



Irregular leadership changes in 2014: Forecasts using ensemble, split-population duration models



Andreas Beger, Cassy L. Dorff, Michael D. Ward*

Department of Political Science, Duke University, Durham, NC 27701, USA

ARTICLE INFO

Keywords:

Government forecasting
Political Instability Task Force

ABSTRACT

We forecast Irregular Leadership Changes (ILCs) – unexpected leadership changes in contravention of a state's established laws and conventions – for six months in mid-2014 using predictions from an ensemble of seven split-population duration regression models. The original forecasts were made in May 2014. Our approach allows us to aggregate models for different mechanisms leading to ILCs in one ensemble forecast, is sensitive to the overwhelming number of non-events (zeros) in the data, and allows us to make real-world forecasts with a lag of approximately five weeks. The data are based on 45 ILCs recorded for the period from March 2001 to March 2014, with monthly observations for up to 168 countries worldwide. The ensemble achieves in- and out-of-sample AUCs of ~ 0.85 , and we present the 10 highest forecasts, which include Thailand.

© 2015 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

In late February 2014, pro-Russian President Viktor Yanukovich of Ukraine fled the capital after mass protests erupted into violence, and the parliament appointed an interim President to rule until the elections in May. The mass protests had originally broken out in November 2013 over Yanukovich's abandonment of an agreement for closer trade ties with the EU. A month earlier, in the Central African Republic, Muslim President Michel Djotodia was forced out of office in January 2014 in the face of escalating violence between the Muslim Séléka regime and the largely Christian anti-balaka coalition. The level of violence verged on genocide. Djotodia and the Séléka coalition had themselves won power in March 2013 through a successful rebellion against the preceding government. Last, in July 2013, the Egyptian military staged a coup and removed democratically-elected Mohammed Morsi from the Presidency, following waves of mass protest against the Muslim

Brotherhood's reign. These three events – a mass protest campaign, a rebellion, and a coup d'état – are treated as different types of events by most of the political science literature.

All of these cases share the same outcome. We propose to call this outcome 'irregular leadership change' (ILC): the unexpected removal of a political leader through means that contravene a state's conventions and laws.

ILCs encompass *successful* coups, mass protests, and rebellions against the central government – mechanisms that are typically associated with violence and upheaval – which can have a significant impact on foreign policy considerations. A major contribution of this article is to shift the focus of inquiry from viewing ILCs as being broken into distinct categories, to analyzing irregular regime change as a general and politically relevant phenomenon.

The goal of this project is to develop a model that can forecast these ILCs at a *monthly* level. It is based on work commissioned by the Political Instability Task Force (PITF), a panel of scholars, methodologists, and practitioners that was formed in 1994 at the request of senior policymakers in the United States Government to assess and explain

* Corresponding author.

E-mail address: michael.d.ward@duke.edu (M.D. Ward).

the vulnerability of states to political instability and state failure. It has since expanded this original task to include research and the provision of advice on general instability, revolutionary and ethnic war, adverse regime changes, genocide, and other events of interest, and provides a forum for exchange between academics, private researchers, and members of the government who support policymaking.

Forecasting ILCs is of interest because they are unexpected, are often associated with violence and political and economic upheaval, and, crucially, can affect foreign policy drastically. Some ILCs are brought about by mechanisms that are inherently violent, e.g., armed rebellions such as the overthrow of Mobutu Sese Seko through the First Congo War. However, violence can also occur as a result of these shifts. For example, although many coups are relatively bloodless, several have sparked significant levels of violence and even civil war (Powell & Thyne, 2011, p. 256), and one of the initial events in the Rwandan Genocide was a military coup following the assassination of Juvénal Habyarimana. The successful mass protests in Ukraine in 2014 led to the annexation of the Krim peninsula by Russia and lingering civil war in the Donbass region.

In addition to the human toll, the political instability associated with ILCs also has negative economic effects (Alesina, Özler, Roubini, & Swagel, 1996). Economic growth is also depressed after coups (Ulfelder, 2013), as well as both during and following civil war (Collier, 1999; Kang & Meernik, 2005).

From a foreign policy perspective, if political leaders play a role in shaping a country's foreign policy, then being able to anticipate an unexpected removal should also be of interest. As an extreme example of this, we might consider the change in relations between Iran and the US after mass protests ousted the Shah. In this sense, the fact that the Shah was ousted was more important than the means – successful mass protests – by which he was ousted.

Critics may question our aggregation of ILCs as combining events with different causal mechanisms, but a literature review shows that coups, revolutions, and rebellions are already treated as multi-causal phenomena and that there is no general model accounting for all of them. Rather than further disaggregate these rare phenomena, we suggest that instead such causal and model heterogeneity be acknowledged and accommodated using an ensemble of thematic models.

2. What are irregular leadership changes?

Irregular leadership changes are transitions between political leaders that occur outside the established rules and conventions of a state, or that follow established conventions procedurally, but under clear outside coercion, e.g., under pressure from protesters or the military. Both our concept and our operationalization draw directly on the Archigos dataset of political leaders (Goemans, Gleditsch, & Chiozza, 2009): ILCs are the composite of “irregular” entry to and exit from office, thus encompassing events where leaders lose power irregularly, a successor gains power irregularly, or both the exit of the previous leader and the entry of the subsequent leader are irregular.

What constitutes the established rules and conventions of a state is variable, and can be established through written, legal means governing succession and accepted convention. It is also established by unwritten but reasonable expectations for leadership transitions. Since the standard for regularity is specific to a regime, a given kind of transition can be regular in one regime and irregular in another. A father-son transition in a monarchy is regular, while it would be irregular in a democracy without an intervening election. Thus, a broad range of transitions can be considered regular in autocratic regimes, e.g., hereditary, election or support by a narrow body as in communist dictatorships, or designation by the preceding rule, depending on the convention of the particular autocratic regime.

For the few ILCs in established democracies, the coercion litmus test is important. Much as the “legal” resignation of a leader at gunpoint by the military does not invalidate a coup (e.g., Sukarno in Indonesia), the resignation of a leader in a democracy in the face of mass protests, without intervening elections or the loss of a parliamentary majority, is irregular, despite following legal procedure. This preserves the element of surprise or unexpectedness that is inherent in ILCs, a factor that contributes to the utility of attempting to forecast them.

The two components that make up ILCs, irregular leader entry and exit, are operationalized and measured by the Archigos dataset on political leaders, which codes the nature of the entry/exit of the “effective” leaders of all countries in the Gleditsch and Ward (1999) state list from 1875 onward (Goemans et al., 2009).¹

Fig. 1 maps all 45 irregular leadership changes that occurred between March 2001 and March 2014, the time period of our study.² They mostly occur in Africa and south and central Asia, with some outlier cases in Europe and Latin America.

ILC encompasses several phenomena that have been studied separately in political science: coups, protests, and armed rebellions. Conceptually, we can think of all of these phenomena as mechanisms that lead to ILC, and empirically they account for the majority of observed ILCs.³ To illustrate the relationship, we discuss three representative cases below.

First, take the case of General Amadou Toumani Touré, the Malian President from 2002 to 2012. In March of 2012, the Malian military, displeased with the government's response to the Tuareg rebellion, took over the presidential palace in the capital city of Bamako. In doing so, the leader of the coup, Amadou Sanogo, and his military collaborators successfully forced the government of Amadou Toumani Touré into hiding. This case qualifies as an ILC, with an irregular exit for Touré and an irregular entry for Sanogo at the same time, where both occurred as a result of internal government conflict and military pressure. This

¹ See the supplementary materials for more information on how Archigos defines irregular entries/exits (see Appendix B).

² The supplementary materials contain a table of all ILCs (see Appendix B).

³ There are also a few residual events that are harder to classify, e.g., non-military coups and independent assassinations that are not part of a coup.

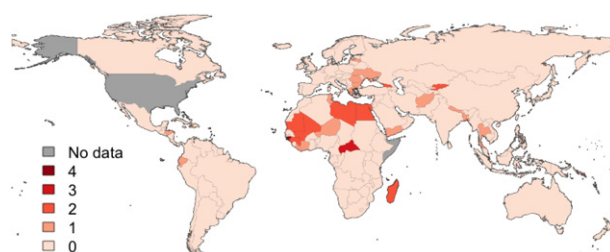


Fig. 1. Archigos irregular leadership exits, March 2001–March 2014. Note that some countries had multiple irregular turnovers.

kind of classic military coup represents 35% of the irregular leadership changes.

Looking at the overlap with coups in more detail, there are 17 coups during the period from March 2001 to December 2013 (Powell & Thyne, 2011), compared with 45 irregular leadership exits. Of those 17 coups, 16 are also coded as irregular leadership changes in our data.

The second example is Romania, where Emil Boc became Prime Minister after the 2008 legislative elections. The government became increasingly unpopular amid corruption issues, economic problems, and austerity policies in the wake of the 2008 financial crisis. He lost a vote of no confidence in October 2009, but was reinstated following a narrow victory by President Băsescu. Changes to health care laws and the dismissal of a critical health minister finally led to open protests against the government, and in early 2012 Boc resigned “to preserve the stability of the country”.⁴ This is an example of the coercion criterion for ILC: although Boc’ resignation was legal, he did so under clear pressure from protesters, and still had confidence in parliament. Irregular leadership changes as a result of mass protests represent a further 38% of irregular exit cases.

The final example is Laurent Gbagbo’s irregular exit in 2011. Gbagbo was president of Côte d’Ivoire from 2000 onward, after winning an election and after street protests forced his reluctant predecessor to recognize the results and leave office. While he was originally elected for a five year term, civil war led to repeated postponements of new presidential elections. When the election finally took place in 2010, Gbagbo lost but refused to leave office. Fighting with the opposition forces broke out, and in April 2011 he was deposed and ultimately arrested by rebels, with some participation by French military forces. Successful armed rebellions, along with coups conducted by non-military actors, make up most of the remaining ILCs. As these cases show, a wide variety of different events can lead to an ILC.

3. Modeling considerations

Although our concept of ILC is new, the literature on the types of ILCs mentioned above is extensive. The work that is relevant for the modeling of ILC includes that on coups, mass protests and protest waves, and civil war, which includes rebellions against the central state. These bodies of work have assembled a large number of empirically-supported arguments and knowledge about the phenomena they study. One of the insights we have adopted in our

modeling is that ILCs are a multi-causal phenomenon, and no single model is likely to capture all contributing forces. However, there are fundamental problems with using the existing models and arguments for the purpose of *ex ante* forecasting of ILCs at the monthly level. We first review and elaborate on these issues, then present the thematic models that we used as the basis for our forecasts.

3.1. Forecasting versus explanatory modeling

The main metric of interest in forecasting is accuracy, and hence, models’ out-of-sample predictions. The political science research on events related to ILCs has focused on the use of explanatory modeling with null hypothesis significance testing to determine whether a theory or argument is supported empirically or not. The ability of a model to predict is rarely evaluated directly; instead, the focus is on discovering statistically significant relationships (Gerber & Malhotra, 2008). However, *p*-values are a poor indicator of a variable’s importance for prediction, and models with large numbers of statistically significant covariates are not guaranteed to predict well (Ward, Greenhill, & Bakke, 2010). A second problem is that when the model fit is evaluated, often it is only evaluated in-sample using the same data that were used originally to estimate the parameters (e.g., Jackman, 1978). Models have a tendency to overfit the peculiar idiosyncrasies of a given data set, and usually predict less well with new data (Beck, King, & Zeng, 2000). Overfitting is usually overcome by an out-of-sample evaluation of the predictions; however, this is rarely done in published work.

There is perhaps no reason to expect research that is focused on theory-building and assessment to evaluate prediction, but this leaves us with little guidance as to which existing models are useful for prediction and forecasting. Where existing arguments and models have been evaluated in retrospect, the conclusions are not positive. Ward et al. (2010) evaluate two highly-cited civil war papers and find that the models lack predictive power, as a result of the model specification being driven by covariate significance values. Similarly, Hill and Jones (2014) evaluate a broad range of factors that were thought to be important for explaining state repression, and find that many of them are not important predictors; conversely, they also identify several factors which do improve prediction but have not been studied well in existing work. In summary, there is little reason to expect that existing models that are designed to produce statistically significant relationships will predict well (Schrodt, 2014; Ward et al., 2010).

⁴ <http://goo.gl/znniYh>.

A compounding problem in our setting is that much of the extant quantitative work focuses conceptually on the mid- to long-term (years) and is based empirically on country-years, while we aim to predict sub-annually. The level of analysis problem is reflected in key concepts that change slowly, over years, such as state strength (Fearon & Laitin, 2003), grievances due to horizontal inequalities between ethnic groups (Cederman, Weidmann, & Gleditsch, 2011), or the counterbalancing of security forces (Powell, 2012; Quinlivan, 1999). As we will show below, such structural factors are useful for distinguishing country risk, but the question of what determines the timing of ILCs at monthly time-scales has not been explored thoroughly in quantitative studies. This is not due to a lack of theorizing or triggering factors, but rather is driven by a lack of data for indicators. For example, Belkin and Schofer (2003) review a range of explanations for coups, explicitly distinguish background from triggering factors, and find that half of the 7–12 full or partial triggering factors cannot plausibly be measured with a large N (we do incorporate several of the measurable factors).

Event data consisting of machine-coded news reports offer a potential solution. Such data have global coverage and near-instantaneous coding, but their use in constructing explanatory indicators is under-explored compared to examinations of their use as a dependent variable (Hendrix & Salehyan, 2012; Raleigh, Linke, Hegre, & Karlsen, 2010) or of their biases and accuracy limits (e.g. Schrodt & Gerner, 1994; Weidmann, 2014).

Another problem arises from the pragmatic consideration that, in order to provide global *ex ante* forecasts, we need variables that are updated regularly and are available with only a short lag. For a variety of reasons, the data in published studies typically end several years prior to the publication of the study. For example, arguments about coup-proofing and counterbalancing are commonly operationalized using either military personnel and spending data from the Correlates of War's National Material Capability dataset, which currently ends in 2007;⁵ or the number and relative sizes of military and paramilitary organizations, for which published data end in 1999 (Powell, 2012).⁶

Without solid knowledge as to which of the existing models of coups, rebellions, and mass protests predict well, or at least which factors are important for prediction, there is little basis on which to choose the best covariates to update for forecasting efforts. Even if this problem could be solved, the models would still have to be altered to incorporate indicators that can be measured at the monthly level. Relying on existing explanatory models without any prior evaluation of their predictions is a difficult and risky strategy for forecasting.

3.2. Multiple mechanisms

Despite the difficulty in adapting existing models directly, the literature provides important insights. Chief

among these is the fact that there are multiple, valid drivers behind each of the majority of ILCs.

This is seen most clearly in the literature on civil war, which is arguably the largest of the three bodies of literature. Civil wars include successful armed rebellions that lead to ILC, and although the focus in the literature has been on the onset, a small amount of work on the termination of civil war has included the direct modeling of rebel victories (e.g. De Rouen & Sobek, 2004). There are at least three factors that influence the occurrence of civil war. The earliest was the grievance-based approach (Gurr, 1970, 1993; Gurr & Moore, 1997), which emphasizes political, economic, and other grievances that might lead to a mobilization for rebellion. This argument is often tied to ethnicity, especially during the wave of ethnic conflicts in the early 1990s, and has had somewhat of a comeback more recently with a more careful analysis of the spatial inequalities among ethnic groups within a country (Cederman et al., 2011). Grievances were later supplanted, in two canonical articles, by arguments stressing the importance of opportunities for rebellion through state weakness and a terrain favorable to insurgency (Fearon & Laitin, 2003), and economic opportunities due to primary commodities like oil and diamonds that could be looted (Collier & Hoeffler, 2004; Lujala, Gleditsch, & Gilmore, 2005).

Similarly, early work in the coup literature focused on grievances within the military, but also considered grievances within the civilian population (e.g. Huntington, 1968; Jackman, 1978; Johnson, Slater, & McGowan, 1984). More recent work has replaced this with the coup analogue of opportunities for rebellion, by studying coup-proofing, the strategies used by rulers to prevent military coups. The main strategies focus on “spoiling” the military with funding and arms purchases, and counterbalancing the military through the creation of multiple competing military and paramilitary organizations (Belkin & Schofer, 2003; Powell, 2012). Interestingly, while counterbalancing reduces the risk of a coup, it also tends to result in a decrease of military effectiveness during instigated wars (Pilster & Böhmelt, 2011). By the same logic, this process should also increase the susceptibility of the regime to armed rebellion. For the larger category of ILCs, coup-proofing can therefore have two opposing effects via the probabilities of coups and armed rebellions, respectively.

Mass protests have received less attention as a general phenomenon, but the protests that led to the fall of Communism in 1989 and the Arab Spring in 2011 have generated a large body of research that tries to explain the apparent suddenness and surprising spread of “protest waves”. At the core of these explanations is the idea that information/protest cascades have tipping points. Under certain conditions, a small initial group of protesters can surpass a threshold beyond which more and more people update their beliefs enough to join the protests themselves, ultimately creating a self-fulfilling prophecy as state institutions are simply overwhelmed (Kuran, 1991).

Inherent in the research on protest waves has been the idea of contagion: unrest in one country may spread to other countries, as it did from Tunisia to the neighboring Arab countries in early 2011. For protest movements, this might occur similarly to the way in which protests spread

⁵ <http://www.correlatesofwar.org/COW2%20Data/Capabilities/nmc4.htm#Overview>.

⁶ Although updates appear to be available up to 2011; see <http://www.jonathanmpowell.com/determinants.html>.

within countries, namely by convincing a sufficient number of people that a revolution can be successful, although other mechanisms have also been proposed (see [Hale, 2013](#)). Contagion has also been identified as a factor in the spread of civil war ([Salehyan & Gleditsch, 2006](#)), and for coups as well ([Belkin & Schofer, 2003](#)).

Notably, contagion does not have to occur strictly through the same type of events. Instead, what seems to be important is the underlying instability, whether it takes the form of protests or armed fighting. The Arab Spring led to widespread protests in all Arab countries, but by 2014, the actual outcomes include overthrown governments (Tunisia, Egypt), civil war (Libya, Syria), a coup (Egypt), another coup in Mali which has been tied to the Libyan conflict, and several movements that faded out with varying degrees of success short of regime change (Jordan, Algeria, Morocco, Kuwait, Iraq, Oman).

At least implicitly, these three bodies of literature also share a common understanding that the key actors include the military, the government, and the larger civilian population. For example, while protest research has focused on how protesters can overcome collective action problems, [Barany \(2011\)](#) notes that, in autocracies faced by mass protests, the military always gets the order to shoot eventually, which it of course often fails to obey. Thus, we can observe variation in outcomes; for example, Tunisia, where the military sided with protesters, Romania in 1989, where the military turned on the secret police, Syria, where it fractured, or Qatar in 2011 and China in 1989, where it ruthlessly suppressed protests.

This presence of a common set of actors provides a basis for the future development of a joint understanding of ILCs. Selectorate theory ([Bueno de Mesquita, Smith, Siverson, & Morrow, 2005](#)) goes some way towards this. It is based on the notions of a selectorate – those in a society who have a say in choosing leadership – and the winning coalition within this group of those who are needed to keep the current leadership in power, in return for public or private goods. The relative ratios of the sizes of the selectorate and the winning coalition to the larger society determine the incentives faced by actors in each group, thereby determining whether the provision of public or private goods is more efficient. In its basic formulation, it does not consider the role of the military explicitly, and, as with the other arguments, it focuses more on structural factors than proximate causes. From a practical perspective, these three key actors and the ways in which they are described in the other related literature suggest the importance of intra-government cohesion and the nature of interactions with protesters as a potentially important predictor.

3.3. Our approach

Although we cannot use existing models directly, we build on the extant literature in three ways. First, rather than settle on one model, we explore several different variants, using an ensemble method to determine the contribution of each to a combined forecast. Second, we account for the distinction between structural risks and proximate triggering factors explicitly using a split-population duration model. Third, when developing these

models, we take several theoretical themes from the literature as initial starting points. To do this, we identify variables that are relevant for a given theme, then choose our final model specification by selecting variables based on whether they improved in- and out-of-sample prediction.

The most straightforward adaption in our models was our attempt to replicate the work of [Goldstone et al. \(2010\)](#), who present an incarnation of a PITF forecasting model that focuses on global instability. Two of the other themes focus on similar outcomes related to ILC: protests and internal conflict. The remaining four are more functional in nature, in the sense that they start with themes that may be common to all types of ILCs: leader characteristics, public discontent (indicating intra-governmental friction), contagion from a nearby conflict, and financial risk. We will discuss these thematic models in some detail in the thematic section. For now, suffice it to say that these models will not satisfy those who are searching for a theory of ILCs. However, that is not our goal, which is prediction.

4. Research design

The considerations that affect the research design for forecasting are different from those in a typical project that is aimed at causal modeling. In the next few sections, we review the key elements of our modeling effort: event-data based indicators from ICEWS, split-population duration regression, EBMA for constructing the final forecast, and partitioning of the data for calibrating the ensemble and test predictions out-of-sample.

4.1. Data

Our data consist of monthly observations for 164 countries from March 2001 to March 2014, with several hundred variables on structural characteristics like wealth and population from sources including the World Bank, Polity ([Marshall & Gurr, 2014](#)), the Political Terror Scale ([Wood & Gibney, 2010](#)), and Ethnic Power Relations ([Wimmer, Cederman, & Min, 2009](#)), behavioral variables based on the ICEWS⁷ event data, and spatial lags of the behavioral variables. Most of the variables were collated for the forecasting component of the ICEWS project, and, due to the large number of variables, we will describe them only superficially here. The supplementary information provides details of the variables included in the models (see [Appendix B](#)). We rely on the ICEWS data because they are updated continuously, with imputation where necessary, and are available with a lag of only a few weeks. The dependent variable is a binary indicator of ILC in a given country-month, constructed using the Archigos dataset's coding for the irregular entry and exit of political leaders.

The behavioral variables are constructed from the ICEWS event data, and record the numbers of certain types of events in a country over the course of a month, e.g., protests directed against the government. The ICEWS

⁷ Integrated Crisis Early Warning System, an operational framework for forecasting several events of interest, funded by the Office of Naval Research.

event data⁸ are based on (machine) coded media reports which are parsed for actors, locations, and actions to create event records, using the CAMEO ontology. We include aggregations of the ICEWS event data, and in particular, so-called quad variables, which capture verbal and material conflict and cooperation within the government and between the government and dissidents. For example, verbal cooperation includes making positive public statements, appeals, or consultations, while verbal conflict captures reports of investigations, public demands, or threats.

The third set of variables includes spatial lags of the behavioral, event-based variables. Spatial lags capture neighborhood effects, e.g., the average level of protests in Egypt's neighbors at the time of the uprisings (Ward & Gleditsch, 2008). There are various different ways to define what constitutes a country's neighborhood, and we include weights constructed on the basis of the four nearest neighboring countries, the distance between country centroids, and finally, Gower distances (Gower, 1971) of one country's similarity to others based on political, economic, and event measures.

4.2. Variable variance and choice of base models

We noted earlier that existing models are difficult to use as a basis for sub-annual forecasting, partly because they rely largely on structural indicators.

Fig. 2 illustrates this point by plotting the ratio of the variance between and within countries to the total variation for the pool of covariates in our data. Variables above the centerline vary more between than within countries.

Variance in the indicators included in a model is a necessary prerequisite for potential predictive importance, although it does not guarantee it. Part of the reason why we cannot rely on structural indicators for monthly prediction is because they are largely measured annually. Beyond the overall amount of variation, the structure of the variance in cross-sectional data is equally important. The kinds of structural variables that form the core of many existing models are shown in red in the plot, and we have labeled a few important ones, such as the Polity Democracy score and GDP per capita. The variation in the structural variables is predominantly between countries, and so while they may be good for distinguishing general risk between countries, they cannot serve as a basis for predicting timing.

Other structural indicators, such as CPI, FDI, tourism receipts, and exchange rates, are outliers that show both a higher overall variation and a higher variation within countries, but there has been little work relating them to the outcomes that we are interested in. The other three groups of variables in our data, behavior indicators constructed from event data, and their distance or Gower distance-based spatial lags, all display similar levels of total variation, but with a much larger fraction within countries, making them useful candidates for predicting timing. This is the reason why we had to develop thematic models that expand from structural characteristics to less well-explored event data aggregations and spatial lags.

4.3. Split-population duration regression

A well-known obstacle when modeling many events of interest in international relations, like war, civil war, and ILCs, is that they are rare events. Country-year observations are the standard in this field, and at this level we have a rate of 18 positives per thousand. The move to country-months reduces the positive rate to 1.7 per thousand. Increasing the number of positives is another, more pragmatic reason for jointly considering coups, revolutions, and rebellions as ILCs, but the resulting rates still reflect an imbalance.

While regular logistic or probit regressions remain typical with conflict data, there have been attempts to develop better suited models (rare events logit; King & Zeng, 2001), or to use case control methods to reduce the class imbalance (Goldstone et al., 2010). An important realization that motivates our choice of model is that, for practical purposes, many polities will not experience an ILC, e.g., Germany, Switzerland, or Canada. Split-population duration models accommodate the fact that many countries are simply not at risk of ILC by estimating the split separately for at risk and not at risk cases, and conditioning the model of duration to the next ILC according to these estimates. This accommodates the excess number of spurious zeros, while acknowledging that there are gray cases in which we can only guess with uncertainty.

Duration models were developed initially in a health context for examining the survival of medical patients, and split-population duration models are useful, for example, for modeling the time to relapse for cancer patients who have gone into remission, where an unknown number have been cured, while in others the cancer is merely below detectable levels. Similarly, in our case, some countries are effectively "cured" of ILC at a given moment, while others are susceptible. By pooling all cases regardless of risk, we would be watering down or understating the hazard of ILC for those countries that are at risk, due to the inclusion of cases that will never fail. Of course, the problem is that we do not know beforehand which countries are at risk at any given time and which are not. For extreme outliers like Canada or Egypt, it is easy to tell, but border cases are harder to distinguish. Taking into account characteristics that are likely to be associated with risk, the split-duration approach allows us to estimate the probabilities of belonging in the risk set and to condition the estimated hazard of ILC – the probability of an event in a country in a given month – on these.

The likelihood function for a split-population regression has been worked out completely, and reflects a mixture of two equations: a first part classifies risk and immunity, while a second part models the expected duration to failure. The likelihood is given as a product of the immunity π and the risk, where δ_i indicates whether a spell ended in failure:

$$\mathcal{L}(\theta | (t_1, \dots, t_n)) = \prod_{i=1}^N \{(1 - \pi)f(t_i)\}^{\delta_i} \times \{\pi + (1 - \pi)S(t_i)\}^{1-\delta_i}.$$

We focus on two quantities generated by these models, the conditional risk and the conditional hazard. The conditional risk estimates the probability that a country is in the

⁸ <http://www.lockheedmartin.com/us/products/W-ICEWS/iData.html>.

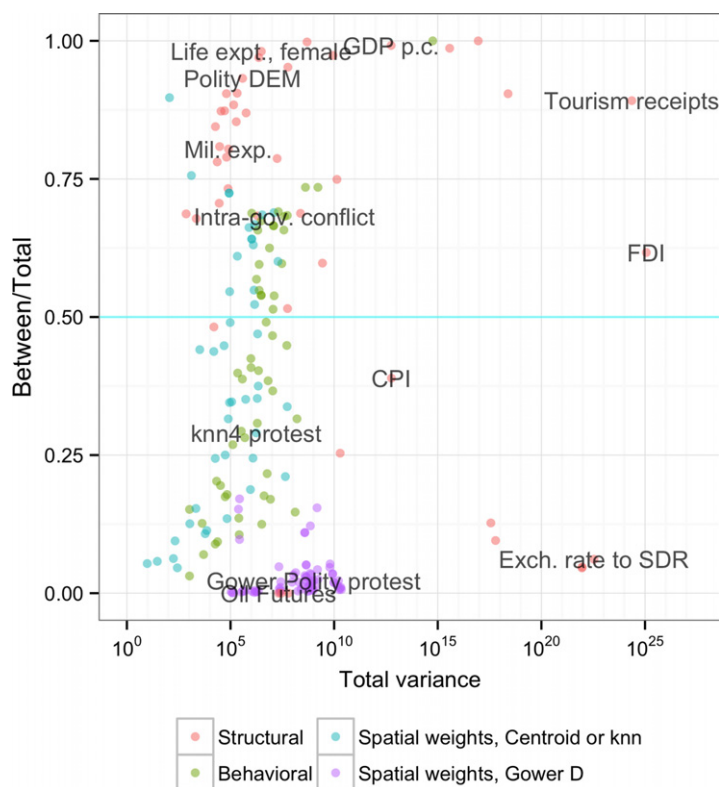


Fig. 2. Ratio of between-country to within-country variance against the total variance for the full set of covariates. A few outliers and commonly used variables are highlighted; for example, note that there are several structural variables like FDI that are rarely used for conflict modeling but have high variances. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

risk set (i.e., not cured) at a given time, given the covariates and how much time has passed since the last ILC. For ease of interpretation, we restricted the covariates in this part of the model to variables that move relatively slowly over time, defined as those in which the variance between countries exceeds that within countries. As a result, the series of risk estimates for a country over the months in our observation period tend to be relatively stable.

The second and main quantity of interest is the conditional hazard. This gives the probability that an ILC will occur in a country in a given month, given the covariates in this part of the model, the time since the last ILC, and the risk estimate for that month. In other words, it is the model's best estimate for ILC in the current country-month, and is analogous to the probabilities produced by logit models. We restricted the variables in this part of the equation to fast-moving indicators, i.e., those that largely change over time within a given country, rather than between countries.

Fig. 3 illustrates the ensemble and thematic model predictions for Thailand over our observed data and forecasting period. We will go through the approach for creating the ensemble below, but for now, note that the risk estimates in the bottom panel are relatively stable over time, while the conditional hazard estimates in the top plot are more variable.

Like all duration models, the split-population duration regression is sensitive to left-censoring, where, for example, a country with an ILC in 2001–02 and another country

with no ILCs for the last 50 years would appear to have the same initial duration counter (time since last event). To address this, we use Archigos data back to 1955 when creating the duration variables required for these models, even though the rest of our data are not included until 2001. Twenty-five countries are still left-censored, but with initial duration values of 555 months not zero.⁹

4.4. Ensemble model averaging

By design, ILCs are multi-causal. Without a general theory to build upon, there are multiple angles from which the problem of prediction could be approached. Even for the three outcomes encompassed by ILCs, there are multiple explanations or factors which the literature argues to be important, and which may explain a particular facet or subset of the event they study. However, at the same time, there is little knowledge of those factors that are most relevant for prediction, since it has not been undertaken previously. The final piece in our design uses ensemble Bayesian model averaging (EBMA) to accommodate the heterogeneity in our dependent variable and the range of plausible models that may predict it.

The concept of ensemble forecasting builds on the basic notion that combining multiple points of view will

⁹ See Section A.5 in the full report available at <http://arxiv.org/abs/1409.7105>.

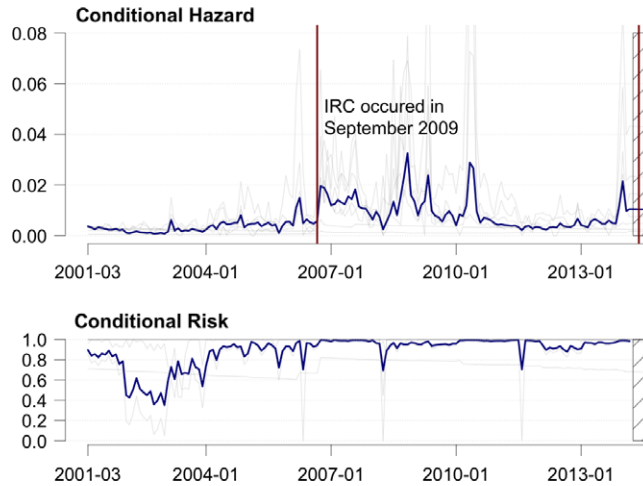


Fig. 3. Conditional hazard and risk evolution in Thailand. Blue indicates the ensemble, constructed from the thematic models shown in grey. The gray box denotes the forecast period. (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lead to a more accurate picture of reality (c.f. [Surowiecki, 2004](#)). Among the more famous demonstrations of this phenomenon was a competition to guess the weight of an ox at the West of England Fat Stock and Poultry Exhibition. [Galton \(1907\)](#) demonstrated that individual entrants were highly inaccurate, yet a simple aggregation of them led to a remarkably accurate estimate.¹⁰ This principle has come to be known as the wisdom of crowds.

In recent years, the advantages of ensembles have meant that they have come to play a particularly prominent role in the machine-learning and nonparametric statistics community ([Hastie, Tibshirani, & Friedman, 2009](#)). A wide range of approaches, including neural nets, additive regression trees, and K -nearest neighbors, fall under the general umbrella of ensemble approaches. Of particular relevance is the success of boosting ([Freund & Schapire, 1997](#); [Friedman, 2001](#)), bagging ([Breiman, 1996](#)), random forests ([Breiman, 2001](#)), and related techniques ([Chipman, George, & McCulloch, 2010](#)) for aggregating so-called “weak learners”. These approaches to classification and prediction have been advertised as the “best off-the-shelf classifier[s] in the world”, and are equally powerful for prediction tasks.

While the advantages of collating information from multiple sources are manifold, it is nevertheless incorrect to assume that more is always better. Not all guesses are equally informative, and naïve approaches to collating forecasts risk overvaluing wild guesses and undervaluing unusual forecasts that are sometimes correct nonetheless. The particular ensemble method that we are using is ensemble Bayesian model averaging (EBMA). First proposed by [Raftery, Gneiting, Balabdaoui, and Polakowski \(2005\)](#), EBMA pools forecasts as a weighted combination of predictive probability distributions. Rather than selecting an individual “best model”, EBMA collects *all* of the insights from multiple forecasting efforts in a principled manner via statistical post-processing. The weight assigned to

each component forecast reflects both its past predictive accuracy and its uniqueness (i.e., the degree to which it makes predictions that differ from those of other component models). Rather than finding the best model, EBMA finds the combination of models that provides the best overall predictions of some quantity of interest.

Assume that the researcher is interested in predicting event $\mathbf{y}^{t*}$ for some future time period $t^* \in T^*$, which we refer to below as the test period. In addition, we have a number of different out-of-sample forecasts for similar events $\mathbf{y}^t$ in some past period $t \in T$, which we term the calibration period. The different predictions were generated from K forecasting models or teams, $M_1, M_2, \dots, M_K$. These predictions might originate from the insights and intuitions of individual subject-experts, traditional statistical models, non-linear classification trees, neural networks, agent-based models, or anything in between. In our case, they are the seven split-population duration thematic models.

For each forecast, there is a prior probability distribution $M_k \sim \pi(M_k)$, and the PDF for $\mathbf{y}^t$ is denoted $p(\mathbf{y}^t|M_k)$. The predictive PDF for the quantity of interest is $p(\mathbf{y}^{t*}|M_k)$, the conditional probability for each model is given as

$$p(M_k|\mathbf{y}^t) = p(\mathbf{y}^t|M_k)\pi(M_k) / \sum_{k=1}^K p(\mathbf{y}^t|M_k)\pi(M_k),$$

and the marginal predictive PDF is

$$p(\mathbf{y}^{t*}) = \sum_{k=1}^K p(\mathbf{y}^{t*}|M_k)p(M_k|\mathbf{y}^t).$$

Thus, the prediction via EBMA is a weighted average of the component PDFs, and the weight for each model is based on its predictive performance on past observations in period T .

4.5. Data partitions

We have complete data from 2001–3 to 2014–3; however, functionally, our data consist of the five partitions shown in [Fig. 4](#). First, as we mentioned above, we used

¹⁰ Draws on [Montgomery, Hollenbach, and Ward \(2015\)](#).

Start for duration	Training period	Calibration	Test	Forecast
1955–01	2001–03	2010–01	2012–05	2014–04
These data are used to calculate the duration counters	Data used to estimate the thematic models	Predictions from thematic models used to calibrate ensemble model	Test ensemble predictions	Obs. data end; 6-month forecast period

Fig. 4. Data scheme. Information on ILCs back to 1955 is used to construct the duration variables, in order to ameliorate left-censoring. The thematic models are then trained using a subset of the complete data (training period), and their predictions over a second period are used to train the ensemble (calibration period). A third portion of the observed data is held back for evaluating the out-of-sample fit (test), and finally, a six-month forecast is made.

Archigos data back to 1955 for the duration variables in order to ameliorate left-censoring in the duration models. We do not have covariates over this time period, but merely use Archigos data on observed ILCs to count the numbers of months between events. The next three partitions are for the period over which we have complete data—ILCs and covariates—from 2001 to 2014. The first two are used to estimate the thematic models (training data) and ensemble parameters (calibration data). The remaining portion is held back as a test set for evaluating out-of-sample predictions. Finally, we forecast a six month period, beginning in April 2014.

5. Results

We include a brief sketch of each thematic model below, together with the conceptual starting point and a few details on the specification. Table A.1 in the Appendix A shows the model specifications and summarizes the results, and the supplementary material includes full results tables for each theme (see Appendix B). Since the goal is prediction, the model fit and ensemble fit presented later on are more important.

5.1. Discussion of thematic models and estimates¹¹

Leader characteristics. Drawing on the literature on leadership tenure (Acemoglu & Robinson, 2006; Bueno de Mesquita et al., 2005; Svobik, 2012), we build a model that captures the leaders' individual characteristics, as well as internal regime cooperation. The literature on leadership survival focuses on a leader's ability to consolidate power over time, but also considers that, as a leader consolidates power, they are more likely to create discontent among those who are not represented politically by the regime. Thus, the risk equation includes a count of the number of months a leader has been in power. To capture the legitimacy of a leader, and, by association, of his or her government, we also include two further variables in the risk equation that indicate whether the current leader of a state entered power through irregular means or by foreign imposition. Leaders who entered through illegitimate, irregular means may be more likely to suffer the same fate themselves. The duration equation captures the timing using the material behavior of dissidents, whether cooperative or conflictual. We use material rather than verbal actions to model the timing of an ILC against illegitimate leaders.

Public discontent. The public discontent model considers verbal interactions as well as protests in order to provide an early warning indicator of ILCs. We also examine verbal cooperation within government, primarily but not exclusively as an indicator of the health of civil–military relations. Since the level of public, verbal interactions in a society is related to the access to media and the ability to voice demands, we also include per capita measures of Internet users and cell subscribers in the risk equation. Many authoritarian governments control the information available to citizens by implementing censorship. As a control, we also include the fraction of a country's population that is excluded, since minority governments facing a large opposition have strong incentives to display unity.

Global instability. Our third model is based on the main components of the model of Goldstone et al. (2010), which was developed to predict general instability for the PITF. In our version, the partial democracy with a factionalism indicator did not perform as well as simply including the Polity participation of competitiveness variable, which captures whether “alternative preferences for policy and leadership can be pursued in the political arena”. Echoing the Goldstone approach, we include GDP and the percentage of the population that is excluded from the political process in the risk equation. Then, to predict the timing of ILC, we include participation competitiveness, a measure of the conflict within the four nearest neighbors, as well as an indicator of the female life expectancy at birth.

Anti-regime protests. This thematic model focuses entirely on protest. Civil resistance campaigns are an effective means of achieving leadership change. The literature on both coup-proofing (Pilster & Böhmelt, 2011; Quinlivan, 1999) and civil resistance campaigns (Chenoweth & Stephan, 2011) describes a key force behind protest movements: their ability to influence the military. A pivotal movement in many civil resistance campaigns is the moment when state forces stop obeying orders from the head of state, and refuse to openly repress protesters. This model captures the basic intuition of this argument by including slower moving structural variables, such as low levels of domestic crises and military expenditure, in the risk equation. Barany (2011) examines the role of the military in countries that experienced unrest during the Arab Spring, and suggests that three factors play a role in the military's decision: professionalization, the role of the military in the current regime relative to other security services, and the potential impact of a successful revolt on the military's own interests. In addition to the factors that may encourage citizens to participate in mass protests, such as poor governance, the military's behavior is a key determinant of a revolution's success. Thus, like coups, revolutions

¹¹ The full estimates for the seven thematic models are available in the supplementary and replication materials (see Appendix B).

are explained from multiple angles, with the arguments and models focusing on tipping points, mass protest, and the state's response. This model is structured based on the argument that the military will be most likely to resist commands to repress when they are least satisfied. In the duration equation, we account for protest and conflict in various different forms: ethnic-religious violence, rebellion, protest events, and nearby rebellion events in other countries.

Contagion. This model captures the possibility of contagion from instability in surrounding areas, which has been mentioned separately as a contributing factor for coups, protests, and rebellions. It uses two spatial weights of opposition resistance and state repression in neighboring countries, weighted by the centroid distance, as key indicators. The risk equation aims to capture the susceptibility to contagion based on both the Political Terror Scale, which captures the overall repressiveness, and the opposition resistance, which counts the number of events conducted by groups associated with armed anti-government groups. Finally, we include the country's population size, as an indicator of the society's inertia and resistance to outside influences. For example, we would expect that, on average, a small country will be more sensitive to events in its neighboring countries than a country with a large population, in which attention is necessarily more domestically oriented.

Internal conflict. The internal conflict model uses GDP per capita, the proximity of the next national election, and the level of autocracy as general indicators of risk, while focusing on intra-governmental conflict and the widespread use of cell technology as duration triggers. Intra-governmental tensions, protests to the government, and the number of cell phones are assumed to interact to influence the duration of leadership tenure and the likelihood of an irregular transfer. The first-order components of this interaction are included in the duration equation as well, but the second-order interactions (e.g., the two-way interactions) are excluded, as they cause instabilities in the likelihood.

Financial risks. This model assumes that financial instability may unseat leaders who are already in a precarious situation. Powell (2012), like others (Galetovic & Sanhueza, 2000; Koga, 2010), suggests that if the status quo is threatened through shocks like economic crises, the military may view coups as favorable even when most satisfied. The baseline risk is determined by the GDP per capita, as a measure of the general prosperity, the looming presence of the next election, and the size of the country, as measured by the population. In addition, the model also includes the Amnesty assessment of terrorism (stability) and the degree of anti-government. For countries in the high risk set, the degree of inflation, as measured by consumer prices, and the health of the country's international financial reserves (taken from the IMF's IFS statistics) affect the duration of leadership most directly.

5.2. Final ensemble and predictive performance

The predictions from the thematic models are combined into an ensemble prediction using the weights

Table 1

Ensemble model, monthly observations.

Model	W	In-sample		Test	
		AUC	F*	AUC	F*
Ensemble		0.86	0.07	0.84	0.11
Leader char.	0.01	0.81	0.04	0.78	0.08
Public disc.	0.01	0.84	0.06	0.82	0.15
Global Inst.	0.01	0.75	0.05	0.73	0.22
Protest	0.01	0.72	0.06	0.75	0.09
Contagion	0.68	0.82	0.05	0.81	0.02
Internal conf.	0.29	0.81	0.08	0.74	0.12
Financial	0.01	0.79	0.03	0.79	0.05

Note: * Maximum *F*-score across possible cutoff values.

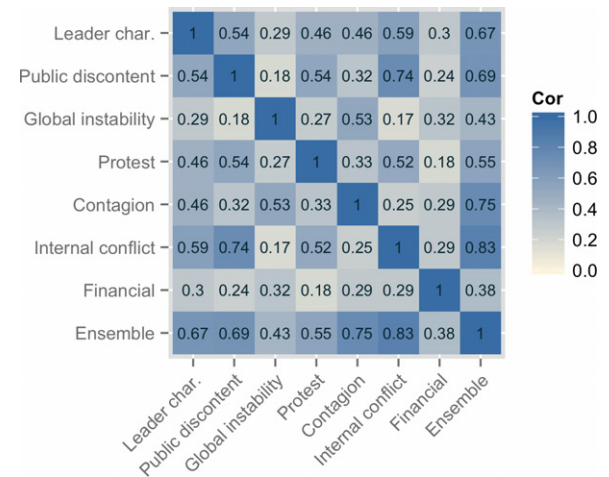


Fig. 5. Thematic and ensemble in-sample prediction correlations.

shown in Table 1. The corresponding AUC – area under the ROC curve – and the *F*-score – a balanced harmonic mean of recall and precision – for the ensemble and thematic models are reported for both the in-sample and out-of-sample data. The ROC curves corresponding to the AUC values for the ensemble are shown in Fig. 6, and Fig. 7 shows the precision–recall tradeoff. A few points about these different pieces of evidence are worth mentioning. The EM algorithm determines the weights that will maximize the fit in the calibration period, and tends to favor models that fit well but also make unique predictions that are not captured by other models. Most of the weight is placed on only two models, the contagion and internal conflict models. The seven thematic models all have similar AUC and *F*-score measures, which implies that these statistics are not good guides for the selection of candidate models that contribute to an ensemble. Rather, the weights must be related to the uniqueness of predictions from the contagion and internal conflict models, something that is more difficult to assess.

The correlations among the seven thematic models' predictions are shown in Fig. 5, and range from 0.17 to 0.74, showing that the models generally separate. This is desirable, insofar as a lack of correlation is necessary for uniqueness. However, like the AUC and *F*-score assessments of model fits, there is no clear relationship with the weights assigned in the final ensemble.

The fit for the thematic models and ensemble has AUC values of around 0.75–0.85. The fit is usually not

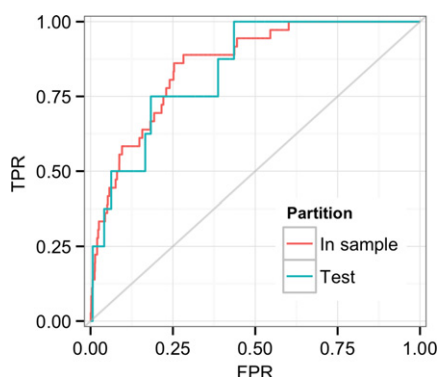


Fig. 6. ROC curves for the ensemble in- and out-of-sample. Values towards the top-left corner indicate better performances, and the gray line reflects the average performance of a random guess model.

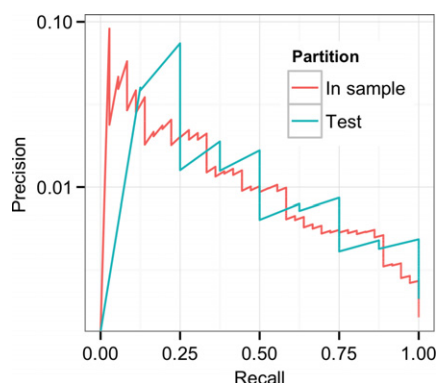


Fig. 7. Precision recall plots for the ensemble in- and out-of-sample. The curves show the tradeoff in these two statistics over the range of cut points for classifying the probabilistic ensemble predictions as binary values. Values closer to the top-right corner indicate better performances. Note that the y-axis is logged.

evaluated well in research on different manifestations of political conflict, but Ward et al. (2010) report AUCs of 0.76 and 0.86 for two canonical models of civil war onset, and we can use these as a rough baseline. Both of these studies use country-year data and dependent variables that are less imbalanced than ours, which suggests that our model's performance at least matches and probably exceeds previous analyses of related outcomes. Indeed, in annualized versions of our predictions, the AUC values increase to between 0.91 and 0.96 for the ensemble.

With data as imbalanced as ours, AUC statistics can be misleading, as their calculation incorporates the overwhelming number of true negatives. Consequently, the maximum F score is drastically lower across models, and reflects precision values that are generally low across all theme models and the ensemble for all values of recall.¹² The precision recall plots in Fig. 7 show that the precision reaches a maximum value of 0.09.

In summary, although the precision is weak and the uneven distribution of ensemble weights suggests that

Table 2

Top ten forecasts for ILC between April and September 2014 (six months) using March 2014 data.

Country	Probability
Ukraine	0.28
Bosnia and Herzegovina	0.19
Yemen	0.10
Egypt	0.07
Thailand	0.06
Guinea	0.05
India	0.04
Turkey	0.04
Libya	0.03
Central African Republic	0.03

the thematic models need to be rebalanced, the overall performance of the ensemble matches or exceeds those of existing modeling efforts for similar forms of political violence, especially after accounting for the inherent loss of precision when disaggregating from country-year to country-month data.

6. Where is irregular leadership change most likely?

Using the ensemble model and data from March 2014, we create forecasts for the probability of ILC over the period from April to September 2014. We aggregate the monthly forecasts produced by this model to an overall probability of ILC during this time period,¹³ and Table 2 shows the ten highest forecasts that result. Fig. 8 maps all forecasts.

These predictions were originally made in early May 2014, about five weeks after our latest data, and thus are largely *ex ante* forecasts. Of course, probabilities are not certainties, and the chance that one or more countries in our top ten list will experience an ILC is no higher than 0.62, which leaves room for uncertainty. Our top five predictions include Ukraine, Bosnia and Herzegovina, Yemen, Egypt, and Thailand.

Ukraine experienced an ILC in February 2014 after President Yanukovych was ousted by pro-EU protesters and replaced with acting President Turchynov. Elections in May brought pro-Western Poroshenko to power. By then, pro-Russian protests and Russian involvement had led to the annexation of the Crimea and an armed revolt in the east. Despite progress by government forces, a *de facto* invasion by Russia in September led to a ceasefire that will probably leave eastern Ukraine autonomous but controlled by Russia. There was no ILC during the forecast window.

Bosnia and Herzegovina was the site of substantial and widespread anti-government protests in early 2014, the so-called Bosnian Spring. These were organized largely because of the frailty of the economy, the high level of unemployment, and the non-payment of pensions. Prime Minister Vjekoslav Bevanda minimized the protests. However, as the leader of a weak central government, the greatest instability may reside in the locally governed regions.

Yemen has been the site of protests, accompanied by the presence of a very powerful Al-Qaeda army. That,

¹² We present the maximum F -score for a given model across all possible threshold values for classifying probabilistic to binary predictions.

¹³ Using $p^* = 1 - \prod_{t=1}^6 (1 - p_t)$, where p_t is the forecast t months ahead.



Fig. 8. Six-month forecasts for ILC between April and September 2014.

alongside a set of rulers who are widely reported to be corrupt, creates an unstable situation. Yemen is now in the throes of (another) reorganization in which central authority seems likely to devolve to regional ruling coalitions.

Egypt has seen an outbreak of protests and violence in February each year to celebrate the resignation of Mubarak and the start of the so-called revolution in the early spring of 2011. In mid-2013, these protests spread and sparked a coup d'état that displaced Morsi. In mid-2013, Mansour was appointed as acting president. In early 2014, a new constitution was overwhelmingly ratified by Egyptian voters, even though roughly 2/3 of the potential voters avoided participating. In May of 2014, a presidential election was won by Abdel Fattah el-Sisi, the Egyptian commander-in-chief. Tension remains high and the legitimacy and popularity of the current regime are tenuous, at best.

Thailand has been a puzzling cauldron of political conflict for over a decade. Thaksin Shinawatra was overthrown by a coup d'état in the fall of 2006, and the resulting junta instituted martial law and forbade many political activities until mid-2007. Yingluck Shinawatra handily won the subsequent elections in mid-2011, but protests heated up towards the end of 2013, resulting in demands for her resignation. After the elections scheduled for February 2014 were not held due to disruptions by anti-government protesters, a court order replaced the prime minister with a caretaker government. In May, several weeks after our original forecasts were made, the caretaker government was overthrown in a military coup d'état.

7. Conclusion

We have used new, temporally disaggregated data which include behavioral variables that are derived from event data. We have also employed split-population duration and ensemble modeling approaches for examining irregular leadership changes over the period from 2001 to the present. Each of these aspects is novel in the study of leadership change. In so doing, we have also developed a suite of new empirical models which are measured monthly. In addition, we have then combined the forecasts of each of these empirical models using ensemble Bayesian model averaging to produce a single probability estimate that benefits from the so-called “wisdom of crowds”. Along the way, we have updated the dependent variable for the past two and a half years.

In our attempt to forecast ILCs accurately, and at a monthly level, we have created *ad hoc* thematic models

that are grounded only loosely in existing arguments. However, one of the advantages of using an ensemble is modularity, and we can replace input models easily as better ones are developed. Two of our major goals for future work are to tie input models to existing arguments more closely, and to explore whether the ensemble performance can be improved by the inclusion of “niche” models that are designed to predict specific cases well, at the expense of the overall fit. We also hope that this study will serve as a foundation for future inquiry and will encourage scholars to conduct similar work with alternative models at the country-month level.

In terms of the broader significance of this study, openly forecasting ILCs provides a public benefit by enabling political actors of all types to be aware of leadership changes and the crises that may potentially stem from such events. Thus, such forecasting efforts become part of a broader accountability effort among policymakers, practitioners, and researchers to prevent the onset of instability and conflict in regions that are experiencing a leadership change. The recent military coup in Thailand serves to demonstrate the types of events – such as the violent repression and censoring of protesters – that often flow from unexpected shifts in power.

A final lesson relates to the level of performance that is needed in order to make credible forecasts. Our ensemble showed a respectable predictive power out-of-sample, with AUCs of above 0.8 in monthly data and an equivalent AUC of above 0.9 with annualized data. Now that our original forecasting period from April to September 2014 has passed, hindsight shows that one of our top ten forecasts, Thailand, experienced an ILC. The disparity between our forecasts and event occurrences is partly a result of low probabilities even at the high end, but also indicates a generally low level of precision. When considering rare events, and as the spatio-temporal resolution increases, the level of predictive power required in order to ensure a high precision is very large, much higher than one might conventionally associate with a “good” model.

Many months pass in each country without an irregular leadership change. They are rare, and our data are very sparse. Our modeling approach has been driven by the goal of forecasting ILCs accurately, and the rarity of these events has led to the novel aspects that we have presented here. Still, we are looking for needles in a haystack. Even our ten highest predictions have low probabilities of irregular leadership change. However, as someone once noted, “reality is a low probability event”.

Table A.1Thematic model estimates: effect directions and *p*-values.

Variable	Thematic model						
	1	2	3	4	5	6	7
Duration equation intercept	+	+	+	+	+	+	+
Material conflict dissident → government	–						
Material cooperation dissident → government	+						
Verbal cooperation within government		–					
Verbal conflict from government → dissident		–					
<i>N</i> anti-government protests		–					
Spatial lag of insurgency events in nearest four neighbors			+				
Female infant mortality, lagged			+				
Factionalism			–				
<i>N</i> low-intensity conflictual deeds: ethnic groups & gov't				+			
<i>N</i> low-intensity conflictual deeds: rebel groups & gov't				+			
<i>N</i> protest events directed against all actors				–			
Low-intensity conflict in countries w/similar pol. struct.				–			
<i>N</i> resistance events in neighboring countries					–		
<i>N</i> repressive events in neighboring countries					+		
Internal tension × protests × cell phones						+	
Intra-governmental tension						–	
Anti-governmental protest						–	
Cell phone users						–	
Inflation, via CPI							–
International reserves							+
In Weibull shape α	–	–	–	–	–	–	–
Risk equation intercept	–	+	+	+	–	+	–
Leader with irregular entrance	–						
Leader who was imposed by foreign actor	–						
Months in power, logged	+						
Internet users		–					
Cell phone users		+					
Excluded population		+	–				
GDP per capita			–			–	–
<i>N</i> high-intensity conflictual deeds				+			
Military spending				+			
Amnesty International terror scale					+		+
Proximity of the next election					–	+	+
Number of acts of resistance					+		
Population					–		+
Level of autocracy						–	
Opposition resistance							+

Note: dark gray $p \leq 0.05$, light gray $0.05 < p \leq 0.10$; themes: 1 = leader characteristics, 2 = public discontent, 3 = global instability, 4 = anti-regime protests, 5 = contagion, 6 = internal conflict, 7 = financial risks.

Acknowledgments

This research was sponsored by the Political Instability Task Force (PITF). The PITF is funded by the Central Intelligence Agency. The views expressed in this report are the authors' alone, and do not represent the views of the US Government.

We would like to thank the participants of the February and August 2014 PITF conferences, the program sponsors, and John Ahlquist for their comments and feedback.

Appendix A. Theme model specification and summary

See Table A.1. The supplementary materials contain tables with the full split-population duration estimates (see Appendix B).

Appendix B. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.ijforecast.2015.01.009>.

References

- Acemoglu, D., & Robinson, J. A. (2006). *Persistence of power, elites and institutions*. Tech. rep. National Bureau of Economic Research.
- Alesina, A., Özler, S., Roubini, N., & Swagel, P. (1996). Political instability and economic growth. *Journal of Economic Growth*, 1(2), 189–211.
- Barany, Z. (2011). The role of the military. *Journal of Democracy*, 22(4), 24–35.
- Beck, N., King, G., & Zeng, L. (2000). Improving quantitative studies of international conflict: A conjecture. *American Political Science Review*, 94(1), 21–35.

- Belkin, A., & Schofer, E. (2003). Toward a structural understanding of coup risk. *Journal of Conflict Resolution*, 47(5), 594–620.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 26, 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Bueno de Mesquita, B., Smith, A., Siverson, R. M., & Morrow, J. D. (2005). *The logic of political survival*. MIT Press.
- Cederman, L.-E., Weidmann, N. B., & Gleditsch, K. S. (2011). Horizontal inequalities and ethnonationalist civil war: A global comparison. *American Political Science Review*, 105(3), 478–495.
- Chenoweth, E., & Stephan, M. J. (2011). *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press.
- Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1), 266–298.
- Collier, P. (1999). On the economic consequences of civil war. *Oxford Economic Papers*, 51(1), 168–183.
- Collier, P., & Hoeffler, A. (2004). Greed and grievance in civil war. *Oxford Economic Papers*, 56(4), 563–595.
- De Rouen, K. R., & Sobek, D. (2004). The dynamics of civil war duration and outcome. *Journal of Peace Research*, 41(3), 303–320.
- Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *The American Political Science Review*, 97(1), 75–90.
- Freund, Y., & Schapire, R. (1997). A decision-theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences*, 55, 119–139.
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189–1232.
- Galetovic, A., & Sanhueza, R. (2000). Citizens, autocrats, and plotters: A model and new evidence on coups d'état. *Economics and Politics*, 12(2), 183–204.
- Galton, F. (1907). Vox populi. *Nature*, 75(1949), 450–451.
- Gerber, A., & Malhotra, N. (2008). Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Quarterly Journal of Political Science*, 3(3), 313–326.
- Gleditsch, K. S., & Ward, M. D. (1999). Interstate system membership: A revised list of independent states since 1816. *International Interactions*, 25, 393–414.
- Goemans, H. E., Gleditsch, K. S., & Chiozza, G. (2009). Introducing Archigos: A dataset of political leaders. *Journal of Peace Research*, 46(2), 269–283.
- Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., Ulfelder, J., & Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science*, 54, 190–208.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857–871.
- Gurr, T. R. (1970). *Why men rebel*. Princeton, NJ: Princeton University Press.
- Gurr, T. R. (1993). Why minorities rebel: A global analysis of communal mobilization and conflict since 1945. *International Political Science Review*, 14(2), 161–201.
- Gurr, T. R., & Moore, W. H. (1997). Ethnopolitical rebellion: A cross-sectional analysis of the 1980s with risk assessments for the 1990s. *American Journal of Political Science*, 41(4), 1079–1103.
- Hale, H. E. (2013). Regime change cascades: What we have learned from the 1848 revolutions to the 2011 Arab uprisings. *Annual Review of Political Science*, 16, 331–353.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer.
- Hendrix, C. S., & Salehyan, I. (2012). Climate change, rainfall, and social conflict in Africa. *Journal of Peace Research*, 49(1), 35–50.
- Hill, D. W., Jr., & Jones, Z. M. (2014). An empirical evaluation of explanations for state repression. *American Political Science Review*, 108(3), 661–687.
- Huntington, S. P. (1968). *Political order in changing societies*. Yale University Press.
- Jackman, R. W. (1978). The predictability of coups d'état: a model with African data. *The American Political Science Review*, 1262–1275.
- Johnson, T. H., Slater, R. O., & McGowan, P. (1984). Explaining African military coups d'état, 1960–1982. *The American Political Science Review*, 78(3), 622–640.
- Kang, S., & Meernik, J. (2005). Civil war destruction and the prospects for economic growth. *Journal of Politics*, 67(1), 88–109.
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2), 137–163.
- Koga, J. (2010). Causes and consequences of coup-proofing strategies. Unpublished.
- Kuran, T. (1991). Now out of never: The element of surprise in the East European revolution of 1989. *World Politics*, 44(1), 7–48.
- Lujala, P., Gleditsch, N. P., & Gilmore, E. (2005). A diamond curse? Civil war and a lootable resource. *Journal of Conflict Resolution*, 49(4), 538–562.
- Marshall, M. G., & Gurr, T. R. (2014). Polity IV project: Political regime characteristics and transitions, 1800–2013. URL <http://www.systemicpeace.org/polity/polity4.htm>.
- Montgomery, J. M., Hollenbach, F. M., & Ward, M. D. (2015). Calibrating ensemble forecasting models with sparse data in the social sciences. *International Journal of Forecasting*, 31(3), 930–942.
- Pilster, U., & Böhmelt, T. (2011). Coup-proofing and military effectiveness in interstate wars, 1967–99. *Conflict Management and Peace Science*, 28(4), 331–350.
- Powell, J. (2012). Determinants of the attempting and outcome of coups d'état. *Journal of Conflict Resolution*, 56(6), 1017–1040.
- Powell, J., & Thyne, C. (2011). Global instances of coups from 1950 to 2010: A new dataset. *Journal of Peace Research*, 48(2), 249–259.
- Quinlivan, J. (1999). Coup-proofing: Its practice and consequences in the Middle East. *International Security*, 24(2), 131–165.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5), 1155–1174.
- Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research*, 47, 651–660.
- Salehyan, I., & Gleditsch, K. S. (2006). Refugees and the spread of civil war. *International Organization*, 60(2), 335–366.
- Schrodt, P. A. (2014). Seven deadly sins of contemporary quantitative political analysis. *Journal of Peace Research*, 51(2), 287–300.
- Schrodt, P. A., & Gerner, D. J. (1994). Validity assessment of a machine-coded event data set for the Middle East, 1982–92. *American Journal of Political Science*, 38, 825–854.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business*. Doubleday, New York: Economics, Society, and Nations.
- Svolik, M. W. (2012). *The politics of authoritarian rule*. New York: Cambridge University Press.
- Ulfelder, J. (2013). Coups slow economic growth. <https://goo.gl/AyzJcF>.
- Ward, M. D., & Gleditsch, K. S. (2008). *Spatial regression models: Vol. 155*. Sage.
- Ward, M. D., Greenhill, B. D., & Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflict. *Journal of Peace Research*, 47(4), 363–375.
- Weidmann, N. B. (2014). On the accuracy of media-based conflict event data. *Journal of Conflict Resolution*, 1–21.
- Wimmer, A., Cederman, L.-E., & Min, B. (2009). Ethnic politics and armed conflict: a configurational analysis of a new global data set. *American Sociological Review*, 74(2), 316–337.
- Wood, R. M., & Gibney, M. (2010). The political terror scale (PTS): a re-introduction and a comparison to CIRI. *Human Rights Quarterly*, 32(2), 367–400.

Andreas Beger is a post-doctoral research associate at Duke University. He received his Ph.D. in 2012 from Florida State University.

Cassy L. Dorff (Duke Ph.D. 2015) is a Post-doctoral Research Fellow at the Sié Chéou-Kang Center for International Security and Diplomacy at the University of Denver. Beginning in Fall of 2016 she will be Assistant Professor of Political Science at the University of New Mexico.

Michael D. Ward got his Ph.D. in the last century. He is professor of political science at Duke University, and director of the eponymous Wardlab, which is found at mdwardlab.com and predictiveheuristics.com.