In-N-Out of Civil War:

A Hidden Markov Model of Civil War

Jeffrey B. Arnold *

June 2, 2011

Abstract

This paper estimates a model of civil war prevalence with a hidden Markov model (HMM) to account for measurement error in classifying civil wars. Since there exist multiple coding rules for civil war with substantial disagreement, each of these coding rules is treated as a noisy indicator of a latent concept. A HMM is used to classify country-years into latent states of civil war and peace and to model the probability of civil war onset and continuation with these latent states. The estimated classification supports an inclusive concept of civil war with a low threshold for what constitutes a civil war. The only statistically significant covariates in the civil war transition models are population and GDP on civil war onset.

1 Introduction

Quantitative models of civil war necessarily require good data on the occurrence of civil war. Yet, there exist multiple lists of civil wars, each with its own coding rules for what is classified as a civil war. Sambanis (2004) compares twelve lists of civil war occurrence and finds that the different coding rules result in "substantial variation" in the values of onset and termination of civil wars.[1]Differences in these coding rules affect the substantive results of models of civil war onset

[*] Author affiliation: University of Rochester, Department of Political Science. Author contact info: jeffrey.arnold@gmail.com. Thanks to Kevin Clarke, Bethany Lacina, Michael Peress, Ian Sulam and Curtis Signorino for their comments.

[1] The coding rules for these civil war occurrence data differ mainly due to the conditions on the threshold of violence, identification of start and end dates of individual wars, and how to distinguish inter-, intra-, and extra-state wars. The threshold of violence condition includes the number of deaths, the inclusion of civilian deaths, and the level of organization of forces (Sambanis 2004).

and continuation, with many parameters not robust to the choice of coding rule for civil war.

The existence of multiple definitions of civil war imply that there is not an agreement about how to define the concept of civil war. This conceptual disagreement poses two problems for the study of civil war. The first problem is a classification problem: what is civil war? Although the definitions disagree, they purport to define the same concept. If so, what are the properties that all these definitions share in common?

The second problem is, given this uncertainty about what things are civil wars, how can a researcher make inferences about the data generating processes of civil war, such as civil war onset and continuation? Since there is conceptual uncertainty about civil war, models of civil war processes must incorporate that uncertainty.

This paper uses a hidden Markov model (HMM) to address both of these problems within a single statistical model.[2]HMMs are models in which the data generating process of the observed data depends on the state of an unobserved (hidden or latent) Markov process. The time-dependency between the latent states make HMMs useful for classification problems when the observed data are a time-series. In these data, the latent state of each observation plausibly depends on the latent state of the previous observation.[3] For this application, whether a country-year is in civil war is unobserved. Instead, the researcher observes multiple noisy indicators of civil war — the existing civil war occurrence data. The transitions between these latent states correspond to models of civil war onset and continuation. Thus, the HMM classifies observations as civil wars, while estimating models of transition into and out of civil war.

The two primary results of the this model are as follows. First, the model classifies more country-years as civil wars than most of the individual lists, supporting an inclusive definition of civil war. Second, when using the latent states as the response variable in models of civil war prevalence, almost all the covariates are statistically insignificant. The only significant covariates are GDP and population with respect to civil war onset; no covariates are significant with respect to civil war continuation.

Methodologically, this paper introduces Hidden Markov Models as a general method for esti-

[2] HMMs go by several different names, including latent Markov models, Markov mixture models, dependent mixture models, and regime switching models (Visser and Speekenbrink 2010; Zucchini and MacDonald 2009, pp. 30-31).

[3] For book length treatments of hidden Markov models, see Cappé, Moulines, and Rydén (2005), Frühwirth-Schnatter (2006) Zucchini and MacDonald (2009). For article-length introductions see Visser and Speekenbrink (2010); Chib (1996); Scott (2002); Rabiner (1989).

mating time series and longitudinal data with latent data. Within political science, changepoint models have been estimated with hidden Markov models (Park 2009; Park 2010; Park 2011; Svolik 2009). The models estimated in those papers are special cases of HMM (Chib 1996; Chib 1998). The model estimated in this paper uses a more general specification and differs from those papers in two ways. The first way is in the restrictions on transitions between latent states. The changepoint model restricts transitions between the latent states to occur sequentially and unidirectionaly, e.g. transitions can only occur from state 1 to state 2 and state 2 to state 3, but not 2 to 1 or 1 to 3. However, there are many applications in which there is a positive probability of transitioning between any states. Examples include civil war (this paper), democratization and regime changes, dyadic rivalry in international conflict, and trade openness. the second difference is in the estimation technique. Those papers use MCMC to estimate the parameters, but Hidden Markov Models can also be estimated with either maximum likelihood (MLE) or expectation maximization (EM) methods. The use of these mode finding methods is fast and avoids the label-switching problem that arises with MCMC estimates of HMM.

Hidden Markov Models (HMMs) have also been used in a purely time-series context. Phillip A. Schrodt (2000), Philip A. Schrodt (2006), Schrodt and Gerner (2004) used HMMs to classify multivariate categorical time-series into periods of conflict and cooperation. The uses of HMMs extend beyond time-series forecasting. This paper includes covariates in the transition equations of the model in order to make inferences about the substantive factors influencing the transitions between latent states.

Hidden Markov models are closely related to two other areas in the political science methods literature. HMMs provide an alternative method for estimating dynamic models of discrete time-series-cross-section data. The HMM has features of both the full transition model and lagged latent variable model of (Jackman 2000; Beck et al. 2001). The HMM can be thought of as a full transition model in which the parameters are conditioned on the lagged latent state rather than the realized response variable. Alternatively, the HMM can be thought of as a lagged latent variable model in which the latent variable is discrete rather than continuous. The HMM also fits into the literature on measurement error and latent variables. Much of that literature focuses one continuous latent variables, for example Treier and Jackman (2008). However, there are many concepts in political science which are more naturally thought of as discrete categories. Many of those models also ignore time-dependency between observations, treating the data as cross-sectional. The HMM provides a method for modeling data with a discrete latent variable in time-series data.

Section 2 provides background of Hidden Markov Models. Section 3 describes the data used in this paper. Section 4 defines the statistical model estimated in this paper. Section 5 presents and interprets the parameter estimates from the model. Section 6 concludes.

2 Hidden Markov Models

Let $Y = Y_1, \dots, Y_{|T|} = Y_{1:|T|}$ be a vector of random variables indexed by an integer vector representing time, $T = 1 : |T|$.[4] Suppose that there exists a finite state space $K = 1 : |K|$, and a sequence of latent states $S = S_{1:|T|}$, where all S_t take values in the state space K. A HMM consists of two key assumptions. The first assumption is that the distribution of Y_t only depends on the current state and not previous states,

$$p(Y_t | S_{1:t}, Y_{1:(t-1)}) = p(Y_t | S_t). \tag{1}$$

The distributions in (1) are called are called the *response* or *state-dependent distributions.*

The second assumption is that the latent states follow a Markov process. Let $p(S_1)$ be the distribution of the first state. Then for all $t \in 2 : |T|$,

$$p(S_t | S_{1:(t-1)}, Y_{1:t-1}) = p(S_t | S_{t-1}). \tag{2}$$

The distributions in (2) are called the *transition distributions*, and $p(S_1)$ is called the the *initial distribution.* This assumption differentiates HMM from mixture models, which also have the first assumption. In a mixture model the latent states of observations are independent, given the data, $p(S_t | S_{1:t-1}, Y) = p(S_t)$. A mixture model can be thought of as a HMM in which each observation is an independent time series of length one.

Figure 1 illustrates (2) and (1) in a directed graph of the dependencies between Y and S.[5]

[4] I will use the convention that

$$t : s = t, t+1, \dots, s-1, s,$$

with the convention that $t : s = \emptyset$ if $t > s$.

[5] The assumptions used in this paper correspond to **Y3** and **S2** in Frühwirth-Schnatter (2006). These are the standard assumptions in the HMM, but they can be weakened in various ways with the cost of increased model and computational complexity (Zucchini and MacDonald 2009, Chapter 8).

The complete data likelihood of this model is the joint sampling distribution of Y and S as a function of parameters θ and covariate data x, and follows from equations (2) and (1),

$$p(Y = y, S = s|\theta, x) = p(S_1 = s_1|\theta, x) \prod_{t \in 2:|T|} p(Y_t = y_t|s_t, \theta, x) \prod_{t \in T} p(S_t = s_t|s_{t-1}, \theta, x). \quad (3)$$

However, (3) cannot be calculated without observing the latent states, which are obviously unobserved, or it would just be a Markov model. Thus, to calculate likelihood $p(y)$, integrate Equation (3) over the $|K|^{|T|+1}$ possible realizations of $S_{1:|T|}$,

$$p(Y = y|\theta, x) = \sum_{s_1 \in K} \cdots \sum_{s_{|T|} \in K} p(S_1 = s_1|\theta, x) \prod_{t \in 2:|T|} p(Y_t = y_t|s_t, \theta, x) \prod_{t \in T} p(S_t = s_t|s_{t-1}, \theta, x). \quad (4)$$

Since the latent states s are unobserved, to estimate (4) I treat the latent states as missing data and use the expectation-maximization (EM) algorithm.The EM algorithm estimates the parameters using the complete data likelihood (3). Take the logarithm of the complete data likelihood in (3),

$$\log p(Y = y, S = s|\theta, x) = \log p(s_1|\mu, x) + \sum_{t \in T} \log p(s_t|s_{t-1}, \tau, x) + \sum_{t=2}^{|T|} \log p(y_t|s_t, \upsilon, x), \quad (5)$$

where $\mu, \tau, \upsilon \subseteq \theta$ are sets of parameters on which the initial, transition, and response distributions depend. Since there are $|K|$ distributions for the transition and response distributions, let τ_k be the parameters in $p(S_t|S_{t-1} = k)$ and υ_k be the parameters in $p(Y_t|S_t = k)$.

In the Expectation step, the latent states in (5) are replaced by their expected values, conditional on the current estimate of the parameters θ. Let θ' be the estimate of θ at the current iteration in the algorithm. The expected value of the log-likelihood with respect to θ' is

$$\begin{aligned} Q(\theta, \theta') = &\sum_{k \in K} \gamma_1(k) \log p(s_1|\mu, x) \\ &+ \sum_{j \in K} \sum_{k \in K} \sum_{t \in T} \xi_t(j, k) \log p(s_t|s_{t-1}, \tau, x), \\ &+ \sum_{k \in K} \sum_{t=2}^{|T|} \gamma_t(j) \log p(y_t|s_t, \upsilon, x) \end{aligned} \quad (6)$$

where $\xi_t(j, k) = p(S_t = k, S_{t-1} = j|y, x, \theta')$ and $\gamma_t(j) = p(S_t = j|y, x, \theta')$. The expected values of γ and ξ are calculated by the forward-backward algorithm in (30) and (31) (see Appendix).

In the Maximization step, (6) is maximized with respect to θ. If μ, τ and υ in (6) are independent, then each of the terms on the right hand side can be maximized individually, and often using existing routines. For example, if Y_t is continuous and $p(Y_t|S_{t-1}=k,\upsilon,x)$ is a normal distribution, then υ_k can estimated by OLS.

While the EM outputs point estimates of θ, the values of the latent states are often substantively interesting. In particular, given the observed sequence of y, what is the most likely sequence of states,

$$\hat{s}_t = \operatorname*{argmax}_{s_{1:|T|}} p(S_{1:|T|} = s_{1:|T|} | Y_{1:|T|} = y_{1:|T|}). \tag{7}$$

The solution to equation (7) is called the *global decoding*. Given values of θ, the solution to this problem is easily calculated; see 7.2 for the details of this process.

While the EM only outputs point estimates of θ, the covariance matrix and confidence intervals can be estimated by a parametric bootstrap as described in Zucchini and MacDonald (2009, p. 55).

Thus far, I have assumed that y is a single time-series, but the results easily extend to time-series cross-section data. suppose that y is time-series-cross-section data, where $y_{t,i}$ is the value of Y for individual $i \in I = \{1 : |I|\}$ at time t.[6] Assuming contemporaneous conditional independence (Zucchini and MacDonald 2009, pp. 122-125), the panel data are modeled as $|I|$ independent time series and the complete data likelihood for these data is

$$p(Y = y, S = s|\theta, x) = \prod_{i \in I} \left(p(s_{0,i}|\mu, x, i) \prod_{t \in T} p(s_{t,i}|s_{t-1,i}, \tau, x, i) p(y_{t,i}|s_{t,i}, \upsilon, x, i) \right). \tag{8}$$

In the previous derivations, the initial, transition, and response distributions have not been assigned any particular functional form. The ability to plug different probability distributions into the HMM for these distributions make the HMM a flexible class of models. The multinomial logit for the transition and initial state distributions. Common choices for the distributions in the response equation are normal for continuous response variables, Poisson for count data, and binomial for binary data.

[6] The notation assumes a balanced panel for simplicity of notation only; the results would follow for an unbalanced panel.

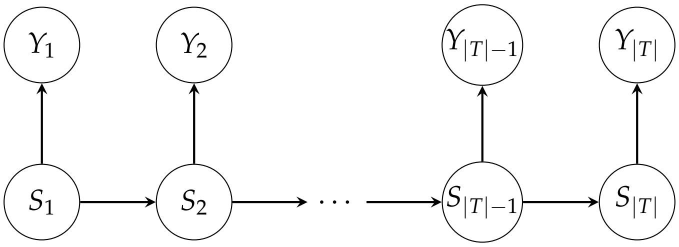

Figure 1: Directed acyclic graph of a simple hidden Markov model, Equations (1) and (2). For $t \in 1 : |T|$, S_t is the latent state, and Y_t is the observed data. An arrow represents a statistical dependency between random variables.

3 Data

The data on civil wars in this paper are a subset of the Sambanis (2004) replication dataset.[7] The Sambanis (2004) replication dataset consists of country-year observations with variables of civil war onset and prevalence from multiple sources and common covariates in models of civil war onset and prevalence. The subset used in this paper consists of the twelve civil war prevalence variables and nine covariates listed in Table 1, as well as the country identifier `cid` and `year` variables which jointly serve as unique identifiers.[8]

The twelve variables of civil war prevalence were those chosen for comparison by Sambanis (2004), and were the existent variables of civil war prevalence at the time at which that paper was written. The civil war prevalence variables are correlated highly enough that *a priori* they appear to be measuring the same concept, yet the correlations show a substantial amount of disagreement between the variables. Figure 2 displays the pairwise Pearson correlation coefficients of the twelve civil war prevalence variables. The mean of pairwise correlation coefficients of the civil war variables is 0.72. The minimum pairwise correlation coefficient is 0.54, between `atwar5` and `atwar9`.

The dataset includes variables commonly used as covariates in regressions of civil war literature: GDP, GDP growth, instability, anocracy, resource dependence, ethnic factionalization, population, mountainous terrain, and a Muslim population.Sambanis (2004) generated this set of explanatory variables based on the models of civil war onset Fearon and Laitin (2003) and Collier and Hoeffler (2001) (see p 835-387 of Sambanis 2004, for a discussion of the choice of covariates and more

[7] Available from the Journal of Conflict Resolution website `http://jcr.sagepub.com.ezp.lib.rochester.edu/content/suppl/2005/11/16/48.6.814.DC1/Sambanis_Data.zip`.

[8] The subset are the response and explanatory variables in Tables 7 of Sambanis (2004).

description).[9] The dataset includes 3708 observations, uniquely identified by country and year. There are 133 unique countries in the dataset, and years have a range of 1954 to 1991. See Table 2 for summary statistics of the data.[10]

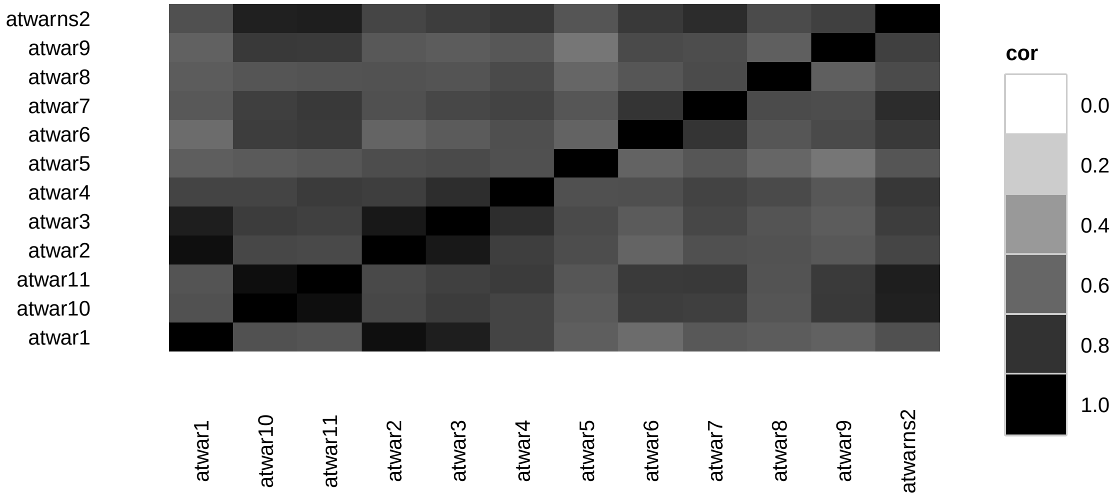

Figure 2: Correlation matrix of civil war prevalence variables. Each rectangle maps the correlation coefficient between the variables on the x and y-axes. The color scale has a domain of $[0, 1]$ since all correlation coefficients are positive, and maps the correlation coefficient to an interval on the color scale.

4 Model

The observed data or responses to be modeled are the civil war prevalence variables from data described in the previous section. Let $J = 1 : 12 = \{\texttt{atwar1} : \texttt{atwar11}, \texttt{atwarns2}\}$ be the set of variables, and y be a $|T| \times |I| \times |J|$ array of the civil war prevalence variables such that $y_{t,i,j}$ is the value of variable j for country i at time t. Let s be a matrix of latent states such that $s_{t,i}$ is the latent state for country i at time t. The state space is $K = \{0, 1\}$ where state 0 is civil war and

[9] All variables use the same names as the Sambanis dataset. The single exception is that the variable `atwarns2` is not in the original dataset; it is `atwarns` corrected for the errors indicated in the readme file distributed with the data. Instability (`inst3l1`) is a yearly change in the Polity score greater than two in the last three years of an observation. Anocracy (`anoc2l1`) is a Polity score between -6 and 6.

[10] Missing data are deleted listwise. The majority of observations dropped due to missing data are due to different year and country coverage of civil war prevalence. All of these observations occur at the beginning or end of each country's time series. The only interior missing values are due to missing values in covariates, and there are very few of those.

variable	description and source
atwar1	Civil war prevalence (Singer and Small 2006, COW 1994)
atwar2	Civil war prevalence (Sarkees and Schafer 2000, COW 2000)
atwar3	Civil war prevalence (Collier and Hoeffler 2001)
atwar4	Civil war prevalence (Licklider 1995)
atwar5	Civil war prevalence (Gleditsch et al. 2002, Wars)
atwar6	Civil war prevalence (Gleditsch et al. 2002, All)
atwar7	Civil war prevalence (Fearon and Laitin 2003)
atwar8	Civil war prevalence (Leitenberg 2006)
atwar9	Civil war prevalence (Regan 1996)
atwar10	Civil war prevalence (Doyle and Sambanis 2000, expanded)
atwar11	Civil war prevalence (Doyle and Sambanis 2000)
atwarns2	Civil war prevalence (Sambanis 2004)
gdp1	GDP
grol1	GDP growth
inst3l1	Instability
anoc2l1	Anocracy
oil2l1	Oil Exporter
ef1	Ethnic Factionalization
lpopnsl1	log Population
mtnl1	Mountainous Terrain
muslim	Muslim population
cid	Country identifier
year	Year

Table 1: Variable names and descriptions. All variables are from the Sambanis (2004) dataset. The postfix 1 indicates that the variable is lagged one year.

variable	mean	sd	median	min	max
`atwar1`	0.08	0.26	0.00	0.00	1.00
`atwar2`	0.08	0.28	0.00	0.00	1.00
`atwar3`	0.09	0.29	0.00	0.00	1.00
`atwar4`	0.11	0.31	0.00	0.00	1.00
`atwar5`	0.07	0.26	0.00	0.00	1.00
`atwar6`	0.18	0.38	0.00	0.00	1.00
`atwar7`	0.14	0.35	0.00	0.00	1.00
`atwar8`	0.12	0.33	0.00	0.00	1.00
`atwar9`	0.16	0.37	0.00	0.00	1.00
`atwar10`	0.14	0.35	0.00	0.00	1.00
`atwar11`	0.14	0.34	0.00	0.00	1.00
`atwarns`	0.14	0.35	0.00	0.00	1.00
`gdpl1`	3.78	4.52	2.05	0.26	51.99
`grol1`	0.02	0.07	0.02	-0.55	0.95
`inst3l1`	0.13	0.34	0.00	0.00	1.00
`anoc2l1`	0.18	0.38	0.00	0.00	1.00
`oil2l1`	0.13	0.33	0.00	0.00	1.00
`ef1`	0.48	0.27	0.51	0.00	1.00
`lpopnsl1`	15.93	1.49	15.84	12.33	20.84
`mtnl1`	17.61	21.07	9.30	0.00	94.30
`muslim`	26.42	37.35	3.00	0.00	100.00
`atwarns2`	0.14	0.35	0.00	0.00	1.00

(a) Summary statistics for data

variable	unique	min	max
`cid`	133	1	172
`year`	38	1954	1991

(b) Unique identifier variables

Table 2: Data summary. See Table 1 for variable descriptions.

state 1 is peace.

To model this data as a HMM, I need to specify functional forms of the two response distributions ($p(Y_t|S_t = k$ for $k = 0, 1$), the two transition distributions ($p(S_t|S_{t-1} = k)$ for $k = 0, 1$), and the initial distribution ($p(S_1)$).

First, consider the response distributions. The response $y_{t,i}$ is a binary vector of length $|J|$, i.e. it has a value for each of the civil war prevalence variables. Thus the response distributions for each state must be multivariate discrete distributions. For computational simplicity, I will assume that the variables are independent, and that the distribution of $y_{t,i}$ is the product of independent Bernoulli distributions,

$$p(y_{t,i}|S_{t,i} = k) = \prod_{j \in J} \mathsf{Bernoulli}(y_{t,i}, v_{k,j}) \qquad \text{for } k \in K, \tag{9}$$

where $v_{k,j}$ is the vector of parameters for the response j when $S_{t,i} = k$, and $\mathsf{Bernoulli}(x, p) = p^x(1-p)^x$ is the Bernoulli probability mass function. Equation (9) means that the distribution from which the observed civil war prevalence variables are drawn is a mixture of two distributions: one when the latent state is peace and one when it is civil war.

Second, consider the transition distributions. These are $p(S_t = k|S_{t-1} = j)$ where $j, k \in \{0, 1\}$. Since the state space is binary, these are Bernoulli distributions. Substantively, $p(S_t = 1|S_{t-1} = 0)$ is the probability of civil war onset, i.e. a transition from peace at $t-1$ to civil war at t. Likewise, $p(S_t = 1|S_{t-1} = 1)$ is the probability of civil war continuation, i.e. a transition from civil war at $t-1$ to civil war at t. Since the transition distributions are models of civil war onset and continuation, I include the covariates from the models in Sambanis (2004, Table 7) and described in the previous section. The transition distributions are thus,

$$\Pr(S_{t,i} = s_{t,i}|s_{t-1,i} = k) = \mathsf{Bernoulli}(s_{t,i}, \Lambda(\tau_k x_{t,i})) \quad \text{for } k \in 0:1,\, t \in 2:|T|\ . \tag{10}$$

where $\Lambda(x) = \frac{1}{1+e^{-x}}$ is the logistic function and x is the design matrix,

$$x = \begin{bmatrix} 1 & \texttt{gdp1} & \texttt{grol1} & \texttt{inst3l1} & \texttt{anoc21l1} & \texttt{oil21l1} & \texttt{lpopnsl1} & \texttt{mtnl1} & \texttt{muslim} \end{bmatrix}. \tag{11}$$

In other words, the transition distributions are simply logit models of civil war onset, when $S_{t-1} = 0$, and continuation, when $S_{t-1} = 1$.

Finally, the initial distribution is

$$\Pr(S_{1,i} = s_{1,i}) = \text{Bernoulli}(s_{1,i}, \mu). \tag{12}$$

Equation (12) makes the simplifying assumption that the probability that the latent state is civil at $t = 1$ is the same for all countries. Later, I include covariates in (12) to allow the distribution of the initial latent state to vary between countries.

The parameters θ of this model are estimated with the EM algorithm as described in Section 4 and 7.1. This paper uses the implementation of this algorithm in the R package **depmixS4** (Visser and Speekenbrink 2010).

The likelihood functions of mixture models, including HMM, often have multiple maxima. Thus the MLE and EM estimates are sensitive to the choice of initial parameter values (Zucchini and MacDonald 2009, pp. 49-50,91). For each HMM I estimate, the EM is seeded with 30 different starting values. The results presented are the parameters set with the highest log-likelihood.

5 Results

The parameter estimates of the model defined in the previous section are tabulated in Table 4. I divide the discussion of the results into two sections. Section 5.1 discusses the classification of observations into latent states. Section 5.2 discusses the results of the transition equations.

5.1 Latent States

The HMM estimates latent states for each country-year; since these latent states correspond in meaning to civil war and peace, the estimated latent states provide a new list of civil war incidence with a definition of civil war that is the latent concept giving rise to the civil war prevalence variables considered. In this section, I summarize the results of the classification of observations as civil wars, compare the classification to the prevalence variables, and compare the model with several variations to understand which assumptions of the HMM are driving the classification results.

As discussed in Section 2, there are several methods to recover estimates of the latent states for each observation. For this discussion, I consider the results of the global decoding, which finds the most likely sequence of states, to be the best estimate of the latent state for each observation.[11]Let $\hat{s}_{t,i}$ be the most likely state of country i at time t, using the global decoding, When $\hat{s}_{t,i} = 1$, the country-year (t, i) is classified as being in a state of civil war, and when $\hat{s}_{t,i} = 0$ the observation is classified as not being in war.

The estimated latent states suggest that the concept of civil war that underlies the civil war prevalence variables is inclusive, considering more observations to be civil wars than most of the individual civil war prevalence variables. The civil war latent state is more similar to a union of the civil war prevalence variables than an intersection of those variables. The model classifies 0.15 of the country-years to be in a state of civil war. This is a large number of civil war observations than all but two of the response variables, (Gleditsch et al. 2002, all conflicts) and `atwar9` and (Regan 1996) `atwar9`; see Figure 3. The fraction classified as wars is one tenth higher than that of the Sambanis variables (`atwar10`, `atwar11`, `atwarns2`) and the Fearon and Laitin (2003) civil war list (`atwar7`). The Correlates of War (COW) variables, `atwar1` and `atwar2`, classify the fewest observations as civil wars; that is almost half the number as the model estimates to be in civil war.

As would be expected by the high fraction of observations classified as wars, the model classifies an observation as a civil war whenever a few of the response variables indicate the observation is a civil war, as shown in Figure 4. A rule that approximates the results of the model is that an observation is certainly not a civil war if less than three variables indicate it was a war, has an even chance of being a civil war if three variables indicate it is a civil war, and is certainly a civil war if four or more variables indicate it is a civil war.

The values of the individual response variables can be compared to the estimated latent states to determine the similarity of each response variable to the latent concept that underlies them. I assess this similarity using six metrics commonly used in binary classification: sensitivity, specificity, negative predictive value, positive predictive value, Jaccard index, and Pearson correlation

[11] In this model, there is little difference between the global and local decodings. They disagree in only 7 observations of 3708 (0.002). In 4 of those disagreements, the global decoding classifies an observation as a war while the local decoding classified it as not a war. The estimated values of $p(s_{t,i})$ are in almost all cases either almost one or almost zero; in only 21 observations is $p(s_{t,i}) \in (0.025, 0.975)$.

coefficient. These metrics are defined as,

$$
\begin{aligned}
\text{Specificity}_j &= \Pr(Y_{t,i,j} = 1 | S_{t,i} = 1) && (13)\\
\text{Sensitivity}_j &= \Pr(Y_{t,i,j} = 0 | S_{t,i} = 0) && (14)\\
\text{Positive predicted value (PPV)}_j &= \Pr(\hat{s}_{t,i} = 1 | y_{t,i,j} = 1) && (15)\\
\text{Negative predicted value (NPV)}_j &= \Pr(\hat{s}_{t,i} = 0 | y_{t,i,j} = 0) && (16)\\
\text{Correlation}_j &= \operatorname{Cov}(\hat{s}_{t,i}, y_{t,i,j}) / \operatorname{Var}(\hat{s}_{t,i}) \operatorname{Var}(y_{t,i,j}) && (17)\\
\text{Jaccard index}_j &= \frac{\sum_{t,i}(\hat{s}_{t,i} = 1 \wedge y_{t,i,j} = 1))}{\sum_{t,i}(\hat{s}_{t,i} = 1 \vee y_{t,i,j} = 1)}. && (18)
\end{aligned}
$$

Specificity and sensitivity follow from the response distributions (9). Specificity is the probability that the response is civil war, given that the latent state is civil war. Sensitivity is the probability that the response variable is peace, given that the latent state is peace. The positive predictive value (PPV) and negative predictive value (NPV) are the converses of specificity and sensitivity. The PPV is the probability that the latent state is civil war, given that the response is one, while the NPV is the probability that the latent state is peace, given the response in zero. The Pearson correlation coefficient and the Jaccard index measure the overall similarity between the response variables and the latent state. While the correlation gives roughly equal weight to agreement in both 0's and 1's, the Jaccard index only considers agreements in 1's, i.e. wars. Given the rarity of wars, the Jaccard index is the preferred similarity metric in this application. However, in the results the correlation coefficient and Jaccard index give qualitatively equal results.

Figure 5 plots the values of the six classification metrics for each civil war prevalence variable. The first thing to note is that the variation in the similarity of response variables to the latent states is primarily due to differences in false negatives, observations classified as wars in the latent state but not by the response variable, which is a result of the latent state classifying more observations as wars. As a result, the specificity of all response variables is almost above 0.95; whenever the latent state is peace, the response variables almost certainly indicate a war. The variation occurs in the probability of indicating a war when the latent state is a war; the sensitivity varies between 0.5 and 0.9.

The values of the Jaccard index and the correlation coefficient show the variables that are most similar to the latent states: the three Sambanis variables (atwar10, atwar11, atwarns2). The variable atwarns2 is the coding rule developed in Sambanis (2004). After he compared the substantive content of the other eleven indicators. Based on these results, atwarns2 does a good job incorporating the concepts of the various coding rules considered in that paper, and by extension this paper. However, it did not do much better at that task than the coding rules used in his previous papers atwarns10 and atwarns11.[12] On the opposite end, the variables that are the least similar to the latent states, are the COW variables (atwar1, atwar2) and Gleditsch et al. (2002, wars only) (atwar5). These are the variables with the most restrictive definition of civil war — requiring at least 1,000 battle deaths total in the case of the COW variables and 1,000 battle related deaths per year in atwar5. While the latent states tend to classifies observations as war when any variable indicates it is a war, a few variables stand out in the specificity and PPV as producing a higher rate of false positives than the other variables: atwar6, atwar9, atawr8, and atwar7.

model	initial	response	transition
$\mathcal{M}_1$	(12)	(9)	(10)
$\mathcal{M}_2$	(12)	(9)	(19)
$\mathcal{M}_3$	(12)	(9)	None

Table 3: Definitions of models compared in Figure 6. The columns specify the equation for the initial, response, and transition distributions for that model.

The HMM contains several moving parts — the transition, response, and initial distributions. To understand the relative importance of these components in producing the results considered so far, I estimate models with different specifications. Let $\mathcal{M}1$ be the model discussed in this section and defined by equations (9), (10), and (12). The first alternative to this model is to remove the covariates from the transition distribution. I will refer to this model as $\mathcal{M}_2$ and it is defined like model $\mathcal{M}1$, except that the transition distribution is

$$\Pr(S_{t,i} = s_{t,i} | s_{t-1,i} = k) = \text{Bernoulli}(s_{t,i}, \tau_k) \quad \text{for } k \in 0:1,\ t \in 2:|T|\ . \tag{19}$$

The second alternative model, $\mathcal{M}3$, drops the transition distribution. Thus, $\mathcal{M}3$, is a standard mixture or latent class model. Now the distribution of the latent state at time t is independent of

[12] Although it may be the case that including all three of these highly correlated variables as responses may be responsible for this result.

the latent state at time $t-1$. I compare the fit of these models the Bayesian Information Criterion (BIC); the BIC of each model is plotted in Figure 6. Model $\mathcal{M}1$ has the lowest BIC, and thus the best fit to the data of the models considered, followed by model $\mathcal{M}2$ which has no covariates in the transition distributions, and finally model $\mathcal{M}3$, which has no transition distribution. There is a large difference in the BIC between $\mathcal{M}2$ and $\mathcal{M}3$, and marginal difference in the BIC between $\mathcal{M}1$ and $\mathcal{M}2$. These results suggest that the time dependency in observations has a large influence on the classification of civil war observations. This result is consistent with the observation that many of the disagreements between the coding rules in these variables is due to how they treat the persistence of civil war. While including a transition model has a large impact on the fit of the model, including covariates in the transition distributions only slightly increases the fit. It is notable that despite the extra parameters, including covariates does influence the fit, meaning that not only would models of civil war onset and continuation benefit from more accurate civil war classifications, but also that better models of civil war onset and continuation could help make more accurate classifications.

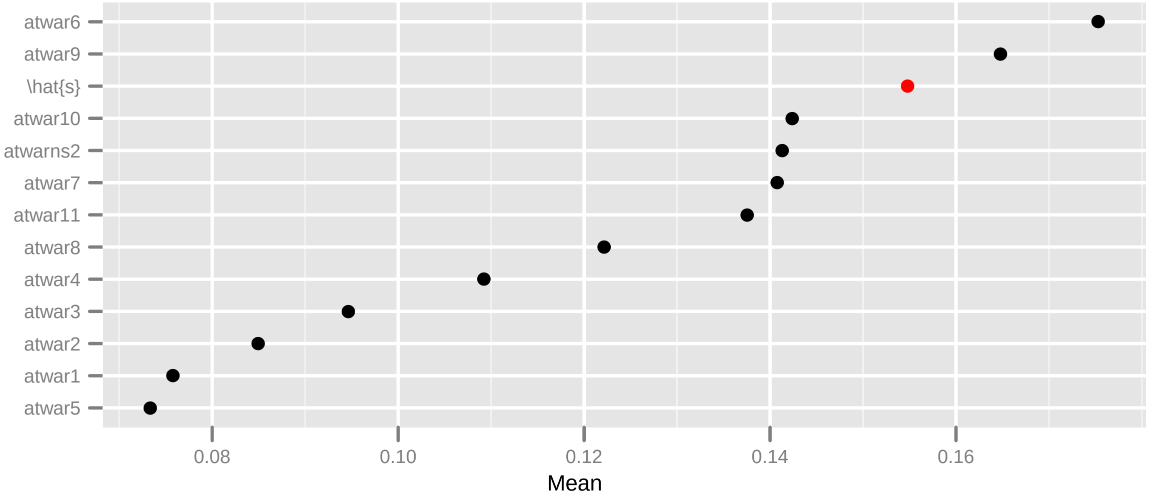

Figure 3: Fraction of wars for all response variables $y_{:,:,j}$ and the latent state $\hat{s}$.

5.2 Transition Distributions

The transition equations are the probability distributions for transitioning between latent states. Since the latent states are civil war and peace, the transition equations in this application are

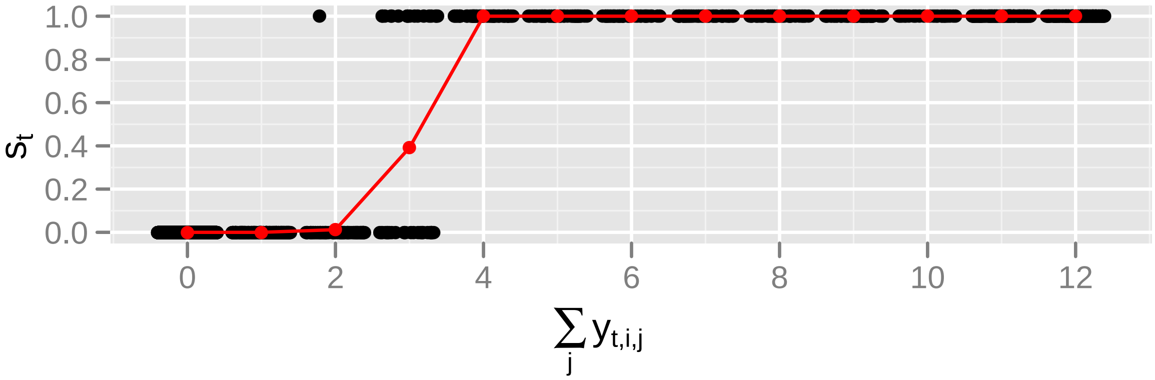

Figure 4: The latent state, $hats_{t,i}$, conditional on the number of response variables indicating war, $\sum_j y_{t,i,j}$. Points are values of $\hat{s}_{t,i}$, jittered horizontally to avoid overplotting. The line is the mean of $\hat{s}_{t,i}$, conditional on $\sum_j y_{t,i,j}$, the number of response variables indicating a civil war.

models of civil war onset, when the previous state is peace, and continuation, when the previous state is also civil war.

Figure 7 plots the parameter estimates of the transition distributions. For civil war onset, $p(S_t = 1|S_{t-1} = 0)$, the only parameters, excluding the intercept, with 95 percent confidence intervals excluding zero are GDP (`gdpl1`) and population (`lnpop1`). For civil war continuation, $p(S_t = 1|S_{t-1} = 1)$, no parameters, including the intercept, have 95 percent confidence intervals excluding zero.

The HMM used in this paper accounts for disagreement between the various civil war lists by estimating latent states from those measures and then estimating transitions between these latent states. The common method in the literature for dealing with this uncertainty is to estimate multiple models which have the same set of covariates but use different coding rules for the outcome variable of civil war. This is the approach taken by Sambanis (2004), which is the source of the data, response variables, and covariates in this paper. To compare the results of the HMM with estimating a transition model for each response variable, I estimate the following full-transition logit models for each response variable $j \in J = \{\texttt{atwar1} : \texttt{atwar11}, \texttt{atwarns2}\}$,

$$p(\{y_{t,i,j}|y_{t,i,j} = 0\}|x, \tilde{\tau}_{0,j}) = \prod_t \prod_i \text{Bernoulli}(y_{t,i,j}, \Lambda(s_{t,i}\tilde{\tau}_{1,j})) \tag{20}$$

$$p(\{y_{t,i,j}|y_{t,i,j} = 1\}|x, \tilde{\tau}_{1,j}) = \prod_t \prod_i \text{Bernoulli}(y_{t,i,j}, \Lambda(x_{t,i}\tilde{\tau}_{0,j})). \tag{21}$$

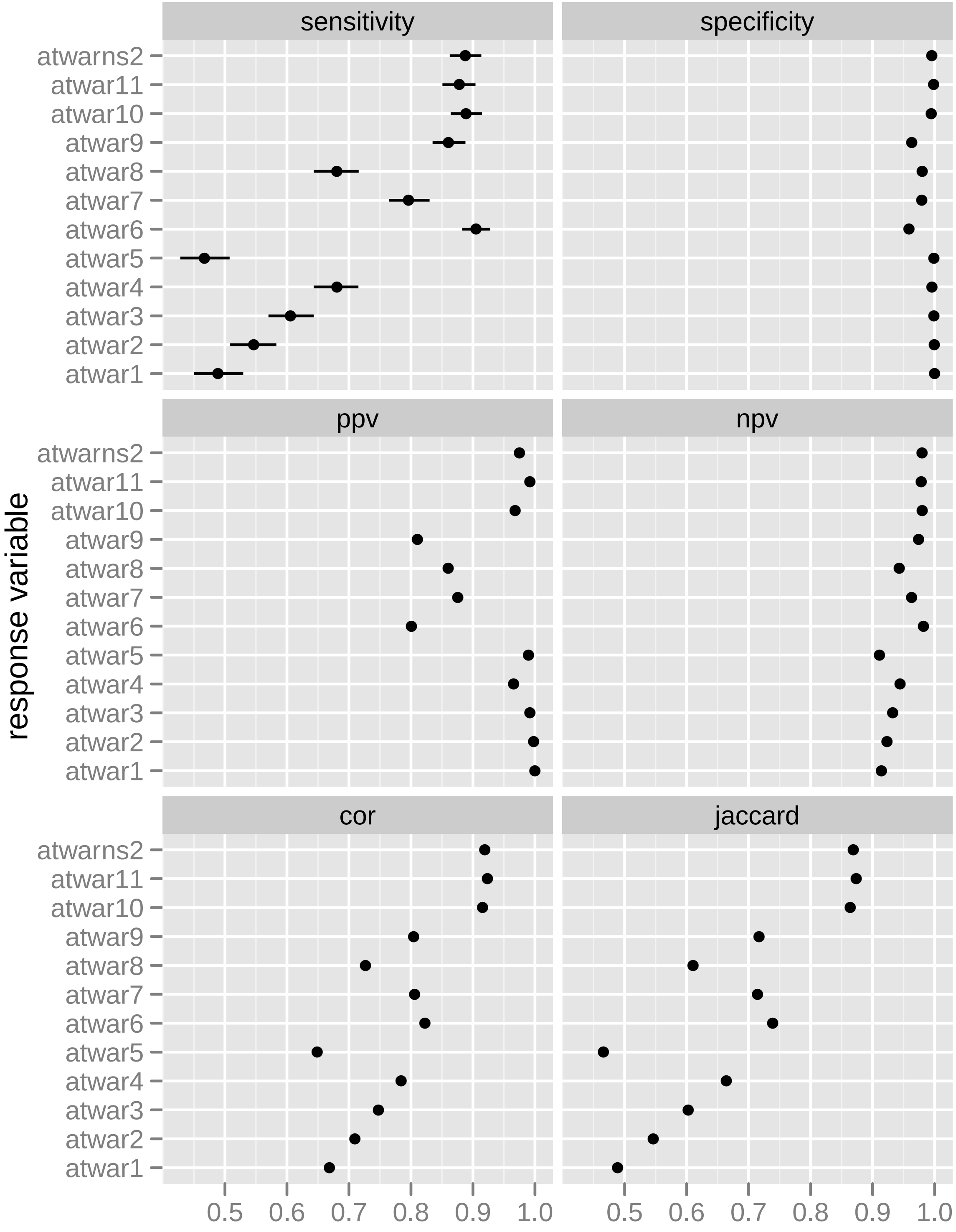

Figure 5: Comparison of response variables to latent states. Response variables are on the y-axis. Values of each metric are on the x-axis. NPV is negative predictive value. PPV is positive predictive value. cor is the Pearson correlation coefficient. See equations (13)-(18) for definitions of the values in these plots. Points are the EM estimate, while the linerange is the 95 percent parametric bootstrap confidence interval.

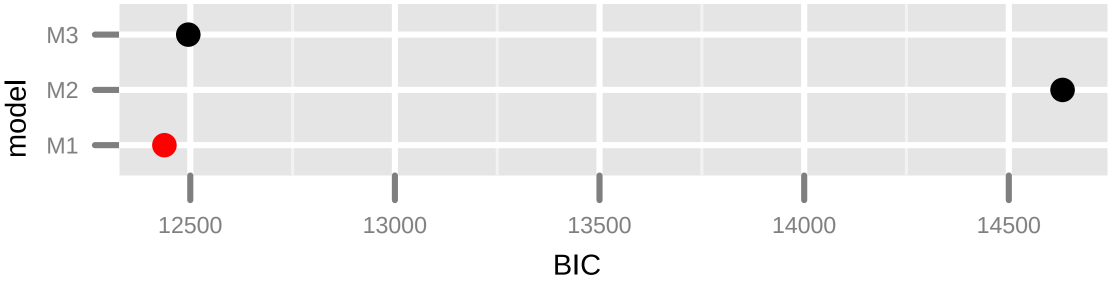

Figure 6: Model comparison with BIC. See 3 for definitions of the models.

In the following discussion, I will collectively refer to these models as the "single response models". The parameter estimates for these models are referenced by the name of the response variable, e.g. `atwar1`, and plotted in Figure 7.[13] The covariate coefficients of the transition distributions of the HMM are compared to the corresponding covariate coefficients for all single response models. For example, $\tau_0(\texttt{gdpl1})$ is the coefficient on `gdpl1` in $p(S_t = 1|S_{t-1} = \tau, x)$. The corresponding parameters in the single response models are $\theta_{0,j}(\texttt{gdpl1})$ for all response variables $j \in J$ from (20).

The results of the HMM are qualitatively similar to the collective results of single response models. In these models, a covariate parameter in the HMM is significant if its statistical significance is robust to changes in the choice of response variable in the single response models.[14] Of the onset parameters, the intercept, `gdpl1`, and `grol1` are significant in the HMM, and are either significant for all or all but one outcome variable. All the other transition parameters are insignificant in the HMM and also insignificant for four or more of the single response models.

However, accounting for the differences in coding rules for civil war using a HMM is not the same as simply taking the average of the estimates of the single response models. Figure 8 and Figure 9 compare the point estimates and the standard errors of the parameters from the HMM and the single response models. In both figures, the values are normalized such that the HMM value equals zero if it is the mean of the values of the single response models. In neither figure are the HMM values always near 0. For the point estimates, the HMM value can actually be quite extreme. For

[13] The parameter estimates reported in this paper differ from those in Hegre and Sambanis (2006, Table 7) because a logit link is used instead of a probit link, and in this paper, all models use the same set of observations, while in Sambanis (2004) the models have different numbers of observations due to missing values.

[14] This is not a general proposition about HMM.

both `grol1` (onset) and `inst3l1` (continuation), the HMM point estimate is more extreme than any point estimate in the single response models. In the case of `inst3l1`, the HMM point estimate has the opposite sign of the point estimates in eleven of the twelve single response models. Although the HMM accounts for the uncertainty in the definition of civil war, the standard errors of the HMM parameters in the transition equations are neither systematically higher nor the mean of those of the equivalent parameters in the single response models; in nine of the twenty parameters estimated, the standard errors of the HMM model is lower than the mean of the standard errors of the single response models. That being said, for several parameters — `oil2l1` in the onset distribution and `oil2l1`, `anoc2l1`, and `inst3l1` in the continuation distribution — the standard error is larger than the standard error of any of the single response models.

6 Conclusion

While this paper uses a HMM to model civil war onset and continuation, HMMs are a general class of models that could be useful for modeling many political science phenomena. HMM could be useful in modeling any data in which the observed outcomes are imperfect measures, the latent concept is discrete, and the data are time-series or longitudinal, for example democratic transitions, state failure, and dyadic rivalry. All of these are discrete concepts with imperfect measures in which the substantive interest revolves around transitions.

As it stands, this paper needs and the statistical model allows it to be extended in several possibly interesting ways. These are a few ideas, but I implore the reader to provide his or her input as to which would be the most valuable to pursue.

- Increase the number of states in the state space. Perhaps there multiple “types” of civil wars.
- Include population as a covariate in the response distributions. This would address Sambanis 2004’s critique use of the absolute values of casualties as a threshold in civil war definitions as biasing civil war prevalence to larger countries.
- Allow transition probabilities to depend on time spent in the latent state. This could be incorporated by expanding the state space. Otherwise, to include the time spent in the latent state would require switching the estimation method to a Gibbs sampler.

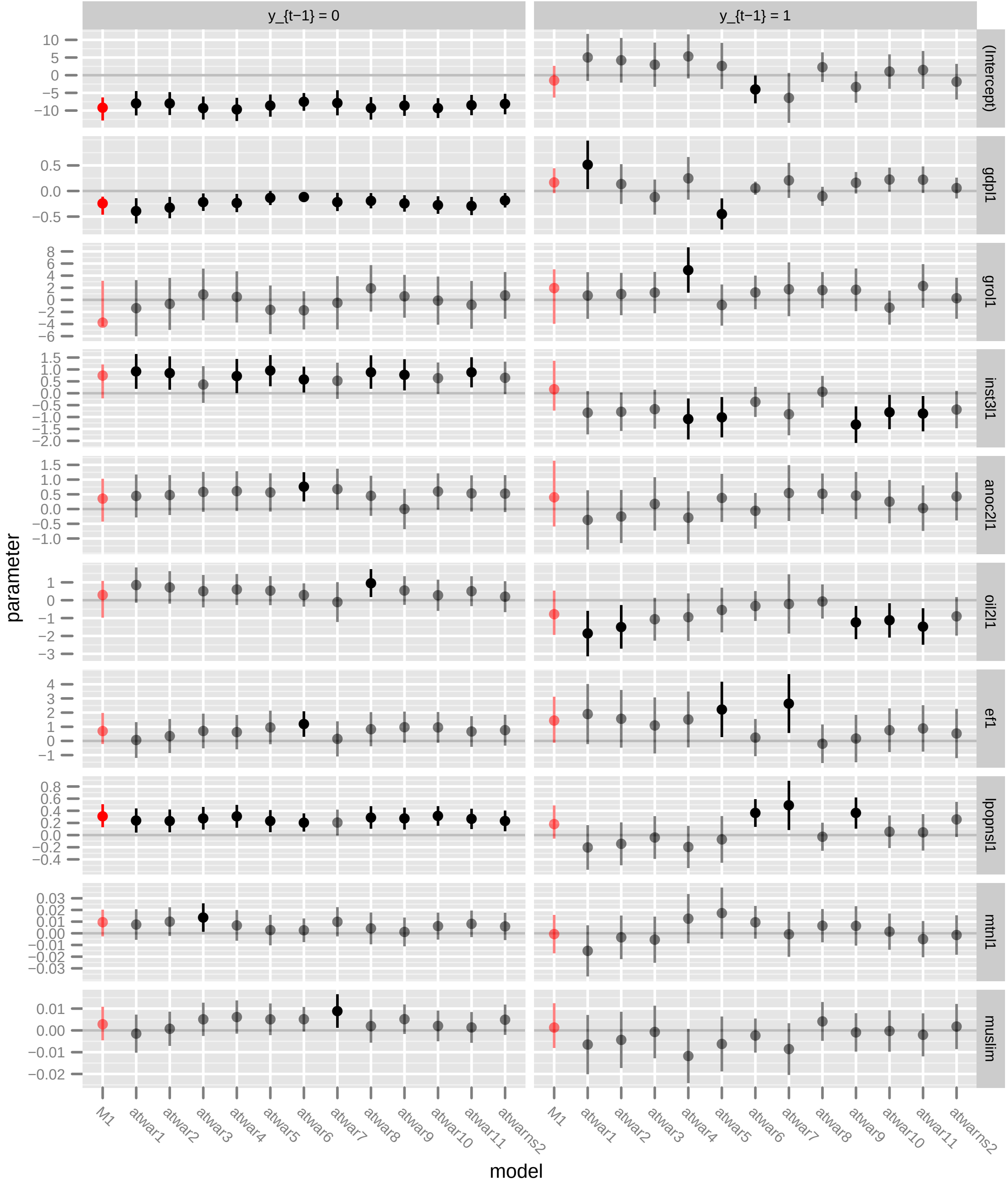

Figure 7: Estimates of transition equation parameters τ from the HMM $\mathcal{M}_1$ and $\tilde{\tau}$ in the logit models for each response variable. The point is the point estimate, the linerange is the 95 percent confidence interval. For the logit models the confidence intervals are calculated from the Hessian; for model $\mathcal{M}_1$ the confidence interval is calculated from a parametric bootstrap. The parameters from $\mathcal{M}_1$ are colored red. Parameters which have 95 percent confidence intervals that cross zero are darker than those which do not.

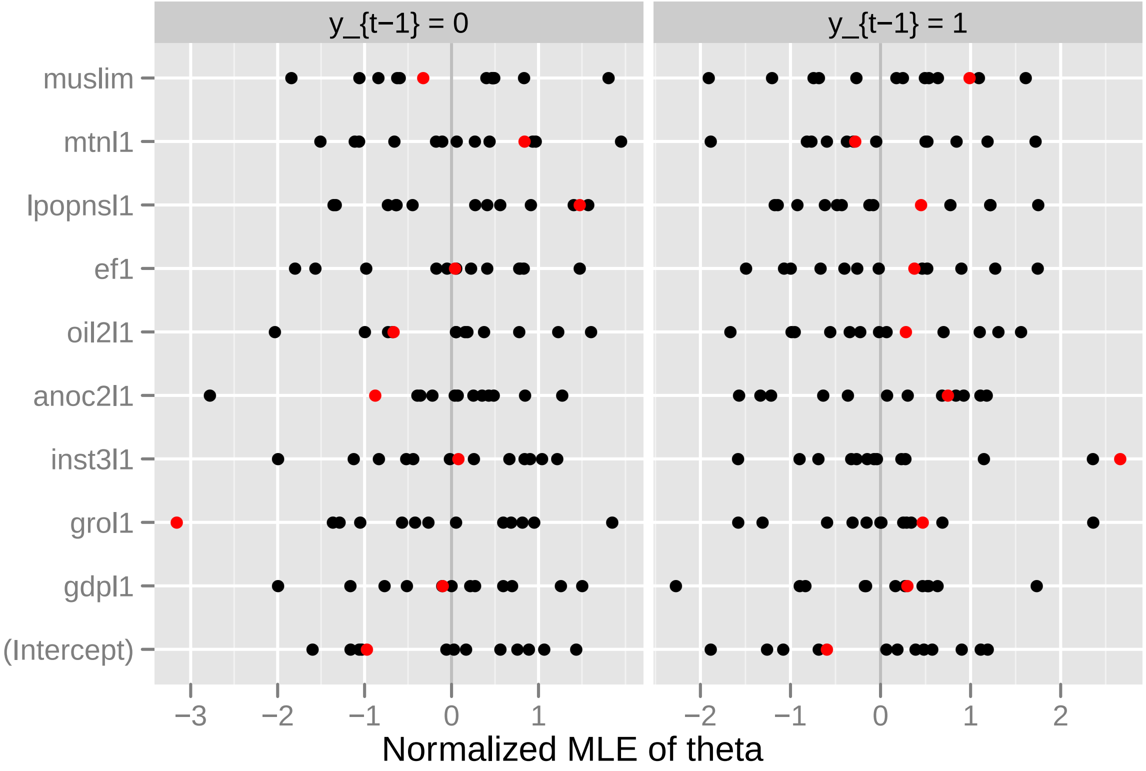

Figure 8: Point estimates of the transition parameters τ in model $\mathcal{M}_1$ and $\tilde{\tau}$ from the logit models of civil war. The value from the HMM variance is colored red, while the values from the logit models are colored black. The values are transformed such that the values of the logit models have a mean of 0 and a standard deviation of 1.

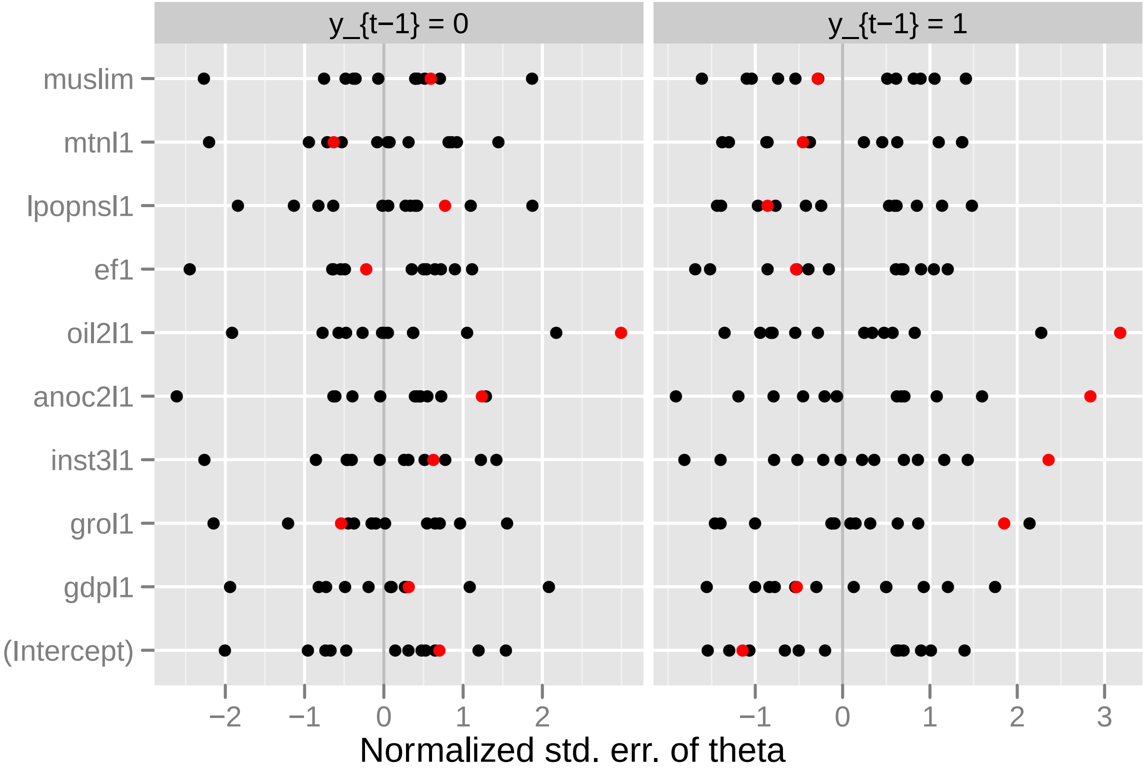

Figure 9: Variance of the parameter estimates of the transition parameters τ in model $\mathcal{M}_1$ and $\tilde{\tau}$ from the logit models of civil war. The value from the HMM variance is colored red, while the values from the logit models are colored black. The values are transformed such that the values of the logit models have a mean of 0 and a standard deviation of 1.

- Do not drop observations that are missing one of more dependent variables. Use either data augmentation or altering the likelihood function.
- Use different set of civil war prevalence variables. The Sambanis (2004) data are a bit dated; should I use a more current set of variables?
- Use a better model of civil war onset and offset?

7 Appendix

7.1 Forward-Backward Algorithm

The EM algorithm in Section 2 requires two intermediate quantities: the forward and backward probabilities. They are so-called because they are calculated by recursions starting at the first and last observation in the data, respectively.[15]

The forward probabilities are the joint probability of the state at time t and the sequence of observations up though time t,

$$\alpha_t(k) = p(Y_{1:t} = y_{1:t}, S_t = k). \tag{22}$$

The forward probabilities are computed using the following recursion,

$$\alpha_1(k) = p(S_1 = k)p(Y_1 = y_1 | S_1 = k) \tag{23}$$

$$\alpha_t(k) = \sum_{j \in K} \alpha_{t-1}(j) p(S_t = k | S_{t-1} = j) p(Y_t = y_t | S_t = k), \tag{24}$$

for all $t \in T$ and $k \in K$.

The backward probabilities are the conditional probabilities of the sequence of observations starting at $t+1$ conditional on the latent state at time t,

$$\beta_t(k) = p(Y_{(t+1):|T|} = y_{(t+1):|T|} | S_t = k) \qquad \text{if } p(S_t = k) > 0\ . \tag{25}$$

[15] See Zucchini and MacDonald (Section 4.1 2009, p. 62) for more detail on these derivations.

The backward probabilities are computed using the following recursion,

$$\beta_{|T|}(k) = 1 \tag{26}$$

$$\beta_t(k) = \sum_{j \in K} p(S_{t+1} = j | S_t = k) p(Y_{t+1} = y_{t+1} | S_{t+1} = j) \beta_{t+1}(j)), \tag{27}$$

for all $t \in T$ and $k \in K$.

Several important results follow from these definitions of the forward and backward probabilities. In particular, the likelihood is a product of the forward and backward probabilities. For any $t \in T$,

$$\begin{aligned}\alpha_t(k)\beta_t(k) &= p(Y_{1:t} = y_{1:t}, S_t = k) p(Y_{(t+1):|T|} = y_{(t+1):|T|} | S_t = k) \\ &= p(S_t = k) p(Y_{1:t} = y_{1:t} | S_t = k) p(y_{(t+1),|T|} | S_t = k) \\ &= p(Y_{1:|T|} = y_{1:|T|}, S_t = k).\end{aligned} \tag{28}$$

Summing over all the latent states gives the likelihood,

$$\sum_{k \in K} p(Y_{1:|T|} = y_{1:|T|}, S_t = k) = p(Y_{1:|T|} = y_{1:|T|}) = p(y). \tag{29}$$

Note that since these derivations work for any t, there are $|T|$ ways to calculate the likelihood using the forward and backward probabilities. In practice, the likelihood is calculated with $t = |T|$, since that calculation only requires calculating the forward probabilities, and thus a single pass through the data.

The E-step of the EM algorithm, Equation (6), requires the expected values $p(S_t = k, S_{t-1} = j | Y = y)$ and $p(S_t = j = k | Y = y)$. Using the forward and backward probabilities,

$$p(S_t = k | Y = y) = \frac{\alpha_t(k)\beta_t(k)}{p(Y = y)} \tag{30}$$

$$p(S_t = k, S_{t-1} = j | Y = y) = \frac{\alpha_{t-1}(j) p(S_t = k | S_{t-1} = j) p(Y_t = y_t | S_t = k) \beta_t(k)}{p(Y = y)}. \tag{31}$$

7.2 Decoding

With HMM, *decoding* is process of finding the most likely latent states to have given rise to the observed data conditional on the model and parameters. There are two definitions of the most likely latent state. *Local decoding* finds the most likely state at a given time. *Global decoding* finds the most likely sequence of states.[16] From (28) it follows that

$$\begin{aligned} p(S_t = k | Y = y) &= \frac{p(S_t = k, Y = y)}{Y = y} \\ &= \frac{\alpha_t(k)\beta_t(k)}{p(Y = y)} \end{aligned} \tag{32}$$

Then most likely state in time t is simply,

$$\hat{s}^*_t = \operatorname*{argmax}_{k \in K} p(S_t = k | Y = y) \tag{33}$$

While (33) finds the most likely state at each time, global decoding finds the most likely sequence of states $s_{1:|T|}$,

$$\hat{s}_t = \operatorname*{argmax}_{s_{1:|T|}} p(S_{1:|T|} = s_{1:|T|} | Y = y). \tag{34}$$

The most likely sequence of states if found using a recursive algorithm known as the Viterbi algorithm. Define the following recursions,

$$\begin{aligned} V_{1,k} &= p(S_1 = k, Y_1 = y_1) \\ V_{t,k} &= \left(\max_k (V_{t-1,j} p(S_t = k | S_{t-1} = j)) \right) p(Y_t = y_t | S_t = k) \quad \text{for } t = 2 : |T|. \end{aligned} \tag{35}$$

The values of V form a $|K| \times |T|$ matrix. To find the most likely sequence given the probabilities from (35), start with the most likely state at time $|T|$ and then work backwards through time finding the state in $t-1$ that is most likely to have transitioned to the most likely state at t,

$$\begin{aligned} \hat{s}_{|T|} &= \operatorname*{argmax}_{k \in K} V_{|T|,k} \\ \hat{s}_t &= \operatorname*{argmax}_{k \in K} V_{t,k} p(S_t = k | S_{t+1} = \hat{s}_{t+1}) \quad \text{for } t = (|T| - 1) : 1. \end{aligned} \tag{36}$$

[16] See Zucchini and MacDonald (2009, Section 5.3).

value	mean	sd	median	p2.5	p97.5
0.90	0.90	0.03	0.90	0.85	0.95
0.10	0.10	0.03	0.10	0.05	0.15

(a) Initial state parameter estimates v (10) from the response distributions, $p(s_1|v)$

states	covariate	value	mean	sd	median	p2.5	p97.5
0	(Intercept)	-9.24	-9.37	1.69	-9.29	-12.82	-6.28
0	gdpl1	-0.24	-0.26	0.09	-0.25	-0.46	-0.11
0	grol1	-3.77	-0.69	1.94	-0.57	-4.40	3.15
0	inst3l1	0.74	0.55	0.37	0.55	-0.22	1.21
0	anoc2l1	0.35	0.37	0.37	0.40	-0.43	1.03
0	oil2l1	0.29	0.20	0.62	0.26	-0.98	1.08
0	ef1	0.70	0.85	0.58	0.84	-0.22	1.97
0	lpopnsl1	0.31	0.31	0.10	0.32	0.13	0.51
0	mtnl1	0.01	0.01	0.01	0.01	-0.00	0.02
0	muslim	0.00	0.00	0.00	0.00	-0.00	0.01
1	(Intercept)	-1.49	-1.70	2.23	-1.56	-6.33	2.65
1	gdpl1	0.17	0.18	0.12	0.17	-0.04	0.44
1	grol1	1.94	0.16	2.20	0.09	-3.96	5.05
1	inst3l1	0.17	0.16	0.51	0.12	-0.73	1.37
1	anoc2l1	0.40	0.43	0.59	0.39	-0.59	1.64
1	oil2l1	-0.78	-0.73	0.95	-0.81	-1.94	0.54
1	ef1	1.44	1.43	0.84	1.41	-0.13	3.13
1	lpopnsl1	0.18	0.20	0.13	0.19	-0.06	0.49
1	mtnl1	-0.00	-0.00	0.01	-0.00	-0.02	0.02
1	muslim	0.00	0.00	0.01	0.00	-0.01	0.01

(b) Transition equation parameter estimates of v for $p(S_{t,i} = 1|s_{t-1,i}, v, x)$ where s_{t-1} takes the value in the "states" column. If "states" is zero, then the parameters are associated with a transition from peace to civil war. If "states' is one, then the parameters are associated with a continuation of civil war.

Table 4: Parameter estimates from the hidden Markov model of civil war prevalence defined by Equations (9), (10), and (12). The "value" column is the EM estimate of the parameter; the other summary statistics are of the posterior distribution estimated from a re-weighted parametric bootstrap.

states	response	value	mean	sd	median	p2.5	p97.5
0	atwar1	0.00	0.00	0.00	0.00	0.00	0.00
0	atwar2	0.00	0.00	0.00	0.00	0.00	0.00
0	atwar3	0.00	0.00	0.00	0.00	0.00	0.00
0	atwar4	0.00	0.00	0.00	0.00	0.00	0.01
0	atwar5	0.00	0.00	0.00	0.00	0.00	0.00
0	atwar6	0.04	0.04	0.00	0.04	0.03	0.05
0	atwar7	0.02	0.02	0.00	0.02	0.02	0.03
0	atwar8	0.02	0.02	0.00	0.02	0.02	0.03
0	atwar9	0.04	0.04	0.00	0.04	0.03	0.04
0	atwar10	0.01	0.01	0.00	0.01	0.00	0.01
0	atwar11	0.00	0.00	0.00	0.00	0.00	0.00
0	atwarns2	0.00	0.00	0.00	0.00	0.00	0.01
1	atwar1	0.49	0.49	0.02	0.49	0.45	0.53
1	atwar2	0.55	0.55	0.02	0.55	0.51	0.58
1	atwar3	0.60	0.61	0.02	0.60	0.57	0.64
1	atwar4	0.68	0.68	0.02	0.68	0.64	0.72
1	atwar5	0.47	0.47	0.02	0.47	0.43	0.51
1	atwar6	0.91	0.91	0.01	0.91	0.88	0.93
1	atwar7	0.80	0.80	0.02	0.80	0.76	0.83
1	atwar8	0.68	0.68	0.02	0.68	0.64	0.72
1	atwar9	0.86	0.86	0.01	0.86	0.83	0.89
1	atwar10	0.89	0.89	0.01	0.89	0.86	0.91
1	atwar11	0.88	0.88	0.01	0.88	0.85	0.90
1	atwarns2	0.89	0.89	0.01	0.89	0.86	0.91

(c) Response equation parameter estimates τ for $p(Y_{t,i,j} = 1|s_{t,i}, \tau)$, where j indexes the different response variables in the “response” column, and $s_{t,i}$ is given in the “states” column.

Table 4: Parameter estimates from the hidden Markov model of civil war prevalence defined by Equations (9), (10), and (12). The “value” column is the EM estimate of the parameter; the other summary statistics are of the posterior distribution estimated from a re-weighted parametric bootstrap.

References

Beck, Nathaniel et al. (July 13, 2001). “Alternative Models of Dynamics in Binary Time-Series-Cross-Section Models: The Example of State Failure”. In: *2001 Annual Meeting of the Society of Political Methodology*. Emory University. (Cit. on p. 3).

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in hidden Markov models*. Springer series in statistics. Springer. ISBN: 9780387402642. URL: http://books.google.com/books?id=-3_A3_l1yssC. (Cit. on p. 2).

Chib, Siddhartha (1996). “Calculating posterior distributions and modal estimates in Markov mixture models”. In: *Journal of Econometrics* 75.1, pp. 79 –97. ISSN: 0304-4076. DOI: 10.1016/0304-4076(95)01770-4. URL: http://www.sciencedirect.com/science/article/B6VC0-3VWT1TW-7/2/b66dae744497c95ff3eddb744f301a6c. (Cit. on pp. 2, 3).

— (1998). “Estimation and comparison of multiple change-point models”. In: *Journal of Econometrics* 86.2, pp. 221 –241. ISSN: 0304-4076. DOI: DOI:10.1016/S0304-4076(97)00115-2. URL: http://www.sciencedirect.com/science/article/B6VC0-3VM1XM5-2/2/469ee3cba827365611dee3677f0babc6. (Cit. on p. 3).

Collier, Paul and Anke Hoeffler (2001). *Greed and Grievance in Civil War*. Policy Research Paper 2355. World Bank. URL: http://go.worldbank.org/A3YEBBVS30. (Cit. on pp. 7, 9).

Doyle, Michael W. and Nicholas Sambanis (2000). “International Peacebuilding: A Theoretical and Quantitative Analysis”. In: *The American Political Science Review* 94.4, pp. 779–801. ISSN: 00030554. URL: http://www.jstor.org/stable/2586208. (Cit. on p. 9).

Fearon, James D. and David D. Laitin (2003). “Ethnicity, Insurgency, and Civil War”. In: *American Political Science Review* 97.01, pp. 75–90. DOI: 10.1017/S0003055403000534. URL: http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=142717&fulltextType=RA&fileId=S0003055403000534. (Cit. on pp. 7, 9, 13).

Frühwirth-Schnatter, Sylvia (2006). *Finite mixture and Markov switching models*. Springer series in statistics. Springer. ISBN: 9780387329093. URL: http://books.google.com/books?id=f8KiI7eRjYoC. (Cit. on pp. 2, 4).

Gleditsch, Nils Petter et al. (2002). "Armed Conflict 1946-2001: A New Dataset". English. In: *Journal of Peace Research* 39.5, pp. 615–637. ISSN: 00223433. URL: http://www.jstor.org/stable/1555346. (Cit. on pp. 9, 13, 15).

Hegre, Håvard and Nicholas Sambanis (2006). "Sensitivity Analysis of Empirical Results on Civil War Onset". In: *Journal of Conflict Resolution* 50.4, pp. 508–535. DOI: 10.1177/0022002706289303. eprint: http://jcr.sagepub.com/content/50/4/508.full.pdf+html. URL: http://jcr.sagepub.com/content/50/4/508.abstract. (Cit. on p. 19).

Jackman, Simon (Jan. 27, 2000). "In and Out of War and Peace: Transitional Models of International Conflict". working paper. URL: http://jackman.stanford.edu/papers/inandout.pdf. (Cit. on p. 3).

Leitenberg, Milton (2006). *Deaths in Wars and Conflicts in the 20th Century.* Peace Studies Program Occasional Paper 29. Cornell University. URL: http://www.einaudi.cornell.edu/PeaceProgram/publications/occasional_papers/Deaths-Wars-Conflicts3rd-ed.pdf. (Cit. on p. 9).

Licklider, Roy (1995). "The Consequences of Negotiated Settlements in Civil Wars, 1945-1993". English. In: *The American Political Science Review* 89.3, pp. 681–690. ISSN: 00030554. URL: http://www.jstor.org/stable/2082982. (Cit. on p. 9).

Park, Jong Hee (2009). "Joint Modeling of Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models". working paper. URL: http://polmeth.wustl.edu/media/Paper/ParkPolmethPaper_Final.pdf. (Cit. on p. 3).

— (2010). "Structural Change in U.S. Presidents' Use of Force". In: *American Journal of Political Science* 54.3, pp. 766–782. ISSN: 1540-5907. DOI: 10.1111/j.1540-5907.2010.00459.x. URL: http://dx.doi.org/10.1111/j.1540-5907.2010.00459.x. (Cit. on p. 3).

— (2011). "Changepoint Analysis of Binary and Ordinal Probit Models: An Application to Bank Rate Policy Under the Interwar Gold Standard". In: *Political Analysis.* DOI: 10.1093/pan/mpr007. eprint: http://pan.oxfordjournals.org/content/early/2011/03/22/pan.mpr007.full.pdf+html. URL: http://pan.oxfordjournals.org/content/early/2011/03/22/pan.mpr007.abstract. (Cit. on p. 3).

Rabiner, Lawrence R. (Feb. 1989). "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recogniation". In: *Proceedings of the IEEE* 77.2. (Cit. on p. 2).

Regan, Patrick M. (1996). “Conditions of Successful Third-Party Intervention in Intrastate Conflicts”. In: *Journal of Conflict Resolution* 40.2, pp. 336–359. DOI: 10.1177/0022002796040002006. eprint: http://jcr.sagepub.com/content/40/2/336.full.pdf+html. URL: http://jcr.sagepub.com/content/40/2/336.abstract. (Cit. on pp. 9, 13).

Richards, D. and D.R. Doyle (2000). *Political complexity: nonlinear models of politics*. University of Michigan Press. ISBN: 9780472109647. URL: http://books.google.com/books?id=TUIVkCnv70EC.

Sambanis, Nicholas (2004). “What Is Civil War? Conceptual and Empirical Complexities of an Operational Definition”. In: *Journal of Conflict Resolution* 48.6, pp. 814–858. DOI: 10.1177/0022002704269355. eprint: http://jcr.sagepub.com/content/48/6/814.full.pdf+html. URL: http://jcr.sagepub.com/content/48/6/814.abstract. (Cit. on pp. 1, 7, 9, 11, 15, 17, 19).

Sarkees, Meredith Reid and Phil Schafer (2000). “The Correlates of War Data On War: an Update To 1997”. In: *Conflict Management and Peace Science* 18.1, pp. 123–144. DOI: 10.1177/073889420001800105. eprint: http://cmp.sagepub.com/content/18/1/123.full.pdf+html. URL: http://cmp.sagepub.com/content/18/1/123.abstract. (Cit. on p. 9).

Schrodt, Philip A. (2006). “Forecasting Conflict in the Balkans using Hidden Markov Models”. In: *Programming for Peace*. Ed. by Robert Trappl. Vol. 2. Advances in Group Decision and Negotiation. Springer Netherlands, pp. 161–184. ISBN: 978-1-4020-4390-1. URL: http://dx.doi.org/10.1007/1-4020-4390-2_8. (Cit. on p. 3).

Schrodt, Philip A. and Deborah J. Gerner (2004). “An Event Data Analysis of Third-Party Mediation in the Middle East and Balkans”. In: *Journal of Conflict Resolution* 48.3, pp. 310–330. DOI: 10.1177/0022002704264137. eprint: http://jcr.sagepub.com/content/48/3/310.full.pdf+html. URL: http://jcr.sagepub.com/content/48/3/310.abstract. (Cit. on p. 3).

Schrodt, Phillip A. (2000). “Pattern Recognition of International Crises Using Hidden Markov Models”. In: *Political Complexity: Nonlinear Models of Politics*. Ed. by Diana Richards. University of Michigan Press. ISBN: 9780472109647. URL: http://books.google.com/books?id=TUIVkCnv70EC. (Cit. on p. 3).

Scott, Steven L. (2002). “Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century”. English. In: *Journal of the American Statistical Association* 97.457, pp. 337–351. ISSN: 01621459. URL: http://www.jstor.org/stable/3085787. (Cit. on p. 2).

Singer, J. David and Melvin Small (2006). *Correlates of War Project: International and Civil War Data, 1816-1992*. DOI: 10.3886/ICPSR09905. URL: http://dx.doi.org/10.3886/ICPSR09905. (Cit. on p. 9).

Svolik, Milas W. (Jan. 2009). “When and Why Democracies Consolidate: Coups, Incumbent Takeovers, and Democratic Survival”. working paper. URL: https://netfiles.uiuc.edu/msvolik/www/research/when.pdf. (Cit. on p. 3).

Treier, Shawn and Simon Jackman (2008). “Democracy as a Latent Variable”. In: *American Journal of Political Science* 52.1, pp. 201–217. URL: http://dx.doi.org/10.1111/j.1540-5907.2007.00308.x. (Cit. on p. 3).

Visser, Ingmar and Maarten Speekenbrink (Aug. 2010). “depmixS4: An R Package for Hidden Markov Models”. In: *Journal of Statistical Software* 36.7, pp. 1–21. ISSN: 1548-7660. URL: http://www.jstatsoft.org/v36/i07. (Cit. on pp. 2, 12).

Zucchini, W. and I.L. MacDonald (2009). *Hidden Markov models for time series: an introduction using R*. Monographs on statistics and applied probability. CRC Press. ISBN: 9781584885733. URL: http://books.google.com/books?id=LDDzvCsdVs8C. (Cit. on pp. 2, 4, 6, 12, 24, 26).