Dynamic Logistic Regression and Variable Selection: Forecasting and Contextualizing Civil Unrest

Abstract

Civil unrest can range from peaceful protest to violent furor, and researchers are working to monitor, forecast, and assess such events to better allocate resources. Twitter has become a real-time data source for forecasting civil unrest because the platform is used by millions of users as a social outlet. Daily word counts are used as model features and predictive terms contextualize the reasons for protest. In order to forecast civil unrest and also infer the reasons for the protest, we consider the problem of Bayesian variable selection for the dynamic logistic regression model and propose using penalized credible regions to select parameters of the updated state vector. This method avoids the need for shrinkage priors, is scalable to high-dimensional dynamic data, and allows the importance of variables to vary in time as new information becomes available. A substantial improvement in both precision and F1-score using this approach is demonstrated by means of simulation. Finally, we apply the proposed model fitting and variable selection methodology to the problem of forecasting civil unrest in Latin America. Our dynamic logistic regression approach shows improved accuracy compared to the static approach currently used in both event prediction and feature selection.

Keywords: Civil Unrest, Dynamic Logistic Regression, Forecasting, Pólya-Gamma Latent Variable, Penalized Credible Regions

1. Introduction

In recent years open source data has become a powerful tool for predicting real-world events. The widespread adoption of the social networking site Twitter has resulted in massive repositories of free information. Researchers mine user-generated content and use daily word counts to forecast diverse outcomes. For example, terms like "flu" and "fever" are leading indicators of influenza like illness rates reported by the Centers for Disease Control and Prevention in the United States Achrekar et al. (2012); Li and Cardie (2013); Culotta (2010). Similarly, Twitter has been used to predict stock price movement for financial markets (Bollen et al., 2011; Nisar and Yeung, 2018),

box office revenue for the film industry (Asur and Huberman, 2010), inner city crime rates for law enforcement (Gerber, 2014), and food supply chains (Singh et al., 2018).

One emerging application for open source data is civil unrest. A nation's citizens may protest for a myriad of reasons, from local economic conditions to national government oppression, and seemingly nonviolent civil unrest can escalate into deadly protests. Both social scientists and government officials have used Twitter to monitor active civil unrest and learn from the events post-unrest. For example, Twitter was leveraged to gain insights on the London riots (Panagiotopoulos et al., 2012), occupy Wall Street (Conover et al., 2013), and the Arab Spring (Eltantawy and Wiest, 2011; Lotan et al., 2011). Twitter is an advantageous data source because the platform allows people to voice their displeasure and discuss their individual motives for engaging in protest. Therefore, it effectively aggregates protest related information and researchers use daily word counts, such as "protest", "racism", and "police", to forecast civil unrest.

To better allocate resources during times of protest, researchers are actively developing models to forecast civil unrest. Qiao and Chen (2016) use Hidden Markov Models to predict the future sequence of anti-government protests in Thailand; Bahrami et al. (2018) compare Naive Bayes, Logistic Regression and Support Vector Machines to predict political protests against the president-elect in 2016; Ertugrul et al. (2019) use a long short-term memory neural network to predict followup protests to the Charlottesville rally in 2017; and Ramakrishnan et al. (2014) use an ensemble method to forecast civil unrest in Latin America. The current performance baseline of forecasting civil unrest at the country level produces F1 scores in the range of 68% to 95% (Korkmaz et al., 2016). Korkmaz et al. (2016) use logistic regression to forecast the probability of civil unrest in six different Latin American countries.

While the primary goal of forecasting civil unrest is prediction accuracy, an emerging secondary goal is variable selection. In the context of civil unrest, variable selection is used to contextualize the reasons for protest and find which social media features must be actively monitored as leading indicators of civil unrest. That is, if the term "police" is predictive of civil unrest, this may require a different allocation of resources than if the term "unemployed" is predictive. Most existing approaches do not consider feature selection for civil unrest forecasting. However, Korkmaz et al. (2016) and Ertugrul et al. (2019) use LASSO and Group LASSO respectively. The regularization allows the researchers to find predictive features in the high dimensional modeling scenario ($p > n$) and infer the reasons for protest.

In this paper, we propose to model civil unrest using a dynamic framework. Dynamic models are used to forecast complex non-stationary time series. The time-varying parameters of these state space regression models allow for greater flexibility in short-term forecasting, and the intrinsic Bayesian framework allows for sequential and efficient updating of model parameters as new information becomes available. Given the inherent dynamic structure of civil unrest data, a dynamic model allows for better short-term forecasting accuracy in comparison to the static logistic regression model baseline. Furthermore, the time-varying structure of dynamic models allows for parameters to evolve along with structural changes in the system over time. We propose using penalized credible regions to choose the nonzero elements of the dynamic parameters each time the model is updated in order to maintain a dynamic variable selection framework. Unlike existing dynamic models, this method is scalable, and we show improved performance over a LASSO baseline, resulting in better inference regarding the reasons for protest.

This paper is organized as follows. In Section 2.1, we discuss fitting dynamic linear and dynamic generalized linear models. In Section 2.2, we review dynamic variable selection methods. In Section 2.3, we propose a dynamic logistic regression model and variable selection technique and review its performance alongside the logistic regression with the LASSO regularization model in Section 3. Finally, we apply the model to the application of forecasting civil unrest in Latin America in Section 4 and conclude with a discussion in Section 5.

2. Methods

2.1. Dynamic Model Fitting

Fitting a dynamic linear model (DLM) is relatively straightforward. The Bayesian paradigm allows for sequential updating of states via the Kalman filter (Ferreira and Gamerman, 2000). Fitting more complex models relies on an extension of the Kalman filter and a Markov chain Monte Carlo (MCMC) Gibbs algorithm (Petris et al., 2009) framework. We first discuss fitting the DLM and later show how the methodology can be extended to fit a dynamic generalized linear model (DGLM).

Consider the univariate DLM specified by the *observation equation* (Equation 1a) and *state equation* (Equation 1b) for $t \geqslant 1$,

$$Y_t = \boldsymbol{X}_t \boldsymbol{\beta}_t + v_t \qquad v_t \sim N(0, \sigma^2) \tag{1a}$$

$$\boldsymbol{\beta}_t = \boldsymbol{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t \qquad \boldsymbol{w}_t \sim N(\mathbf{0}, \boldsymbol{W}) \tag{1b}$$

where Y_t is the observed scalar at time t, $\boldsymbol{\beta}_t$ is the p-dimensional parameter vector (also called the *state vector*), $\boldsymbol{X}_t$ is a p-dimensional vector of known covariates, and $\boldsymbol{G}_t$ is a known $p \times p$ transition matrix governing the system disturbances, or the changes in the true underlying model. In addition, v_t and $\boldsymbol{w}_t$ are two independent sequences of independent Gaussian errors with mean zero and known variance components.

To compute the posterior distribution, $\pi(\boldsymbol{\beta}_{0:T}, \sigma^2, \boldsymbol{W}|\boldsymbol{y}_{1:T})$, we must sample from the full conditional densities of $\pi(\sigma^2|\boldsymbol{\beta}_{0:T}, \boldsymbol{W}, \boldsymbol{y}_{1:T})$, $\pi(\boldsymbol{W}|\boldsymbol{\beta}_{0:T}, \sigma^2, \boldsymbol{y}_{1:T})$, and $\pi(\boldsymbol{\beta}_{0:T}|\sigma^2, \boldsymbol{W}, \boldsymbol{y}_{1:T})$. Sampling from the full conditionals for the variance components is problem specific. To sample from the unobservable states however, Carter and Kohn (1994) developed the FFBS algorithm to draw from the distribution of all states $\boldsymbol{\beta}_{0:T}$. The FFBS algorithm proceeds by running the Kalman filter and then recursively drawing realizations from $\boldsymbol{\beta}_{0:T}$. By coupling FFBS and the Gibbs sampler, samples from the joint posterior distribution, $\pi(\boldsymbol{\beta}_{0:T}, \sigma^2, \boldsymbol{W}|\boldsymbol{y}_{1:T})$, can be drawn.

For this paper, we focus solely on the dynamic generalized linear model (DGLM) with a Bernoulli response. For dynamic logistic regression, the *observation equation* is defined as

$$y_t|\boldsymbol{\beta}_t \sim \text{Bernoulli}(\pi_t), \quad \pi_t = \frac{e^{\boldsymbol{X}_t\boldsymbol{\beta}_t}}{1 + e^{\boldsymbol{X}_t\boldsymbol{\beta}_t}} \quad t = 1, \ldots, T\,, \tag{2}$$

and the *state equation* is the same as in the DLM scenario,

$$\boldsymbol{\beta}_t = \boldsymbol{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t \qquad \boldsymbol{w}_t \sim N(\mathbf{0}, \boldsymbol{W}). \tag{3}$$

In this case, the response $y_t \in \{1, 0\}$ represents a "success" or "failure," respectively, at time t and the probability of success at time t is linked to the p-dimensional state vector $\boldsymbol{\beta}_t$ and known vector of covariates $\boldsymbol{X}_t$. The matrices $\boldsymbol{G}_t$ and $\boldsymbol{W}$ maintain the same purpose and interpretation as before.

In the case of DLM's, the one-step-ahead predictive distribution of $\boldsymbol{\beta}_t|\boldsymbol{y}_{1:t-1}$, the one-step-ahead predictive distribution of $Y_t|\boldsymbol{y}_{1:t-1}$, and the filtering distribution of $\boldsymbol{\beta}_t|\boldsymbol{y}_{1:t}$ are all calculated online because their Gaussian parameters are completely and sequentially determined by the Kalman filter. The same technique is, of course, not possible in the DGLM framework.

However, a variety of techniques have been proposed to sequentially update the state vector $\boldsymbol{\beta}_t$.

For example, West et al. (1985) describe an approach using Linear Bayes estimation. Other methods to sequentially update the state vector generally rely on linearly approximating the *observation equation* to allow for the assumption of normality. The state vector can then be sequentially updated using the Kalman filter (Ferreira and Gamerman, 2000). More modern DGLM fitting techniques rely on MCMC and can be applied in a similar fashion as the DLM. Sampling from the full conditional densities for the evolution matrix $\pi(\boldsymbol{W}|\boldsymbol{\beta}_{0:T}, \boldsymbol{y}_{1:T})$ is problem specific. Sampling from the states $\pi(\boldsymbol{\beta}_{0:T}|\boldsymbol{W}, \boldsymbol{y}_{1:T})$ is accomplished via the Metropolis-Hastings algorithm within the Gibbs sampler using a pseudo FFBS algorithm (Gamerman, 1998). One drawback of the MCMC approach is that convergence can be slow (Carter and Kohn, 1994). Sampling each element of the state vector for each time period is burdensome, as the number of states grows linearly with time. A novel approach to efficiently update the *state vector* in the DGLM scenario is described below.

In the static logistic regression scenario, Polson et al. (2013) proposed a data augmentation technique for fully Bayesian inference when considering binomial likelihoods. The authors introduce the Pólya-Gamma random variable to be used as a latent variable. The random variable X has a Pólya-Gamma distribution, denoted $X \sim PG(b, c)$, if

$$X = \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/4\pi^2} \ , \tag{4}$$

where $b > 0$, $c \in \mathbb{R}$, $g_k \sim Gamma(b, 1)$ are independent Gamma random variables. The main result from Polson et al. (2013) is that likelihoods from binomial response data can be represented as mixtures of Gaussians according to the Pólya-Gamma distribution. This provides for an efficient Gibbs sampler where the full conditional for $\boldsymbol{\beta}$ is multivariate normal and the full conditional for the latent variable is a Pólya-Gamma distribution.

Windle et al. (2013) extend the Pólya-Gamma data augmentation approach to dynamic logistic regression. Starting with the likelihood of observed data $\boldsymbol{y}_{1:T}$ and distribution of hidden states $\boldsymbol{\beta}_{1:T}$,

$$\pi(\boldsymbol{\beta}_{1:T}|\boldsymbol{y}_{1:T}) = \left[\prod_{t=1}^{T} \frac{\exp(\boldsymbol{X}_t\boldsymbol{\beta}_t)^{y_t}}{1 + \exp(\boldsymbol{X}_t\boldsymbol{\beta}_t)} \right] \pi(\boldsymbol{\beta}_{1:T}) \ , \tag{5}$$

they introduce the Pólya-Gamma latent variable, $\omega_t \sim PG(1, \psi_t)$, for $t = 1, \ldots, T$ to construct the

joint distribution,

$$\pi(\boldsymbol{\beta}_{1:T}, \boldsymbol{\omega}_{1:T}|\boldsymbol{y}_{1:T}) = \left[\prod_{t=1}^{T} \frac{\exp(\boldsymbol{X}_t\boldsymbol{\beta}_t)^{y_t}}{1+\exp(\boldsymbol{X}_t\boldsymbol{\beta}_t)} \pi(\omega_t|\boldsymbol{X}_t\boldsymbol{\beta}_t)\right] \pi(\boldsymbol{\beta}_{1:T}) \ . \tag{6}$$

Using properties of the Pólya-Gamma distribution, the joint posterior can be rewritten as

$$\pi(\boldsymbol{\beta}_{1:T}, \boldsymbol{\omega}_{1:T}|\boldsymbol{y}_{1:T}) \ \propto \ \left[\prod_{t=1}^{T} \exp\left(-\frac{\omega_t}{2}\left(\frac{k_t}{\omega_t} - \boldsymbol{X}_t\boldsymbol{\beta}_t\right)^2\right)\right] \pi(\boldsymbol{\beta}_{1:T}) \ , \tag{7}$$

where $k_t = y_t - 1/2$. The addition of the latent variable provides pseudo data $z_t = k_t/\omega_t$ where $z_t \sim N(\boldsymbol{X}_t\boldsymbol{\beta}_t, 1/\omega_t)$. If we specify $\pi(\boldsymbol{\beta}_{1:T})$ such that the states vary according to a random walk, then sampling from the conditional distribution for $\boldsymbol{\beta}_{1:T}$ is equivalent to sampling from the DLM *observation equation* and *state equation*

$$z_t = \boldsymbol{X}_t\boldsymbol{\beta}_t + v_t \qquad v_t \sim N(0, 1/\omega_t) \tag{8a}$$

$$\boldsymbol{\beta}_t = \boldsymbol{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t \qquad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}) \tag{8b}$$

As a result, the posterior simulation can now be implemented using the FFBS algorithm to sample state vectors. The combination of the Pólya-Gamma latent variable augmentation approach and FFBS is highly advantageous and more computationally efficient than other DGLM fitting methods because it avoids analytic approximations, numerical integration, and use of the Metropolis-Hastings algorithm. For these reasons we use the Pólya-Gamma latent variable augmentation approach to fit the dynamic logistic regression model. The full posterior simulation for dynamic logistic regression is detailed in Appendix 6.1.

2.2. Dynamic Variable Selection Methods

Since fitting dynamic models has become somewhat straightforward with modern computing power and MCMC techniques, more recent work has focused on effect selection, or determining which elements of $\boldsymbol{\beta}_t$ are nonzero at each successive time point t. With applications in finance, such as equity premium and inflation forecasting (Kalli and Griffin, 2014), it is common to have a high-dimensional state vector where many elements are unrelated to the target. Erroneous effects

may reduce prediction accuracy and hinder model inference. This has spurred efforts to intelligently remove irrelevant predictors from hypothesized models.

The variable selection methods for dynamic models are predominantly extensions of techniques from the static model literature and focus almost exclusively on linear models. Given the intrinsic Bayesian framework of dynamic models, the approaches are consequently extensions of Bayesian Model Averaging (Hoeting et al., 1999), latent variable augmentation (Dellaportas et al., 2002; George and McCulloch, 1993), and shrinkage priors (Park and Casella, 2008; Griffin et al., 2010). However, variable selection for dynamic models is necessary over time, effectively allowing the importance of features to change as new information becomes available.

Extending the concept of BMA to the dynamic setting, Raftery et al. (2010) propose dynamic model averaging (DMA) and Koop and Korobilis (2012) apply the methodology to forecast inflation. The key difference between BMA and DMA is that DMA allows the true model to vary over time. The model *observation equation* and *state equation* are specified as

$$y_t = \boldsymbol{X}_t^{(k)}\boldsymbol{\beta}_t^{(k)} + v_t^{(k)} \qquad v_t^{(k)} \sim N(0, \sigma^{2(k)}) \;, \tag{9a}$$

$$\boldsymbol{\beta}_t^{(k)} = \boldsymbol{G}_t^{(k)}\boldsymbol{\beta}_{t-1}^{(k)} + \boldsymbol{w}_t^{(k)} \qquad \boldsymbol{w}_t^{(k)} \sim N(\boldsymbol{0}, \boldsymbol{W}^{(k)}) \;, \tag{9b}$$

where k denotes the model index. DMA requires calculating the probability of each model being the true model at time t and averaging forecasts using posterior model probabilities,

$$E[\hat{y}_t|\boldsymbol{y}_{1:t-1}] = \sum_{k=1}^{K} \hat{y_t}^{(k)}\pi(M_k|\boldsymbol{y}_{1:t-1}) \;. \tag{10}$$

Similar to the static model literature, considering all possible 2^p models is not feasible, and as a result, both BMA and DMA typically consider only a small set of candidate models. However, extensions of DMA have been developed to be more computationally efficient. Onorante and Raftery (2016) propose a dynamic form of Occams window to consider larger model spaces which they call the FEAR (Forecast, Expand, Asses Reduce) algorithm. In four iterative steps, the authors initialize a subset of all possible models at the outset to obtain a forecast distribution, expand the number of models to a larger population by considering models similar to the current population, assess the models by computing weights once a new observation is available, and then finally reduce the

number of models to those that are in Occams window. The computational advantage of this method is that only a subset of models are considered and models are allowed to evolve slowly over time. Risse and Ohl (2017) go one step further with the Dynamic Occam's Window (DOW) method and do not assume the models chosen in successive iterations are similar. In this scenario the parameters are less stable but mimic DMA more closely for a large number of features. Another novel approach to scaling DMA is the technique of using Google probabilities (Koop and Onorante, 2019). This approach uses Google web queries to determine inclusion probabilities of model features. The realized volume of proxy terms for model features effectively acts as a dynamic model selection tool. An overview of DMA techniques is discussed in Nonejad (2021) and R code for fitting DMA with DOW is available in the fDMA pacakge (Drachal, 2020).

Other recent and efficient Bayesian approaches rely on fitting the model with all parameters and shrinking elements of $\boldsymbol{\beta}_t$ toward zero in accordance with the data. Motivated by the "spike-and-slab" approach, Nakajima and West (2013) propose shrinking parameters to zero if their absolute value falls below a threshold at any point in time t. This latent threshold modeling (LTM) approach introduces a matrix of latent variables, $\boldsymbol{I}_t = diag(I_{1t}, \ldots, I_{pt})$, into the *observation equation*,

$$y_t = \boldsymbol{X}_t(\boldsymbol{I}_t\boldsymbol{\beta}_t) + v_t \qquad v_t \sim N(0, \sigma^2) \; , \tag{11}$$

where $I_{jt} = I(|\beta_{jt}| \geqslant d_j)$ and $d_j \geqslant 0$ for all p. The degree of sparseness is controlled by tuning the elements of $\boldsymbol{d} = (d_1, \ldots, d_p)$.

A few adaptive shrinkage priors have been discussed in the DLM literature as well. These prior distributions shrink elements of $\boldsymbol{\beta}_t$ toward zero if supported by the data and avoid shrinkage otherwise. The priors are adaptive in the sense that regularization adheres to data-driven evidence and the degree of regularization can be tuned by altering prior distribution parameters, a concept similar to altering the threshold parameter λ in non-Bayesian regularization. For example, Caron et al. (2012) place the following multivariate hierarchical prior on $\boldsymbol{\beta}_t$,

$$\boldsymbol{\beta}_t|\tau \sim N(\boldsymbol{\mu}, \tau\boldsymbol{\Sigma}) \qquad \tau \sim GiGauss(\nu, \delta, \gamma) \; , \tag{12}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma}$ is a $p \times p$ covariance matrix, and $GiGauss(\cdot)$ is the generalized inverse Gaussian distribution (Barndorff-Nielsen and Shephard, 2001). Letting $\boldsymbol{\beta}_j = (\beta_{j1} \ldots, \beta_{jT})$ denote the evolution of the j^{th} parameter, then $\boldsymbol{\beta}_j \in \mathbb{R}^T$ follows the multivariate generalized hyperbolic distri-

bution that simplifies to static model adaptive shrinkage priors under specific parameterizations. For example, if $\delta = 0$ and $\nu = 1$ the prior reduces to the Laplace prior of Park and Casella (2008). Likewise, if $\delta = 0$, $\nu \neq 1$, and $\nu > 0$ the prior reduces to the Normal-Gamma prior of Griffin et al. (2010).

Another adaptive shrinkage prior in the DLM literature is referred to as the Normal-Gamma Autoregressive (NGAR) process (Kalli and Griffin, 2014). As its name implies, this prior is motivated by the Normal-Gamma prior of Griffin et al. (2010) with an extension to the DLM. The process is written $\boldsymbol{\beta}_j \sim NGAR(\lambda_j, \mu_j, \varphi_j, \rho_j)$. The authors show, similar to the static model with Normal-Gamma prior, the parameter λ_j controls the degree of sparseness. Small values of λ_j place more prior mass at zero and cause heavier shrinkage for the j^{th} coefficient. The autocorrelation parameters ρ_j and φ_j control the dependence between state parameters. Thus, ρ_j and φ_j control the ability for the importance of parameters to vary in time.

Although a powerful and flexible modeling tool for non-Gaussian time series data, DGLMs have received far less attention in the literature than their predecessor, the DLM. Currently, the only DGLM variable selection method reported in the literature uses a variation of Bayesian model averaging (BMA) techniques, which does not scale efficiently as the number of predictors grows. McCormick et al. (2012) extend the DMA approach of Raftery et al. (2010) to the dynamic logistic regression model. Given a set of K candidate models, $(M_1, \ldots, M_K)$, and letting L_t be the model indicator at time t, the observation equation becomes

$$y_t | L_t = M_k \sim \text{Bernoulli}(p_t^{(k)}), \qquad \text{and} \qquad \text{logit}(p_t^{(k)}) = \boldsymbol{X}_t^{(k)} \boldsymbol{\beta}_t^{(k)} \ . \tag{13}$$

The forecasts from each of the K models are then averaged using $\pi(L_t = M_k | \boldsymbol{y}_{1:t-1})$ as weights. A key contribution lies in the sequential updating of model parameters in the dynamic logistic regression model.

2.3. Scalable Variable Selection in Dynamic Logistic Regression

An ideal variable selection method for dynamic models is one that can scale to models with many predictors and can efficiently conduct variable selection each time the model is updated. Here we describe a method originally developed for Bayesian linear models and show that it can be extended to dynamic models.

Bondell and Reich (2012) proposed Bayesian variable selection via penalized credible regions for the traditional, static linear model. This approach separates the model fitting and variable selection process by constructing credible regions from the posterior distribution of $\boldsymbol{\beta}$ and Σ. Given a $(1-\alpha)\times 100\%$ credible region, any point within the region is a feasible estimate of $\boldsymbol{\beta}$. The authors suggest selecting the sparsest solution to accomplish variable selection. Thus, the proposed estimate is

$$\tilde{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \|\boldsymbol{\beta}\|_0 \qquad \text{subject to } \boldsymbol{\beta} \in C_\alpha \tag{14}$$

where $\|\boldsymbol{\beta}\|_0$ is the L_0 norm of the vector $\boldsymbol{\beta}$, i.e. the number of nonzero elements, and C_α is the $(1-\alpha)\times 100\%$ credible region. The selected model excludes predictors where $\beta_j = 0$ is included in the credible region.

As the coverage increases, the credible region expands leading to a sparser model. The authors suggest creating a sequence of credible sets that correspond to creating a sequence of selected models. The sequence of p models are the only p possible models to be considered. The best model, given the solution path, can then be chosen by a goodness-of-fit metric such as the Akaike information criterion or the Bayesian information criterion. This is a drastic reduction in the comparisons required for all possible 2^p models.

Figure 1 demonstrates the joint credible region approach to variable selection for a linear model with only two parameters, β_1 and β_2. Starting from the largest ellipse, referring to the 95% credible region, the sparsest solution would be the null model as $\beta_1 = \beta_2 = 0$ is a feasible solution, marked by the $\times$. Next, the 90% credible region indicates only $\beta_2 = 0$, marked by a thick line along the x-axis. Finally, the sparsest solution within the smallest credible region includes both parameters in the model as nonzero coefficients. Therefore, the solution path in this example is $\{\beta_1, \beta_2\}$.

In some cases, the posterior of $\boldsymbol{\beta}$ is elliptical with density of the form $H[(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T\hat{\Sigma}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})]$ where H is a monotone decreasing function. Therefore, the highest density region is of the form $\{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T\hat{\Sigma}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}) = K_\alpha\}$ for some K_α. In general, the posterior distribution does not maintain elliptical contours but credible sets still can be created. In addition, the solution to Equation (14) is not unique and requires searching over a possibly high dimensional region. To overcome these challenges, Bondell and Reich (2012) apply several alterations to Equation (14). First, the authors replace the L_0 norm by a smooth combination of L_0 and L_1 (Lv and Fan, 2009) which leads to

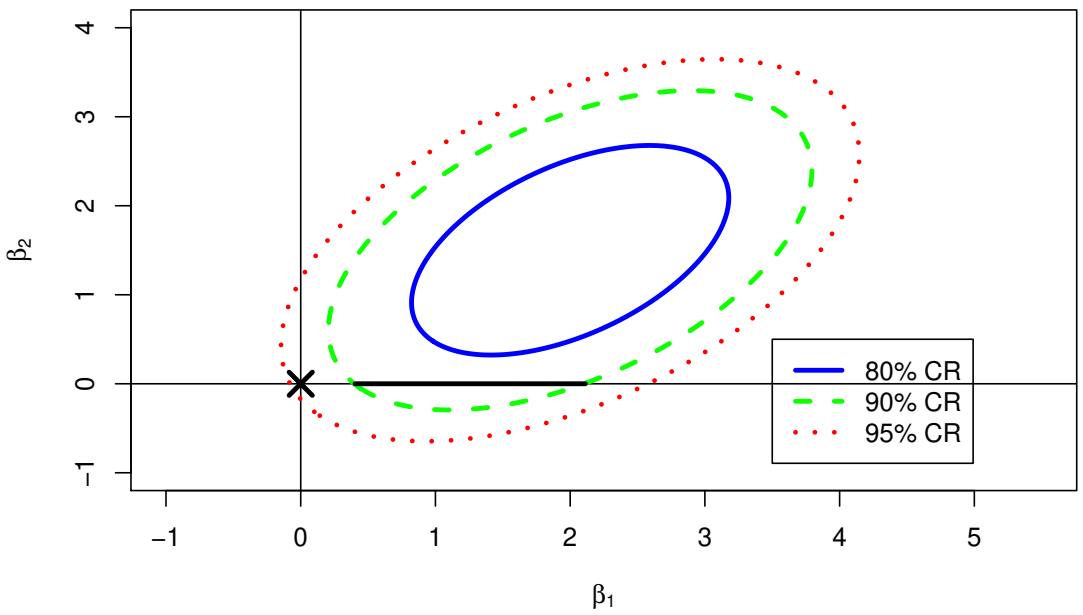

Figure 1: Example of the joint credible region variable selection approach for a linear model with two parameters. Here the solution path is $\{\beta_1.\beta_2\}$.

a non-convex optimization problem. Next, it is converted to a convex optimization problem by applying a local linear approximation which is equivalent to the Lagrangian optimization problem

$$\tilde{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \lambda \sum_{j=1}^{p} \frac{|\beta_j|}{|\hat{\beta}_j|^2} \right], \tag{15}$$

where λ is the tuning parameter and has one-to-one correspondence to α. Therefore, the proposed sequence of selected models is given by the solution to Equation 15 as a function of λ. For a given λ, regressing $\boldsymbol{Y}^* = \hat{\boldsymbol{\Sigma}}^{-1/2}\hat{\boldsymbol{\beta}}$ on $\boldsymbol{X}^* = \hat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{D}$ with an L_1 penalization, where $\boldsymbol{D} = diag(\hat{\beta}_1^{\,2}, \ldots, \hat{\beta}_p^{\,2})$, returns the solution $\boldsymbol{\beta}^*$. The solution to Equation (15) is then calculated as $\tilde{\boldsymbol{\beta}} = \boldsymbol{D}\boldsymbol{\beta}^*$. One can readily find the entire solution path using the Least Angle Regression algorithm (Efron et al., 2004).

The joint credible region variable selection approach can be extended to the dynamic logistic regression model by altering the inputs to the optimization problem in Equation (15). In order to create the solution path for a linear model detailed above, the technique only requires posterior distribution estimates of the parameter vector $\hat{\boldsymbol{\beta}}$ and model covariance $\hat{\Sigma}$. To extend this approach to dynamic logistic regression we can replace these inputs with posterior estimates of the *state*

vector $\hat{\boldsymbol{\beta}}_t$ and its covariance $\hat{\boldsymbol{W}}$. Thus, the optimization problem becomes

$$\tilde{\boldsymbol{\beta}}_t = \underset{\boldsymbol{\beta}_t}{\operatorname{argmin}} \left[(\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_t)^T \hat{\boldsymbol{W}}^{-1} (\boldsymbol{\beta}_t - \hat{\boldsymbol{\beta}}_t) + \lambda \sum_{j=1}^{p} \frac{|\beta_{jt}|}{|\hat{\beta}_{jt}|^2} \right]. \tag{16}$$

There are two advantages of the joint credible region variable selection approach. First, the method is scalable because we are able to include all p predictors into the model fitting process unlike the DMA approach in McCormick et al. (2012). Second, the method allows for dynamic variable selection, meaning, we are able to construct joint credible regions and conduct variable selection at every time point for every *state vector* $\hat{\boldsymbol{\beta}}_t$. As new information becomes available and the underlying process generating the model changes, the dynamic joint credible region approach can determine nonzero elements of the state vector sequentially in time.

3. Simulation

In this section, we evaluate the performance of the dynamic logistic regression and variable selection approach described in Section 2.3 against dynamic model averaging and the static logistic regression with the LASSO regularization model. The latter approach is used in the motivating example of predicting civil unrest in South America (Korkmaz et al., 2016). We compare both the prediction and variable selection results for the three models.

3.1. Setup

In each case of the simulation, data sets are generated from the dynamic logistic regression *observation equation*

$$y_t|\boldsymbol{\beta}_t \sim \text{Bernoulli}(\pi_t), \quad \pi_t = \frac{e^{\boldsymbol{X}_t\boldsymbol{\beta}_t}}{1 + e^{\boldsymbol{X}_t\boldsymbol{\beta}_t}} \quad t = 1, \ldots, T \tag{17}$$

and *state equation*

$$\boldsymbol{\beta}_t = \boldsymbol{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t \qquad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}) \tag{18}$$

The number of observations for each time series is varied in $T \in \{50, 100\}$ and the covariates are standard normal. We let the parameters freely vary according to a random walk by setting both the transition matrix and the *state equation* covariance matrix to the identity, $\boldsymbol{G}_t = \boldsymbol{W} = \boldsymbol{I}_p$.

The number of candidate predictors is varied in $p \in \{5, 10, 15, 25, 50, 100\}$ and the true number of predictors used to generate $\boldsymbol{y}_{1:T}$ is approximately 10% of p, $p^* \in \{2, 2, 2, 3, 5, 10\}$. For each data set, p^* indices are randomly chosen as the nonzero locations of the state vectors $\boldsymbol{\beta}_t$ to avoid a potential ordering bias.

To simulate the nonzero elements of the state vectors, we consider two scenarios for generating dynamic coefficients: 1) structural break parameters and 2) completely dynamic parameters. First, completely dynamic parameters are simulated as seen in the *state equation* above. That is, the nonzero parameters vary according to a random walk by adding white noise to those parameters at every single time point. Structural break parameters are a relaxed version of the completely dynamic scenario. In this case, parameters are assumed static for $t > 1$ until the model experiences a shift. For this simulation, we shock the true p^* parameters every 10 time points by simply adding white noise to the parameters. For structural break parameters the *state equation* is written

$$\boldsymbol{\beta}_t = \boldsymbol{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t\mathbb{1}\big(0 \equiv t \mod 10\big) \qquad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}) \tag{19}$$

Figure 2 shows how a single parameter may evolve over time under the two scenarios. Clearly, it should be easier for the model to track structural break parameters as there are periods of stationarity and the parameters lack extreme movement during the shock. The completely dynamic parameters add an increased layer of complexity. For example, Figure 2 shows that the completely dynamic parameter is negative at first and then increases to a relatively large positive coefficient until it moves back to zero at the end of the graph. The goal is to see if adding complexity to the model reduces performance given that is more difficult to track completely dynamic parameters.

3.2. Metrics

For each time series length $T \in \{50, 100\}$, we fit the model on the first 50% of the observations, $\{25, 50\}$, and then forecast one time period ahead and five time periods ahead. For the LASSO regularization of the static model, the regularization parameter is chosen via five-fold cross-validation. For dynamic model averaging we forecast using the ensemble of fit models (2^p) and choose the best model according to posterior probabilities for variable selection. For the dynamic logistic regression variable selection, we apply the penalized credible region approach described in Section 2.3 on the terminal state vector. We then move one unit of time ahead, make a forecast, and conduct variable

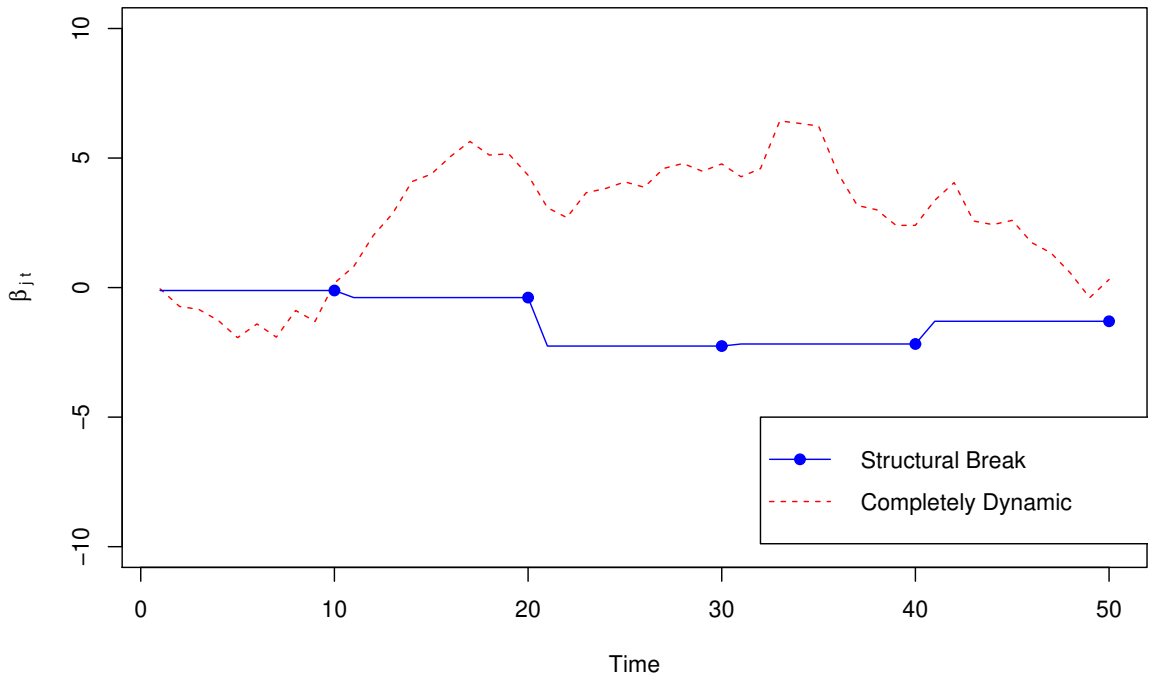

Figure 2: Example of a single structural break and completely dynamic parameter.

selection again. The process is repeated until the end of the times series. At each forecast and variable selection period, we measure the accuracy of predictions with the described metrics.

Prediction performance is measured by the standard confusion matrix for a binary response. Let true positives (TP) be a correctly predicted event, true negatives (TN) a correctly predicted nonevent, false positives (FP) an incorrectly predicted event, and false negatives (FN) an incorrectly predicted nonevent. Overall model accuracy can be measured as the proportion of correct predictions, $\frac{TP+TN}{TP+TN+FP+FN}$. For each of the four prediction performance possibilities, we use a cutoff rate of 0.50.

We also examine the trade-off between power and the false discovery rate by measuring both the *recall* and *precision* for each model. Recall, also known as the TP rate, is the ratio of TPs to the total number of events, $\frac{TP}{TP+FN}$. Hence, it measures the model's ability to correctly predict an event when there is in fact an event. Precision, on the other hand, is the ratio of TPs to the total number of predicted events, $\frac{TP}{TP+FP}$. These two metrics must be measured in combination because, for example, recall could be inflated by allowing the model to simply predict events in all cases. To

aggregate precision and recall, we report the F1 scores, or the harmonic average of precision and recall, for each case in the simulation.

Variable selection performance is measured using the same metrics as prediction. For each prediction, we count the number predictors correctly included in the model (TP), correctly excluded from the model (TN), incorrectly included (FP), and incorrectly excluded (FN). In addition, the precision, recall, and F1 scores for variable selection are also measured for the same reasons above.

3.3. Simulation Results

Table 1: Forecast results for completely dynamic parameters, $T = 50$, and $p < 25$.

Model	Forecast	Predictors	TP	TN	FP	FN	Precision	Recall	F1
LASSO	1	5	50.83	3.12	45.83	0.21	52.59	99.59	68.83
	5	5	50.53	2.93	46.25	0.29	52.21	99.43	68.47
DMA	1	5	44.38	42.71	6.04	6.88	88.02	86.59	87.30
	5	5	39.09	38.08	10.91	11.92	78.17	76.63	77.39
DGLM	1	5	49.58	47.50	1.46	1.46	97.14	97.14	97.14
	5	5	47.26	46.15	3.03	3.56	93.98	93.00	93.49
LASSO	1	10	46.67	4.79	48.33	0.21	49.12	99.56	65.79
	5	10	47.31	4.28	48.41	0.00	49.42	100.00	66.15
DMA	1	10	51.04	42.29	3.12	3.54	94.23	93.51	93.87
	5	10	40.48	34.42	10.67	14.42	79.14	73.73	76.34
DGLM	1	10	46.67	52.92	0.21	0.21	99.56	99.56	99.56
	5	10	46.78	51.92	0.77	0.53	98.38	98.88	98.63
LASSO	1	15	47.50	4.17	48.12	0.21	49.67	99.56	66.28
	5	15	47.36	3.99	48.46	0.19	49.42	99.60	66.06
DMA	1	15	42.92	48.12	4.38	4.58	90.75	90.35	90.55
	5	15	29.47	33.75	18.65	18.12	61.24	61.92	61.58
DGLM	1	15	47.29	52.29	0.00	0.42	100.00	99.13	99.56
	5	15	45.82	50.91	1.54	1.73	96.75	96.36	96.56

For each combination of time series length $T \in \{50, 100\}$, candidate predictors $p \in \{5, 10, 15, 25, 50, 100\}$, and dynamic parameter scenario, we average the metrics across 100 data sets. For each data set we use 5000 MCMC iterations to approximate the posterior distribution after a 500 iteration burn-in period. Note, results for the completely dynamic parameters and the structural break parameters

are markedly similar. For this reason we primarily refer to the completely dynamic parameter results presented in Tables 1-5. The structural break parameter results are displayed in Appendix 6.2. Also note, dynamic model averaging was only considered for a small number of predictors, $p \in \{5, 10, 15\}$. As described in Section 2.2, DMA for binary outcomes is currently not scalable for even a moderate number of predictors, $p \in \{25, 50, 100\}$.

For the prediction results, perhaps the first thing to notice is the accuracy of the dynamic logistic regression model is higher than the DMA model and much higher than the LASSO model. That is, the TP and TN rates for the DGLM are almost 50% each, and therefore, false predictions are extremely low. For the LASSO model, it appears that a majority of the predictions are events as the TP and FP rates are near 50% each. As a result, the LASSO model rarely makes the correct prediction of a nonevent. The TN rates for the DMA are better than that of the LASSO model but slightly worse than for the DGLM model.

Due to the high FP rate for the LASSO model, the precision is kept low, ranging between 50% and 53%. On the other hand, recall is high in the upper 90% range for the LASSO model because the FN rate is low. The low precision and high recall again indicate that the LASSO model over predicts the rate of events in each data set. Averaging out the low precision and high recall for the LASSO model gives F1 scores in the $67\% - 69\%$ range. For DMA, the false prediction rates are similar for each forecast scenario, resulting in higher precision than the LASSO model and lower recall. Averaging out these effects, DMA outperforms the LASSO model with F1 scores in the $61\% - 94\%$ range. Because of the high accuracy of the DGLM model the precision, recall, and F1 scores are all in the range of 95% to 100%.

Moving from a times series length of $T = 50$ in Table 2 to $T = 100$ in Table 3 for the LASSO and DGLM, it does not seem to impact the prediction results when forecasting either one time period ahead or five periods ahead. F1 scores remain high in either case. The same is true when T is held constant and the number of candidate predictors varies in $p \in \{25, 50, 100\}$. F1 scores remain high even moving into the $p > T$ scenario. For $p \in \{5, 10, 15\}$ in Table 1, the DMA performs considerably better forecasting one time period ahead than five periods ahead causing the difference in F1 scores for DMA to increase as p increases.

The variable selection results (Table 4 and Table 5) also show that the DGLM outperforms the other two models. In all cases except for $p = 5$, the TP and TN rates for the DGLM are higher and the FP and FN rates are lower than both LASSO and DMA. As a result, the precision, recall,

Table 2: Forecast results for completely dynamic parameters, $T = 50$, and $p \geqslant 25$.

Model	Forecast	Predictors	TP	TN	FP	FN	Precision	Recall	F1
LASSO	1	25	50.13	2.83	46.60	0.43	51.82	99.14	68.07
	5	25	50.31	2.98	46.15	0.56	52.16	98.89	68.29
DGLM	1	25	50.46	49.33	0.10	0.10	99.80	99.80	99.80
	5	25	49.65	48.16	0.90	1.21	98.22	97.62	97.92
LASSO	1	50	49.23	1.93	48.23	0.60	50.51	98.79	66.84
	5	50	49.29	2.26	47.85	0.59	50.74	98.82	67.05
DGLM	1	50	49.73	50.07	0.10	0.10	99.79	99.79	99.79
	5	50	49.38	49.64	0.48	0.50	99.03	98.99	99.02
LASSO	1	100	50.20	1.30	47.83	0.67	51.21	98.68	67.43
	5	100	50.23	1.57	47.46	0.74	51.41	98.54	67.58
DGLM	1	100	50.87	49.13	0.00	0.00	100.00	100.00	100.00
	5	100	50.83	48.91	0.12	0.14	99.76	99.73	99.74

and F1 scores are all higher for the DGLM, except for $p = 5$ where DMA has the largest TN rate. On average the DGLM is better able to pick out the true predictors and reject the noise variables. Furthermore, for both the DGLM and LASSO models, the TP rate and F1 scores increase as the number of observations increase from T=50 to $T = 100$, holding p constant. Intuitively this makes sense; the more data each model has, the better they are equipped to pick the correct variables. Note that the variable selection performance values in Table 5 are much lower than the prediction results in Tables 2 and 3. This illustrates that although identifying the active predictors is a challenging problem, the prediction results are fairly robust against model misspecification.

The dynamic logistic regression model outperforms the DMA and LASSO models in both prediction and variable selection as measured by precision, recall, and the F1 score. For prediction, the LASSO model consistently over classifies predictions as events, DMA had better precision than LASSO due to fewer false positives, and the DGLM was consistently accurate with F1 scores in the high 90% range. Regarding variable selection, the DGLM overall outperformed the other two models in all four categories of the confusion matrix (TP,TN,FP,FN) and as a result, the precision, recall, and F1 scores were all consistently higher as well. The DGLM model was better able to track the underlying distribution, leading to better rates of inclusion of true predictors and exclusion of noise variables.

Table 3: Forecast results for completely dynamic parameters, $T = 100$, and $p \geqslant 25$.

Model	Forecast	Predictors	TP	TN	FP	FN	Precision	Recall	F1
LASSO	1	25	50.54	2.34	47.02	0.10	51.80	99.80	68.21
	5	25	50.48	2.29	47.10	0.13	51.73	99.74	68.13
DGLM	1	25	50.62	49.28	0.08	0.02	99.84	99.96	99.90
	5	25	50.01	48.68	0.71	0.51	98.60	98.99	98.79
LASSO	1	50	48.64	2.90	48.36	0.10	50.14	99.79	66.75
	5	50	48.65	2.80	48.37	0.17	50.14	99.65	66.72
DGLM	1	50	48.72	51.26	0.00	0.02	100.00	99.96	99.97
	5	50	48.64	50.99	0.19	0.18	99.61	99.63	99.62
LASSO	1	100	49.40	2.02	48.36	0.22	50.53	99.56	67.04
	5	100	49.53	2.01	48.27	0.19	50.64	99.62	67.15
DGLM	1	100	49.62	50.38	0.00	0.00	100.00	100.00	100.00
	5	100	49.72	50.22	0.06	0.04	99.88	99.92	99.89

Table 4: Variable selection results for completely dynamic parameters, $T = 50$, and $p < 25$.

Model	Predictors	TP	TN	FP	FN	Precision	Recall	F1
LASSO	5	44.75	21.79	18.21	15.25	71.08	74.58	72.79
DMA	5	42.25	25.83	14.17	17.75	74.89	70.42	72.58
DGLM	5	52.29	22.04	17.96	7.71	74.44	87.15	80.29
LASSO	10	21.38	40.90	29.10	8.62	42.34	71.25	53.12
DMA	10	20.08	39.40	30.60	9.92	39.62	66.94	49.78
DGLM	10	26.62	54.71	15.29	3.38	63.52	88.75	74.04
LASSO	15	13.42	54.96	25.04	6.58	34.89	67.08	45.90
DMA	15	11.39	36.82	43.18	8.61	20.87	56.94	30.55
DGLM	15	17.38	66.01	13.99	2.62	55.40	86.88	67.66

Table 5: Variable selection results for completely dynamic parameters and $p \geqslant 25$.

Model	Predictors	T	TP	TN	FP	FN	Precision	Recall	F1
LASSO	25	50	7.16	67.97	20.03	4.84	26.33	59.67	36.54
DGLM	25	50	9.12	75.77	12.23	2.88	42.72	76.00	54.69
LASSO	25	100	8.65	68.91	19.09	3.35	31.18	72.08	43.53
DGLM	25	100	11.29	69.64	18.36	0.70	38.08	94.16	54.23
LASSO	50	50	4.06	76.83	13.17	5.94	23.56	40.60	29.82
DGLM	50	50	5.77	79.95	10.05	4.22	36.47	57.76	44.71
LASSO	50	100	6.66	72.85	17.14	3.33	27.98	66.67	39.42
DGLM	50	100	8.26	76.30	13.69	1.73	37.63	82.68	51.72
LASSO	100	50	2.18	82.08	7.92	7.82	21.58	21.80	21.69
DGLM	100	50	3.07	83.71	6.29	6.92	32.79	30.73	31.73
LASSO	100	100	4.63	77.99	12.00	5.37	27.84	46.30	34.77
DGLM	100	100	5.86	79.04	10.96	4.14	34.84	58.60	43.69

4. Application

We now compare the logistic regression with LASSO regularization baseline method to the dynamic logistic regression and variable selection approach applied to civil unrest in South America. Dynamic model averaging does not scale to a data set of this magnitude. We model civil unrest in Argentina, Brazil, Colombia, Mexico, Paraguay, and Venezuela from Novermber 2012 to August 2014, using data collected by Korkmaz et al. (2016).

Ground truth of the binary outcome, daily civil unrest, is produced by social scientists within the region and is reported in the Gold Standard Report. For model features we include only Twitter terms. A 10% sample of all generated tweets were collected during the period in Latin America from Datasift, equating to approximately 500 million tweets. The tweets were filtered using a dictionary of $p = 962$ protest-related Twitter terms for each country. The dictionary was created by subject matter experts in Latin America and contains words such as "revolution," phrases such as "walk for peace," and political individuals such as "Henrique Capriles," who is the leader of the Venezuelan opposition party.

The number of protests in each country over the two year period is displayed in Figure 3. On the

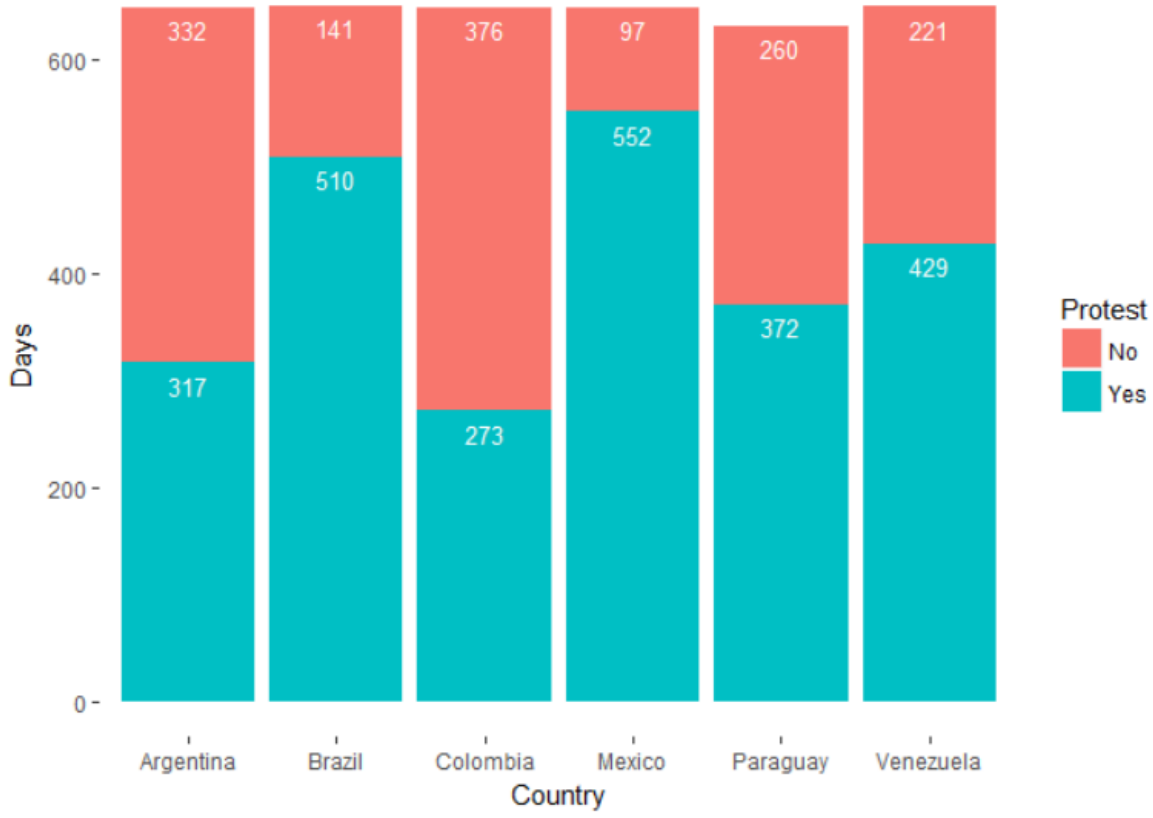

Figure 3: The number of protest and non-protest days for Argentina, Brazil, Colombia, Mexico, Paraguay, and Venezuela from November 2012 to August 2014 according the the Gold Standard Report.

per-country level, Argentina, Brazil, Colombia, Mexico, Paraguay, and Venezuela all experienced civil unrest at least 40% of the days throughout the 2-year period. Furthermore, in Brazil and Mexico, protests occurred on more than 500 days of the approximately 600-day period.

For more efficient implementation via parallelization, we model the first 600 observations in 50 day increments. In each 50 day period, $(1-50, 51-100, \ldots, 551-600)$, we fit each model on the first 25 observations and then forecast one day and five days ahead. We then move one day ahead in time, fit the model to the past data, and forecast again. The process is repeated until the last day is reached. For each of the six countries, we use 1000 MCMC iterations to approximate the posterior distribution after a 100 iteration burn-in period. Forecast results are shown in Table 6 and the top five most commonly selected variables are provided in Table 7.

The recall of each model is markedly similar to the simulation results discussed in Section 3.3, with ranges of 97% to 100% for each model and country. Both the static and dynamic models rarely forecasted no civil unrest when there was in fact a future protest looming. The precision for the dynamic model also is high for each country with ranges of 98% to 100% indicating few false prediction of future protest. For the LASSO model, however, the precision varies widely among countries. For example, the precision for Brazil and Mexico are both in the lower to mid 80% range and the Precision for Argentina is in the mid 50% range. As discussed in the simulation results of Section 3.3, the LASSO model tends to over forecast the probability of an event; therefore, the inconsistency of the precision for the civil unrest application can best be explained by Figure 3. Brazil and Mexico have the highest proportion of protests at 0.78 and 0.85 respectively. Countries where the proportion of protests is closer to 0.50, such as Argentina and Colombia, the precision for the LASSO model is closer to 50%.

Averaging the precision and recall results for the dynamic model, the F1 scores remain in the 97% to 100% range, providing evidence of a highly accurate civil unrest forecasting model. In both Argentina and Venezuela, the one day head forecasts were perfect over the entire 600 day period. In all cases, moving from a one day ahead to five days ahead forecast reduced the forecast accuracy as expected, indicating there is in fact a dynamic component to the unknown forces generating civil unrest. For the LASSO model, the F1 scores mimic the behavior of the precision, as the recall is invariably high. The F1 scores for Brazil and Mexico, the two countries most likely to experience civil unrest, are in the 89% to 91% range. F1 scores for the other four countries are in the 60% to 81% range. Overall the dynamic model outperforms the static model in forecasting civil unrest. The F1 scores for each country and forecast period are all higher for the dynamic logistic regression model than that of the baseline model.

Understanding the reasons people protest and civil unrest inference in general is achieved via variable selection. Table 7 reports the most common terms selected by each model over the two year forecasting period. For example, the most predictive terms for civil unrest in Colombia according to the dynamic model are reform (reforma), judgment (sentencia), traditional (tradicional), environment (ambiental), and wages (salarial). In addition, the dynamic model selected these variables for model inclusion 25%, 23%, 23%, 19%, and 18% of the time respectively. Thus, for approximately 6 months of the 2-year time period, Colombians were presumably protesting reform of traditional values, environmental impact, and low wages. In fact, in 2013 the Colombian government deployed

Table 6: Civil unrest forecast results for each country, model, and forecast period.

Country	Model	Forecast Days	Precision	Recall	F1 Score
Argentina	LASSO	1	54.06	98.16	69.72
		5	55.18	97.85	70.56
	DGLM	1	100	100	100
		5	98.64	98.26	98.45
Brazil	LASSO	1	80.33	100	89.09
		5	80.43	99.70	89.04
	DGLM	1	99.17	99.59	99.38
		5	99.32	99.51	99.41
Colombia	LASSO	1	43.10	100	60.24
		5	43.04	99.44	60.08
	DGLM	1	100	98.43	99.21
		5	99.78	97.77	98.77
Mexico	LASSO	1	83.85	99.59	91.04
		5	84.29	99.15	91.11
	DGLM	1	99.46	99.94	99.70
		5	98.99	99.90	99.45
Paraguay	LASSO	1	53.69	98.76	69.56
		5	52.17	98.64	68.24
	DGLM	1	99.72	99.65	99.69
		5	98.31	98.24	98.27
Venezuela	LASSO	1	67.33	100	80.47
		5	64.76	99.63	78.49
	DGLM	1	100	100	100
		5	98.27	97.42	97.85

50,000 troops to stymie the violent protests rooted in its citizens demanding reform of Colombian agricultural business and unfair trade agreements that were forcing farmers out of business (CNN, 2013). The protests resumed in 2014 when the Colombian government failed to keep promises that quelled the riots of 2013 (BBC, 2014).

Most of the predictive terms reported in Table 7 provide some degree of inference regarding the reason for protest. For example, in Mexico, which is a highly traditional and religious country, the term homophobia (homophobia) is the third most commonly selected term according to the dynamic model, which indicates groups protesting for and against gay rights. On the other hand, in Argentina, the dynamic model finds terms such as warning (advertencia) and assembly (asamblea) to be predictive of civil unrest. These terms do not necessarily provide context to the protest, rather they are simply words that must be monitored within the Twitter feed to forecast the likelihood of future protests. Furthermore, Figure 4 displays the dynamic parameters for the three most predictive terms in Argentina. The graphic not only shows the periods which the terms are predictive but also how critical the terms are for prediction long term. For example, the term "advertencia" appears to be predictive in the first 150 days and the term "asamblea" is only predictive around the 200 day mark. The term "salud" shows periods where its associated parameter is relatively large in magnitude for more than half the entire period. The term "salud" is a more consistent predictor of civil unrest in Argentina and the other two terms represent the need for a dynamic model as the reasons for protest are fluid.

As expected, the most predictive terms between countries are different given a specific model. Civil unrest occurs for a multitude of reasons, and one nation's predictive terms or motives for protest do not necessarily correlate or influence another. The variable selection results also show the most frequently selected terms are completely different for each model within a given country. This is counterintuitive given the variable selection simulation results discussed in Section 3.3. Under controlled conditions, the two models performed relatively similar in terms of the TP and TN rates. For the civil unrest application, the five most commonly selected variables for the dynamic model are selected between 16% and 25% of the time and for the LASSO model the range is reduced to only 6% to 14%. In this setting, the LASSO model appears to select different terms for each 50 observation period. Conversely, the dynamic model identifies terms that are predictive over several months. Although we expect the active features to change over time (hence the use of a dynamic generalized linear model), it is also reasonable to expect a certain set of words to repeat as highly

Table 7: Top five most commonly selected variables for each country and model.

Country	Model	Twitter Features (% of Time Selected)
Argentina	LASSO	Seguridad(9) Iniciativa(8) Efectivo(7) Proyecto(7) Legalizacia(6)
	DGLM	Salud(22) Advertencia(21) Asamblea(19) Prohibir(18) Resultados(17)
Brazil	LASSO	Atrocidad(14) Incremento(8) Laa(8) Legalizacia(7) Damnificado(7)
	DGLM	Accidente(18) Arara(17) Ministro(16) Masacre(16) Convenios(15)
Colombia	LASSO	Effectivo(10) Electricidad(9) Justificacia(8) Directivos(8) Universitarios(7)
	DGLM	Reforma(25) Sentencia(23) Tradicional(23) Ambiental(19) Salarial(18)
Mexico	LASSO	Rumores(9) Humillar(8) Plantear(7) Dictadura(7) Rancheros(7)
	DGLM	Excesos(17) Marcha(17) Homofobia(16) Caro(16) Agua(16)
Paraguay	LASSO	Contaminar(8) Procesados(8) Derechos(7) Afectados(7) Corrupcia(7)
	DGLM	Embargo(25) Encontrar(25) Huelga(23) Destrozar(22) Compromiso(20)
Venezuela	LASSO	Prejuicios(10) Aumento(9) Violar(8) Perseguir(7) Transportistas(7)
	DGLM	Realizar(25) Hidroela(21) Derivar(17) Recuerdos(17) Lucha Armada(16)

predictive features within a country. We believe this suggests that the dynamic logistic regression model is not only more predictive of civil unrest, but also more accurately identifies the issues that drive its occurrence.

5. Conclusion

In this paper we present a combined model fitting and Bayesian variable selection methodology for dynamic logistic regression. We include the Pólya-Gamma latent variable into the joint posterior distribution to more efficiently sample draws of state vectors using the FFBS algorithm. After model fitting, we use the estimated state vector at time t and its covariance to create penalized credible regions for variable selection. This method provides an entire solution path for the modeler to select the best of only p possible models. Furthermore, one can do variable selection dynamically using joint credible regions, or simply, at each time point a new observation becomes available and the state vector is updated.

Through simulation, we show that this approach significantly improves the precision of predicting an event and, thus, the F1-score as well, consistently reaching over 95% and often over 99%.

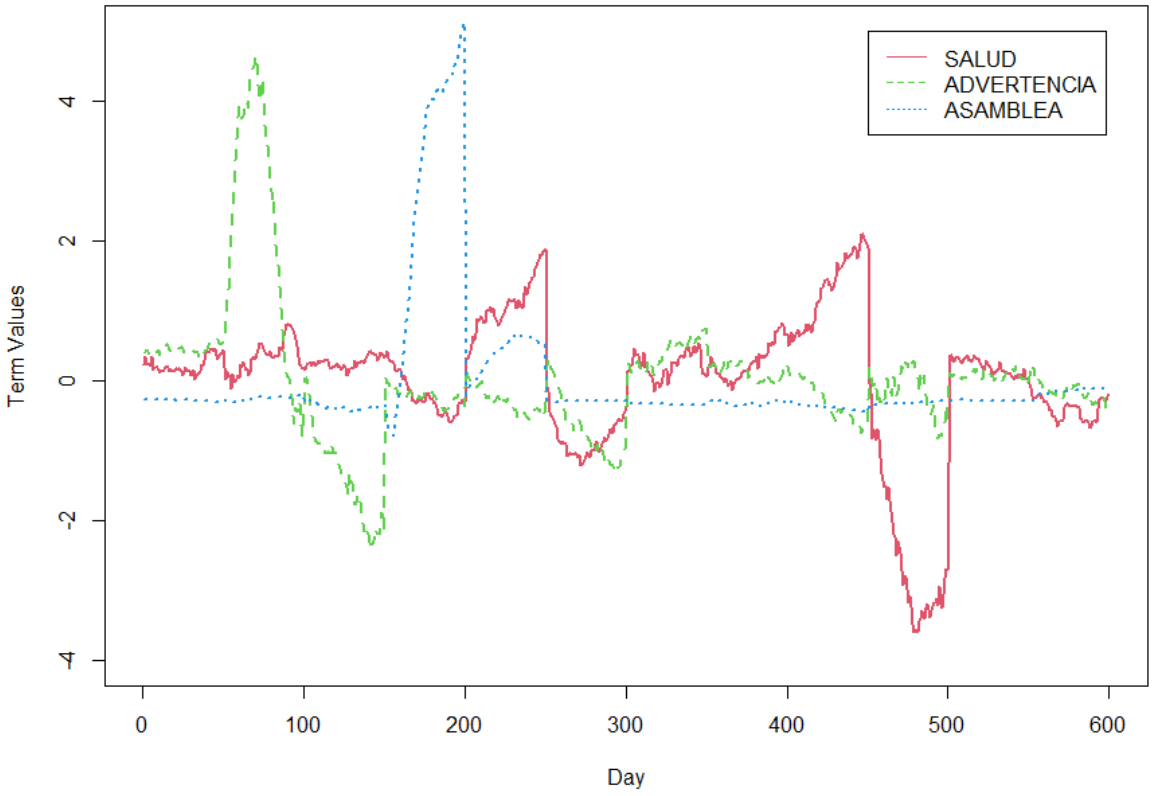

Figure 4: Normalized coefficients for the three most commonly selected variables in Argentina.

We also show that the active features were correctly identified more frequently with this approach, particularly when the parameters change in a completely dynamic fashion. However, even in a situation where the parameter values periodically shift, the dynamic logistic regression with the variable selection approach still outperforms the static LASSO.

The proposed methodology is applied to the problem of civil unrest in Latin America. We forecast the probability of future protest one day ahead and five days ahead in Argentina, Brazil, Colombia, Mexico, Paraguay, and Venezuela. Using only protest-related terms as model features extracted from Twitter, we show improved accuracy compared to the baseline static logistic regression using the LASSO regularization model. The F1 scores improved from a range of 60% to 92% for the baseline model to 97% to 100% for the dynamic logistic regression model. The dynamic model is able to forecast using the most recent information and account for dependencies between

successive observations. Furthermore, the flexibility of the model captures the inherent dynamic nature of Twitter and allows protest predictive terms to vary in time. The proposed variable selection technique dynamically selects predictors and captures the fluid reasons for civil unrest.

6. Appendix

6.1. Dynamic Logistic Regression Posterior Simulation

The joint posterior density for the general exponential family dynamic model here is written to include the latent variable,

$$\pi(\boldsymbol{y}_{1:T}, \boldsymbol{\beta}_{0:T}, \boldsymbol{W}, \boldsymbol{\omega}_{1:T}) = \pi(\boldsymbol{\beta}_0) \times \pi(\boldsymbol{W}) \times \prod_{t=1}^{T} \pi(y_t|\boldsymbol{\beta}_t) \times \prod_{t=1}^{T} \pi(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{W}) \prod_{t=1}^{T} \pi(\omega_t|\boldsymbol{\beta}_{t-1}) \quad (20)$$

For dynamic logistic regression, specify $\pi(y_t|\boldsymbol{\beta}_t)$ as the Bernoulli *observation equation*, $\pi(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{W})$ as the normally distributed *state equation*, and $\pi(\omega_t|\boldsymbol{\beta}_{t-1})$ as the Pólya-Gamma distributed latent variable. Let $\pi(\boldsymbol{\beta}_0)$ be the normally distributed prior for the initial state vector, which is now a conjugate prior due to the latent variable, and let $\pi(\boldsymbol{W})$ be the prior for the *state equation* covariance. For computational purposes, assume the covariance structure is $\boldsymbol{W} = diag\left(\frac{1}{\tau_1}, \ldots, \frac{1}{\tau_p}\right)$, where τ_i represents the i^{th} inverse variance component. Therefore, the prior can be written as a product due to its diagonal form, $\prod_{i=1}^{p} \pi(\tau_i)$. Let the inverse variance components be gamma distributed, $\pi(\tau_i) \sim gamma(\alpha, \gamma)$, to provide a conjugate prior. Given that the posterior density is completely specified, now draw posterior samples of $\{\boldsymbol{\beta}_{1:T}, \boldsymbol{W}, \boldsymbol{\omega}_{1:T}\}$ from the joint posterior distribution in the Equation above.

1. Initialize the latent variable vector $\boldsymbol{\omega}_{1:T}^{(0)}$, the states $\boldsymbol{\beta}_{1:T}^{(0)}$, and the state covariance $\boldsymbol{W}^{(0)}$.
2. For iterations $k = 1, \ldots, N$:
   (a) Sample $\boldsymbol{\beta}_{1:T}^{(k)}$ using the FFBS algorithm using pseudo Gaussian data from Equation 8a.
   (b) Sample the components of $\boldsymbol{W}^{(k)}$ individually from the updated Gamma distribution,

$$\pi\left(\tau_i^{(k)}|\cdot\right) \sim Gamma\left(\alpha + \frac{T}{2}, \gamma + \sum_{t=1}^{T} (\beta_{ti}^{(k)} - \beta_{(t-1)i}^{(k)})^2\right)$$

   for $i = 1, \ldots, p$.

Table 8: Forecast results for structural break parameters and $T = 50$.

Model	Forecast	Predictors	TP	TN	FP	FN	Precision	Recall	F1
LASSO	1	25	49.73	1.60	47.97	0.70	50.90	98.61	67.14
	5	25	49.45	1.85	47.98	0.73	50.75	98.54	67.00
DGLM	1	25	50.13	49.10	0.47	0.30	99.07	99.41	99.24
	5	25	48.38	47.60	2.22	1.79	95.61	96.43	96.02
LASSO	1	50	50.43	1.93	46.53	1.10	52.01	97.87	67.92
	5	50	50.60	1.90	46.51	0.98	52.11	98.10	68.06
DGLM	1	50	51.40	48.40	0.06	0.13	99.88	99.74	99.82
	5	50	50.70	47.51	0.90	0.88	98.26	98.29	98.27
LASSO	1	100	50.73	1.33	47.33	0.60	51.73	98.83	67.92
	5	100	50.71	1.04	47.72	0.53	51.51	98.96	67.76
DGLM	1	100	51.23	48.43	0.23	0.10	99.55	99.81	99.68
	5	100	50.61	48.04	0.72	0.63	98.59	98.77	98.68

(c) Sample each element of the latent variable vector $\omega_t^{(k)}$, from the Pólya-Gamma distribution conditioned on the states, $PG(1, \boldsymbol{X}_t\boldsymbol{\beta}_t^{(k)})$, for $t = 1, \ldots, T$.

Note, R code to fit the dynamic logistic regression model and conduct variable selection using penalized credible regions is located at `https://github.com/jordanbakerman/Dynamic`.

6.2. Structural Break Parameters Simulation Results

7. Acknowledgements

The research at North Carolina State University was funded by the Consortium for Nonproliferation Enabling Capabilities (CNEC) Grant DE-NA0002576, which is sponsored by the U.S. Department of Energy (DOE), National Nuclear Security Administration Office of Defense Nuclear Nonproliferation, and also National Science Foundation Grant DGE-1633587. The research at Pacific Northwest National Laboratory (PNNL) was part of the Analysis in Motion Initiative and was funded by the Laboratory Directed Research and Development Program (DE-AC05-76RL01830). PNNL is a multi-program national laboratory operated by Battelle for the DOE. This article has been cleared by PNNL for public release as PNNL-SA-135434.

Table 9: Forecast results for structural break parameters and $T = 100$.

Model	Forecast	Predictors	TP	TN	FP	FN	Precision	Recall	F1
LASSO	1	25	51.98	0.62	47.16	0.24	52.43	99.54	68.68
	5	25	51.91	0.63	47.21	0.24	52.37	99.54	68.63
DGLM	1	25	51.60	47.28	0.50	0.62	99.04	98.81	98.92
	5	25	49.60	45.25	2.59	2.56	95.04	95.09	95.06
LASSO	1	50	50.40	0.96	48.24	0.40	51.09	99.21	67.45
	5	50	50.47	1.03	48.16	0.34	51.17	99.33	67.54
DGLM	1	50	50.74	49.16	0.04	0.06	99.92	99.88	99.90
	5	50	49.95	48.50	0.69	0.85	98.64	98.33	98.48
LASSO	1	100	51.60	0.68	47.32	0.40	52.16	99.23	68.38
	5	100	51.67	0.70	47.26	0.37	52.22	99.29	68.45
DGLM	1	100	51.96	48.00	0.02	0.02	99.96	99.96	99.96
	5	100	51.86	47.61	0.35	0.18	99.32	99.65	99.49

Table 10: Variable selection results for structural break parameters.

Model	Predictors	T	TP	TN	FP	FN	Precision	Recall	F1
LASSO	25	50	6.02	68.07	19.93	5.98	23.19	50.17	31.73
DGLM	25	50	7.81	67.36	20.64	4.19	27.45	65.08	38.62
LASSO	25	100	8.70	69.22	18.78	3.29	31.66	72.56	44.08
DGLM	25	100	10.62	68.42	19.58	1.37	35.17	88.57	50.34
LASSO	50	50	3.65	77.69	12.31	6.35	22.87	36.50	28.12
DGLM	50	50	4.79	78.68	11.32	5.21	29.73	47.90	36.69
LASSO	50	100	5.95	73.99	16.00	4.05	27.12	59.50	37.24
DGLM	50	100	7.92	72.21	17.79	2.07	30.81	79.28	44.37
LASSO	100	50	1.87	82.91	7.09	8.13	20.87	18.70	19.73
DGLM	100	50	2.48	82.22	7.78	7.52	24.17	24.80	24.48
LASSO	100	100	3.61	80.28	9.71	6.39	27.10	36.10	30.96
DGLM	100	100	5.39	80.05	9.95	4.61	35.14	53.90	42.54

References

Harshavardhan Achrekar, Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu, and Benyuan Liu. Twitter improves seasonal influenza prediction. In *HEALTHINF*, pages 61–70, 2012.

Sitaram Asur and Bernardo A Huberman. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 492–499. IEEE, 2010.

Mohsen Bahrami, Yasin Findik, Burcin Bozkaya, and Selim Balcisoy. Twitter reveals: using Twitter analytics to predict public protests. *arXiv preprint arXiv:1805.00358*, 2018.

Ole E Barndorff-Nielsen and Neil Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001.

BBC. Colombian farmers protest against government's 'broken promises'. `http://www.bbc.com/news/world-latin-america-27198890`, 2014. Accessed 28-October-2017.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

Howard D Bondell and Brian J Reich. Consistent high-dimensional Bayesian variable selection via penalized credible regions. *Journal of the American Statistical Association*, 107(500):1610–1624, 2012.

François Caron, Luke Bornn, and Arnaud Doucet. Sparsity-promoting Bayesian dynamic linear models. *arXiv preprint arXiv:1203.0106*, 2012.

Chris K Carter and Robert Kohn. On Gibbs sampling for state space models. *Biometrika*, 81(3): 541–553, 1994.

CNN. Colombian president deploys 50,000 troops after violent protests. `http://www.cnn.com/2013/08/30/world/americas/colombia-protests/index.html`, 2013. Accessed 28-October-2017.

Michael D Conover, Emilio Ferrara, Filippo Menczer, and Alessandro Flammini. The digital evolution of occupy Wall Street. *PLOS One*, 8(5):e64679, 2013.

Aron Culotta. Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics*, pages 115–122. ACM, 2010.

Petros Dellaportas, Jonathan J Forster, and Ioannis Ntzoufras. On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12(1):27–36, 2002.

Krzysztof Drachal. Dynamic model averaging in economics and finance with fDMA: A package for R. *Signals*, 1(1):47–99, 2020.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of Statistics*, 32(2):407–499, 2004.

Nahed Eltantawy and Julie B Wiest. Social media in the Egyptian revolution: reconsidering resource mobilization theory. *International Journal of Communication*, 5:1207–1224, 2011.

Ali Mert Ertugrul, Yu-Ru Lin, Wen-Ting Chung, Muheng Yan, and Ang Li. Activism via attention: interpretable spatiotemporal learning to forecast protest activities. *EPJ Data Science*, 8(1):1–26, 2019.

Marco AR Ferreira and Dani Gamerman. Dynamic generalized linear models. In DK Dey, SK Ghosh, and BK Mallick, editors, *Generalized linear models: A Bayesian perspective*, pages 57–72. Marcel Dekker, 2000.

Dani Gamerman. Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, 85(1):215–227, 1998.

Edward I George and Robert E McCulloch. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

Matthew S Gerber. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014.

Jim E Griffin, Philip J Brown, et al. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010.

Jennifer A Hoeting, David Madigan, Adrian E Raftery, and Chris T Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, pages 382–401, 1999.

Maria Kalli and Jim E Griffin. Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, 178(2):779–793, 2014.

Gary Koop and Dimitris Korobilis. Forecasting inflation using dynamic model averaging. *International Economic Review*, 53(3):867–886, 2012.

Gary Koop and Luca Onorante. Macroeconomic nowcasting using Google probabilities. In I Jeliazkov and J Tobias, editors, *Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A*, pages 17–40. Emerald Publishing Limited, 2019.

Gizem Korkmaz, Jose Cadena, Chris J Kuhlman, Achla Marathe, Anil Vullikanti, and Naren Ramakrishnan. Multi-source models for civil unrest forecasting. *Social Network Analysis and Mining*, 6(50), 2016.

Jiwei Li and Claire Cardie. Early stage influenza detection from Twitter. *arXiv preprint arXiv:1309.7340*, 2013.

Gilad Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, et al. The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication*, 5:1375–1405, 2011.

Jinchi Lv and Yingying Fan. A unified approach to model selection and sparse recovery using regularized least squares. *Annals of Statistics*, pages 3498–3528, 2009.

Tyler H McCormick, Adrian E Raftery, David Madigan, and Randall S Burd. Dynamic logistic regression and dynamic model averaging for binary classification. *Biometrics*, 68(1):23–30, 2012.

Jouchi Nakajima and Mike West. Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*, 31(2):151–164, 2013.

Tahir M Nisar and Man Yeung. Twitter as a tool for forecasting stock market movements: A short-window event study. *Journal of Finance and Data Science*, 4(2):101–119, 2018.

Nima Nonejad. An overview of dynamic model averaging techniques in time-series econometrics. *Journal of Economic Surveys*, 35(2):566–614, 2021.

Luca Onorante and Adrian E Raftery. Dynamic model averaging in large model spaces using dynamic Occam's window. *European Economic Review*, 81:2–14, 2016.

Panagiotis Panagiotopoulos, Alinaghi Ziaee Bigdeli, and Steven Sams. 5 days in August–how London local authorities used Twitter during the 2011 riots. In HJ Scholl, M Janssen, MA Wimmer, CE Moe, and LS Flak, editors, *International Conference on Electronic Government*, pages 102–113. Springer, 2012.

Trevor Park and George Casella. The Bayesian LASSO. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

Giovanni Petris, Sonia Petrone, and Patrizia Campagnoli. Dynamic linear models. In G Petris, S Petrone, and P Campagnoli, editors, *Dynamic Linear Models with R*, pages 31–84. Springer, 2009.

Nicholas G Polson, James G Scott, and Jesse Windle. Bayesian inference for logistic models using Pólya–gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.

Fengcai Qiao and Kedi Chen. Predicting protest events with hidden Markov models. In *2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pages 109–114. IEEE, 2016.

Adrian E Raftery, Miroslav Kárnỳ, and Pavel Ettler. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1):52–66, 2010.

Naren Ramakrishnan, Patrick Butler, Sathappan Muthiah, Nathan Self, Rupinder Khandpur, Parang Saraf, Wei Wang, Jose Cadena, Anil Vullikanti, Gizem Korkmaz, et al. 'Beating the news' with EMBERS: forecasting civil unrest using open source indicators. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1799–1808. ACM, 2014.

Marian Risse and Ludwig Ohl. Using dynamic model averaging in state space representation with dynamic Occam's window and applications to the stock and gold market. *Journal of Empirical Finance*, 44:158–176, 2017.

Akshit Singh, Nagesh Shukla, and Nishikant Mishra. Social media data analytics to improve supply chain management in food industries. *Transportation Research Part E: Logistics and Transportation Review*, 114:398–415, 2018.

Mike West, P Jeff Harrison, and Helio S Migon. Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83, 1985.

Jesse Windle, Carlos M Carvalho, James G Scott, and Liang Sun. Efficient data augmentation in dynamic models for binary and count data. *arXiv preprint arXiv:1308.0774*, 2013.