Research Article

Research and Politics
July-September 2014: 1–10
© The Author(s) 2014
DOI: 10.1177/2053168014544586
rap.sagepub.com
SAGE

Predicting the duration of the Syrian insurgency

Ulrich Pilster and Tobias Böhmelt

Abstract

While there were several relatively short uprisings in Northern Africa and the Middle East during the Arab Spring, the dispute between the rebels and government forces in Syria has evolved into a full-scale civil war. We try to predict the length of the Syrian insurgency with a three-stage technique. Using out-of-sample techniques, we first assess the predictive capacity of 69 explanatory variables for insurgency duration. After determining the model with the highest predictive power, we categorize Syria according to the variables in this final model. Based on in-sample approaches, we then predict the duration of the Syrian uprising for three different scenarios. The most realistic point prediction is 5.12 years from the insurgency's start, which suggests an end date between the end of 2016 and early 2017.

Keywords

Civil war, event history models, insurgency, prediction, Syria

Introduction

Since riots and demonstrations began across the Arab world in December 2010, dictators and their regimes have fallen. Zine El Abidine Ben Ali in Tunisia, Hosni Mubarak in Egypt, Muammar Gaddafi in Libya, and Ali Abdullah Saleh in Yemen were all forced to resign from power, albeit in different ways. Despite their differences, a common characteristic of the uprisings that brought down those dictators was their relatively short durations. Internal and external observers alike have been surprised by the speed with which old regimes fell, even as the stability of their successor regimes is far from being secured.

The insurgency in Syria constitutes a special case in this context. After nationwide demonstrations began in March 2011, the dispute between the protest movement and the government did not lead to a quick end, either by opposition victory or Bashar al-Assad securing his power. Instead, the conflict evolved into what fulfils all the criteria of a full-scale civil war or insurgency[1] (Gleditsch et al., 2002; see also Lyall, 2010): it takes place within a state; it involves the central government as one of the principal actors; it results in a minimum of 25 battle-related combatant fatalities within a 12-month period; and all parties involved demonstrate military activity that causes battle-related deaths suffered by either side. In total, Price et al. (2013) estimate 92,901 casualties (both battle-related and civilian) between March 2011 and April 2013. The most recent figures released in April 2014 by the Syrian Observatory for Human Rights suggest that at least 150,000 people have been killed since the beginning of the uprising.

At the time of writing this article, it seems unclear when the fighting will come to an end, and estimates have varied widely. On one hand, some experts have suggested that "the Syrian civil war may well turn out to be shorter than generally anticipated" (Balcells and Kalyvas, 2012a) and that "the regime appears to have only a few weeks left [as of December 2012] before it collapses" (White and Tabler, 2012). On the other hand, the International Crisis Group has indicated that "the conflict's evolving dynamics do not suggest a quick denouement" (International Crisis Group, 2012a) and that the conflict will continue, as a decisive military victory for either side is unlikely (International Crisis Group, 2013). The group's most recent report (International Crisis Group, 2014) underlines this again.

To the best of our knowledge, however, there are no systematic predictions of the length of the Syrian insurgency,

University of Essex, Colchester, UK

Corresponding author:
Dr Ulrich Pilster, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK.
Email: ulrich.pilster@gmail.com

despite the importance of such an exercise for policymakers and scholars alike.[2] For example, insights on the duration of the Syrian insurgency could help us anticipate whether we will see military intervention by Syria's neighbours (Biddle et al., 2012), or whether neighbouring countries will be destabilized through continuous refugee flows or ethno-sectarian polarization along Sunni–Shia lines (see International Crisis Group, 2012b).

We address this uncertainty by systematically predicting the length of the Syrian insurgency. This is in line with Ward et al. (2010: 364) who highlight that scholars should "be willing to make predictions explicitly – and plausibly be wrong, even appear foolish – because our policy prescriptions need to be undertaken with results that are drawn from robust models that have a better chance of being correct. The whole point of estimating risk models is to be able to apply them to specific cases." Methodologically, we follow and expand on Bennett and Stam (2006) who predicted the length of the 2003 war between the US and Iraq.[3] To this end, using out-of-sample techniques, we first assess the predictive capacity of 69 explanatory variables on insurgency duration, employing insurgency data from Lyall and Wilson (2009) and Lyall (2010) as a starting point. After determining the final model with the highest predictive power, Syria is categorized according to those variables in this model. Based on in-sample approaches, we then predict the length of the Syrian uprising for three different scenarios. The most realistic point prediction for the length of the Syrian insurgency is 5.12 years since its beginning, which suggests an end date between the end of 2016 and early 2017.

Predicting insurgency durations: Data setup

To identify the most important predictors, we use data from Lyall and Wilson (2009) that cover 286 insurgencies between 1800 and 2005 (see also Lyall, 2010). The advantage of these data is that they offer a wider variety of variables relating to the combatants' military capabilities and strategies than other comparable sources. Note that despite our interchangeable treatment of the terms 'civil war' and 'insurgency', it may seem that Lyall and Wilson (2009) focus exclusively on insurgencies in which rebels resorted to guerrilla warfare as defined through 'hit-and-run' tactics and the attempt to win the allegiance of the civilian population (Lyall, 2010: 175). However, Kalyvas and Balcells (2010) show that the data do in fact include a significant number of conventionally fought civil wars.

We first dropped all civil war cases that: (1) were fought by foreign counterinsurgents or as anti-colonial wars;[4] and (2) started in the pre-1945 period[5] in order to create a more homogenous data set of cases. We also exclude Lyall's (2010) 21 right-censored observations, i.e. all insurgencies that were ongoing by the end of 2005. Ultimately, after accounting for missing values, we obtained a data set of 98 insurgencies that started and ended between 1945 and 2005.

We use Lyall and Wilson's (2009) coding of insurgency start and end dates (see also Lyall, 2010). The former is defined as the first large-scale insurgent attack or open declaration of hostilities, while the latter is coded once the government or the insurgents have achieved their objectives. This also includes the last day of war-related fatalities or once a peace deal has been reached. On average, in our data an insurgency lasted about eight years.

Theoretical considerations and out-of-sample predictions

Which factors influence the duration of civil wars and insurgencies? Studies dealing with this question primarily rely on the idea that conflicts continue if at least one set of belligerents believe a dispute is more beneficial than peace (e.g. Balch-Lindsay and Enterline, 2000; Brandt et al., 2008; Collier et al., 2004; Cunningham, 2006, 2010; Fearon, 2004; Fearon and Laitin, 2003; Hegre and Sambanis, 2006; Lyall, 2010; Mason and Fett, 1996; Regan, 2002). We therefore deal with issues of willingness and opportunity (Most and Starr, 1989; Starr, 1978). First, willingness is generally "concerned with the processes and activities that lead men to avail themselves of the opportunities to go to war" (Starr, 1978: 364). In the civil war duration literature, one can find this willingness aspect particularly when scholars discuss existing (or new) grievances. Second, Starr defines opportunity as the 'possibility of interaction': "interaction exists between individuals of one nation state so that it is possible for conflict to arise – and to arise over values potentially important enough to warrant the utilization of violent coercive action" (Starr, 1978: 368). Put differently, 'opportunity' refers to a set of possibilities and constraints in the context of conflict duration and describes those factors that increase the likelihood that belligerents *can* sustain fighting. Most scholars emphasize the importance of 'greed' factors in terms of the opportunity to continue fighting.

After setting up our data and consulting the relevant literature on insurgency and civil war duration, in light of the willingness and opportunity framework, we identified 69 explanatory variables that were likely to affect the duration of these types of conflicts. These variables fall into the clusters of: (1) physical terrain and geography; (2) cultural terrain; (3) insurgency power; and (4) state power.[6] Note that Lyall (2010: 187) actually examines the duration of insurgencies and civil wars, as well; thus, his core model might constitute an obvious choice for predicting the length of the Syrian insurgency. In fact, we also considered his core model and the explanatory variables therein as the base for our predictions. However, all our calculations demonstrate that the predictions based on Lyall's (2010) explanatory variables are less accurate than those stemming from our final model that (partly) differs from his setup. We therefore concentrate on our results, but point to Lyall (2010) for comparison.

In order to assess the predictive accuracy and strength of our 69 variables for the duration of insurgencies, we conducted out-of-sample predictions via a cross-validation exercise of 69 bivariate models (Ward et al., 2010: 364). Cross-validation relies on dividing existing data into subsets, using random assignment of the cases to the different sets. All except one of the subsets are then pooled together and routinely estimated by applying the preferred model specification.[7] The remaining subset, called the 'test set' (Ward et al., 2010: 370), subsequently serves to assess the predictive power of the model estimated on the pooled subsets – in our case, only one explanatory variable at each time.

Specifically, we performed a five-fold cross validation exercise for each variable, which we repeated ten times (Ward et al., 2010). To this end, we randomly divided our sample into five equally-sized subsets, pooled four in order to 'train the model' (i.e. to estimate the coefficients and parameters of interest), and kept one as the test set. After this exercise, we assessed the out-of-sample predictive power of each variable by following Bennett and Stam (2009: 259), who rely on Costner's (1965) measure of the proportional reduction of error (PRE) for duration models, which is defined as

$$\text{PRE} = \frac{\begin{array}{c}\text{(Sum of absolute months}\\ \text{of error from naive model} -\\ \text{sum of absolute months}\\ \text{of error from improved model)}\end{array}}{\begin{array}{c}\text{Sum of absolute months of}\\ \text{error from naive model}\end{array}}. \qquad (1)$$

We repeated this procedure five times for 10 different partitions of the sample, yielding 5 × 10 = 50 different PRE values for each variable, from which we finally calculated the average PRE.[8]

Out of the 69 variables, we then: (1) identified those 10 predictors with the highest PRE values; and (2) calculated the PRE measures when dropping each of these 10 variables from the model separately and combinations thereof. Ultimately, we obtained the highest average out-of-sample PRE with a value of 0.2369 (SD of 0.0305) when estimating a model that relies on eight covariates.[9] This PRE value of 0.24 of our final model can be interpreted as an average 24% reduction in total error in the 10 rounds of out-of-sample prediction when compared to an empty model. In comparison, note that the model based on Lyall (2010) yields a negative PRE-measure of −0.005 (with a SD of 0.031), indicating an out-of-sample performance that is on average worse than that of an empty model.

Our final eight variables are defined theoretically and empirically as follows. First, there is *Non-contiguous territory*. Countries experiencing an insurgency with non-contiguous territory are "countries with territory holding at least 10,000 people and separated from the land area containing the capital city either by land or by 100 km of water" (Fearon and Laitin, 2003: 81). Non-contiguous territory makes it easier for rebels to find sanctuaries into which the central government is unable to project power. This is likely to prolong civil wars.

Second, there is *Size of second largest ethnic group*, i.e. "a measure of the share of population belonging to the second largest ethnic group" (Fearon and Laitin, 2003: 78). This variable operationalizes how equal demographic power is distributed between the major ethnic groups. In countries where ethnicity is politically salient, roughly equally-sized ethnic groups are likely to translate into combatants of similar strength, thereby prolonging insurgencies.

Third, insurgents might rely on contraband to finance their campaigns. Fearon (2004), among others, argues that "cocaine, precious gems, or opium" constitute another source of rebel income (next to, for example, foreign states or diasporas) that can be used to purchase weapons, other resources, draft and train fighters, etc. Given that contraband then raises the capabilities of the insurgents, it is likely to increase the duration of civil wars. We denote this binary variable as *Contraband*.

Fourth, we employ the dichotomous variable *Cold war* in order to capture insurgencies that took place during the period 1949–1989 (*Cold war*= 1) or after the Cold War (*Cold war*= 0). According to Strauss (2012: 196ff; Fearon, 2004: 280f), insurgencies and civil wars during the Cold War lasted longer, primarily due to the higher support that insurgents could expect from the superpowers.

Fifth, there is *External support*, which is "a scaled variable that measures whether insurgents received two critical types of assistance: material economic and military aid, and the ability to use a neighboring country as a sanctuary" (Lyall and Wilson, 2009: 84). Rebels capable of drawing on external sanctuaries or support from third parties become more powerful in relation to the government and, hence, are able to sustain their military effort for a longer time.

Sixth, in terms of governmental support, Regan (2002) coded a variable that captures economic interventions in favour of the government by a third party. The underlying theoretical rational for the *Economic intervention government* item is that, given a third-party intervention, governments can rely on more resources in their counterinsurgency campaigns, which may lengthen civil wars. Regan (2002) finds empirical support for this claim. Somewhat surprisingly, the military intervention variables from Regan (2002) scored comparatively low predictive power and, thus, are excluded from our final model.

Seventh, the government might make use of helicopters. This variable *Rotary-wing assets* measures whether an

Table 1. Final model for prediction: Weibull regression model.

	Final model (1)	Substantive effects (2)
Non-contiguous territory	–0.76	109.43
	(0.30)**	[21.67; 237.34]
Size of second largest ethnic group	–2.27	112.77
	(1.00)**	[23.20; 238.28]
Contraband	–0.39	29.14
	(0.20)*	[–0.82; 62.14]
Economic intervention government	–0.41	44.39
	(0.17)**	[7.02; 88.53]
Cold war	–0.60	42.58
	(0.22)***	[11.51; 81.12]
External support	–0.34	65.58
	(0.11)***	[19.34; 131.24]
Rotary-wing assets	–0.34	38.17
	(0.22)	[–9.05; 92.04]
Conventional insurgency	0.83	–52.20
	(0.22)***	[–91.22; –21.64]
Constant	–6.95	
	(0.74)***	
Observations	98	
Log pseudolikelihood	–148.98	
p	1.08 (0.10)	
Wald χ^2	84.08***	

Table entries are coefficients in column (1) or the average simulated change in predicted duration (in months = 30 days/month) in column (2) when moving from the minimum towards the maximum of a respective variable while holding all other variables at their means; robust SE in parentheses; 95% confidence intervals in brackets; *significant at 10%; **significant at 5%; ***significant at 1% (two-tailed).

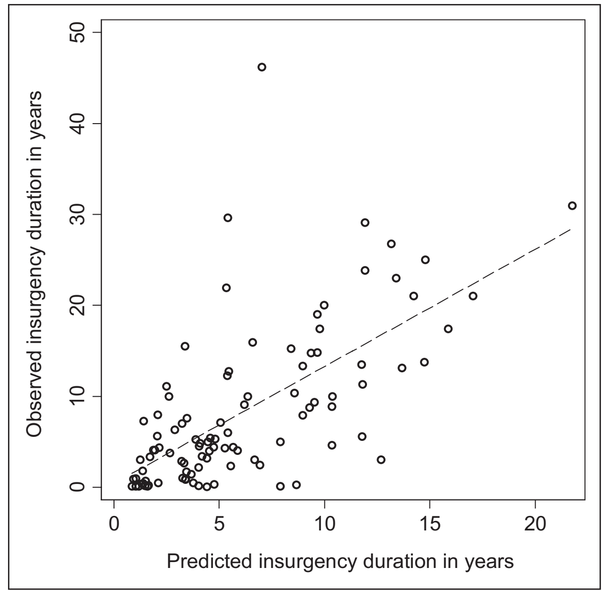

Fig 1. Observed and predicted insurgency durations. Dashed line pertains to estimated linear fit.

incumbent deployed 25 or more helicopters during an insurgency (Lyall and Wilson, 2009: 84). In general, the usage of helicopters gives rise to governmental strategies aimed at suppressing insurgents rather than strategies intended to establish the more permanent, on-the-ground presence in rebel-influenced areas that would be necessary to destroy the insurgents' organization. This is likely to allow the insurgents to continue their military efforts.

Finally, we include *Conventional insurgency*, which codes whether rebels and governments in the conflict predominantly use heavy armour and weaponry, resulting in clashes along defined front lines and/or between armed columns (Kalyvas and Balcells, 2010: 419). Conventionally fought insurgencies tend to be significantly shorter in their duration, as opponents see direct and decisive confrontation rather than the attrition through indirect confrontation that is typical of guerrilla warfare (Arreguin-Toft, 2001).

In-sample predictions

Based on our results from the out-of-sample predictions, we now turn to the in-sample predictions that bring us closer to estimating the length of the Syrian insurgency. We estimated the full model using the eight strongest predictors along substantive quantities of interest and calculated the PRE measure of this Weibull regression model to illustrate its in-sample predictive power. In this estimation, which is summarized in Table 1, the sign of the coefficient indicates how a covariate affects the hazard rate: a positive coefficient increases the hazard rate and, therefore, reduces the expected duration, while a negative coefficient decreases the hazard rate and therefore increases the expected duration. While all covariates have the expected impact on the duration of insurgencies, our calculation of this model's in-sample PRE yields a value of 0.295: i.e. we achieve a 30% reduction in total error from the full model relative to an empty model. This result seems promising in comparison to other studies that seek to predict the duration of wars. For example, Bennett and Stam (2009: 266) achieved a slightly lower PRE of 0.20 for their final model, while the model estimation based on Lyall (2010) achieves an in-sample PRE of 0.073.

Figure 1 plots the predicted insurgency duration against the observed durations of insurgencies. Evidently, the relationship is highly positive with a pairwise correlation of 0.65 ($p < 0.000$). Thus, this figure essentially mirrors our assessment according to the PRE above. Our model also constitutes an improvement over Lyall (2010): the pairwise correlation between observed and predicted durations is only 0.43 ($p < 0.000$) in Lyall's work. Nonetheless, there are absolute differences between predicted and actual durations in our setup: the average absolute error is 1585.83 days (about 4.34 years), and the model also seems to underpredict the duration of insurgencies, with an average difference of −794.77 days between predicted and actual

Table 2. Predicted durations of scenarios.

Scenario	point Prediction	Adjusted point prediction according to model's tendency to underpredict
Irregular insurgency with economic support government	6.34 years (1.83)	8.52 years
Conventional insurgency without economic support government	2.01 years (0.69)	4.19 years
Conventional insurgency with economic support government	2.94 years (1.01)	5.12 years

SE in parentheses.

durations. For comparative purposes, the model based on Lyall (2010) shows an average absolute error of 2069.71 days and underpredicts the duration of insurgencies on average by −997.32 days. Thus, we adjust our point predictions for the length of the Syrian insurgency below according to an average difference of −794.77 days between predicted and actual durations.

Predicting insurgency durations: How long will the Syrian insurgency last?

What does this model predict for the case of Syria? We calculated the predicted duration of the current Syrian civil war for three different scenarios that are based on a general uncertainty among experts and scholars with regard to two factors: (1) an irregularly fought insurgency with economic support for the government by a third party; (2) a conventional insurgency without governmental economic support; and (3) a conventional insurgency in which the regime receives economic support from a third-party state. Covering all possible scenarios in this regard, instead of ignoring current controversies and opting for a somewhat deterministic assessment, would increase the value of our research. Put differently, the values for six out of our final model's eight variables are held constant throughout the calculations, reflecting their invariable character in the context of the Syrian insurgency. First, the Syrian territory is contiguous. Second, Shia (i.e. Alawites, Ismailis, and other Shia combined), as the second-largest ethno-religious group, comprise about 14% of the population. Third, the insurgents cannot and do not use finances from contraband. Fourth, the insurgency takes place after the Cold War. Fifth, Assad can still rely on rotary-wing assets and the regime's command of the skies is given – particularly since a no-fly zone has not been imposed at the time of writing this article. Finally, the insurgents are externally supported in the form of sanctuary and material support (e.g. Holliday, 2012).

Our scenarios consequently focus on variation in two variables: whether the Syrian insurgency is fought irregularly or conventionally, and whether the government receives economic support from a third party. Table 2 summarizes the predictions for these scenarios. First, when assuming an irregular insurgency, i.e. guerrilla warfare in which the rebels and the government do not predominantly use heavy armour and weaponry (Kalyvas and Balcells, 2010: 419), our model suggests a point estimate of 6.34 years in length. Adjusting for the systematic underprediction in our model, we obtain a point estimate of 8.52 years. Hence, depending on the possible codings for the start date of the Syrian insurgency, which could be in March 2011 (when the demonstrations started), on 29 July 2011 (when the former Syrian Army Colonel Riad al-Asaad announced the formation of the Free Syrian Army), or in early September 2011 (when the first major clashes between the Syrian Army and the armed opposition occurred in the Idlib province), our model predicts that the fighting might even continue until early 2020.

However, there are reasons to believe that the situation might not turn out that badly. While the Syrian civil war may have started as an irregularly fought insurgency, it seems to have turned into a conventional conflict due to growing rebel capabilities (Balcells and Kalyvas, 2012a), which are (partly) caused by the foreign support they receive (Borger, 2012). Moreover, after about three years of fighting, the economic capabilities of the Syrian regime are decreasing, with international support increasingly vanishing as well. Put differently, if the Syrian insurgency is actually a conventionally fought civil war and Assad's regime cannot rely on (much) economic support from an outside party, our model gives a point prediction of about 2.01 years, which leads to a prediction of 4.19 years when adjusted for the model's tendency to underpredict durations. According to this scenario, we thus might see the end of this dispute at the end of 2015 or the beginning of 2016.

Still, even if there is evidence that the Syrian insurgency is now a conventionally fought conflict, it seems more realistic to assume that Assad can still rely on economic support from a third party. Although it is rarely acknowledged or confirmed, several sources suggest that the Syrian regime receives economic support from the Russian government in the form of grants, loans, and non-military equipment or expertise, etc. (Grove, 2013; Peel and Clover, 2012). Therefore, a more realistic scenario from our vantage point at the present time is a conventional insurgency

Table 3. Weighted average of durations.

	Weighted average of durations
Point prediction	3.56 years
Adjusted point prediction	5.74 years

in which the regime can rely on the economic support from a third-party state. According to our model, we obtain a point estimate of 2.94 years in length for the Syrian insurgency in this scenario, which leads to an adjusted point prediction of 5.12 years. Hence, this finding suggests an end date of the Syrian civil war between the end of 2016 and early 2017.

In light of this, finally, if we explicitly address the probabilities we assign to each of the three scenarios for which we estimated the duration of the Syrian insurgency, in our view scenarios 1 (irregular insurgency) and 2 (conventional insurgency without economic support) are equally likely, but less so than scenario 3 (conventional insurgency with economic support). Hence, we assign probabilities of 0.25 to scenarios 1 and 2, while the third scenario receives a value of 0.50, which in turn allows us to calculate a weighted average of the resulting predictions. Table 3 summarizes the results. While the weighted average of the point prediction is fairly close to the non-weighted point prediction of scenario 3 (Table 2), the weighted average of the adjusted point prediction suggests a duration of 5.74 years. Depending on the start date of the Syrian civil war, this could imply an end in late 2017. The prediction is fairly similar to our most realistic scenario even when using a weighted average of predicted durations across all scenarios.

Conclusions

We built our work on out-of-sample and in-sample prediction techniques, which allowed us to identify the strongest predictors of insurgency durations and to assess the predictive power of a model that uses these variables. Ultimately, we predict the length of the current Syrian insurgency for three different scenarios. The most realistic scenario suggests a length of 5.12 years (when adjusted for the model's tendency to underpredict durations).

Aside from providing an estimate of the duration of the Syrian insurgency, this article makes two additional contributions to the scientific discourse. Methodologically, we further develop Bennett and Stam's method to enable the out-of-sample assessment of duration models' predictive capacities (Bennett and Stam, 2006; see also Slantchev, 2004). This approach also helps us to build a model whose predictions outperform recent work on insurgency durations (Lyall, 2010). The model that we summarize in Table 1 might consequently be used in future research as a baseline model against which to assess new variables of theoretical interest.

Substantively, the majority of variables contained in our final model support a recent finding by Balcells and Kalyvas (2012b) that "technologies of rebellion," i.e. variables describing relative capabilities of combatants and their type of interaction, are central to understanding insurgencies and civil wars.

There are two potential limitations of this research, however. First, the two variables we varied for the three prediction scenarios, *Conventional insurgency* and *Economic support for the government*, are essentially time-invariant in the data. This could possibly limit the conclusions we derived. Although we cannot address this issue in this research, future work should pay closer attention to those dynamics that are characteristics of *most* insurgencies and civil wars (see Bennett and Stam, 2006: 109; Enterline et al., 2013).

Second, and more importantly, conflict begets conflict. Walter (2004), for instance, estimates that about 38% of civil wars or insurgencies between 1945 and 1996 recurred. Even if the current insurgency were to end with the overthrow of the Assad regime in late 2014, this act is unlikely to indicate the end of the Syrian crisis.

Acknowledgements

We thank David Cunningham for sharing his data, and Anjali Dayal, the anonymous reviewers, and the journal's editor-in-chief, Erik Voeten, for helpful comments. For this project, we relied on the data by Regan (2002), Fearon and Laitin (2003), Fearon (2004), Collier et al. (2004), Cunningham (2006, 2010), Hegre and Sambanis (2006), Lyall and Wilson (2009), and Lyall (2010). These data are all publicly available. We used the statistical software package Stata 13 for the estimations, figures, and tables. (StataCorp LP, 4905 Lakeway Drive,College Station, Texas 77845-4512USA).

Declaration of conflicting interest

The authors declare no conflict of interest.

Funding

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

Supplementary material

The replication files are available at: http://thedata.harvard.edu/dvn/dv/researchandpolitics

Notes

1. We use the terms 'civil war' and 'insurgency' interchangeably.
2. Note that other 'predictions' have been made on the outcome of the Syrian insurgency, however. According to Walter (2012) or the Senate Select Intelligence Committee (2012), the evidence is overwhelming that Assad will be

unable to win. However, the crucial question remains *when* this will be.

3. Bennett and Stam (2006: 113) highlight the importance of the techniques we use in this article, although they have not been employed in their study. See also Slantchev (2004) for an earlier study using duration models on interstate wars and predicting exercises.
4. See Lyall (2010) for the differences between wars by foreign occupiers in comparison to domestic insurgencies. See Fearon (2004) for the uniqueness of anti-colonial wars.
5. See Johnston and Urlacher (2012) on how the duration of insurgencies varies between the pre- and the post-Second World War periods.
6. We list these items in the appendix.
7. Point predictions, which are necessary for estimating the proportional reduction of error as defined below, require a parametric duration model (Bennett and Stam, 2009: 260). We use a Weibull regression model as it seemed to provide the best fit when comparing the sum of absolute months of error from empty models using other distributions (e.g. log–logistic, exponential, or Gompertz). Moreover, all our predictions are based on median durations. Note, however, that median durations tend to underpredict durations. Thus, we also considered logged median durations, regular mean durations, or logged mean durations as these might provide ways in addressing the underestimation. When replacing our reported median durations with logged median durations, regular mean durations, or logged mean durations, we obtain models with predictors that are virtually identical to the eight covariates we analyse in Table 1.
8. We summarize the results for the different indicators in the appendix.
9. The ‘progressive elimination of variables’ in Appendix 1, Table 6, shows that our final model is: (1) nested within the model that incorporates the ten best predictors; but (2) yields a substantially improved PRE score over the former model. Thus, independent from the fact that we (only) focus on the ten best predictors in Appendix 1, Table 5 as a second step, we would have ended up with that model with those eight predictors as it scores the highest out-of-sample PRE. Put differently, those covariates that do not contribute to increasing an overall model’s PRE, even if they might have had a positive PRE in Appendix 1, Table 4, would have eventually been eliminated, leading to the final model with the eight predictors.

References

Arreguin-Toft I (2001) How the weak win wars: A theory of asymmetric conflict. *International Security* 26(1): 93–128.

Balcells L and Kalyvas SN (2012a) Endgame in Syria? Foreign policy. Available at: http://mideast.foreignpolicy.com/posts/2012/12/19/endgame_in_syria (accessed 11 June 2014).

Balcells L and Kalyvas SN (2012b) Does warfare matter? Severity, duration, and outcomes of civil wars. Unpublished manuscript.

Balch-Lindsay D and Enterline A (2000) Killing time: The world politics of civil war duration, 1820–1992. *International Studies Quarterly* 44(4): 615–642.

Bennett DS and Stam AC (2006) Predicting the length of the 2003 US–Iraq War. *Foreign Policy Analysis* 2(2): 101–116.

Bennett DS and Stam AC (2009) Revisiting predictions of war duration. *Conflict Management and Peace Science* 26(3): 256–267.

Biddle S, Freidman JA and Long S (2012) Civil war intervention and the problem of Iraq. *International Studies Quarterly* 56(1): 85–98.

Borger J (2012) Arms and the Manpads: Syrian rebels get anti-aircraft missiles. Available at: http://www.guardian.co.uk/world/2012/nov/28/syria-middleeast (accessed 11 June 2014).

Brandt P, Mason D, Mehmet Gurses M, et al. (2008) When and how the fighting stops: Explaining the duration and outcome of civil wars. *Defence and Peace Economics* 19(6): 415–434.

Collier P, Hoeffler A and Söderbom M (2004) On the duration of civil war. *Journal of Peace Research* 41(3): 253–273.

Costner H (1965) Criteria for measures of association. *American Sociological Review* 30(3): 341–353.

Cunningham D (2006) Veto players and civil war duration. *American Journal of Political Science* 50(4): 875–892.

Cunningham D (2010) Blocking resolution: How external states can prolong civil wars. *Journal of Peace Research* 47(2): 115–127.

Enterline AJ, Stull E and Magagnoli J (2013) Reversal of fortune? Strategy change and counterinsurgency success by foreign powers in the twentieth century. *International Studies Perspectives* 14(2): 176–198.

Fearon JD (2004) Why do some civil wars last so much longer than others? *Journal of Peace Research* 41(3): 275–302.

Fearon JD and Laitin D (2003) ethnicity, insurgency, and civil war. *American Political Science Review* 97(1): 75–90.

Gleditsch NP, Wallensteen P, Eriksson M, et al. (2002) Armed conflict 1946–2001: A new dataset. *Journal of Peace Research* 39(5): 615–637.

Grove T (2013) Insight: Russia’s Syria diplomacy, a game of smoke and mirrors. Available at: http://www.reuters.com/article/2013/06/06/us-syria-crisis-russia-insight-idUSBRE95505W20130606 (accessed 11 June 2014).

Hegre H and Sambanis N (2006) Sensitivity analysis of the empirical literature on civil war onset. *Journal of Conflict Resolution* 50(4): 508–535.

Holliday J (2012) *Syria’s Maturing Insurgency*. Middle East Security Report 5 (March 2012). Washington DC: Institute for the Study of War.

International Crisis Group (2012a) *Syria’s mutating conflict. Middle East Report N°128* (August 2012). Brussels: International Crisis Group.

International Crisis Group (2012b) *A precarious balancing act: Lebanon and the Syrian conflict. Middle East Report N°132* (November 2012). Brussels: International Crisis Group.

International Crisis Group (2013) *Syria’s metastasising conflicts. Middle East Report N°143* (June 2013). Brussels: International Crisis Group.

International Crisis Group (2014) *Crisis watch report N°129* (May 2014). Brussels: International Crisis Group.

Johnston PB and Urlacher BR. (2012) Explaining the duration of counterinsurgency campaigns. Unpublished manuscript.

Kalyvas S and Balcells L (2010) International system and technologies of rebellion: How the end of the cold war shaped internal conflict. *American Political Science Review* 104(3): 415–429.

Lyall J (2010) Do democracies make inferior counterinsurgents? Reassessing democracy’s impact on war outcomes and duration. *International Organization* 64(1): 167–192.

Lyall J and Wilson W (2009) Rage against the machines: Explaining outcomes in counterinsurgency wars. *International Organization* 63(1): 67–106.

Mason D and Fett P (1996) How civil wars end: A rational choice approach. *Journal of Conflict Resolution* 40(4): 456–568.

Most BA and Starr H (1989) *Inquiry, Logic and International Politics*. Columbia: University of South Carolina Press.

Peel M and Clover C (2012) *Syria and Russia's "Special Relationship*." Available at: http://www.ft.com/cms/s/0/e64a3076-c9b2–11e1-a5e2–00144feabdc0.html#axzz32q3kMem3 (accessed 11 June 2014).

Price M, Klingner J, Qtiesh A, et al. (2013) *Updated Statistical Analysis of Documentation of Killings in the Syrian Arab Republic*. New York: United Nations Office of the High Commissioner for Human Rights.

Regan P (2002) Third-party interventions and the duration of intrastate conflicts. *Journal of Conflict Resolution* 46(1): 55–73.

Senate Select Intelligence Committee. (2012) Transcript of hearing on worldwide threats. Available at: https://www.hsdl.org/?view&did=699254 (accessed 11 June 2014).

Slantchev B (2004) How initiators end their wars: The duration of warfare and the terms of peace. *American Journal of Political Science* 48(4): 813–829.

Starr H (1978) 'Opportunity' and 'willingness' as ordering concepts in the study of war. *International Interactions* 4(4): 47–67.

Strauss S (2012) Wars do end! Changing patterns of political violence in Sub-Saharan Africa. *African Affairs* 111(443): 179–201.

Walter B (2004) Does conflict beget conflict? Explaining recurring civil war. *Journal of Peace Research* 41(3): 371–388.

Walter B (2012) Why Assad will fight to the death. Available at: http://politicalviolenceataglance.org/2012/06/18/why-assad-will-fight-to-the-death/ (accessed 11 June 2014).

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.

White J and Tabler AJ. (2012) Is the end near in Damascus? Available at: http://www.washingtoninstitute.org/policy-analysis/view/is-the-end-near-in-damascus (accessed 11 June 2014).

Appendix 1

Table 4. Variables assessed in out-of-sample prediction.

Variable group	Variable	Average PRE
Physical terrain and geography	Elevation	−0.0032722 (0.0060865)
	Distance	0.0098763 (0.0075878)
	Mountains	−0.0005474 (0.002349)
	Non-contiguous territory	0.0225718 (0.0081879)
	Land area	0.0035975 (0.003115)
Cultural terrain	Number of languages	0.0001732 (0.0020759)
	Number of languages and number of languages (square)	0.0078821 (0.0083242)
	Number of languages (square)	−0.0007054 (0.0009195)
	Ethnic fractionalization	0.006761 (0.0033684)
	Ethnic fractionalization and ethnic fractionalization (square)	0.0026253 (0.0125996)
	Ethnic fractionalization (square)	0.0031383 (0.0024121)
	Size largest ethnic group	0.003511 (0.0045756)
	Size largest ethnic group and size of second largest ethnic group	0.020937 (0.0085101)
	Size of second largest ethnic group	0.0258608 (0.0091791)
	Religious fractionalization	0.0073836 (0.004578)
	Religious fractionalization and religious fractionalization (square)	0.0038901 (0.0027951)
	Religious fractionalization (square)	0.0067465 (0.0023353)
	Size largest religious group	0.0048344 (0.0037365)
	Size largest religious group and size second largest religious group	0.0009933 (0.0040933)
	Size second largest religious group	0.0009006 (0.0054689)
	Excluded groups	−0.0029415 (0.0056506)
	Excluded population	−0.003388 (0.0058354)
	Strict veto players	−0.0047353 (0.0147295)
	Number of original groups	0.0051407 (0.0010528)
	Number of splinter factions	0.0002879 (0.0002445)

Table 4. (Continued)

Variable group	Variable	Average PRE
	Number of external states	−0.0021709 (0.0021134)
	Neutral military intervention	0.0030321 (0.0041873)
	Any intervention	0.0068944 (0.0032269)
	Excluded groups and population	−0.0037593 (0.0060004)
	Excluded groups, population, and interaction	−0.0037593 (0.0060004)
	Sons of the soil insurgency	0.0100837 (0.0085133)
	Eastern Europe	0.0098187 (0.0032172)
	Dummy Latin America	−0.0010724 (0.002178)
	Dummy Sub-Sahara	−0.0006032 (0.005402)
	Dummy Northern Africa and Middle East	–0.0034668 (0.0039797)
	Dummy Asia	−0.0014174 (0.0102341)
Insurgency power		
	External support	0.01467 (0.0033422)
	Cold war	0.0160783 (0.0126929)
	Contraband	0.0101434 (0.007635)
	Military intervention on insurgents' side	0.0026752 (0.0027975)
State power		
	GDP per capita	0.0082486 (0.0037382)
	GDP per capita (log)	0.006796 (0.0021959)
	Population	−0.00064 (0.0013317)
	Population (log)	0.0078632 (0.0078393)
	Regime type (polity IV)	−0.0008878 (0.0039057)
	Regime type (polity IV) and squared regime type (polity IV)	0.0085294 (0.0094659)
	Squared regime type (polity IV)	0.0149086 (0.0040673)
	Democracy dummy	0.0111061 (0.00769)
	Anocracy dummy	0.0180738 (0.0108059)
	Democracy dummy and anocracy dummy	0.0142623 (0.0085948)
	Instability	0.0030432 (0.0040705)
	Oil revenue	0.0017107 (0.0036406)
	Military personnel	0.0013935 (0.0014662)
	Military personnel (log)	0.0052514 (0.0028808)
	Force to population ratio	0.0000663 (0.0000914)
	Force to space ratio	0.0007514 (0.0015806)
	Rotary-wing assets	0.0390018 (0.0057526)
	Mechanization index	−0.0027025 (0.0046273)
	Mechanized vehicles	0.0013849 (0.0024433)
	State military-industrial-demographic power index	−0.0011649 (0.0029505)
	State military-industrial-demographic power (log)	0.0094872 (0.0069025)
	Energy per capita	−0.0020641 (0.0071142)
	Energy per capita (log)	0.0016802 (0.0024093)
	Trade over GDP	−0.0034268 (0.010102)
	Trade over GDP (log)	−0.0025337 (0.0058002)
	Strategy ordered	0.0650261 (0.0109498)
	Conventional insurgency	0.067737 (0.0100478)
	Military intervention on government Side	0.0058166 (0.0048748)
	Economic intervention on government side	0.0353493 (0.003695)

SD in parentheses. Variables have been derived from Regan (2002), Fearon and Laitin (2003), Fearon (2004), Collier et al. (2004), Cunningham (2006, 2010), Hegre and Sambanis (2006), Lyall and Wilson (2009), and Lyall (2010). The detailed definitions of the variables can be found in these sources, as we only describe our final model's variables in detail in the main text. 'Top-10' variables are shaded in grey. Average PRE values and their standard deviations are based on the 50 different PRE values we obtained for each variable due to the five-fold cross-validation exercise – we first calculated the mean for the 50 data points (average PRE) and then computed the difference of each data point from the mean before we squared the result of each. Finally we calculated the average of these values and then took the square root in order to obtain the SD. The table's estimates can be replicated with the replication files.

Table 5. Dropping 'Top-10' variables successively.

Variable	Average PRE
Non-contiguous territory (5. Rank)	0.1999705 (0.0209535)
Size of second largest ethnic group (1. Rank)	0.1723613 (0.0234619)
Sons of the soil insurgency (10. Rank)	0.2331758 (0.0320679)
External support (7. Rank)	0.2076225 (0.0162367)
Cold war (3. Rank)	0.1941528 (0.0293113)
Contraband (8. Rank)	0.2111903 (0.0201458)
Anocracy dummy (9. Rank)	0.2320173 (0.0250284)
Rotary-wing assets (4. Rank)	0.1958788 (0.0224344)
Conventional insurgency (6. Rank)	0.20015 (0.0201423)
Economic intervention on government side (2. Rank)	0.189485 (0.0345911)

SD in parentheses. Average PRE values and their standard deviations are based on the 50 different PRE values we obtained for each variable due to the five-fold cross-validation exercise, i.e. we first calculated the mean for the 50 data points (average PRE) and then computed the difference of each data point from the mean before we squared the result of each. Finally we calculated the average of these values and then took the square root in order to obtain the SD. The table's estimates can be replicated with the replication files.

Table 6. Progressive elimination of variables and PRE assessment.

Specification (variables dropped)	Average PRE
Full 'top-10' model – no variables dropped	0.2195739 (0.0272849)
Sons of the soil insurgency, anocracy dummy	0.2369405 (0.0305187)
Sons of the soil insurgency, anocracy dummy, contraband	0.2262971 (0.0203109)
Sons of the soil insurgency, anocracy dummy, contraband, external support	0.2129432 (0.0263627)
Sons of the soil insurgency, anocracy dummy, contraband, external support, conventional insurgency	0.1593038 (0.0188108)
Sons of the soil insurgency, anocracy dummy, contraband, external support, conventional insurgency, non-contiguous territory	0.1327485 (0.0164156)
Sons of the soil insurgency, anocracy dummy, contraband, external support, conventional insurgency, non-contiguous territory, rotary-wing assets	0.0834405 (0.013856)
Sons of the soil insurgency, anocracy dummy, contraband, external support, conventional insurgency, non-contiguous territory, rotary-wing assets, cold war	0.0541267 (0.0086808)

SD in parentheses. Average PRE values and their standard deviations are based on the 50 different PRE values we obtained for each variable due to the five-fold cross-validation exercise, i.e. we first calculated the mean for the 50 data points (average PRE) and then computed the difference of each data point from the mean before we squared the result of each. Finally we calculated the average of these values and then took the square root in order to obtain the SD. The table's estimates can be replicated with the replication files.