What to do about Atheoretic Lags

ABSTRACT

We examine a problem that is confronted frequently by political science researchers seeking to model longitudinal data: what to do when one suspects a lag between the realization of a regressor and its effect on the outcome variable, but one has no theoretical reason to suspect a particular lag length. We examine the theoretical challenges posed by atheoretic lags, review existing methods for atheoretic lag analysis – most notably distributed lag specifications – and their shortcomings, and present an alternative approach for atheoretic lag analysis based on Bayesian model averaging (BMA). We demonstrate the use and utility of our approach with two examples: the litigant signal model in American politics and modernization theory in political economy. Our examples show the increasing difficulty of analyzing models with atheoretic lags as the set of possible specifications increases, and demonstrate the effectiveness of Bayesian model averaging for the modal type of specification in time-series cross-sectional applications.

1. INTRODUCTION

Temporally lagged regressors are commonly used in political science, and other social sciences as well, for two main reasons: the first is to accommodate the basic requirement that the hypothesized cause must occur temporally prior to the expected effect.[1] Most commonly, but particularly in the fields of international relations and comparative politics, this involves a lag of one period, usually a year. The second usage of lags is to test a relationship for which there is an expected delay between a change in one exogenous variable and change in another. In this case, the lag may be more than one period.

Within this second class of lags, there are some instances in which the length of the lag effect can readily be anticipated. For example, a change in the distribution of the population of the United States, as measured by the Census in 2010, can be anticipated to have an effect on the distribution of House seats two and (perhaps) four years later. More troubling, however, are research problems in which the expected time for the effect of change in a variable to be observed is either unknown or subject to some degree of uncertainty. We call these *atheoretic lags*. Consider the following from the path-breaking research by Ghobarah, Huth and Russett (2003, p. 193) on the effect of politics on Disability Adjusted Life years (DALY).

> "Using civil war deaths in the years 1991-97 gives us a lag to the DALY rates for 1999: Theory does not tell us that there is a single correct lag. For most infectious diseases–which we hypothesize as the principal cause of indirect civil war deaths–the lag time would seem short (less than five years) while the effects of damage to the health care system would probably last longer (between five and ten years). The lag for some cancers could be so long that we cannot reasonably test for many of them."

The uncertainty they face with respect to the appropriate lag is what we call an *atheoretic lag* problem.[2]

[1] One should note that this discussion pertains to the evaluation of causal theories through correlational analysis, by far the most common means of making inferences in political science.

[2] To their credit Ghobarah, Huth and Russett (2003) are well aware of the difficulty they face and they report exploring the sensitivity of their results to various lags. Our point is not that they, or anyone else, who recognizes the problem and explores various lags has

One common approach to an atheoretic lag problem is to specify a lag that seems reasonable. This represents a "best guess" procedure. Without a substantive or theoretical reason, such an approach cannot help but be arbitrary. Two related problems accompany this procedure. First, the reader of the research report cannot know if the results presented are robust or if they break down when different, but also reasonable, lag lengths are specified. At the very least, this practice poses problems for the transparency of results, and at the worst could lead a field to accept non-robust results as reliable knowledge. Second, even if the results are more or less robust to different lag lengths, presenting only one statistical model and acting as if that model generated the data underestimates uncertainty and produces overconfident inferences. Even if the author explicitly acknowledges these issues and exercises caution in interpreting their results, the reader is provided no empirical method by which to adjust their own interpretation, and thus the research breaks from typical standards of research transparency and statistical rigor. In all, the "best guess" as well as other problematic approaches present important implications for political science. One survey of articles appearing between 1995 and 2005 in the field's top three journals identified more than 70 articles invoking lagged effects in time series regression models, with *none* of those articles reporting tests for the included lag (DeBoef and Keele 2008). With research failing to justify lag choices, the robustness, transparency, and overconfidence issues are potentially profound.

Our solution is a simple one, but one that has long gone unnoticed and under-utilized in political science. We posit that uncertainty regarding the length of a lag is a special case of uncertainty in the structure of the statistical model. Because we must appropriately reflect the degree of uncertainty in the model specification, we propose the use of Bayesian model averaging (BMA) over comparable models with varying lag lengths. This approach allows us to average across multiple candidate lags and appropriately capture our degree of uncertainty with respect to the data generating process. Despite its limitation to relatively simple regression models (e.g. those without elaborate temporal dynamics), this process produces a statistically sound and easily interpretable set of results, sheds light on the temporal process at work in the lag, and provides a solution for the modal category of political analyses in which time-series cross-sectional data are modeled with a lagged outcome variable and possibly some fixed effects. We illustrate the atheoretic lag problem and how BMA can solve it through two applications: the litigant signal model in American politics, and the

necessarily done anything incorrect, but only that their analysis may be inefficient and less revealing of the underlying data generation process than could be the case.

modernization model in political economy.

2. PREVIOUS APPROACHES TO ATHEORETIC LAGS

We, of course, are not the first to be concerned with the problem of atheoretic lags. Consternation over the atheoretic lag problem dates back more than 50 years. A class of models called *distributed lag models* was developed for atheoretic lags, primarily within the field of econometrics (Almon 1965; Amemiya and Fuller 1967; Dufour and Kiviet 1998; Geweke 1978; Koyck 1954; Lütkepohl 1980; Martin 1967; Nelson and Schwert 1974; Solow 1960; Trivedi 1985) and later introduced to political science (Beck 1985, 1991). These models are called distributed lag models in reference to the fact that the effect of the regressor is distributed across all the lag lengths thought to be feasible. For example, a simple linear model with a single explanatory variable and a set of feasible lag lengths defined generally as $\mathbf{k} = \{0, 1, \ldots, K\}$ would take the following form:

$$y_i^t = \beta_0 + \beta_1 x_i^t + \beta_2 x_i^{t-1} + \ldots + \beta_N x_i^{t-K} + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

where time is denoted with superscripts and units are denoted with subscripts. As can be seen in equation (1), the implementation of distributed lag models is straightforward: one need only create a series of lagged x variables corresponding to the varying lags in the feasible set and include those lags as regressors in a regression model.

As explained by DeBoef and Keele (2008), there are three major advantages to the use of distributed lag models. First, if the regressor has effects at several lag lengths, including each of those lags produces models that fit the data well. Second, researchers can produce intuitive descriptions of the lagged effects based on the coefficients of the lagged variables. For example, if the effect of the regressor was decreasing as lag lengths increase, this would be shown clearly by a monotonic reduction in point estimates as lag lengths increase. Third, researchers can potentially evaluate both short *and long-term* effects of the predictors, though this potentially useful avenue for theory-testing has been severely underutilized in research to date. In all, the models serve as a useful tool for the modeling of time dynamics in political and social science research.

Yet while simple and useful in a variety of atheoretic lag situations, distributed lag models suffer from several unresolvable shortcomings. First, if the set of feasible lag lengths is considerable and/or if more than one variable is specified as a distributed lag, researchers

will often find themselves with a large number of regressors in their specification. This is particularly problematic when the number of observations in a dataset is relatively small, but even with large datasets interpretation of the many coefficients becomes cumbersome and arduous. Second, in the common case where the variable being lagged has a high degree of autocorrelation, the inclusion of its lags often generates substantial collinearity, thus biasing the standard errors upward; in extreme situations, this can even prevent the estimation of the model. Finally, distributed lag models require the effect of the lag to follow either a polynomial or geometric distribution; meaning that, as lag lengths increase, the effect must increase monotonically to a peak and then decrease monotonically to zero (polynomial) or decrease monotonically towards zero from the initial lag (exponential). Note that none of these issues are resolved by using the Akaike Information Criterior [AIC] for variable selection (Akaike 1974). When these shortcomings are not present, then distributed lags can readily be and have been used and with great advantage. However, if they are present, the researcher will likely be better served by alternative model specifications.

Before introducing our approach to atheoretic lag analysis, we must first define a concept common to all atheoretic lag techniques, including distributed lag models: that of the feasible set of lags. When undertaking an analysis with atheoretic lags, the researcher must begin by defining the set of lags ($\mathbf{k}$) at which the effect of the regressor of interest is feasible. This set must be finite and the substance of the research problem will determine the range of the set. For example, if a researcher is analyzing the effect of changes in per capita income on levels of democracy, the specification might be $\mathbf{k} = \{1, 2, \ldots, 15\}$, meaning that the researcher deems it reasonable to think the effect of a change in per capita income may take as long as 15 time periods (years in this case) to be realized in the outcome variable or that the effect of income could be felt for as long as 15 periods after its realization. The specification of the feasible set is a theoretically important step that dictates the lag space to be addressed, and as such, should be tied as closely to the substance of the model as possible.

A cautionary note is also necessary before proceeding: researchers implementing the lag techniques described here should note that the inclusion of lags reduces the time span of the analysis by the number of periods of the lag. For example, if one has (as we do for the modernization data analyzed in Section 6) 55 years of data (1945-2000), and specifies $\mathbf{k} = \{1, 2, \ldots, 15\}$, a good deal of missing data is generated as the lags get longer. When $k = 1$, the time span of the analysis is 1946-2000 because $t - 1$ for 1945 (the lowest time value in the dataset) is unknown. For $k = 2$, the time span is 1947-2000 and so on until for $k = 15$, the time span being analyzed is 1960-2000. The reduction of the period of analysis

is perfectly tied to the span of the feasible set, and this tradeoff should be considered by researchers as they specify the feasible set.

3. ADDRESSING ATHEORETIC LAGS THROUGH BAYESIAN MODEL AVERAGING

A simple alternative to the distributed lag model is to address atheoretic lags through a process known as Bayesian model averaging (BMA). BMA, as the name suggests, is a Bayesian technique for making a single inference from a finite set of models, essentially by averaging over the results from the models. BMA has the benefit of appropriately accounting for uncertainty in the model specification through a process and framework well grounded in statistical theory, and affords analysts the ability to make powerful statements about the probabilities of certain lags and even certain models being part of the data generating process.

BMA traces its history back to Barnard (1963), who first suggested the combination of multiple models, but the technique only came into significant usage in the 1970's by economists forecasting financial markets; Clemen (1989) provides an excellent overview of this early literature. Though the basic framework for BMA was developed by Leamer (1978), like many elements of Bayesian statistics, it was not until the 1990s – with the invention of Markov chain Monte Carolo and the computational power to implement it – that BMA became practical and used more widely, both in statistics and applied fields (see e.g. Madigan and Raftery (1994), Raftery (1995), and Draper (1995)). In these fields, BMA is most typically used in specification searches, to model uncertainty in which regressors should be included, or for forecasting, because BMA tends to produce better predictions than single models (Hoeting et al. 1999). In political science, BMA has been used less often than it has in perhaps most fields in which statical analysis is widespread. The technique was first presented by Bartels (1997), received a technical review by Gill (2007), and has more recently been advocated by Montgomery and Nyhan (2010). Still, substantive applications have been limited: Bartels and Zaller (2001), Erikson, Bafumi and Wilson (2001), Zaller (2004), Imai and King (2004), and Geer and Lau (2006).

BMA is designed specifically to address uncertainty in the model specification. Most typically, this takes the form of what regressors should be included. However, we posit that atheoretic lags are a special case of model uncertainty: where the uncertainty pertains not to what variables to include, but to what lags should be included. BMA can thus be used to average over uncertainty in the feasible set of lags and appropriately reflect lag uncertainty

in the same way that it would reflect model uncertainty.

BMA does not require particularly strong assumptions, but it does require a Bayesian framework. Such a framework implies several key differences from the likelihood-based statistics that are most commonly used in political science. Careful exposition of these differences can be found in work such as Gill (2007) and Jackman (2009) and is beyond the scope of this discussion. It will suffice to point out here that, because Bayesian models treat the data as constant and the parameters as random, they produce distributions for the parameters, called posterior distributions, that are described simply by measures of central tendency and spread (e.g. means and standard deviations). Though results may be interpreted in a similar fashion, this is a notable departure from the likelihood framework in which the data are treated as random (drawn from the sampling distribution) and the parameters are treated as fixed.

The mechanics of BMA are relatively straightforward and require only two concepts that are new to most analysts familiar with basic Bayesian methods: assigning priors to *models* (rather than simply to parameters) and producing posterior distributions that are averaged across a finite set of models. Using notation based on Hoeting et al. (1999) and Gill (2007), denote each of the candidate models $M_1, M_2, \ldots, M_K$. While in the general case, these models may differ substantially in their specification and structure, when applying BMA to the atheoretic lag problem, model M_1 will have a different lag length than model M_2 and so on, where K is the limit of the feasible set. Note too that this does not imply that the feasible set must begin at a one-period lag, M_1 is merely the first model in the feasible set.

Now, for each k, one must determine a prior for the model $p(M_k)$. This is the subjective prior probability of the model M_k itself. These priors are generally scalars which sum to one. Though one need not have low information or uniform priors – indeed, some lag lengths may be seen by the researcher as substantially more likely than others – a common prior that is natural for the researcher without strong prior beliefs about model probabilities is $1/K$ for all models (Hoeting et al. 1999).

Consider then the formula for Bayes law (and thus the basic structure of Bayesian models):

$$p(\boldsymbol{\theta}|\mathbf{D}) = \frac{p(\mathbf{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{D})}, \tag{2}$$

where θ are the parameters and $\mathbf{D}$ represents the data, $p(\mathbf{D}|\boldsymbol{\theta})$ is the sampling density (i.e. the likelihood specified identically to non-Bayesian models), $p(\boldsymbol{\theta})$ is the vector of priors for the parameters, and $p(\mathbf{D})$ is the unconditional probability of the data. Recall too that the

denominator of Bayes law, commonly referred to as the integrated likelihood, is

$$p(\mathbf{D}) = \int p(\boldsymbol{\theta}, \mathbf{D}) p(\boldsymbol{\theta}) d\theta. \tag{3}$$

Notice though, that writing the integrated likelihood as such, is equivalent to writing it as

$$p(\mathbf{D}|M_k) = \int p(\boldsymbol{\theta}_k|M_k, \mathbf{D}) p(\boldsymbol{\theta}_k|M_k) d\boldsymbol{\theta}_k. \tag{4}$$

if only one model M_k is considered and the probability of that model is 1. However, when multiple models are considered, one still articulates the integrated likelihood as in (4), only now each M_k will not be equal to 1.

Writing the integrated likelihood conditional on the model M_k, as in (4), is the conceptual leap necessary to compute the posterior model probability for every model in the feasible set $k = 1, \ldots, K$. One must simply apply Bayes' law:

$$\pi(M_k|\mathbf{D}) = \frac{p(\mathbf{D}|M_k)p(M_k)}{\sum_{j=1}^{K} p(\mathbf{D}|M_j)p(M_j)}, \tag{5}$$

where the denominator is summed rather than integrated because the feasible set is discrete. Note too that these probabilities are implicitly conditional on the feasible set of lags.

While $\pi(M_k|\mathbf{D})$ is a useful quantity in and of itself – indeed, it tells the analyst directly about the probability of any particular model M_k (again implicitly conditional on the feasible set) – it can also be used to compute other quantities of interest. For example, in the variable selection literature, it is usually used to compute the posterior probability that a given regressor is in the model:

$$P(\boldsymbol{\theta}_j \neq 0|\mathbf{D}) = \sum_{\boldsymbol{\theta}_j \in M_k} \pi(M_k|\mathbf{D}), \tag{6}$$

though this quantity is less useful in the case of atheoretic lags.

In general, we will be interested in computing the posterior mean and variance for each parameter. For example, given a simple linear model, these quantities would be

$$E[\boldsymbol{\theta}_j|\mathbf{D}] \approx \sum_{i=1}^{K} \boldsymbol{\theta}_j(k)\pi(M_k|\mathbf{D}) \tag{7}$$

and,

$$Var[\boldsymbol{\theta}_j|\mathbf{D}] \approx \sum_{k=1}^{K} \left[(Var_{\boldsymbol{\theta}_j}(k) + \boldsymbol{\theta}_j(k)^2)\pi(M_k|\mathbf{D}) - E[\boldsymbol{\theta}_j|\mathbf{D}]^2 \right], \tag{8}$$

see Raftery (1995). The computation of such quantities is, however, simplified considerably because BMAs are estimated using MCMC and, for example, the computation of a posterior mean is as simple as taking the empirical mean over the MCMC draws. BMA is also well-implemented in freely available software and can be used rather simply for the majority of common approaches in political science.

4. LIMITATIONS OF THE BMA APPROACH

We wish to make clear, prior to showing our approach in application, that using BMA to address atheoretic lags has a few important limitations. Yet we also wish to explicate why, despite these limitations, we find the approach compelling for the majority of situations encountered by political scientists. There are a few major points/limitations that warrant specific attention.

First, BMA is generally limited by its need to compute the integrated likelihood described in equation (3). This is in contrast to most Bayesian models, which can use proportionality to the posterior to skip computation of the integrated likelihood. This need makes using BMA progressively harder as the structure of the statistical model becomes more complicated. While BMA is not particularly difficult for, say, a logit model with several regressors and a lagged dependent variable, it will be substantially more difficult once greater structure (e.g., multilevel structure or a more nuanced incorporation of temporal dynamics) is added, possibly prohibitively difficult. This is a limitation without remedy. However, we believe that BMA is still broadly useful as the majority of regression models in political science, including those with time series / panel data, do not include structure much more complicated than a generalized linear model with a lagged dependent variable (e.g., in general keeping with the advice given by Beck and Katz (1996)). In other words, BMA is comparatively easy to compute and well-implemented for the modal category of empirical analyses in political science, though it does suffer from a substantial difficulty with more complicated models.

Second, BMA is often used in the context of variable selection problems (e.g., Valentino, Brader and Jardina 2013; Warren 2014). When the true specification of the model is unknown, BMA may help the analyst choose, or at least justify, a specification based on the

posterior model probabilities and/or the probabilities assigned to specific regressors. Substantial progress in variable selection has been made in recent decades, particularly in the machine learning literature, and it is fair to say that BMA is not an especially powerful or flexible technique when compared to selection algorithms such as Boosting (Freund and Schapire 1996, 1997) and the Lasso method (Tibshirani 1996, 1997). The reason alternative variable selection algorithms may be considered superior to BMA is because they work quickly and efficiently, even with very complicated model structures or in high dimensional situations, whereas BMA becomes increasingly difficult as the complexity of the model structure increases. Indeed, as discussed above, the more complicated the model structure, the more difficult it is to compute the integrated likelihood. On the other hand, Boosting, the Lasso, and other variable selection techniques do not require this quantity and can easily produce a well-fitting specification with even very complicated models.

Why then would we suggest that BMA is fruitful here? The reason is that we are not searching for *a* good specification of the model, nor do we seek solely to reduce a high dimensional space. Instead, we seek to properly reflect uncertainty in the lag length across all lags in the feasible set, which is something alternative variable selection techniques cannot and do not intend to do. To make this point more concrete, consider the case where the feasible set of lags for some variable of interest is five. Boosting or the Lasso would produce a specification with the best fitting lag or set of lags (suppose lag 3 in this case). In other words, it selects one or a few lags from the feasible set and then assumes the model specification is correct, *an approach that underestimates uncertainty in the model specification*. Alternatively, BMA averages over all lags in the feasible set – a set defined by theory even though a specific lag length is not – and thus appropriately reflects our uncertainty in the specification. It is true that the BMA approach will include and average over models with lag lengths that may be unlikely compared to some others within the feasible set, but, by virtue of the fact that they are in the feasible set, they should be included. Viewed as such, the interest here is not in variable selection, but in fully and appropriately accounting for model uncertainty. In this respect, better, faster, and more flexible variable selection techniques cannot compare to BMA.

What is more, BMA is flexible enough to consider a multiplicity of model specification combinatorials across lags of the independent variables beyond what we present below, and is applicable to a wide variety of time series models with atheoretic lagged regressors and an outcome variable. In the former case, within a BMA framework, the researcher may consider an extensive configuration of different variables lagged by different periods within the feasible

set. This is of particular note because, while in what follows the BMA is computed across only k models, with k equal to the length of the feasible set, the approach generalizes to the context when the researcher would instead wish to consider all 2^k models. In the latter case, BMA is a method for averaging over many different model specifications, and has been employed in an extensive number of different research settings. Thus, the approach also generalizes to other time series models with atheoretic lags.

Finally, one may lament that the analyst loses information on the temporal dynamics of the lagged effects that a distributed lag model could recover. For example, does the magnitude of the coefficients across feasible lags follow a parabolic pattern, or one of decay? Alternatively, what are the long-term effects versus the short-term effects of a one-unit changed in the lagged variable (DeBoef and Keele 2008)? While, in the case of the temporal process, some information is recoverable by looking at the posterior model probabilities and their pattern as lag lengths increase (see examples below), we do lose information on long-term effects. Yet we do not believe the loss of such fine-grained information on individual lag effects from BMA is particularly problematic for two reasons. First, the reason an atheoretic lag is atheoretic is because the analyst does not have specific expectations about what the temporal process looks like; instead, the analyst has a potentially great deal of uncertainty about the estimates and thus should hesitate to conclude a relationship from the inclusion of multiple lags in a single model. Second, though long-term effects may be substantively useful for theory development (DeBoef and Keele 2008), their accuracy is contingent on the proper specification of the model. As highlighted above, and in DeBoef and Keele (2008), there are a number of reasons to suspect specification problems, or at the least a significant understatement of uncertainty. Moreover, these long-term effects are rarely reported (DeBoef and Keele 2008). Therefore, the drawback of potential lost information is minimized by the trade-off of rarely reported statistics for a better reflection of model uncertainty.

In all, while we readily acknowledge the limitations of BMA, we argue for its utility in atheoretic lag situations in so far as it is readily-implemented for the majority of political science research applications, and provides a more accurate reflection of model uncertainty in these applications.

Now that we have seen, theoretically, how BMA can be used to address atheoretic lag problems, we illustrate this solution with two real political science examples drawn from the literature.[3] The first example – the litigant signal model from American politics – offers a

[3]In the models we discuss, we assume a generally consistent relationship between the lagged regressor and the outcome variable across subgroups in the data. As in general

relatively uncomplicated distributed lag model, and provides evidence of the validity of BMA in recovering substantively similar results as distributed lag models while also accounting for model uncertainty in ways that distributed lag models do not. The second example, from international politics/political economy, offers evidence of the utility of BMA in settings where theoretical models are implied but distributed lag specifications are inestimable due to intractable methodological issues.

5. APPLICATION I: THE LITIGANT SIGNAL MODEL.

Our first example considers a model in American politics that overlaps a number of subfields including interest group, agenda-setting, and institutional research. As the Supreme Court can choose to hear only those cases brought to it, studies of the Court have predominantly focused on case-specific characteristics and the *certiorari* process whereby the Court can choose from the many cases brought to it those cases the Justices wish to hear (e.g., Black and Owens 2009; Brenner and Krol 1989; Caldeira 1981; Ulmer 1984; Caldeira, Wright and Zorn 1999). Taking a broader institutional perspective, the litigant signal model of Supreme Court agenda-setting (Baird 2004, 2007; Baird and Jacobi 2009) suggests that the Court's behavior shapes the pool of available cases through signals to litigants, whether intentional or not. As opposed to a view of a passive institution awaiting litigation (Epp 1998), the model indicates an institution with powerful agenda-setting capabilities.

Built on extensive interviews with justices (Perry 1991) and a rich literature on the role of interest groups as policy-minded litigants (Epstein and Rowland 1991; Epstein and Kobylka 1992; Caldeira and Wright 1988; McGuire and Caldeira 1993), the model holds that sitting justices signal litigants and members of the interest group community to mobilize or reframe arguments in policy areas the sitting Justices consider important or of some priority. Subsequently, additional and better-framed cases are available in the pool from which the justices will choose cases to be heard. These cases then stand a better chance of being chosen from the pool of available Court cases, increasing the subsequent attention the Court pays to a policy area.

Evidence for the litigant signal model is found in an increase in the Supreme Court's

with time series cross-sectional data, the author should explicitly model instances where the dynamics across subgroups vary in significant and important ways (Green, Kim and Yoon 2001).

docket within a policy area five years after politically salient decisions in that very same policy area (Baird 2004; Peters 2007) and four and five years after the Court exerts unusual amounts of influence in that area (Baird 2007). While subsequent research (Peters 2007) has questioned the underlying mobilization of litigants, the methodological approach has been carried over with each author utilizing an autoregressive distributed lag specification. Therefore, all approaches to the study of the litigant signal model and the dynamic processes underlying it have employed an approach that suggests a number of the potential issues identified above.

The data for this application are taken directly from Baird (2004) and are maintained as a panel dataset with eleven policy areas of Supreme Court agenda attention covering 1953 through 1995; these are identical to Baird's analyses.[4] The outcome variable, a measure of the Court's agenda, is simply the count of Supreme Court cases within a policy area during a year. The model we specify of the Court's agenda is a hybrid of theoretical models from Baird (2004) and Baird (2007) and not an exact replication. Specifically, the theoretical model[5] estimated here is as follows:

$$\begin{aligned} CASELOADS = {} & POLICY\ \ PRIORITY + LEGIS.\ ATTENTION \\ & + COURT\ COMPOSITION + DECISION\ IDEOLOGY \\ & + CHIEF\ JUSTICE. \end{aligned} \tag{9}$$

Existing research has primarily utilized two measures of the Court's signals to litigants: a composite measure of policy area activity (Baird 2007) or a count of salient decisions (Baird 2004). Here, we utilize the composite measure of policy area priority (Baird 2007), which is an index of (1) salient decisions (Epstein and Segal 2000), (2) declarations of unconstitutionality, (3) lower court reversals, and (4) formal alterations of precedent. In line with previous research, we expect this to have a significant, positive relationship with the number of cases

[4]These eleven policy areas are discrimination, free speech, privacy, criminal, labor, environment, economic regulation, taxation, due process, judicial, and federalism.

[5]In presenting our theoretical models, we aim to provide a brief and gentle introduction to the underlying theories which motivate our applications before estimating a series of equations for comparison. Particularly given that the temporal (t) subscripts vary across subsequent specifications (reduced form, distributed lag, and BMA), we omit subscripts from the theoretical models.

	Mean	Std. Deviation	2.5%	97.5%
Lag(Supreme Court Cases)[LOV]	0.60	0.05	0.50	0.70
Lag(Index of Policy Priority)	0.30	1.14	-1.96	2.56
Lag(Legislative Attention)	-0.53	1.54	-3.55	2.51
Mean Ideology Output	-1.15	0.51	-2.16	-0.17
Median Ideology Change	1.59	1.25	-0.87	4.03
(Intercept)	3.25	0.76	1.76	4.75
N=435				

Results for Lags

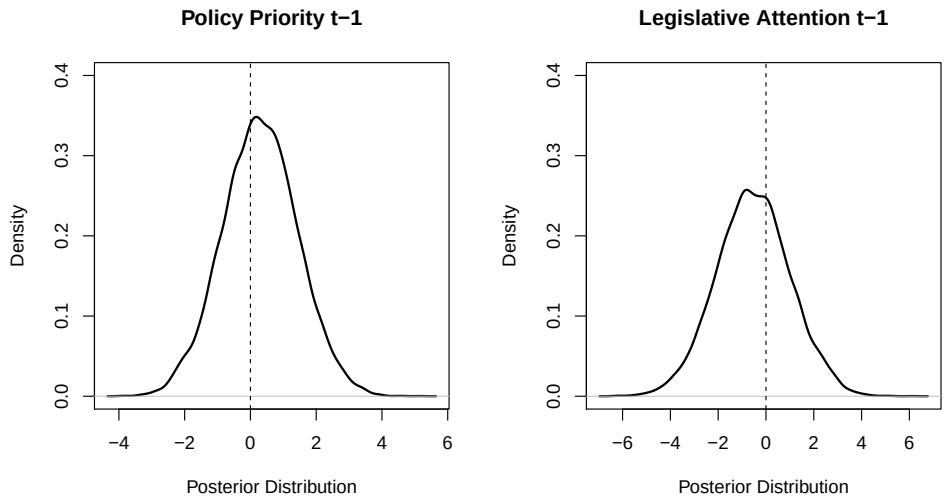

Table 1: *The baseline litigant signal model.* The outcome variable in this analysis is the total number of Supreme Court cases within a policy area during a year. The specification includes a lagged outcome variable, and one-year lags on index for Policy Priority and the variable for Legislative Attention. Fixed effects for the Chief Justice and policy area were included but are not reported.

in a policy area at four or five year lags.

Additional variables accounting for legislative attention, the Court's ideology, and the sitting Chief Justice are also included. Legislative attention is again an index, this time comprised of attention devoted to a policy area in *Congressional Quarterly* and legislative hearings in that policy area. In previous analyses, this variable has not attained standard levels of statistical significance at any of the reported lags. The Court's ideology is measured as the change in the Martin and Quinn (2002) ideology score of the median justice from one time period to the next. Finally, indicator variables for the sitting Chief Justice are included, with Warren serving as the excluded category.

We implement this specification as a Bayesian linear regression and estimate it via MCMC (Gibbs sampling).[6] In this model, we also include a lagged outcome variable and fixed effects for policy area. The results in Table 1, which we consider the baseline results for the theoretical model,[7] are reported for a model with policy priority and legislative attention each lagged one year. In this instance, the first period lag on the primary regressor of interest—the policy priority index—reveals little evidence of systematic influence on the future number of Supreme Court cases. This is unsurprising given the results of Baird (2004, 2007) and Peters (2007), where the significant impact came in later years though the exact time period was not known or specified *a priori*. It is this dynamic process and the impact of specification we hope to model below.

5.1. *Distributed Lag Model.*

In previous examinations of the litigant signal model (2004; 2007), Baird has utilized a distributed lag approach with six lags of two variables: indexes of the Court's policy priorities and indexes of legislative attention. The model has therefore included six lags of the index of Supreme Court policy priorities and six lags of a measure of legislative attention, used as controls. Baird defines the feasible set, though the length of these lags is acknowledged as atheoretic: "I chose to include the first six lagged years of the index because it would be difficult to claim that the effect of the signal of policy priorities would last longer than six years" (Baird 2007, p. 94).

The results of our distributed lag model based on the theoretical model in equation 1 and Baird's (2004) specification of the feasible set are presented in Table 2. As before, dichotomous variables for each policy area and sitting Chief Justice are included in the model but are not reported here.

The distributed lag analysis in Table 2 shows the strongest relationships at the fourth and fifth lags of the index of Policy Priority; not only are the posterior means greatest for these lags, but they are the only lags for which the 95% credible intervals do not include zero. We can draw no such conclusions from any of the lags of Legislative Attention. In the

[6]In all models, we discard a 1,000 iteration burn-in sample, and utilize a subsequent 10,000 iteration sample. Convergence diagnostics show no indication of possible non-convergence.

[7]The results reflect a baseline to the extent that the variables of interest are lagged one year, thus satisfying the need for causes to occur before effects, but no effort has been made to find the appropriate lag lengths.

	Mean	Std. Deviation	2.5%	97.5%
Index of Policy Priority$_{t-1}$	-1.09	1.13	-3.33	1.11
Index of Policy Priority$_{t-2}$	0.94	0.93	-0.88	2.80
Index of Policy Priority$_{t-3}$	0.98	0.93	-0.84	2.80
Index of Policy Priority$_{t-4}$	3.87	0.92	2.09	5.71
Index of Policy Priority$_{t-5}$	2.87	0.93	1.01	4.67
Index of Policy Priority$_{t-6}$	0.61	0.91	-1.21	2.38
Legislative Attention$_{t-1}$	0.57	2.02	-3.36	4.49
Legislative Attention$_{t-2}$	0.82	2.14	-3.41	5.00
Legislative Attention$_{t-3}$	-1.79	2.18	-6.10	2.49
Legislative Attention$_{t-4}$	-0.37	2.18	-4.72	3.84
Legislative Attention$_{t-5}$	-0.23	2.12	-4.29	4.00
Legislative Attention$_{t-6}$	1.31	2.02	-2.57	5.29
Median Ideology Change	1.50	1.19	-0.81	3.80
Mean Ideology Output	-1.10	0.51	-2.12	-0.11
Supreme Court Cases$_{t-1}$	0.45	0.06	0.34	0.56
Intercept	1.12	0.81	-0.45	2.73
N=387				

Trends in Lag Coefficients

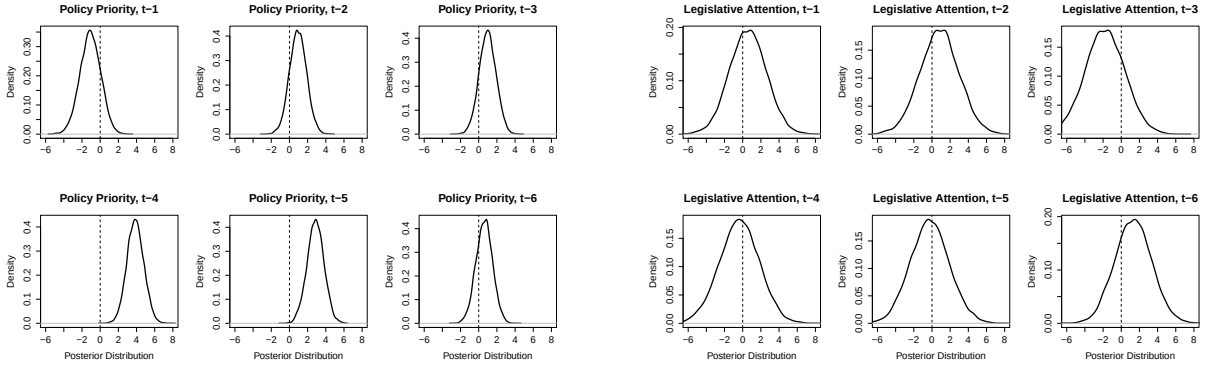

Table 2: *Distributed Lag Linear Model of Supreme Court Caseloads.* The outcome variable in this analysis is the total number of Supreme Court cases within a policy area during a year. The specification includes distributed lags $\mathbf{k} = \{1, 2, \ldots, 6\}$ for legislative attention and the index of policy priorities. Controls for the Chief Justice and policy area were included but are not reported above.

lower cell of the table, we present density plots of the posterior distributions of the index of Policy Priority and Legislative Attention. We can see yet more clearly that only four and five year lags for Policy Priority exhibit a marked relationship with subsequent Supreme Court attention.

The results in Table 2 providing evidence that four and five years after the Supreme Court indicates a policy area as a priority, there is a significant increase in the total number of cases the court hears in that policy area. The only other effect on which we can draw any conclusion is the measure of the ideological output of the Court.

Why did this model not fit better or produce more in the way of clear results? With two variables each lagged six times as well as a lagged outcome variable, the model has a large amount of potential collinearity which biases standard errors upwards and leads to type I errors. The inclusion of such a large number of collinear lagged variables also leads to overfit models, which may present problems for future refinements in litigant signal models of Supreme Court agenda-setting. As but one example, the model includes no indication of Presidential agenda priorities, despite the fact that executive branch attention influences the attention of other institutions (Flemming, Wood and Bohte 1999) and — through the office of the solicitor general — exerts strong and consistent influence on the Court's behavior (e.g., Nicholson and Collins 2008).

5.2. *BMA and Atheoretic Lags.*

We now adapt the specification of the litigant signal model from its distributed lag specification above in order to average across multiple models with varying lags. Such an approach better reflects our uncertainty over the appropriate lag length, while mitigating collinearity and avoiding overfit models. Recall that the two regressors for which we have atheoretic lag problems are the index of Policy Priority and the measure of Legislative Attention. Baird (2007) argued that it did not make sense for these effects to take longer than six years to affect the outcome (p. 94), but did not have a theoretical reason for selecting a particular lag. Given the strengths of BMA, we can include additional lags; as documented below, changing the choice of the number of lags has little substantive effect on the conclusions we draw. Therefore, we estimate eight different models with Policy Priority and Legislative Attention both lagged between one and eight years, then average across the model space. The results are shown in Table 3.

	Mean	Std. Deviation	2.5%	97.5%
Lag(Supreme Court Cases)[LOV]	0.51	0.06	0.39	0.62
Lags(Index of Policy Priority)	5.08	2.33	0.52	9.65
Lags(Legislative Attention)	0.12	1.58	-2.98	3.22
Mean Ideology Output	-1.22	0.51	-2.21	-0.23
Median Ideology Change	1.66	1.21	-0.71	4.04
(Intercept)	1.83	1.04	-0.21	3.87
N = 387; feasible set = 8.				

Results for Lags

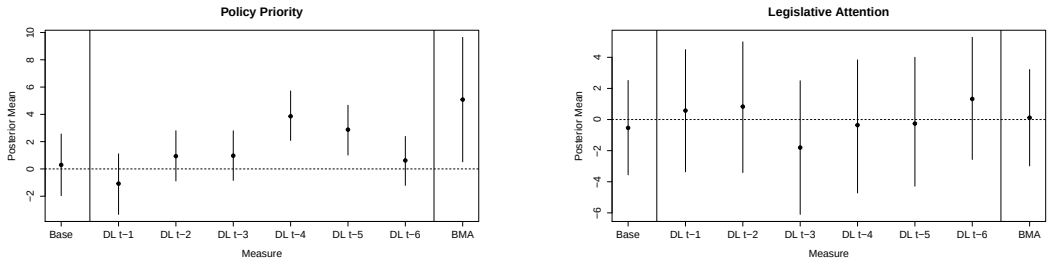

Table 3: *Bayesian model averaging across feasible set: Litigant Signal model.* The outcome variable in this analysis is the total number of Supreme Court cases within a policy area during a year. The specification includes a lagged outcome variable, and atheoretic lags on index for Policy Priority and the variable for Legislative Attention. Fixed effects for the Chief Justice and policy area were included but are not reported.

Table 3 displays the posterior means of the lagged coefficients from the model averaging. Note first that the substantive results do not change from the distributed lag specification; only the index of Policy Priority and the measure of the Court's ideological output have statistically reliable effects on future Supreme Court caseloads, as judged by their 95% credible intervals not including zero. In other words, the conclusions of the literature persist under our approach. Yet, the BMA approach does not lead to over-confident conclusions in particular time-series dynamics. Rather, in taking seriously the researcher's *a priori* uncertainty over lag length, we conclude only that the amount of priority the Court accords some issue area has some subsequent influence on caseloads within that issue area.

This dynamic is particularly evident in the parameter plots of Table 3. Each plot in the Figure corresponds to one of the lagged regressors in the models. The points correspond to posterior means and the bars to 95% credible intervals across specifications. Take first the index of Policy Priority, where the posterior mean from model averaging (far right) is positive but with considerable associated uncertainty. This can be compared with the false negative identified in a baseline model (far left) and the six posterior means from the distributed lag model (middle). Whereas the researcher estimating distributed lag models without theory as to the lag length often engages in post-hoc justifications, there is no similar necessity under the BMA approach. Rather, the average posterior mean for the regressor of interest captures the effect as it relates to that variable without attempting to identify a particular lag length for which the researcher *can only* offer post-hoc justification.

One potential concern of researchers considering using the BMA approach as compared to distributed lag models could be the loss of information as it relates to the length of the lagged effect, particularly where that may be of substantive interest to the scholar. Yet in those instances, BMA offers a further advantage. Rather than relying on estimates derived from models that are likely afflicted with the particular issues identified above, we can examine posterior model probabilities as derived in equation (5), the precise entity with which some prior researchers have been concerned in proposing BMA (e.g., Montgomery and Nyhan 2010). To that end, we plot the posterior model probabilities for each of the six estimated models in Figure 1. Note the high posterior model probability for the model featuring four year lags of policy priority and legislative attention, consistent with prior work on the litigant signal model.

We turn now to demonstrating the robustness of the results across these alternatives, starting with the choice of the maximum lag length for the feasible set. Could the choice of the maximum lag length influence posterior model probabilities such that it biases any

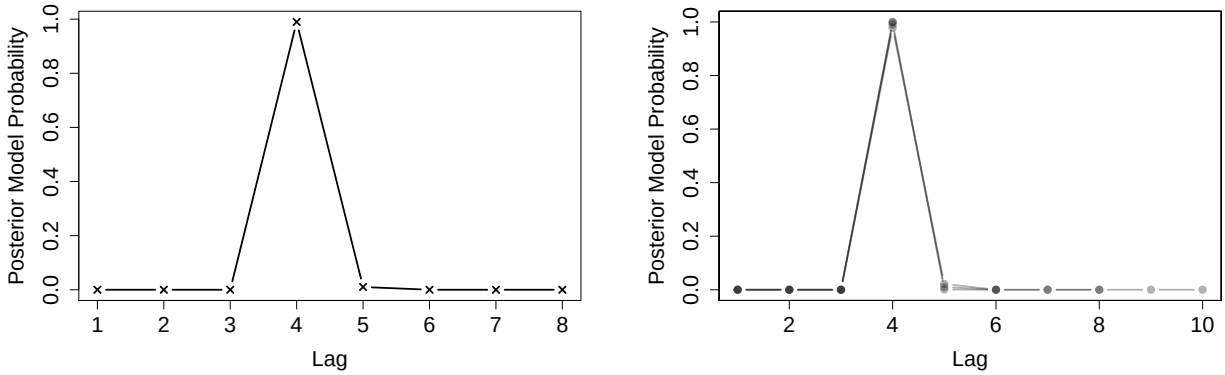

Figure 1: *Posterior Model Probabilities from Bayesian Model Averaging of Litigant Signal Model.* The left panel shows posterior model probabilities for the eight-lag model. The right panel shows posterior model probabilities across different choices of the maximum number of lags. The probabilities line up so closely, that differences can only barely be distinguished at lags 4 and 5.

inference one may draw from the structure of the relationship? To test this, we estimate posterior model probabilities for maximum lags of four, six, eight and ten periods and plot the results in the right panel of Figure 1. The lines are transparent gray, thus darker shades of gray indicate areas where the lines overlap. Such an approach is necessary as the lines are almost entirely overplotted; a function of the fact that there is little to no differentiation in posterior model probabilities across different selections of the maximum lag length.

Similarly, one could be concerned that the choice of the maximum lag in the feasible set influences the Bayesian model average of the posterior means for the regressors. To examine this, in Figure 2 we plot the magnitude of the average posterior means of the regressor of interest, the index of Policy Priority, across choices of the maximum lag in the feasible set. Note that across choices of maximum lag lengths, posterior means are consistent with 95% credible intervals that do not include zero.

Finally, we examine the sensitivity of the results to the choice of prior mean and prior precision. In the left panel of Figure 3, we plot posterior model probabilities for each lag length across the choice of the prior mean.[8] Note that across specifications, the four-lag model retains the highest model probability. Similarly, in the right panel of Figure 3, we plot the posterior model probabilities, but across varying levels of prior precision. Of particular note is that, once again, across all possibilities considered, the posterior model probabilities consistently identify a four-lag model as best.

Thus, the results of Bayesian model averaging are broadly consistent with a distributed lag approach while avoiding the concomitant complications of distributed lag models. Moreover, in the under-theorized environment of atheoretic lags, Bayesian model averaging offers an approach that better reflects the uncertainty of the causal dynamics and, by virtue of

[8]In each of these iterations, we set prior precision to 0.01. Subsequent analyses revealed that the substantive results persist across alternative choices to prior precision <5, at which point "large" choices of the prior (i.e.,>5) in concert with increased precision impugn the results. In the context of atheoretic lags, however, choosing a strong prior is, at best, theoretically questionable.

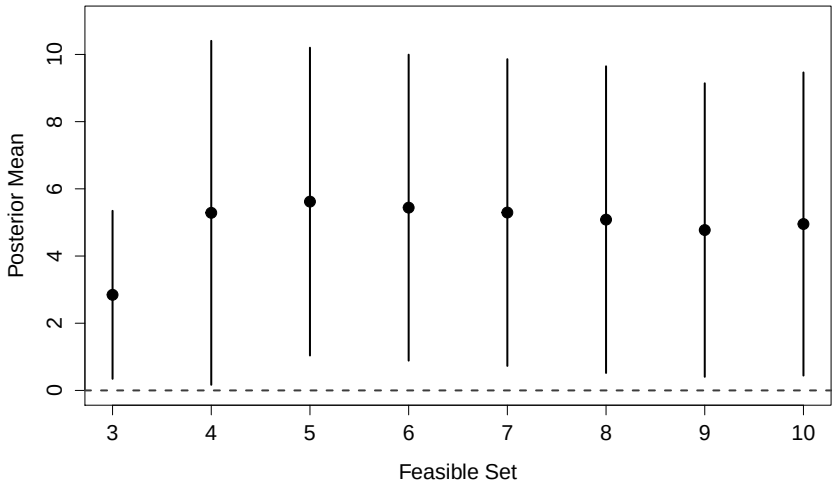

Figure 2: *Bayesian Model Averaged Posterior Means for Index of Policy Priority Across Different Specifications of Maximum Lag Length.* The plot shows the averaged posterior mean and 95% credible intervals for the index of Policy Priority across choices of the maximum lag in the feasible set.

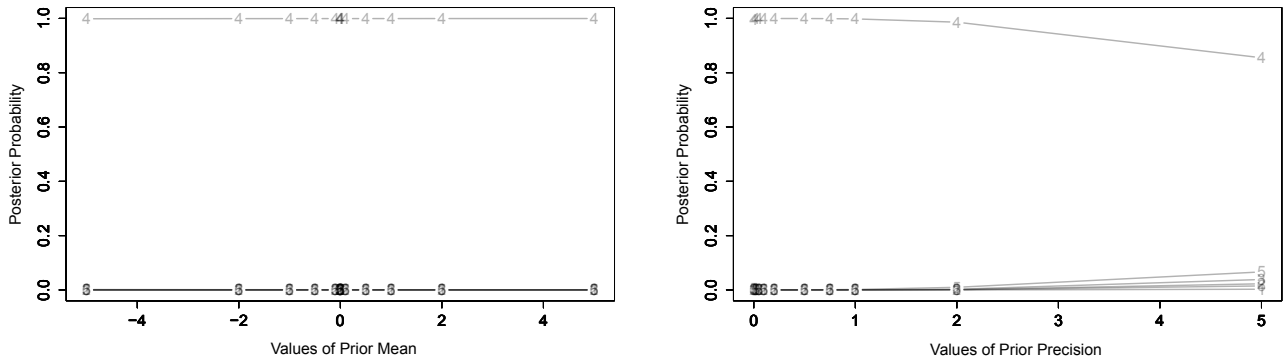

Figure 3: *Posterior Model Probabilities Across Choices of Prior and Prior Precision.* The left panel shows posterior model probabilities by lag for choices of the prior. The right panel shows posterior model probabilities by lag for choices of prior precision.

averaging across a multitude of specifications, is less susceptible to the type of "p-hacking" which recent research has identified as carrying broad implications for scientific studies (Gerber et al. 2010; Simonsohn, Nelson and Simmons 2013). We now move to a second application in order to demonstrate the broader utility of Bayesian model averaging to atheoretic lags in seemingly intractable research settings.

6. APPLICATION II: MODERNIZATION THEORY

In the field of political economy, and at the intersection of comparative politics and international relations, modernization theory postulates a relationship between democracy and economic performance. While modernization theory dates back to seminal works on economic modernization (Lerner 1958; Rostow 1960; Kuznets 1966; Chenery and Taylor 1968), the link between democratization and economic growth was first reported by Lipset (1959). Lipset demonstrated a correlation between democracy and economic performance, and hypothesized that citizens become less tolerant of authoritarian regimes as they become more prosperous. Lipset's results spurred the establishment of a substantial literature on the relationship between prosperity and democracy.

The literature on democratization is particularly robust in that it has examined the effects of economic growth using a wide variety of sampling frames. Large N statistical studies are common in this literature and yield consistent support for the importance of per capita GDP on the establishment and maintenance of democracy (Cutright 1963; Deutsch 1961;

Dahl 1971; Burkhart and Lewis-Beck 1994; Londregan and Poole 1996). Further, several well executed small N studies also find support for the effect of income on democratization (most notably Moore (1966) and Rueschemeyer, Stephens and Stephens (1992)).

An important contemporary contribution to this literature is a series of works by Adam Przeworski and his collaborators (Prezeworski 1990; Przeworski and Limongi 1993; Przeworski et al. 1996; Przeworski and Limongi 1997) that culminated in the result that democracies seem to emerge at random, but that democratic endurance is correlated to income (Przeworski et al. 2000). Since the Przeworski et al. (2000) result, the literature on democracy and income has progressed more slowly than before. Notable challenges to Przeworski et al. have come from Boix (2003), Boix and Stokes (2003), and Epstein et al. (2006). Epstein et al. (2006) challenge Przeworski et al. (2000) by using a trichotomous rather than dichotomous measure of democracy, thus capturing the difference between non-democracies, partial democracies, and full democracies. This slightly sharper coding scheme is perhaps more appropriate because the distribution of the Polity measure, on which the democracy codes are based, is severely bimodal. With their new outcome variable, Epstein et al. (2006) re-establish the effect of income on democracy.

Because analyses of the relationship between income and democracy are generally conducted using panel data, it is possible that the reported results are either incomplete or nonrobust because the atheoretic lag problems in these analyses have not been sufficiently addressed. Our goal is to re-examine the income-democracy relationship with special attention to the atheoretic lag problems. We do so in the context of slightly different data than such analyses have typically used. We look at the income-democracy relationship in the context of internal antagonisms that could inhibit the formation and perpetuation of democracy.

The data for this analysis are drawn from Fearon and Latin (2003) who have gathered a country-year dataset covering the period 1945 to 1999. The data include a number of variables expected to predict internal political stability and allow the post-war democracy-income relationship to be examined while controlling for precipitators of domestic instability.

We specify a theoretical model as follows:

$$\begin{aligned} DEMOCRACY = {} & INCOME + POPULATION + ETHNIC\ FRACT. \\ & + RELIGIOUS\ FRACT. + POLITICAL\ INSTABILITY \\ & + OIL + NON-CONTIGUITY. \end{aligned} \tag{10}$$

We use a dichotomous coding of democracy in which countries with Polity 2 scores lower than 7 are coded as non-democracies, and countries with Polity 2 scores greater than or equal to 7 are coded as democracies. The regressor of interest is income, measured as per capita GDP in the country year. As controls for other sources of domestic stability/instability, we include measures for population, ethnic fractionalization, religious fractionalization, political instability, and indicators for whether the country produces oil and is non-contiguous. We use Fearon and Latin's (2003) coding and measures for all controls.

	Mean	Std. Deviation	2.5%	97.5%
Lagged Democracy[LOV]	4.023	0.082	3.863	4.184
Lag(Income)	0.035	0.008	0.020	0.050
Lag(Population)	0.042	0.029	-0.014	0.099
Lag(Eth. Frac.)	-0.321	0.153	-0.626	-0.027
Lag(Rel. Frac.)	-0.267	0.201	-0.665	0.126
Lag(Instability)	0.079	0.101	-0.123	0.276
Oil	-0.312	0.129	-0.557	-0.064
Non-contiguous	0.341	0.110	0.122	0.555

N = 6449.

Table 4: *The baseline modernization theory model.* The specification lags the relevant regressors at the "default" value of one time period. The analysis includes a lagged outcome variable and year fixed effects (not reported).

We implement the above specification as a Bayesian probit model of democracy (dichotomous scale) and the predictors described above, as well as a lagged outcome variable. Consider the results of the basic model, with the "default" one period lags shown in Table 4. We see a small but significant effect for the one period lag of income, significant negative effects for ethnic fractionalization, significant negative effects from oil exports, and evidence that non-contiguous states has a strong, significant, and positive effect on democracy. Each of these is consistent with our theory. However, the 95% credible intervals for the remainder of the variables contain zero, with log population, religious fractionalization, and political instability all failing to attain results from which we can confidently draw any conclusions. This baseline model will serve as the jumping off point for an exploration of this model's atheoretic lag problem.

6.1. *Distributed Lag Model.*

While the litigant signal model worked reasonably well as a distributed lag model, the modernization model does not. Here, beginning with a broad, inclusive model and testing restrictions – as (DeBoef and Keele 2008) suggest – is impossible. Such a distributed lag model breaks down for the modernization application because of a larger feasible set, the need to estimate many parameters, and high autocorrelation in the variables for which we specify atheoretic lags.

We begin with consideration of the feasible set. The process of democratization is often a slow one. States are slow to change their governments and slow to react to the will of their people; this is particularly true with autocratic regimes (e.g., Cranmer and Siverson 2008). As we consider the regressors for which we have atheoretic lags (i.e., income, the log of population, ethnic fractionalization, religious fractionalization, and instability) we would not expect any of these variables to have particularly early effects on the democratic nature of a government. More likely, the effects of changes in these variables will take years to be observed. As such, we require a longer set of feasible lags than we had for the litigant signal model. In this case, we will specify a feasible set of lags ranging from one year to ten years. While this feasible set may seem a bit short, particularly for variables like population, if the regressors are indeed affecting the outcome, it seems reasonable to think that their effects would be apparent within the decade. That said, if our purpose was analytic rather than expository, a longer set of lags might well be used.

When we attempt to specify a distributed lag model based on the specification above and the atheoretic lags with ten-year feasible sets, we are quickly confronted with a problem. All but one of the parameters for the lags of religious fractionalization are not estimable. An examination of the autocorrelations for the variables with atheoretic lags in Figure 4 reveals the cause: including several lags of religious fractionalization results in such high collinearity in the model, that parameters are inestimable. Indeed, religious fractionalization is not the only variable about which we should be concerned.

The situation is not improved greatly if we are willing to drop religious fractionalization from our model; something an analysis for purposes other than exposition should not consider because the variable belongs in the model specification by theory. Nonetheless, if we are willing, for the purposes of exposition, to drop this variable entirely, we still have problems with estimation: high autocorrelations in ethnic fractionalization result in an inestimable parameter for its tenth lag. Furthermore, we know that high collinearity will still be present

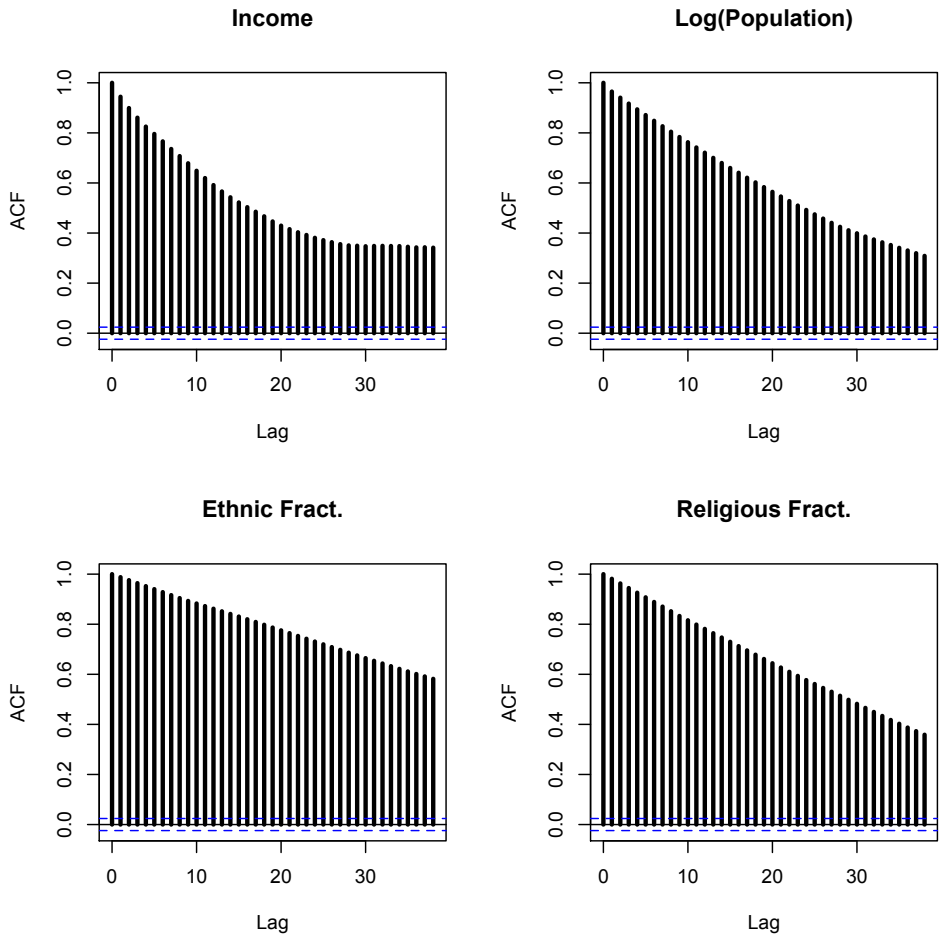

Figure 4: Autocorrelation functions for four regressors in the modernization application for which atheoretic lags must be specified. Such high autocorrelations pose a problem for the estimation of distributed lag models.

in the model because of the high autocorrelations of the remaining variables; even if the collinearity is not so high that the software is incapable of estimating parameters, such collinearity will bias uncertainty measures. The result of this poorly fitting specification can be seen in Table 5. Note that, according to this specification, income lagged one year is *negatively* related to democracy, while income lagged two years is *positively* related to democracy, both with credible intervals that do not contain zero. The same "switching" dynamic also appears with political instability and offers stark evidence of the ill effects of highly collinear specifications. Moreover, the posterior means and credible intervals of the non-continuous indicator no longer provides any evidence of a relationship with democracy. Finally, notice that the trends in the posterior means for the variables with distributed lags

generally do not conform to either polynomial or exponential assumptions about their form, assumptions required for the distributed lag model.

	Estimate	Std. Devation	2.5%	97.5%
Lagged Democracy [LOV]	4.269	0.108	4.062	4.481
Oil	-0.321	0.159	-0.639	-0.017
Non-contiguous	0.145	0.140	-0.139	0.421
N = 5,032				

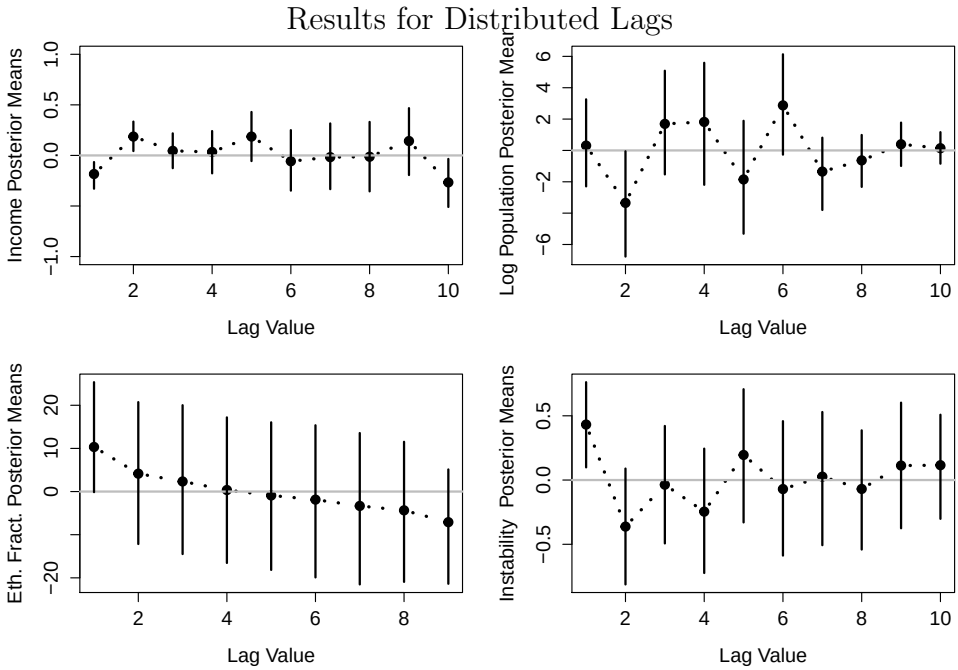

Table 5: *Distributed Lag Probit Results for Modernization Model.* The outcome variable is democracy. Results for the regressors *not* specified as distributed lags are in the upper table, results for the distributed lags are presented graphically in the lower cell. The distributed lags are specified as $\mathbf{k} = \{1, 2, \ldots, 10\}$ for income, log of population, ethnic fractionalization, and instability. Note that a tenth lag for ethnic fractionalization could not be estimated due to collinearity.

In summary, following the standard advice of DeBoef and Keele (2008) is unworkable. We cannot estimate a distributed lag model with all the variables that theoretically ought to be in our model, a problem which is compounded by the fact that ten years is a short feasible set for this model but a longer feasible set is computationally impossible. One consequence of these problems are results that suggest peculiar dynamics. In all, the general distributed lag specification in the context of the modernization application is troublesome and illustrative of the limitations of this class of model.

6.2. *Modernization Theory and BMA.*

The advantages of BMA in the context of atheoretic lag problems are seen even more clearly here. Recall that we have five regressors with atheoretic lags: income, instability, log of population, ethic fractionalization, and religious fractionalization. We also have a process for which we expect to have a longer feasible set of lags than was the case in the litigant signal model. In the above approach, the high degrees of collinearity prevented us from successfully modeling these data as distributed lags. Now, we specify a series of models that lags each of the athoeretic lag variables by up to fifteen periods (more than was possible for the distributed lag model, even omitting the religious fractionalization variable), and compute posterior means of the lagged coefficients through model averaging. The results are shown in Table 6.

	Mean	Std. Deviation	2.5%	97.5%
Lagged Democracy[LOV]	4.146	0.120	4.065	4.227
Lag(Income)	0.063	0.021	0.049	0.077
Lag(Population)	0.061	0.037	0.037	0.086
Lag(Eth. Frac.)	-0.229	0.210	-0.370	-0.088
Lag(Rel. Frac.)	-0.628	0.265	-0.807	-0.449
Lag(Instability)	-0.191	0.292	-0.388	0.006
Oil	-0.385	0.166	-0.497	-0.273
Non-contiguous	0.265	0.158	0.159	0.371
N = 6449; feasible set = 15.				

Table 6: *Results of Bayesian model averaging across feasible set.* The outcome variable in each of the analyses is democracy. For each of $\mathbf{k} = \{1, 2, \ldots, 15\}$, the model is estimated, then we compute the Bayesian model averages for the coefficients from each of the 15 models.

Whereas the posterior means of the distributed lag were counter to expectations, the posterior means from BMA offer results that fit with the literature. First, results provide evidence that both non-lagged controls are significantly related to democracy. Second, in the case of the regressor of interest, we see that lagged income has a positive relationship with democracy, consistent with modernization theory. Third, the average posterior means and credible intervals for the other lagged variables provide evidence suggestive of theoretically intuitive patterns with credible intervals that do not include zero for all but political instability. Finally, notice also that religious fractionalization, which could not even be included in the distributed lag specification, is now demonstrated to be negatively related with

democracy.

Figure 5 provides a comparison of the results from the several model approaches which further reveals the advantage of Bayesian model averaging for atheoretic lag problems. We see that primary effect of interest, income, has a positive and significant effect. This is in keeping with the positive results from the baseline model but offers a clearer picture of the effect than the inconsistent "switching" result of the distributed lag model. Lagged values of log population have a significant and positive effect across models, a dynamic not identified at the first lag, and opposite to the significant and *negative* result seen at the first lag of the distributed lag model. Ethnic fractionalization and instability are never statistically significant at traditional thresholds, while a distributed lag model suggested that instability at one year lags positively predicted predicts democracy. Finally, some may be concerned that the structure of the relationship at it relates to time is lost when using BMA. This, however, is not the case. The bottom panel of Figure 5 shows the posterior model probabilities (i.e. those from eq. (5)) across the feasible set. We see that model probabilities peak at a four-year lag, and drop of quickly thereafter.

We can see clearly in this application that Bayesian model averaging works particularly well for model specifications with atheoretic lags on many regressors and relatively long feasible lag sets. A researcher could gain further purchase on effects that differ over time (as opposed to *just* over lags) by creating subsets of the data for particular time periods of interest (i.e. during and after the Cold War) and re-running the multiple fit comparison analyses. See Cranmer and Siverson (2008) for an example involving conflict pre- and post-WWII.

7. CONCLUDING THOUGHTS

Social scientists frequently encounter research questions that feature temporal dynamics. To account for these dynamics, researchers include temporally lagged regressors in model specifications. While in some cases the choice of the temporal lag is clear, in many instances there is no definitive *a priori* reason to suspect a particular lag or set of lags. We term these situations *atheoretic lags*, and have presented a set of techniques based on Bayesian model averaging to address the unique methodological concerns such situations invoke. These techniques, while based on the simple concept of averaging over repeated estimations across lags, appropriately account for model uncertainty in ways that common extant approaches do not. Moreover, our techniques overcome two prominent problems with one particularly

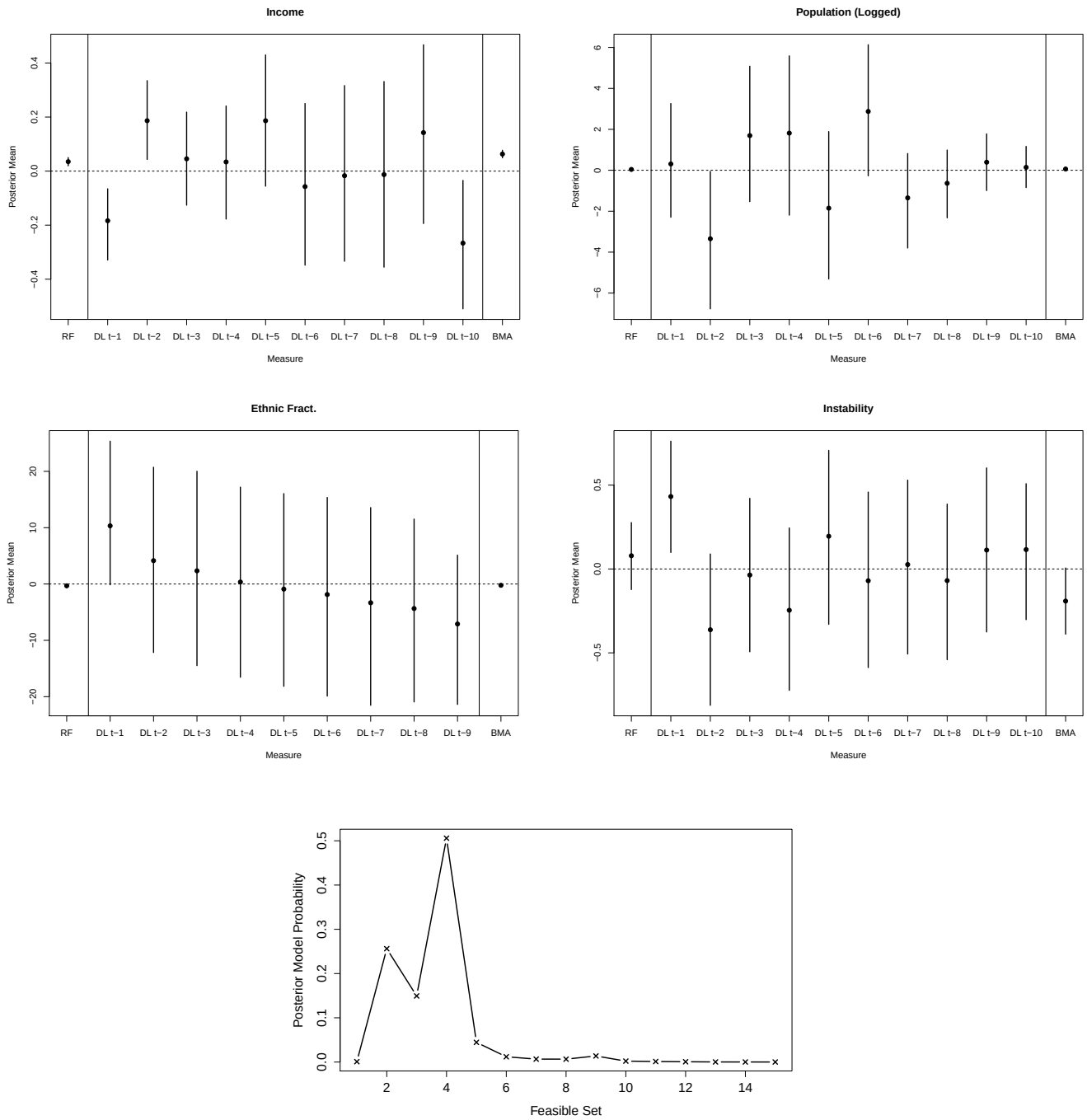

Figure 5: *Comparison of Model Results for Modernization Application.* The outcome variable in each of the analyses is democracy. Each point represents the posterior mean of the coefficient, while the vertical bars represent 95% credible intervals. The lower panel presents posterior model probabilities across models estimated with each variable lagged $\mathbf{k}$ times, $\mathbf{k} = \{1, 2, \ldots, 10\}$ for the distributed lag application, and $\mathbf{k} = \{1, 2, \ldots, 15\}$ for the Bayesian model averaging application.

common approach, distributed lag models: an excessively large number of regressors in a single model specification, and the introduction of high degrees of collinearity among the regressors.

We have shown, in the context of two examples in the discipline, the utility of the Bayesian model averaging approach across distributed lag specifications with atheoretic lags. Where the distributed lag specification works fairly well, as in the litigant signal model with only two variables with atheoretic lags and a feasible set of only six years, Bayesian model averaging recovers substantively identical results while mitigating concerns over the selective presentation of model specifications or p-hacking. Where the distributed lag approach breaks down, as in the modernization example with five atheoretic lag variables and a feasible set of fifteen years, the Bayesian model averaging approach recovers estimates that accurately reflect the state of the world while avoiding the arbitrary selection of lag lengths or the omission of relevant variables. Therefore, we provide evidence that Bayesian model averaging offers an avenue by which scholars facing model specifications with atheoretic lags may reflect their underlying model uncertainty in a statistically rigorous way.

This simple solution comes with little cost, as the results are more easily presented than in distributed lag specifications, and the structure of the relationship across time can be recovered by examining posterior model probabilities. The relatively minor additional effort — estimating additional models — associated with this approach is one all too frequently incurred by researchers anyhow, though they then present only one specification. In sum, BMA offers researchers the means to gain and communicate substantive insight beyond what is possible with distributed lag models or the arbitrary choice of a particular lag length.

REFERENCES

Akaike, Hirotugu. 1974. "A new look at the statistical model identification." *IEEE Transactions on Automatic Control* 19(6):716–723.

Almon, S. 1965. "The Distributed Lag Between Capital Appropriations and Expenditures." *Econometrica* 33(1):178–196.

Amemiya, Takeshi and Wayne Fuller. 1967. "A Comparative Study of Alternative Estimators in a Distributed Lag Model." *Econometrica* 35(3/4):509–529.

Baird, Vanessa. 2004. "The Effect of Politically Salient Decisions on the U.S. Supreme Court's Agenda." *The Journal of Politics* 66(3):755–722.

Baird, Vanessa. 2007. *Answering the Call of the Court: How Justices and Litigants Set the Supreme Court Agenda*. Charlottesville: University of Virginia Press.

Baird, Vanessa and Tonja Jacobi. 2009. "Judicial Agenda-Setting through Signaling and Strategic Litigant Responses." *Washington University Journal of Law and Policy* 29:215–239.

Barnard, George A. 1963. "New Methods of Quality Control." *Journal of the Royal Statistical Society, Ser. A* 126(2):255–258.

Bartels, Larry M. 1997. "Specification Uncertainty and Model Averaging." *American Journal of Political Science* 41(2):641–674.

Bartels, Larry M. and John Zaller. 2001. "Presidential Vote Models: A Recount." *PS: Political Science and Politics* 34(1):8–20.

Beck, Nathaniel. 1985. "Estimating Dynamic Models Is Not Merely a Matter of Technique." *Political Methodology* 11:71–89.

Beck, Nathaniel. 1991. "Comparing Dynamic SSpecification: The Case Of Presidential Approval." *Political Analysis* 3:27–50.

Beck, Neal and Jonathan N. Katz. 1996. "Nuisance versus substance: specificying and estimating time-series-cross-section models." *Political Analysis* 15:182–195.

Black, Ryan C. and Ryan J. Owens. 2009. "Agenda Setting in the Supreme Court: The Collision of Policy and Jurisprudence." *The Journal of Politics* 71(3):1062–1075.

Boix, Carles. 2003. *Democracy and Redistribution*. New York: Cambridge University Press.

Boix, Carles and Susan Stokes. 2003. "Endogenous Democratization." *World Politics* 55(4):517–549.

Brenner, Saul and John F. Krol. 1989. "Strategies in Certiorari Voting on the United States Supreme Court." *The Journal of Politics* 51(4):828–840.

Burkhart, Ross E. and Michael S. Lewis-Beck. 1994. "The Economic Development Thesis." *American Political Science Review* 88:903–910.

Caldeira, Gregory A. 1981. "The United States Supreme Court and Criminal Cases, 1935-1976: Alternative Models of Agenda Building." *British Journal of Political Science* 11(4):449–470.

Caldeira, Gregory A. and John R. Wright. 1988. "Organized Interests and Agenda Setting in the U.S. Supreme Court." *American Political Science Review* 82(4):1109–1127.

Caldeira, Gregory A., John R. Wright and Christopher J.W. Zorn. 1999. "Strategic Voting and Gatekeeping in the Supreme Court." *Journal of Law, Economics and Organization* 15(3):549–572.

Chenery, Hollis B. and Lance Taylor. 1968. "Development Patters: among Countries and Over Time." *Review of Economic Statistics* 50(4):391–416.

Clemen, Robert T. 1989. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting* 5(4):559–583.

Cranmer, Skyler J. and Randolph M. Siverson. 2008. "Demography, Democracy and Disputes: The Search for the Elusive Relationship Between Population Growth and International Conflict." *Journal of Politics* 70(3):794–806.

Cutright, Phillips. 1963. National Political Development: Social and Economic Correlates. In *Politics and Social Life: An Introduction to Political Behavior*, ed. Nelson W. Polsby, Robert A. Dentler and Paul A. Smith. Boston: Houghton Mifflin pp. 569–581.

Dahl, Robert A. 1971. *Polyarch: Participation and Opposition*. New Haven: Yale University Press.

DeBoef, Suzanna and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science* 52(1):184–200.

Deutsch, Karl W. 1961. "Social Mobilization and Poltical Development." *American Political Science Review* 55(3):493–510.

Draper, David. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society, Series B* 57(1):45–97.

Dufour, Jean-Marie and Jan F. Kiviet. 1998. "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models." *Econometrica* 66(1):79–104.

Epp, Charles. 1998. *The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective*. Chicago: University of Chicago.

Epstein, David L., Robert Bates, Jack Goldstone, Ida Kristensen and Sharyn O'Halloran. 2006. "Democratic Transitions." *American Journal of Political Science* 50(3):551–569.

Epstein, Lee and C.K. Rowland. 1991. "Debunking the Myth of Interest Group Invincibility in the Courts." *American Political Science Review* 85(1):205 – 217.

Epstein, Lee and Jeffrey A. Segal. 2000. "Measuring Issue Salience." *American Journal of Political Science* 44(1):66–83.

Epstein, Lee and Joseph F. Kobylka. 1992. *The Supreme Court and Legal Change: Abortion and the Death Penalty*. Chapel Hill: University of North Carolina Press.

Erikson, Robert S., Joseph Bafumi and Bret Wilson. 2001. "Was the 2000 Presidential Election Predictable." *PS: Political Science and Politics* 34(4):815–819.

Fearon, James D. and David D. Latin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1):75–90.

Flemming, Roy, B. Dan Wood and John Bohte. 1999. "Attention to Issues in a System of Separated Powers: The Macrodynamics of American Policy Agendas." *Journal of Politics* 61(1):76–108.

Freund, Yoav and Robert Schapire. 1996. Experiments with a new boosting algorithm. In *Machine Learning: Proceedings of the Thirteenth International Conference*. pp. 148–156.

Freund, Yoav and Robert Schapire. 1997. "A decision-theoretic generalization of on-line learning and an application to boosting." *Journal of Computer and System Sciences* 55(1):119–139.

Geer, John and Richard R. Lau. 2006. "Filling in the Blanks: A New Method for Estimating Campaign Effects." *British Journal of Political Science* 36(2):269–290.

Gerber, Alan, Neil Malhotra, Conor Dowling and David Doherty. 2010. "Publication Bias in Two Political Behavior Literatures." *American Politics Research* 38(4):591–613.

Geweke, John. 1978. "Temporal Aggregation in the Multiple Regression Model." *Econometrica* 46(3):643–661.

Ghobarah, Hazem A., Paul Huth and Bruce Russett. 2003. "Civil Wars Kill and Maim People –Long after the Shooting Stops." *American Political Science Review* 97(2):189–202.

Gill, Jeff. 2007. *Bayesian Methods*. Boca Raton: Chapman & Hall/CRC.

Green, Donald, Soo Yeon Kim and David Yoon. 2001. "Dirty Pool." *International Organization* 55(2):441–468.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery and Chris T. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14(4):382–417.

Imai, Kosuke and Gary King. 2004. "Did Illegal Overseas Absentee Ballots Decide the 2000 U.S. Presidential Election?" *Perspectives on Politics* 2(3):537–549.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. New York: Wiley.

Koyck, Leendert M. 1954. *Distributed Lags and Investment Analysis*. Amsterdam: North-Holland.

Kuznets, Simon. 1966. *Modern Economic Growth: Rate, Structure, and Spread*. New Haven: Yale University Press.

Leamer, Edward E. 1978. *Specification Searches*. New York: Wiley.

Lerner, Daniel. 1958. *The Passing of Traditional Society*. New York: Free Press of Glencoe.

Lipset, Seymour M. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Reivew* 53(1):69–105.

Londregan, John B. and Keith T. Poole. 1996. "Does High Income Promote Democracy." *World Politics* 49(1):1–30.

Lütkepohl, Helmut. 1980. "Approximation of Arbitrary Distributed Lag Structures by a Modified Polynomial Lag: An Extension." *Journal of the American Statistical Association* 75(370):428–430.

Madigan, David and Adrian E. Raftery. 1994. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." *Journal of the American Statistical Sssociation* 89(428):1535–1546.

Martin, Andrew D. and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court." *Political Analysis* 10(2):134–153.

Martin, James E. 1967. "Isolation of Lagged Economic Responses." *Journal of Farm Economics* 49(1):160–168.

McGuire, Kevin T. and Gregory A. Caldeira. 1993. "Lawyers, Organized Interests, and the Law of Obscenity: Agenda-Setting in the Supreme Court." *American Political Science Review* 87(3):715–726.

Montgomery, Jacob M. and Brendan Nyhan. 2010. "Bayesian Model Averaging: Theoretical Developments and Practical Applications." *Political Analysis* 18(2):245–270.

Moore, Barrington. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Boston: Beacon Press.

Nelson, Charles R. and G. William Schwert. 1974. "Estimating the Parameters of a Distributed Lag Model from Cross-Section Data: The Case of Hospital Admissions and Discharges." *Journal of the American Statistical Association* 69(347):627–633.

Nicholson, Chris and Paul Collins. 2008. "The Solicitor General's Amicus Curiae Strategies in the Supreme Court." *American Politics Research* 36:382–415.

Perry, H.W. 1991. *Deciding to Decide: Agenda Setting in the United States Supreme Court*. Cambridge: Harvard University Press.

Peters, C. Scott. 2007. "Getting Attention: The Effect of Legal Mobilization on the U.S. Supreme Court's Attention to Issues." *Political Research Quarterly* 60(3):561–572.

Prezeworski, Adam. 1990. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. New York: Cambridge University Press.

Przeworski, Adam and Fernando Limongi. 1993. "Political Regimes and Economic Growth." *Journal of Economic Perspectives* 7(3):51–69.

Przeworski, Adam and Fernando Limongi. 1997. "Modernization: Theories and Facts." *World Politics* 49(2):155–183.

Przeworski, Adam, Michael E. Alvarez, Jose A. Cheibub and Fernando Limongi. 1996. "What Makes Democracies Endure?" *Journal of Democracy* 7(1):39–55.

Przeworski, Adam, Michael E. Alvarez, Jose A. Cheibub and Fernando Limongi. 2000. *Democracy and Development: : Political Institutions and Well-Being in the World, 1950-1990*. New York: Cambridge University Press.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111–163.

Rostow, Walt Whitman. 1960. *The Stages of Economic Growth: A Non-Communist Manifesto*. Cambridge: Cambridge University Press.

Rueschemeyer, Dietrich, Evelyne H. Stephens and John D. Stephens. 1992. *Capitalist Development and Democracy*. Chicago: University of Chicago Press.

Simonsohn, Uri, Leif Nelson and Joseph Simmons. 2013. "P-Curve: A Key to the File Drawer." *Journal of Experimental Psychology* Forthcoming.

Solow, Robert M. 1960. "On a Family of Lag Distributions." *Econometrica* 28(2):393–406.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society, Series B* 58:267–288.

Tibshirani, Robert. 1997. "The Lasso Method for Variable Selection in the Cox Model." *Statistics in Medicine* 16:385–395.

Trivedi, Pravin K. 1985. "Distributed Lags, Aggregation and Compounding: Some Econometric Implications." *The Review of Economic Studies* 52(1):19–35.

Ulmer, S. Sidney. 1984. "The Supreme Court's Certiorari Decisions: Conflict as a Predictive Variable." *American Political Science Review* 78(4):901–911.

Valentino, Nicholas, Ted Brader and Ashley Jardina. 2013. "Immigration Opposition Among U.S. Whites: General Ethnocentrism or Media Primining of Attitudes About Latinos?" *Political Psychology* 34(2):149–166.

Warren, T. Camber. 2014. "Not by the Sword Alone: Soft Power, Mass Media, and the Production of State Sovereignty." *International Organization* 68(1):111–141.

Zaller, John R. 2004. Floating voters in U.S. presidential elections, 1948-2000. In *Studies in public opinion: Attitudes, nonattitudes, measurement error, and change*, ed. Willem Saris and Paul M. Sniderman. Princeton: Princeton University Press.