where more serious weaponry or better training was required (Fujii, 2009: 87; Straus, 2015: 311).

Second, by allowing killers to easily identify victims, *Interahamwe* networks facilitated the perpetration of violence (Fletcher, 2007; Fujii, 2009), while looting and appropriating the property of victims reduced the necessity to provide logistic support to *Interahamwe* members (Mueller, 2000). *Interahamwe* troops were able to rely on the local population for information, with the locals pointing out—as well as independently killing—victims. They would also loot the property of local civilians and burn down their houses, with or without killing

them (Fujii, 2009: 160–164). Moreover, members of the *Interahamwe* were more likely to share the extremist government's aims, which meant that the costs of cooptation and recruitment were low, as many troops joined willingly, either for ideological (Straus, 2015: 304–313) or pragmatic (Mueller, 2000) reasons.

Interestingly, Rwanda could be considered a relatively “strong” state, compared with other countries, such as Zaire/the DRC, meaning that its government possessed a relatively high degree of control over the nation's territory and its population (Straus, 2006). This meant that the central government had a relatively high capability to orchestrate the violence—especially as the state issued IDs that included a person's ethnic designation—although not to directly control and manage it (Ahram, 2014). This divergence produced variation across different regions, where in some areas orders were passed through direct government channels while in others violence was carried out through actions of the *Interahamwe* (Ahram, 2014; Fletcher, 2007). Evidently, at least in the Rwandan case, this combination of an intermediate level of bureaucratic capability and PGM availability produced a highly efficient mass killing campaign. At the onset of the conflict with the RPF, the Rwandan government had a specific security repertoire, produced by a combination of tradition and necessity (Ahram, 2014; Straus, 2015), of which the *Interahamwe* were a crucial part. Especially in the earlier phases of violence, when massacres were more the result of spontaneous outrage than an organized policy of extermination (Straus, 2015: 306–309), the availability of the *Interahamwe* militia for this purpose might have produced some incentive to “[t]urn to the population as a last resort to defeat the RPF and impose massive collective punishment on Tutsis” (Straus, 2015: 308).

The example of the instrumental use of the *Interahamwe* corresponds to a large body of research about mass killing and political violence,⁵ heretofore known as “rationalist”. Proponents of this approach focus on the logical use of violence against civilians during national crises. From this perspective, state-led mass killing is the result of a search for a “scapegoat”, the aspirations of a specific interest group to take advantage of the crisis and consolidate power (Gagnon, 2004; Harff and Gurr, 1998: 553–556; Krain, 1997), or because a specific group is seen as a threat or an enemy supporter and must therefore be removed (Esteban et al., 2010; Kalyvas, 2006; Valentino, 2004; Valentino et al., 2004; Wood, 2010). Hence, leaders use large-scale violence against civilians “when they perceive it to be both necessary and effective” (Valentino, 2004: 67).

The rationalist approach is helpful in explaining why some regimes should be more likely to use PGMs for mass killing, that is, in explaining the *willingness* of leaders to use PGMs. Proponents of the rationalist body of scholarship tend to focus on violence against noncombatants perpetrated specifically during civil wars, and in doing so neglect events with lower numbers of combatant casualties, although some of the latter might involve a high number of intentional noncombatant casualties (Koren, 2014). Therefore, I have included cases from the entire spectrum of internal conflict, ranging from nonviolent civil disobedience campaigns to violent civil wars, all of which I term “internal threat.” Understandably, owing to other related factors, mass killing might be more likely during full-blown civil wars, while some events—e.g. coups d'état—might be less likely to involve widespread violence. To account for this, I also include indicators denoting event type.

Proponents of the rationalist approach also tend to minimize the role that government security repertoires and agent-specific characteristics play in influencing the likelihood of violence, treating perpetrators as obedient henchmen (Valentino, 2004). Although the focus on the structural factors that influence the willingness of a government to perpetrate mass

killing has proven useful, the Rwandan case, at least, suggests that, by facilitating its occurrence, PGMs can influence the likelihood of state-led mass killing by making it less costly to the regime. Namely, PGMs are a type of agent that—by reducing the costs incurred by the regime for perpetrating a specific military action—might make it more likely. This argument is similar to the claim that having a Predator drone in a regime’s national security repertoire might make targeted assassinations more likely, because it significantly reduces the costs incurred from this action in comparison with, say, a F-18 jet (Cronin, 2013). Several studies have previously argued that a relationship between a given regime’s security repertoire and the likelihood of violence against civilians exists (Koren, 2014; Mitchell et al., 2014). This suggests that taking the type of perpetrator, for example PGMs, into account can improve our ability to predict, if not explain, violence against civilians.

A second relevant body of research emphasizes the role of ethnic, social, religious or economic cleavages within society and communal pressures on generating inter-group violence (Chirot and McCauly, 2010; Horowitz, 1985: 95–140; Staub, 1989: 13–34; Varshney, 2002), or the importance of democracy and democratic institutions in preventing mass killing and other forms of violence (Lutz and Sikkink, 2000; Poe and Tate, 1994; Poe et al., 1999; Rummel, 1995). This approach highlights conditions under which violence against civilians, as well as the use of PGMs, might be more likely, that is, where the *opportunity* to perpetrate violence is more likely to exist.

The research of several militia scholars supports this argument by pointing out that, under certain conditions—for example, a leadership’s fear of a military coup d’état—the state is more likely to rely on such organizations for internal security purposes (Ahram, 2011; Carey et al., 2013; Dowdle, 2007). Correspondingly, PGMs are more likely to exist in contexts where the state is weakened or challenged, such as during internal conflict (Mueller, 2000); in cases where national military tradition does not exist (e.g. Indonesia; see Ahram, 2011: 25–55); or when the country is geographically fragmented (e.g. Indonesia, The Philippines, Papua New Guinea). This suggests that, in these contexts, the government might have both a higher incentive for co-opting PGMs and a higher number of militias on which to draw.⁶ However, although this militia-centric perspective is helpful in illustrating why PGMs might operate in countries where the government exercises little control over its military or where the military poses a danger to the regime, such as Zaire/the DRC, it does not fully account for cases where the state possesses a higher degree of control over its security forces, for example, in Yugoslavia (Mueller, 2000). I argue that in cases that resemble the latter, militias are used because they reduce the costs resulting from a large-scale targeting of civilians, and can hence be a useful indicator of this form of violence.

The process that leads to a regime considering mass killing goes as follows: governments react rationally to the rise of some form of an internal threat—for example, a coup d’état, a civil disobedience campaign, a “sons-of-soil” conflict, or a civil war—which I treat as exogenous for the purpose of analysis.⁷ The government might first try to contain this threat by other means, because rationalist state-led mass killing is generally a strategy of last resort (Valentino, 2004). However, if it fails to contain the threat by conventional means, and especially if the rebel/opposition group is gaining strength and legitimacy (both domestically and internationally), the government might decide to resort to more drastic means—mass killing—and remove this threat by targeting those viewed as its supporters or corroborators, either by violently expelling or completely annihilating them.

Building on the aforementioned scholarship, I argue that PGMs can be a useful indicator in state-led mass killing forecasting models because they can reduce the costs of perpetration

in such situations in at least three respects. First, current literature about militias and violence suggests that by allowing the government to shift responsibility for the crimes to groups that are known (or quickly gain a reputation) for being violent, PGMs provide the government with plausible deniability in relation to these crimes. For example, Mitchell et al. offer one explanation as to why PGMs, specifically, are a useful tool for perpetrating violence: “[t]he principal may knowingly recruit those with a reputation for violence (for example, criminals) and then refuse to control these agents—rather than actually lose control over them” (Mitchell et al., 2014: 818). This allows the government to avoid, or at least postpone, naming-and-shaming and international sanctions and, at the same time, prevents the formation of a large domestic opposition against its actions. A government bent on strategic mass killing might hence hide behind a veil of ignorance in relation to the actions perpetrated by its agents, especially if those acts are perpetrated by informal groups whose connection to the regime is easy to conceal (Mitchell et al., 2014: 818–820).

A second cost-related advantage of PGMs is that their use for mass killing allows official military forces to concentrate their efforts on containing more serious military threats. As illustrated by the *Interahamwe* example, this might be especially true during full-blown civil wars or other military campaigns where a rebel group or an external military force poses an immediate existential threat to the regime (Valentino et al., 2004). If the government must choose between allocating the official military to fight off an invading enemy force or employing some of its troops to remove a specific noncombatant population, it is more likely to choose the former (assuming the military threat is severe enough), simply because an invading military force requires more specialized and better trained forces. However, if a PGM can be easily formed or co-opted, the government can have its cake and eat it, too; it can allocate its better trained official military force to fend off the military threat, while sending its less trained militias to remove the enemy’s potential base of support (as happened during the Rwandan Genocide).

The third cost-related advantage provided by PGMs is that their price of formation or co-optation is lower than those of other groups, for at least four reasons. First, many PGMs are composed of criminals, hooligans, and other negative elements unleashed on the population without any form of regulation. Hence, as the *Interahamwe* example illustrates, the troops of many PGMs do not require payment, as they are perfectly capable of sustaining themselves through looting, pillaging, and preying on the local population (Mueller, 2000). Second, in many cases the government invests little or nothing in training these groups (Ahram, 2014; Koren, 2014), and even in cases where training is provided, it is likely to be shorter and involve less sophisticated equipment than that provided to the official military forces. Third, some (e.g. tribal) PGMs are more likely to share the nationalist views of the regime and support its aims, which—combined with their ability to independently sustain themselves—might make these groups more likely to willingly participate in state-led mass killing. The official military, in contrast, might have higher costs of cooptation, either because it finds these duties demeaning—especially if it has a strong military tradition—or because diverting necessary resources from fighting to committing mass killing is more costly (Koren, 2014). Lastly, PGMs might be more likely to possess local knowledge, which lowers learning costs when operating in distant regions (Carey et al., 2013).

I argue that these cost-related advantages can play an important role in influencing the decision to perpetrate strategic mass killing, and that, although PGMs should not be considered the cause of state-led mass killing (in contrast to, perhaps, other forms of abuses), the

availability of PGMs is more likely to “tip the balance” in favor of mass killing in response to an internal threat.

A potential critique of this argument is that it assumes total government control over its PGMs, and implies that the latter are obedient henchmen rather than agents with their own agendas. Although my argument, like other social scientific arguments, is a simplification for the purpose of generalization—in numerous cases militias have demonstrated they can pursue their own interests, even against government aims—this simplification is valid for at least two reasons.

First, none of the advantages mentioned above negates the principal–agent perspective on militias, and in fact the first and third points are quite in tune with it (Ahram, 2014; Mitchell et al., 2014). Second, because the use of militias for mass killing is more likely when the government orchestrates the violence, as the Rwandan example illustrates, one would expect mass killing to be most likely in countries that possess some—albeit limited—level of bureaucratic capacity, not in countries that experience a complete breakdown of state institutions.⁸ As I show in the empirical section, this argument is indeed supported by the data. This perspective also helps explain the use of PGMs in different contexts, including where the state maintains a relatively effective hold on the monopoly of violence. Especially where large-scale mass killing is concerned, I argue that violence is not the result of agency loss, but rather the result of a well-calculated strategic use of PGMs by the regime. This suggests the highest likelihood of strategic mass killing to be in regimes that can exercise some level of control (with or without plausible deniability) over these organizations, which accordingly suggests the following hypothesis:

H1: The likelihood of state-led mass killing should be positively related to the existence of pro-government militias.

Although this logic suggests that regimes with PGMs will be more likely to engage in mass killing than regimes without PGMs, the existence of militias alone might not necessarily be a good determinant of mass killing. Referring back to the first cost-related advantage discussed above, one might expect groups whose connection to the regime is easier to establish, for example, semi-official militias, to provide a lower degree of plausible deniability. This does not suggest that these groups will not be used for strategic mass killing. As I argued, PGMs provide cost-related advantages other than plausible deniability, which means that perpetrating regimes would still favor using PGMs, even if regime accountability is easier to establish. However, plausible deniability and the ability to shirk responsibility for the crimes is still a significant advantage, so groups that provide it will be favored as a means for perpetrating mass killing. The availability of informal PGMs suggests, in turn, that the tipping point for perpetrating strategic mass killing can be more easily reached, which means that these groups specifically might be more likely to be involved in such extreme violence. These groups (e.g. tribal militias) might also be more likely to self-support and have local knowledge of their areas of operation, which again decreases the costs associated with strategic mass killing. This suggests the following hypothesis:

H2: The likelihood of state-led mass killing should be positively related to the existence of informal or clandestine PGMs, specifically.

Lastly, and most importantly, recall that this paper is mainly concerned with the use of PGMs as an indicator in forecasting models of state-led mass killing. If, as I claim, PGMs make reaching the mass killing tipping point more likely, then more than just statistical and substantive significance, one would expect PGMs to produce a significant *predictive* improvement in forecasting models, when other factors are taken into account. This perspective does not mean that structural factors emphasized by the bodies of research concerned with political violence do not matter. It simply means that PGMs matter as well. This suggests the following hypothesis:

H3: The predictive strength of state-led mass killing forecasting models should significantly increase with the addition of PGM indicators.

Variable operationalization and research design

In this article I do not attempt to account for all civilian casualties caused by different types of political violence, both because some civilian casualties can be caused by independent PGM actions and because many civilians are unintentionally killed in most conflicts. Victims of mass killing may be members of any group, regardless of their ethnicity, political affiliation, religion, gender or other categorical characteristic. At the same time, because I examine a wide range of internal threat events with different degrees of intensity and duration, I avoid a high-casualty threshold for mass killing (which generally ranges from 25,000 to 50,000 intentional deaths). Hence, I choose a threshold of 1000 intentional civilian deaths for a given mass killing campaign, which I believe is high enough to capture motivated killing of civilians, yet low enough to capture many cases of intentional killing that are neglected by other studies. Lastly, because the focus here is on state-led mass killing, I am only interested in mass killing perpetrated by groups that support the state, and not by organizations defined as a rebel group for a given campaign-year.

The data for coding this dependent variable were obtained from the report “Assessing risks of state-sponsored mass killing” written by Ulfelder and Valentino. This dataset covers all mass killing campaigns (as defined above) perpetrated by governments against their own civilians between 1945 and 2008. Ulfelder and Valentino define mass killing as “any event in which the actions of state agents result in the intentional death of at least 1000 noncombatants from a discrete group in a period of sustained violence ... [i]f fewer than 100 total fatalities are recorded annually for any three consecutive years during the event, the event was considered to have ended during the first year within that three-year period in which fatalities dropped below 100 per year (even if killing continues at levels in later years)” (Ulfelder and Valentino, 2008).⁹ Note that, because they are primarily interested in actions perpetrated by governments against their own civilians, Ulfelder and Valentino do not code any mass killing perpetrated as a part of an interstate conflict. It is therefore important to emphasize that in this study I examine only internal threat episodes, and that my findings regarding the effectiveness of PGMs as a predictive indicator of mass killing might not be applicable to interstate conflicts, although they might be suggestive of such a relationship. To code the dependent variable—whether mass killing occurred during a conflict-year or not—a value of 1 was assigned to all conflict years coded in the report as a year during which mass killing took place, and a value of 0 otherwise.

To test my hypotheses, I constructed a dataset of 177 different internal threat campaigns, divided into 270 episodes and 848 campaign-years.¹⁰ A campaign-year is defined as any year during which an internal threat episode was ongoing. An episode is defined as a part of the campaign, either the only part (if the campaign was never renewed) or any period during which the campaign was renewed after having subsided. So, for example, the internal threat campaign in Rwanda is coded as having begun in 1990 and ended in 2002. However, as no combatant casualties were recorded for the years 1995 and 1996, this campaign has two episodes: one lasting from 1990 to 1994 (the RPF invasion), and another lasting from 1997 to 2002 (in which the new government under RPF leadership fought the Democratic Forces of the Liberation of Rwanda, or FDLR). The campaign-years analyzed in the dataset are 1990–1994 and 1997–2002.

The dataset records episodes that began during or after 1981 or ended/were ongoing before or during 2007. I chose 1981 and 2007 as my temporal markers, because they were the same ones used in the Pro-Government Militias Database (PGMD), on which I relied for coding PGMs (Carey et al., 2013). Because I focused on internal threat episodes for modeling temporal dependency, episodes belonging to campaigns that began before 1981 were included in the dataset; renewal episodes began in or after 1981. All state-led mass killing cases mentioned in the “Assessing risks of state-sponsored mass killing” report occurred during one of the internal threat campaigns coded in my data.

Internal threat campaigns were obtained from two different datasets. The first dataset included conflict episodes in Kreutz’s (2010a) “UCDP Conflict Termination Dataset”. Kreutz relies on UCDP/PRIO armed conflict data to code conflict-years, including the years during which each episode began and ended (Kreutz, 2010b: 2).¹¹ To separate each episode, Kreutz codes seven different types of episode/conflict termination: peace agreements, cease-fire agreement with conflict regulation, ceasefire agreement, victory, low activity, other, and joining alliance (Kreutz, 2010b: 2–3). Based on these criteria, one conflict can have numerous renewal episodes over time. Again, because Ulfelder and Valentino report only intra-state mass killing episodes, I coded only conflicts whose Type 2 score was three, defined as “[i]nternal armed conflicts between the government of a state and one or more internal opposition group” (Kreutz, 2010b: 6).¹²

In addition to the UCDP Conflict Termination Dataset, I relied on the dataset composed by Stephan and Chenoweth (2008) for coding “nonviolent” internal campaigns between 1981 and 2007.¹³ These researchers define a campaign as “a series of observable, continuous, purposive mass tactics or events in pursuit of a political objective”, which can last “anywhere from days to years, distinguishing it from one-off events or revolts” (Chenoweth, 2008: 1). Nonviolent resistance is defined as “a civilian based method used to wage conflict through social, psychological, economic, and political means without the threat or use of violence” (Stephan and Chenoweth, 2008: 9). For both databases, an observation was defined as a campaign-year. For each internal threat episode, I included the year the episode began, ended, and any year in-between.¹⁴ Using both datasets allowed me to take into account a wider variety of internal threat campaigns—in terms of violence levels and duration—than is examined in extant literature.

Explanatory variables were obtained from various data sources.¹⁵ I relied on the PGM Dataset created by Carey et al. (2013) to code PGM indicators. The authors define a PGM as a group identified by primary sources as pro-government or sponsored by a national or sub-national government; that is not considered part of the security force; that is armed; and that has some level of organization (Mitchell and Carey, 2013: 5). The PGM country-year

dataset codes both the type of PGM (semi-official or informal) and their activity or presence for a given year, with “presence” referring to situations where no information about the PGM’s disbandment was found.

Because the focus of this paper is state-led mass killing, it is important to stress that PGMs have been mentioned in primary and secondary sources in connection with perpetrating violence against different groups of both combatants and noncombatants.¹⁶ However, this evidence does not necessarily suggest that PGMs are systemically more likely to be involved in mass killing, which—as mentioned above—might require a significantly higher degree of state capability. Although the case study of Rwanda presented in the previous section shows that the *Interahamwe* PGM was directly involved in much of the killing, providing extensive cross-national anecdotal evidence is not feasible within the scope of this paper. However, the question of how much of the killing is actually perpetrated by PGMs, especially when considering that Ulfelder’s and Valentino’s death estimates are conservative compared with other official sources, is of secondary importance for the purpose of forecasting mass killing. In other words, PGM availability increases the likelihood of state-led mass killing, not necessarily its scope.¹⁷

To test hypothesis H1—regarding the role of PGMs—and hypothesis H2—regarding the role of informal militias—I coded four binary variables. The first two indicators code the existence of semi-official PGMs. A PGM was coded as “semi-official” if it was mentioned in primary sources as having “a formally and/or legally acknowledged status” (Mitchell and Carey, 2013: 11). A value of 1 was assigned to campaign-years during which a semi-official PGM was coded as present (*Semi-official militias present*¹⁸) or active (*Semi-official militias active*¹⁹), and a value of 0 otherwise. I also coded two additional binary variables to account for the existence of informal PGMs. A pro-government militia was specifically coded as “informal” if “the link ... is not officially or formally acknowledged” (Mitchell and Carey, 2013) by the government. A value of 1 was assigned to campaign-years during which an informal PGM was coded as present (*Informal PGM present*²⁰) or active (*Informal PGM active*²¹), and a value of 0 otherwise.²²

Additional indicators were included to account for other explanations the literature provides on state-led violence and mass killing discussed above. As both regime-centric approaches emphasize the importance of democracy in decreasing the likelihood of violence against civilians and mass killing specifically, I included a democracy indicator. I relied on the “democracy” score included in the “Democracy–Dictatorship” dataset composed by Cheibub et al. (2010) to code a binary variable (*Democracy*) denoting whether the regime was a democracy or not. This dataset was preferred to the more widely used Polity IV dataset (Marshall et al., 2014) because it has been suggested that the latter might be susceptible to some biases (e.g. Trier and Jackman, 2008).²³

Likewise, proponents of the rationalist body of scholarship have suggested a relationship between mass killing and economic prosperity (e.g. Esteban et al., 2010). To account for the potential effect of income per capita, I coded a continuous variable (*Log GDPpc*) denoting the log of the average GDP per capita (in current US Dollars) for a given conflict-year. Because my dependent variable is based on quantitative standards, the possibility exists that countries with larger population might naturally be more likely to experience state-led mass killing as defined above. Hence, I coded a continuous variable (*Log Pop*) denoting the log of the size of the country’s population for a given conflict-year to account for this possibility. Data for both population and GDP per capita were obtained from the “GDP and population data updated to 2011” composed by Gleditsch (2013).²⁴

Because the rationalist body of scholarship strongly suggests that strategic mass killing is more likely during civil wars, I included a binary variable (*War.ep*) denoting whether a given episode experienced more than 1000 combatant casualties (based on minimal estimates; Lacina and Gleditsch, 2009). Alternatively, as a result of being relatively short in duration (lasting from days to at most weeks), coups d'états might be less likely to generate strategic mass killing as an instantaneous response. I relied on data from Powell and Thyne (2011) to code a binary variable (*Coup d'état*) denoting whether a coup d'état occurred in a given year.²⁵ I also created a binary variable (*Nonviolent.campaign*) measuring whether a campaign was primarily nonviolent (assigned a value of 1) or not, according to the Stephan and Chenoweth dataset. To account for the possibility that episodes in campaigns that previously experienced mass killing might experience it again, I coded a binary variable (*MKConf*). Episodes in campaigns that previously experienced mass killing were given a score of 1; 0 otherwise. To account for potential regional variation in violence, I coded binary indicators for Africa (*Africa*), Asia and Oceania (*Asia.Ocean*), Europe (*Europe*) and the Middle East (*ME*).²⁶

Lastly, I argued earlier that if PGMs are more likely to be used rationally for strategic mass killing, in comparison with other forms of less severe violence, then one might expect the highest likelihood of mass killing using PGMs to be observed in states that do not experience a complete breakdown of state institutions. This allows their governments to exercise some degree of control over PGMs and suggests a curvilinear relationship, with the most dangerous cases (i.e. cases that are most likely to perpetrate mass killing using PGMs) involving intermediate state capability. To proximate bureaucratic capacity, I used the percentile rank for a given country's mean rule-of-law estimates for a given year (*Rule-of-law*) obtained from the World Bank's World Development Indicators (2015). A squared polynomial term (*Rule-of-law2*) is also included to account for the expected negative curvilinear relationship, as regimes with a high rule-of-law score are more likely to respect human rights and, therefore, less likely to perpetrate mass killing.²⁷

Empirical analysis and results

To test my hypotheses, I utilize four different logistic regression (logit) models, the results of which are reported in Table 1. I relied on the method of modeling time dependence in binary data using cubic time polynomials presented by Carter and Signorino (2010). The first year in each episode (either new or renewed) was assigned a value of 0, with every consecutive year assigned a correspondingly increasing value.²⁸ In addition, I provide the first difference change in probability for statistically significant variables ($p < 0.05$, two-tailed test) to highlight the substantive impact of variables whose effect is unlikely to be random in Online Appendix I, owing to space limitations. Taken together, statistical significance and substantive effects tell us how likely PGM indicators are to impact state-led mass killing during internal threat episodes in in-sample data, but they tell us little about their strength as a *predictive* indicator. To assess the improvement in the predictive strength of state-led mass killing forecasting models when PGM indicators are included, I evaluate the change in the area under the curve (AUC) for receiver-operator characteristic (ROC) curves of these four models (Ward et al., 2010).

The improvement in AUCs provided by PGM indicators is evaluated in three phases. In the first phase, the AUC of the ROC curve of each model is compared with those of

Table 1. Potential indicators of state-led mass killing during internal threat episodes, 1981–2007

Explanatory variable	Model I	Model II	Model III	Model IV
Democracy	—	0.1355 (0.4668)	-1.1180 * (0.2427)	-1.1530 * (0.2442)
Rule-of-law	—	0.0552 (0.0337)	—	—
Rule-of-law2	—	-0.0020 * (0.0006)	—	—
Log GDPpc	—	-0.1651 (0.4366)	-0.4935 (0.3010)	-0.5032 (0.3016)
Log pop	—	-0.7936 * (0.3717)	-0.9170 * (0.1838)	-0.8726 * (0.1884)
Coup d'état	—	—	-1.0820 * (0.4866)	-1.0780 * (0.4870)
Nonviolent.campaign	—	—	-1.2270 * (0.3788)	-1.2320 * (0.3787)
War.ep	—	—	2.0500 * (0.2338)	2.0540 * (0.2341)
MKConf	—	—	0.6106 * (0.2053)	0.6508 * (0.2077)
Africa	—	—	0.6886 (0.4310)	0.7117 (0.4314)
Asia.Ocean	—	—	1.150 * (0.4243)	1.1750 * (0.4251)
Europe	—	—	0.9042 (0.5183)	0.8906 (0.5181)
ME	—	—	-0.8960 (0.5633)	-0.8645 (0.5688)
Semi-Official PGM present	0.1838 (0.1587)	1.2907 * (0.4116)	0.7502 * (0.2598)	0.8179 * (0.3434)
Informal PGM present	0.7352 * (0.1581)	1.5652 * (0.4136)	0.7561 * (0.2366)	0.5598 (0.3021)
Semi-Official PGM active	—	—	—	-0.0851 (0.3070)
Informal PGM active	—	—	—	0.2746 (0.2674)
Time	0.1891 * (0.0805)	0.4821 (0.3602)	0.1521 0.1074	0.1500 (0.1080)
Time ²	-0.0129 (0.0114)	-0.0212 (0.1120)	-0.0134 (0.0143)	-0.0130 (0.0143)
Time ³	0.0002 (0.0004)	-0.0009 (0.0085)	0.00005 (0.0005)	-0.00004 (0.0005)
Constant	-1.0081 * (0.1536)	1.9753 (1.9911)	3.5100 * (1.4580)	3.3470 * (1.4650)
N (campaign-years)	822	255	822	822
AIC	1088.4	254.83	772.41	775.32

Coefficient estimates for each variable are reported, with clustered standard errors in parentheses; * $p < 0.05$.

identical model specifications that do not include PGM indicators for in-sample data. In the second phase, I conduct a leave-one-out cross-validation analysis (e.g. Beck, 2001) to highlight the fact that including PGM indicators in the model reduces forecast error. For the second phase, I repeat the process used in the first phase, only this time each observation's prediction is computed from a re-estimated model using a dataset that does not include that observation. In the third phase, I test the coefficient estimates obtained from in-sample data (presented in Table 1) on independently coded, out-of-sample data for the years 2008–2013 (i.e. observations that were not used in the estimation of the coefficients initially) to highlight the strength of PGMs as an overall predictive indicator. I show that in almost every case, PGM indicators significantly increase the predictive strength of a given model and reduce forecast error, which confirms hypothesis H3. Owing to space limitations, the ROCs of non-PGM inclusive models are provided in Online Appendix I.

Table 1 presents the statistical results of four different models. These results lend support to hypotheses H1 and, to a lesser extent, H2. Model I includes only PGM-centric variables, as well as cubic-time polynomials, to account for duration related effects. As the results show, the existence of informal PGMs alone is statistically significant, which supports hypothesis H2, but not necessarily H1. Model II provides a robustness test to my theoretical argument by accounting for the expected curvilinear relationship between state capability (approximated by the rule-of-law indicator) and strategic mass killing when PGM and some structural variables are included.²⁹ Most importantly, the existence of informal PGMs and semi-official PGMs shows a significant and positive association with state-led mass killing, while *rule-of-law2* shows a significant and negative association, which underscores the significant advantages provided by PGMs to regimes that are able to exploit these groups. The log population variable is statistically significant and negatively associated with mass killing, which suggests that—as countries with smaller populations are more likely to experience mass killing—the quantitative standard used for coding the latter does not bias these results.³⁰

In model III, political, economic, conflict type and regional indicators—which the literature on political violence and mass killing points to as potential predictors of violence—are added to the model. Importantly, once these structural factors are taken into account, the existence of semi-official and informal PGMs is again significant, which supports hypothesis H1, and suggests that PGMs are important in “tipping the balance” toward mass killing. In addition to PGM-centric variables, democracy, nonviolent campaigns and coups d'état are significant and show a negative association with state-led mass killing, as expected. The occurrence of mass killing in previous episodes, a civil war and a regional designation of Africa or Asia (in comparison to the Americas) shows a significant positive association with state-led mass killing. The log population variable is again statistically significant and negatively associated with mass killing.

Model IV is identical to model III, only in this case the reported activity of both semi-official and informal PGMs is included, in addition to their presence indicators.³¹ On the whole, all the significant coefficients from model III maintain their significance and sign, with the exception of informal PGMs. The Akaike Information Criterion (AIC) score of model III is lower, which suggests that model III provides a slightly better fit for the data than model IV, that is, adding PGM activity measures does not improve this model. More important, however, is the fact that the AIC in both models III and IV is significantly lower than the AIC score of a similar model that does not include PGM-related variables, which suggests that including PGMs in these models improves their fit.³² Note that the lower AIC

scores of model II are mainly the result of a smaller N and do not necessarily indicate better model fit in relations to models III and IV. On the whole, AIC scores suggest that the best explanatory models should include both structural (i.e. regime-centric, conflict type, and regional indicators) and PGM indicators. Owing to space limitations, the first difference change in probability for statistically significant variables ($p < 0.05$, two-tailed test) is provided in Table AI-I, Online Appendix I.

As mentioned above, statistical and substantive estimates tell us little about the strength of PGMs as predictive indicators of state-led mass killing. To evaluate whether including PGM indicators improves the *predictive* strength of a given model, I also provide the AUC for the ROC curve of each model with and without PGM indicators. I do so, as discussed above, in three phases, moving from in-sample data, through cross-validation, to out-of-sample data. The ROC of a given model shows the ratio of state-led mass killing cases correctly classified (true positives rate) to that of mass killing cases incorrectly classified (false positives rate). To evaluate whether a change in the AUC of a given model produced by including PGM indicators is significant, I use the DeLong et al. (1988) test.

In the first phase, I evaluate the change in prediction strength of the four models presented in Table 1 using in-sample data, that is, the data used to produce these models' estimates. Figure 1 shows the ROC curves for the four models presented in Table 1 with 95% confidence intervals for the AUC of each model. Table 2 presents the AUC of each model and whether it was significantly (to a $p < 0.05$ level) different from that of an identical model that does not include PGM indicators based on the DeLong et al. test, as well as the difference in model fit based on AIC scores. As Table 2 shows, in each case adding PGM indicators provides a significant improvement in predictive strength, in comparison to a baseline model, which lends strong support to hypothesis H3.

Although these estimates suggest that PGM indicators improve the predictive strength of state-led mass killing models, they do not account for the forecast error that might result by including or omitting these measures. Moreover, although AIC scores are useful in model selection, they are dataset specific. This means that although these scores inform us about a given model's ability to explain the observed data, they do not provide a measure of forecasting accuracy. Hence, in the second phase, I provide an estimate of forecast error of each model, as well as whether the difference in forecast error produced by including PGM indicators is preferred, by way of leave-one-out cross-validation. Specifically, I repeat the process utilized in the first phase using the same coefficient estimates presented in Table 1, only this time each observation's prediction is computed from a re-estimated model using a dataset that does not include that observation. This process is repeated (in a randomized fashion) for each observation in the data.

Figure 2 shows the ROC curves for the four models with 95% confidence intervals for the AUC of each model based on leave-one-out cross-validation data simulations. Table 3 presents the AUC of each model and whether this area was significantly (to a $p < 0.05$ level) different than that of an identical model that does not include PGM indicators based on the DeLong et al. test. However, instead of AIC scores, I provide each model's cross-validation forecast error estimates, which report the ratio of times a given model failed to correctly predict the outcome of the n th observation using a dataset composed of $n - 1$ observations to all prediction attempts. A better model is therefore a model with a smaller prediction error. As Table 3 illustrates, PGM-inclusive models produce a significantly larger AUC, even when cross-validated data are used. Moreover, the leave-one-out cross-validation estimates show

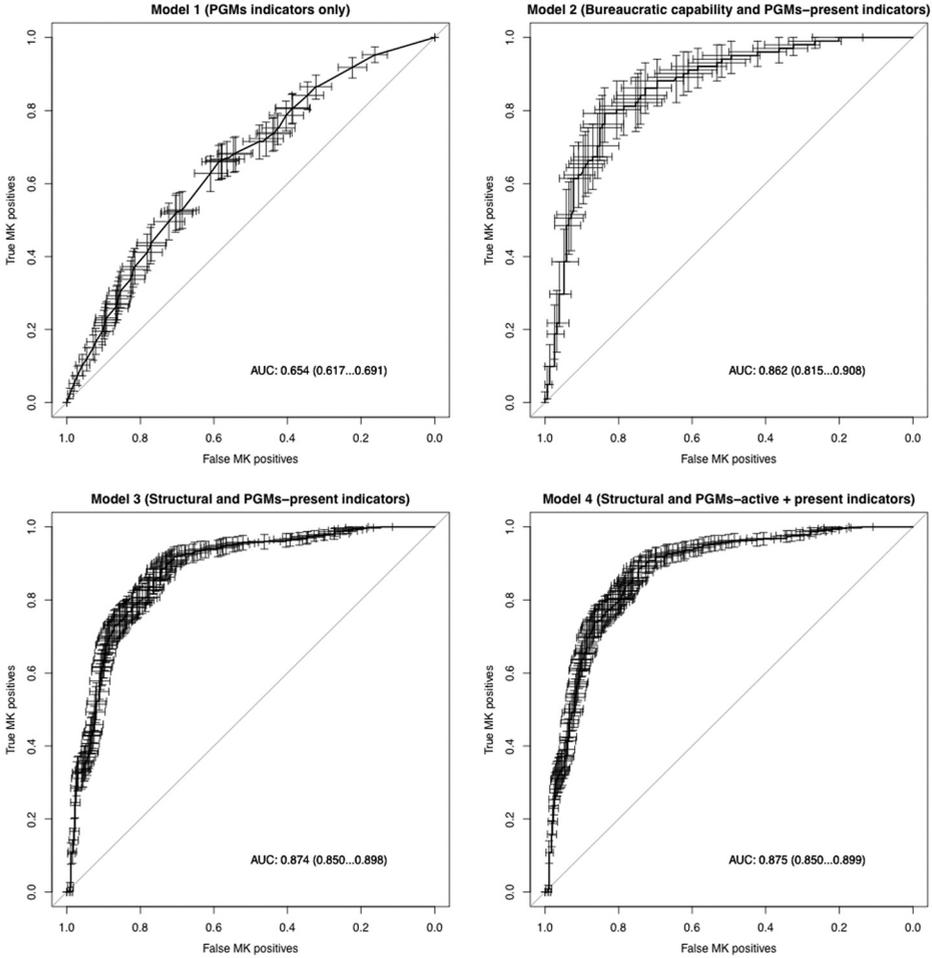


Figure 1. In-sample ROC curves state-led mass killing logit models, 1981–2007.

that, in all but model IV, including PGM indicators also reduces the estimated forecast error. These findings lend strong support to hypothesis H3.

Having shown the improvement in the predictive strength in state-led mass killing models when PGM indicators are included using both in-sample data and leave-one-out cross-validation, I repeat the process used in the first phase with out-of-sample data as an auxiliary robustness test in the third phase. Namely, I utilize the coefficient estimates obtained from in-sample data (presented in Table 1) on an independently coded out-of-sample dataset that was not used in the estimation of these coefficients in the first place to produce new AUC estimates. This out-of-sample dataset includes 33 internal threat episodes that began after 2008 and ended or were ongoing before or during 2013 (a total of 50 campaign-years), based on conflict (Themnér and Wallensteen, 2014) and civil disobedience campaigns data. Out-of-sample mass killing data and PGM indicators (both activity and presence) were coded using the same guidelines used by the authors of the original datasets on which I drew for

Table 2. In-sample predictions, state-led mass killing logit models, 1981–2007

	Model I	Model II	Model III	Model IV
AUC (with PGM indicators)	0.654 [0.617–0.691]	0.862 [0.815–0.908]	0.874 [0.850–0.898]	0.875 [0.850–0.899]
AUC (without PGM indicators)	0.614 [0.577–0.652]	0.813 [0.762–0.864]	0.864 [0.839–0.890]	0.864 [0.839–0.890]
DeLong et al. test	$p = 0.0065$	$p = 0.0049$	$p = 0.0062$	$p = 0.0040$
Favored AUC	AUC with PGM indicators			
AIC (with PGM indicators)	1088.4	254.83	772.41	775.32
AIC (without PGM indicators)	1110.6	273.88	785.38	785.38
Favored AIC	Model with PGM indicators			
N (campaign-years)	822	255	822	822

Confidence intervals in brackets; null hypothesis for DeLong et al. (1988) test for two correlated ROC curves: true difference in AUC is equal to 0.

Table 3. Leave-one-out cross-validation predictions, state-led mass killing logit models, 1981–2007

	Model I	Model II	Model III	Model IV
AUC (with PGM indicators)	0.621 [0.582–0.659]	0.835 [0.784–0.885]	0.862 [0.836–0.887]	0.861 [0.836–0.887]
AUC (without PGM indicators)	0.537 [0.496–0.578]	0.782 [0.727–0.837]	0.853 [0.826–0.879]	0.853 [0.826–0.879]
DeLong et al. test	$p = 6.599e-08$	$p = 0.0044$	$p = 0.0097$	$p = 0.0272$
Favored AUC	AUC with PGM indicators			
Forecast error (with PGM indicators)	0.383	0.203	0.209	0.213
Forecast error (without PGM indicators)	0.451	0.290	0.213	0.213
Favored forecast error	Model with PGM indicators	Model with PGM indicators	Model with PGM indicators	No significant difference
N (campaign-years)	822	255	822	822

Confidence intervals in brackets; null hypothesis for DeLong et al. test for two correlated ROC curves: true difference in AUC is equal to 0.

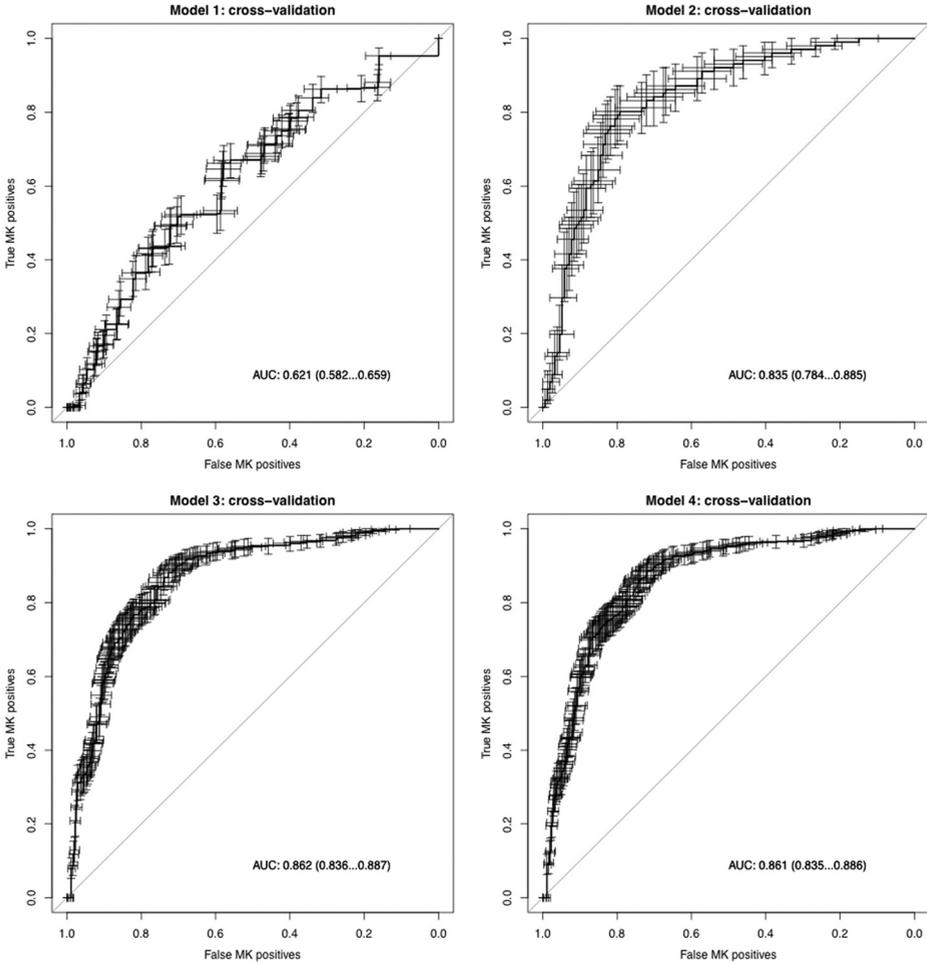


Figure 2. Leave-one-out cross-validation ROC curves of state-led mass killing logit models, 1981–2007.

the in-sample analysis. Owing to temporal limitations with the Democracy–Dictatorship data, I draw on the Polity IV dataset (Marshall et al., 2014) for coding a post-2008 binary indicator of democracy, with a campaign-year with a Polity2 score of 7 or higher, given a democracy score of 1, 0 otherwise. In addition, because the population and GDP data used for in-sample data were updated only until 2011, I used World Bank data (2015) to code the out-of-sample log GDP per capita indicator and US Census Bureau data (2015) to code the out-of-sample log population indicator.

The differences between the AUC of models I–IV with and without PGM indicators for out-of-sample data and the DeLong et al. test results are presented in Table 4, with ROC curves presented in Figure 3. Because the out-of-sample dataset includes only 50 observations, I treated a DeLong et al. test of $p < 0.1$ as significant. Once more, PGM indicators provide a significant improvement over baseline models in almost every case (excluding

Table 4. Out-of-sample predictions, state-led mass killing logit models, 2008–2013

	Model I	Model II	Model III	Model IV
AUC (with PGM indicators)	0.763 [0.635–0.891]	0.850 [0.744–0.956]	0.857 [0.755–0.960]	0.857 [0.755–0.960]
AUC (without PGM indicators)	0.550 [0.383–0.717]	0.781 [0.656–0.905]	0.783 [0.653–0.912]	0.783 [0.653–0.912]
DeLong et al. test	$p = 0.0109$	$p = 0.1664$	$p = 0.0568$	$p = 0.0676$
Favored AUC	AUC with PGM indicators	No significant difference	AUC with PGM indicators	AUC with PGM indicators
N (campaign-years)	50	50	50	50

Confidence intervals in brackets; null hypothesis for DeLong et al. test for two correlated ROC curves: true difference in AUC is equal to 0.

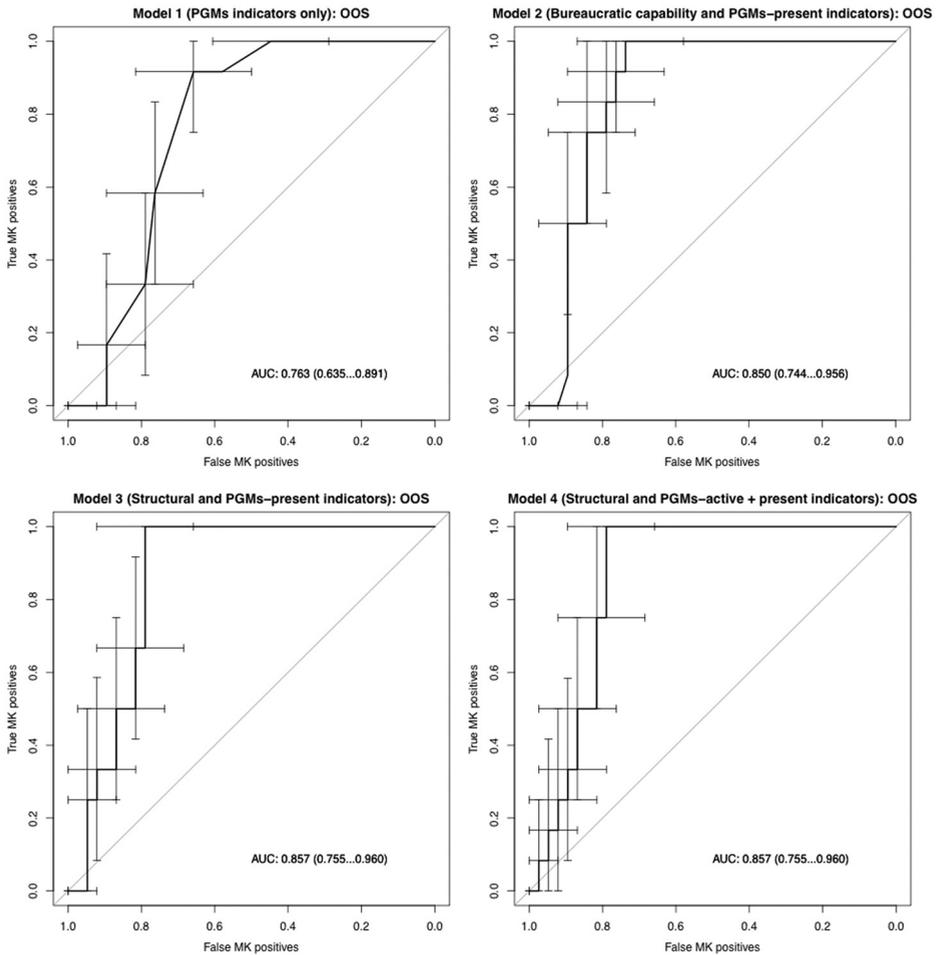


Figure 3. Out-of-sample ROC curves of state-led mass killing logit models, 2008–2013.

model II), even when out-of-sample data are used, which lends additional support to hypothesis H3. On the whole, these results strongly support the argument that PGM indicators indeed improve the strength of state-led mass killing prediction models with only regime-centric, conflict type and regional location variables, and that, as such, they can be used as predictive indicators of strategic mass killing.

To summarize the empirical section, I have shown that both semi-official and informal PGMs are significantly and positively associated with state-led mass killing. Starting from a fairly basic PGM-centric model, I demonstrated how these results hold even as a large number of regime-centric, conflict type, and region indicators were added to the model. Hence, my analysis found strong support for hypotheses H1 and, to a lesser extent, H2. I then proceeded to test whether PGM indicators produce a significant improvement in the prediction strength of state-led mass killing models based on the AUC of ROC curves for each model in three phases, using both in-sample, leave-one-out cross-validation and out-of-sample data. I found that in almost every case, including PGM indicators produced a significant and positive change in the AUC of model ROCs in in-sample data, cross-validation, and out-of-sample data, and reduced leave-one-out cross-validation estimates of forecast error. These findings lend strong support to hypothesis H3.

I conclude this section by listing two interesting findings of my analysis: first, and perhaps most important, I found that including PGM-centric indicators in state-led mass killing forecasting models provides a significant improvement in their predictive strength. This suggests that models using only structural indicators can benefit from taking agent-specific characteristics into account. Second, and somewhat related, I found that PGMs—as part of a regime's security repertoire—appear to be as strong a correlative factor as some regime-centric and conflict type variables, at least when state-led mass killing is concerned.

Conclusion

My main contention in this paper is that PGMs can produce a significant improvement in the predictive strength of state-led mass killing models. In particular, I argued that when a regime reacts rationally to some form of an internal threat—a civil war, a coup d'état, a “sons-of-soil” conflict, or a civil disobedience campaign—it might consider carrying out a mass killing campaign against one or more civilian groups associated with it, assuming other attempts to contain the threat have failed. In such contexts, PGMs provide at least three cost-reducing advantages to the regime. First, because the government can shirk responsibility for the violence, blaming it on the individual actions of “bad apples”, PGMs provide a plausible deniability mechanism. Second, using PGMs allows the official military forces to concentrate their efforts on containing more serious military threats, especially when the campaign involves an armed insurgency or an invasion by a foreign military force. Third, the costs of co-opting, or even forming, a PGM are many times lower than the costs of allocating a portion of the official military force for these purposes. A regime faced with an internal threat can take these three advantages into consideration, and so the availability of PGMs for these purposes can “tip the balance” in favor of mass killing. As a consequence, I expect that the presence and activity of PGMs will increase the occurrence of state-led mass killing. After reviewing some historical and anecdotal evidence in support of these expectations, this paper then empirically establishes that PGMs (both semi-official and informal) are indeed a consistent predictive indicator of state-led mass killing for the years 1981–2013.

The above findings have important implications for the study of both state-led mass killing and militias. In recent years, our collective understanding of mass killing has greatly benefited from numerous exceptional theories of the strategic dynamics underlying regimes' decisions to employ mass killing as a rational means (Esteban et al., 2010; Kalyvas, 2006; Valentino, 2004; Valentino et al., 2004; Wood, 2010). These studies—which treat mass killing as more of a deliberate policy directed toward civilians than as a phenomenon of wartime “collateral damage”—explain a large and critical part of the mass killing puzzle. In addition, the role played by independent actors operating under the auspices of unscrupulous regimes in different contexts has also been emphasized (Ahram, 2011; Koren, 2014; Mitchell et al., 2014; Nordås and Cohen, 2012; Raleigh, 2012). These studies highlight the important effect of agency on different forms of violence against civilians. My article has demonstrated the predictive advantages provided by bringing closer together these two bodies of scholarship, even if the correlative associations between PGMs and state-led mass killing cannot necessarily represent a causal relationship.

Given the immense social costs of mass killing, as well as the efforts and funding invested in early warning endeavors, such as the Political Instability Task Force, my identified linkages between PGM indicators and the prediction of state-led mass killing also have important policy relevance. Indeed, my study not only pinpointed a new potential explanatory variable for state-led mass killing, but also highlighted the manners in which PGM indicators can be used to improve our ability to reduce and ideally prevent it. More broadly, this study highlights the potential advantages of taking security repertoires into account. Because including PGM indicators in state-led mass killing forecasting models significantly improves their predictive strength, it is not implausible to conjecture that taking into account other characteristics of the security repertoire—such as ethnically exclusive militaries or the availability of police/paramilitary groups—can yield additional improvements. I hence believe that a better understanding of the effects of agent-related characteristics on global practices of violence is an especially compelling area for future research.

Acknowledgment

The author would like to thank John R. Freeman, Ronald R. Krebs, Holly Dunn, Kristen Senz, the anonymous reviewers of *Conflict Management and Peace Science*, and especially Benjamin E. Bagozzi for their invaluable input.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. Replication data and Online Appendix for this study are available at: <https://apps.cla.umn.edu/directory/profiles/koren044>.
2. I use the same term as utilized by Carey et al. (2013) because I rely on their database to code militia-centric variables. Understandably, there are theoretical and empirical costs and benefits incurred by hewing to this definition. However, because I focus on state-led mass killing I believe that the PGM dataset provides the best currently available pool of observations for coding this indicator. In this article I use the terms PGMs and militias interchangeably.

3. Mass killing was defined as the intentional killing of at least 1000 noncombatants during a campaign. I further elaborate the coding standards in the fourth section of this article.
4. For the purpose of analysis “internal threats” encompass the entire spectrum of civil wars, “sons-of-soil” wars, coups d’états and civil disobedience campaigns.
5. In order to draw on a wider theoretical framework, I also discuss literature related to state-led violence against civilians.
6. A robustness test that includes an indicator coding the number of PGMs in a country is presented in Online Appendix II.
7. Granted, in many or even most cases conflict begins owing to endogenous processes occasioned by the regime over time (e.g. Young, 2012). Nevertheless, the necessity to contain an internal threat can serve as a catalyst that might push the government to perpetrate mass killing against its citizens.
8. This argument is at least partly supported by the findings that anocracies might be especially likely to experience atrocities (Ulfelder, 2012).
9. The reader can refer to the original report for further elaboration on the coding criteria utilized by Ulfelder and Valentino and their definition of a “noncombatant”.
10. Note that only 822 campaign-years are analyzed owing to missing data for Somalia. In addition, I provide an alternative analysis, which utilizes a cross-sectional analysis of cases by regime as a robustness test, presented in Online Appendix II.
11. A conflict is defined as campaign with at least 25 battle-related deaths (Kreutz, 2010b: 2). Episodes that continued after 2007 were coded as having ended 31 December 2007.
12. Within the temporal framework employed in this article (1981–2007), 36 out of a total of 884 campaign-years and 14 international campaigns were removed.
13. Ten nonviolent civil disobedience campaigns were already included in the UCDP Conflict Termination Dataset and not recoded twice. One additional campaign (1986 Argentina) was dropped because it was not coded as a coup in the Powell and Thyne dataset (2011).
14. Campaigns coded by Chenoweth and Stephan were treated as containing a single episode.
15. The correlation matrix of all explanatory variables is provided in Online Appendix I.
16. This statement is based on both PGMD documentation (see: <http://www.sowi.uni-mannheim.de/militias-public/data/target/>) and because most media sources used for coding PGM indicators for out-of-sample data accessed via LexisNexis mention violence against different noncombatant groups.
17. As Table AII-I of the Online Appendix shows, PGMs are still significantly more likely to be associated with higher orders of magnitude of mass killings.
18. Semi-official militias were coded as present in 513 out of a total of 822 internal threat-years during which PGM presence was coded.
19. Semi-official militias were coded as active in 416 out of a total of 822 internal threat-years during which PGM activity was coded.
20. Informal PGMs were coded as present in 476 out of a total of 822 internal threat-years during which informal PGM presence was coded.
21. Informal PGMs were coded as active in 279 out of a total of 822 internal threat-years during which PGM activity was coded.
22. For robustness purposes, I also used cross-sectional data with militias coded by the region in which they operated. See Online Appendix II.
23. Owing to the temporal limitations of the Democracy–Dictatorship dataset, a dichotomous democracy variable based on the Polity2 score was used for out-of-sample data. Additional robustness models that utilize the Polity IV dataset are provided in Online Appendix I.
24. Owing to the temporal limitations of the GDP and Population dataset, population and GDP data for out-of-sample analysis were obtained from the US Census Bureau (2015) and the World Bank’s World Development Indicators (2015), respectively.

25. Owing to temporal limitations, I coded the coup d'état variable for the out-of-sample dataset from different media sources using LexisNexis, based on Powell and Thyne's coding standards.
26. The Americas were used as a reference category. These regional coding standards follow those used by Kreutz (2010b).
27. The primary reason the WDI Rule-of-Law index was preferred over other measures, such as the Polity2 scale, is that it was less likely to be correlated with either the dependent variable or the democracy variable. Robustness models that utilize the Polity2 scale are provided in Online Appendix I.
28. Robust standard errors (RSEs) were not used as there was no theoretical justification to include them (especially as including RSEs might cover other misspecification issues; see King and Roberts, 2014), and because the results of the RSEs models do not differ significantly from the models with regular standard errors. RSE model estimates are provided in Online Appendix I.
29. Note that, because rule-of-law scores go back only to 1996, the remaining data contain 255 campaign-years and 99 episodes.
30. A fully specified model that includes both the rule-of-law and rule-of-law2 indicators (model V) is presented in Online Appendix I.
31. Note that this can inflate standard errors, which might serve as an additional robustness measure.
32. See Table 2 and Online Appendix I.

References

- Ahram AI (2011) *Proxy Warriors: The Rise and Fall of State-Sponsored Militias*. Stanford, CA: Stanford University Press.
- Ahram AI (2014) The role of state-sponsored militias in genocide. *Terrorism and Political Violence* 26(3): 488–503.
- Alvarez A (2006) Militias and genocide. *War Crimes, Genocide and Crimes Against Humanity* 2: 1–33.
- Beck N (2001) Time-series–Cross-section data: What have we learned in the past few years? *Annual Reviews of Political Science* 4: 271–293.
- Brandt PT, Freeman JR and Schrodt PA (2011) Real time, time series forecasting of inter- and intra-state political conflict. *Conflict Management and Peace Science* 28(1): 41–64.
- Carey SC, Mitchell NJ and Lowe W (2013) States, the security sector, and the monopoly of violence: A new database on pro-government militias. *Journal of Peace Research* 50(2): 249–258.
- Carter DB and Signorino CS (2010) Back to the future: Modeling time dependence in binary data. *Political Analysis* 18(3): 271–292.
- Cheibub JA, Gandhi J and Vreeland JR (2010) Democracy and dictatorship revisited. *Public Choice* 143(1–2): 67–101.
- Chenoweth E (2008) Nonviolent and violent campaigns and outcomes: Codebook to accompany summer 2008 article in *International Security*. Available from: <http://www.ericachenoweth.com/research/>
- Chirot D and McCauly C (2010) *Why Not Kill Them All? The Logic and Prevention of Mass Political Murder*. Princeton, NJ: Princeton University Press, 2010.
- Cronin AK (2013) Why drones fail? When tactics drive strategy. *Foreign Affairs* 92: 44–54.
- DeLong ER, DeLong DM and Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44: 837–845.
- Dowdle AJ (2007) Civil wars, international conflicts, and other determinants of paramilitary strength in Sub-Saharan Africa. *Small Wars and Insurgencies* 18(2): 161–174.
- Esteban J, Morelli M and Rohner D (2010) Strategic mass killings. Institute for Empirical Research in Economics, University of Zurich Working Paper Series no. 486. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1615375.

- Fletcher L (2007) Turning interahamwe: Individual and community choices in the Rwandan genocide. *Journal of Genocide Research* 9(1): 25–48.
- Fujii LA (2009) *Killing Neighbors: Webs of Violence in Rwanda*. Ithaca, NY: Cornell University Press.
- Gagnon VP Jr (2004) *The Myth of Ethnic War: Serbia and Croatia in the 1990s*. Ithaca, NY: Cornell University Press.
- Gleditsch KS (2013) Updated GDP data to 2011 (6.0 beta), last updated November 2013. Available from: <http://privatewww.essex.ac.uk/~ksg/exptradegdp.html>.
- Goldhagen DJ (2009) *Worse Than War: Genocide, Eliminationism, and the Ongoing Assault on Humanity*. New York: PublicAffairs, 2009.
- Goldstone JA, Bates RH, Epstein DL, Gurr TR, Lustik MB, Marshall MG, Ulfelder J and Woodward M (2010) A global model for forecasting political instability. *American Journal of Political Science* 54(1): 190–208.
- Harff B and Gurr TR (1998) Systemic early warning of humanitarian emergencies. *Journal of Peace Research* 35(5): 551–579.
- Horowitz DL (1985) *Ethnic Groups in Conflict*. Berkeley/Los Angeles, CA: California University Press.
- Kalyvas SN (2006) *The Logic of Violence in Civil War*. Cambridge: Cambridge University Press, 2006.
- King G and Roberts ME (2014) How robust standard errors expose methodological problems they do not fix, and what to do about it. *Political Analysis* 23(2): 159–179.
- Koren O (2014) Military structure, civil disobedience, and military violence. *Terrorism & Political Violence* 26: 688–712.
- Krain M (1997) State-sponsored mass murder: The onset and severity of genocide and politicides. *Journal of Conflict Resolution* 41(3): 331–360.
- Kreutz J (2010a) How and when armed conflicts end: Introducing the UCDP Conflict Termination Dataset. *Journal of Peace Research* 47(2): 243–250.
- Kreutz J (2010b) UCDP Conflict Termination Dataset Codebook. Available from: http://www.pcr.uu.se/digitalAssets/124/124924_1ucdp_conflict_termination_dataset_v_2010-1.pdf
- Lacina B and Gleditsch NP (2005) Monitoring trends in global combat: A new dataset of battle deaths. *European Journal of Population* 21(2–3): 145–166.
- Lutz EL and Sikkink K (2000) International human rights law and practice in Latin America. *International Organization* 54(3): 633–659.
- Mamdani M (2001) *When Victims Become Killers: Colonialism, Nativism, and the Genocide in Rwanda*. Princeton, NJ: Princeton University Press, 2001.
- Marshall MG, Jaggers TR and Gurr K (2014) Polity IV project: Political regime characteristics and transitions, 1800–2013. Technical Report. Available from: <http://www.systemicpeace.org/>.
- Mitchell N and Carey SC (2013) Pro-Government Militias Database (PGMD) Codebook. Available from: http://www.sowi.uni-mannheim.de/militias/?page_id=11
- Mitchell NJ, Carey SC and Butler CK (2014) The impact of pro-government militias on human rights violations. *International Interactions* 40: 812–826.
- Mueller J (2000) The banality of “ethnic war”. *International Security* 25(1): 42–70.
- Nordås R and Cohen DK (2012) Sexual violence by militias in African conflicts: Not a question of “delegation” by states. Center for the Study of Civil War Policy Brief, January 2012. 2012. Available from: http://file.prio.no/publication_files/cscw/Nordas-Cohen-Sexual-Violence-Militias-African-Conflicts-CSCW-Policy-Brief-01-2012.pdf
- Poe SC and Tate N (1994) Repression of human rights to personal integrity in the 1980s: A global analysis. *American Political Science Review* 88(4): 858–872.
- Poe SC, Tate N and Keith LC (1999) Repression of the human right to personal integrity revisited: A global cross-national study covering the years 1976–1993. *International Studies Quarterly* 43: 291–313.
- Powell JM and Thyne CL (2011) Global instances of coups from 1950 to 2010: A new dataset. *Journal of Peace Research* 48(2): 249–259.

- Raleigh C (2012) Violence against civilians: A disaggregated analysis. *International Interactions* 38(4): 462–481.
- Rummel RJ (1995) Democracy, power, genocide and mass murder. *Journal of Conflict Resolution* 39(1): 3–26.
- Staub E (1989) *The Roots of Evil: The Origins of Genocide and Other Group Violence*. Cambridge: Cambridge University Press.
- Stephan MJ and Chenoweth E (2008) Why civil resistance works: The strategic logic of nonviolent conflict. *International Security* 33(1): 7–44.
- Straus S (2006) *The Order of Genocide: Race, Power, and War in Rwanda*. Ithaca, NY: Cornell University Press, 2006.
- Straus S (2015) *Making and Unmaking Nations: War, Leadership, and Genocide in Modern Africa*. Ithaca, NY: Cornell University Press, 2015.
- Themnér L and Wallensteen P (2014) Armed conflict, 1946–2013. *Journal of Peace Research* 51(4): 541–544.
- Treier S and Jackman S (2008) Democracy as a latent variable. *American Journal of Political Science* 52(1): 201–217.
- Ulfelder J (2012) Forecasting onsets of mass killing. Paper prepared for presentation at the *Annual Northeast Political Methodology Meeting*.
- Ulfelder J and Valentino B (2008) Assessing risks of state-sponsored mass killing. Report written for the Political Instability Task Force, February.
- US Census Bureau (2015) International Data Base. Available from: <http://www.census.gov/population/international/data/idb/informationGateway.php> (accessed 2 February 2015).
- Valentino BA (2004) *Final Solutions: Mass Killing and Genocide in the Twentieth Century*. Ithaca, NY: Cornell University Press.
- Valentino B, Huth P and Balch-Lindsay D (2004) “Draining the sea”: Mass killing and guerilla warfare. *International Organization* 58(2): 375–407.
- Varshney A (2002) *Ethnic Conflict and Civil Life: Hindus and Muslims in India*. New Haven, CT: Yale University Press, 2002.
- Ward MD, Metternich NW, Dorff CL, Gallop M, Hollenbach FM, Schultz A and Weschle S (2013) Learning from the past and stepping into the future: The next generation of crisis forecasting models. *International Studies Review* 15: 473–490.
- Ward MD, Greenhill BD and Bakke KM (2010) The perils of policy by *P*-value: Predicting civil conflicts. *Journal of Peace Research* 47(4): 363–375.
- Wood RM (2010) Rebel capability and strategic violence against civilians. *Journal of Peace Research* 47(5): 601–614.
- The World Bank (2015) World Development Indicators, 1960–2013. Last updated 30 January 2015. Available from: <http://data.worldbank.org/data-catalog/world-development-indicators>
- Young JK (2012) Repression, dissent, and the onset of civil war. *Political Research Quarterly* 66(3): 516–532.