

Adaptive Variable Selection for Sequential Prediction in Multivariate Dynamic Models

Isaac Lavine^{*}, Michael Lindon[†], and Mike West[‡]

Abstract. We discuss Bayesian model uncertainty analysis and forecasting in sequential dynamic modeling of multivariate time series. The perspective is that of a decision-maker with a specific forecasting objective that guides thinking about relevant models. Based on formal Bayesian decision-theoretic reasoning, we develop a time-adaptive approach to exploring, weighting, combining and selecting models that differ in terms of predictive variables included. The adaptivity allows for changes in the sets of favored models over time, and is guided by the specific forecasting goals. A synthetic example illustrates how decision-guided variable selection differs from traditional Bayesian model uncertainty analysis and standard model averaging. An applied study in one motivating application of long-term macroeconomic forecasting highlights the utility of the new approach in terms of improving predictions as well as its ability to identify and interpret different sets of relevant models over time with respect to specific, defined forecasting goals.

Keywords: Bayesian forecasting, decision analysis, dynamic dependency network models, dynamic linear models, Gibbs model probabilities, macroeconomic forecasting, model averaging, model structure uncertainty, shotgun stochastic search.

1 Introduction

Model structure uncertainty lies at the heart of much of scientific modeling but remains a central challenge to statistical methodology. Specific problems of variable selection and model weighting or averaging are central in Bayesian analysis and have seen enormous development to date. However, more recent literature has increasingly emphasized the need for broader Bayesian views of model structure uncertainty. In particular, the proscribed nature of standard Bayesian model uncertainty and issues faced in realistic \mathcal{M} -open settings have led to growing recent interest in new Bayesian approaches to scoring, weighting and combining models. We are concerned with these general questions in contexts of sequential analysis for forecasting and decision-making using dynamic state-space models for time series. Here the issues of model uncertainty are exacerbated by the potential for relevant, data-respecting models to change in structure over time, as well as for parameters within a model to be time-varying. This, together with broader challenges, has been highlighted across a class of dynamic model contexts in West (2020). The current paper picks-up the theme and addresses the challenges with a

^{*}SIG, Philadelphia PA 19004, U.S.A., lavine.isaac@gmail.com

[†]Optimizely, San Francisco CA 94105, U.S.A., michael.s.lindon@gmail.com

[‡]Duke University, Durham NC 27708, U.S.A., mike.west@duke.edu

novel Bayesian approach to weighting and, over time, adaptively reweighting structures from large model classes, guided by specific forecasting and/or decision goals.

Critical limitations of model probabilities are significantly highlighted in the sequential time series setting. First, model marginal likelihoods score only 1-step ahead forecasting accuracy. The marginal likelihood value on a model from n observations is the product of realized values of 1-step forecast densities. This clearly demarks the applied relevance of this score. Models are built with forecasting and decision goals, and 1-step forecast accuracy is rarely the only motivating goal. A model scoring highly in that sense may be hopeless for multi-step ahead forecasting, or define poor forecasts for resulting decisions. More broadly, the need to consider explicit goals in model structure assessment has been recognized at least implicitly in recent literature on model weighting and combination (e.g. Clyde and Iversen, 2013; Amisano and Geweke, 2017; McAlinn et al., 2018; McAlinn and West, 2019; McAlinn et al., 2020; Yao et al., 2018), and explicitly in some areas related to multi-step forecasting and decisions when comparing and combining models (e.g. Nakajima and West, 2013a; Kapetanios et al., 2015; West, 2020). A model that forecasts well on one subset of outcomes in a multivariate setting may be poor in other dimensions. A model scoring highly on one purely statistical metric may be inferior to other models in a decision problem or in terms of contextually relevant forecast accuracy measures (e.g. Berry and West, 2020; Berry et al., 2020). We argue for a more explicit, core focus on integrating forecasting and decision goals as arbiters of model assessment to advance practically relevant methodology.

A further concern is that practical interests in model uncertainty rarely include identifying “true models”; rather, model structure is often a nuisance parameter and not of inherent interest otherwise. In variable selection, identifying a model, or a few models, that are useful for prediction is typically the goal. The academic enterprise of treating increasingly large sets of models defined by many subsets of potential predictors quickly runs into the well-known– and intractable– problems of model multiplicities, redundancies and collinearities: many models with differing structures generate similar predictions, collinearities drive complications in interpretation, and model averaging induces increased noise in resulting predictions (e.g. Hans et al., 2007a,b; George, 2010; Giannone et al., 2018). Practically, interest often lies in “good choices” in terms of forecast and decision outcomes (e.g. Gruber and West, 2016, 2017; West, 2020). Then, the increasing dimension of model spaces argues against the traditional Bayesian view of maintaining interest in all possible models. In sequential analysis of time series this is particularly highlighted, as monitoring and updating scores on many models over time quickly raises the computational stakes. As many models will be of little or no interest, coupled with the common issue of huge redundancy of model classes, this argues for selective analysis of smaller numbers of models and a concern to– at selected points over time– review and refresh selected sets of models under consideration.

We address the above issues with a new Bayesian approach to adaptive (over time) model uncertainty analysis. The ideas are general while being presented in the motivating context of multivariate time series forecasting with specific forecast goals, and in which the model structure in question is the specification of sets of predictor variables in dynamic linear models for the multivariate series. The example context uses

flexible classes of dynamic dependency network models for a vector time series, and explores analysis in a topical macro-economic forecasting context. Section 2 defines the time series setting. Section 3 opens with explicit desiderata underlying the perspective on sequential analysis and forecasting in the context of predictor variable uncertainty, responding to the issues and challenges discussed above. This section then defines both the conceptual basis and technical/computational details of the novel adaptive variable selection strategy. A simulation study in Section 4 is followed by results from a macroeconomic case study in Section 5. The application focuses on the relevance of model structure uncertainty with respect to multi-step ahead and path forecasting, i.e., the specific and key goal in monetary policy-related contexts of forecasting trajectories of economic indicators over a path of time points into the future. Summary comments appear in Section 6, with supporting material in Appendices.

2 Time Series Setting and Perspectives

2.1 Multivariate Time Series: Notation and Models

The $m \times 1$ vector \mathbf{y}_t comprises a set of m univariate time series $y_{j,t}$ in equally-spaced time. The class of Dynamic Dependency Network Models (DDNMs) is a flexible framework for modeling and forecasting, and is increasingly exploited due to the ability to customize univariate series, sensitively model cross-series relationships and their dynamics over time, and to scale with m . DDNMs couple together sets of univariate dynamic linear models (DLMs) and exploit the well-known, analytic forward filtering and forecasting results of DLMs (e.g., chapt. 4 in each of West and Harrison, 1997; Prado and West, 2010). Full details can be seen in Zhao et al. (2016), with a recent, relevant example in Irie and West (2019). The cross-series structure of DDNMs is also intimately related to other popular multivariate models applied in economics, finance and related areas (e.g. Primiceri, 2005; Nakajima and West, 2013a,b; Zhou et al., 2014; Nakajima and West, 2015, 2017; Shirota et al., 2017; Lopes et al., 2018).

A DDNM is defined by a set of univariate dynamic models

$$y_{j,t} = \mathbf{F}'_{j,t} \boldsymbol{\theta}_{j,t} + \nu_{j,t}, \quad \mathbf{F}_{j,t} = \begin{pmatrix} \mathbf{x}_{j,t} \\ \mathbf{y}_{pa(j),t} \end{pmatrix}, \quad \boldsymbol{\theta}_{j,t} = \begin{pmatrix} \phi_{j,t} \\ \gamma_{j,t} \end{pmatrix}, \quad \nu_{j,t} \sim N(0, 1/\lambda_{j,t}), \quad (1)$$

where $j = 1, \dots, m$ indexes series and $t = 1, 2, \dots$ indexes time. In each series j , the state vector and volatility $(\boldsymbol{\theta}_{j,t}, \lambda_{j,t})$ evolve via a linear state equation coupled with a discount volatility model, assumedly independently across series. Observation errors $\nu_{j,t}$ are independent across j and over t . The regression vector $\mathbf{F}_{j,t}$ involves: (a) a subvector $\mathbf{x}_{j,t}$ of exogenous predictors and/or selected lagged values of some of the m series—giving opportunity for sparse and time-varying vector autoregressive components as well as external predictor variables; (b) a subvector $\mathbf{y}_{pa(j),t}$ of *parental predictors*—here $pa(j) \subseteq \{j+1, \dots, m\}$ is an index set selecting some of the contemporaneous values of other series ordered higher than j in the vector. The triangular structure of the parental sets defines the multivariate model of \mathbf{y}_t by a series of conditional relationships. The conformably partitioned state vector $\boldsymbol{\theta}_{j,t}$ includes subvectors of dynamic coefficients

$\phi_{j,t}$ on exogenous and lagged predictors, and $\gamma_{j,t}$ on parental predictors, viz $E(y_{j,t}|\ast) = \mathbf{F}'_{j,t}\boldsymbol{\theta}_{j,t} = \mathbf{x}'_{j,t}\phi_{j,t} + \mathbf{y}'_{pa(j),t}\gamma_{j,t}$ where \ast indicates all relevant terms.

The joint distribution can theoretically be decomposed into any series ordering; in practice the series order matters for interpretation and will impact on variable selection for each of the decoupled univariate models. In our economic time series examples below we follow prior authors in choosing a contextually relevant ordering (e.g. Nakajima and West, 2013b; Eickmeier et al., 2015). For broader commentary on the ordering in applications, we refer to the particularly germane discussion and reply to discussion in Zhao et al. (2016), as well developments in related models in Nakajima and West (2015) and Crespo Cuaresma et al. (2019). Common empirical experiences are that the impact of the ordering is typically very limited from the viewpoint of forecasting accuracy. Otherwise, with interests in justification a modeller is free to define her/his own ordering based on application-specific rationales, and then explore and evaluate multiple possible choices. While these are important applied considerations, they are not of main interest in connection with the primary contributions of this paper and we proceed with a chosen order.

At each time t , denote by \mathcal{D}_t the current information set. This includes initial information \mathcal{D}_0 , all data $\mathbf{y}_1, \dots, \mathbf{y}_t$ up to time t , and all other information used in the modeling process— including the $\mathbf{x}_{j,t}$, interventions or changes to model structure, etc. Implicitly in what follows, \mathcal{D}_t also includes information needed or relevant in forecasting multiple steps ahead, including future values of exogenous predictors.

2.2 Sequential Learning and Forecasting

The models of (1) are standard DLMS amenable to analytic computation for forward filtering and 1-step ahead forecasting. In our example context, the evolution of $\boldsymbol{\theta}_{j,t}$ is a simple (linear, conditionally normal) random walk, and is coupled with a discount/random walk volatility model for $\lambda_{j,t}$. This standard framework allows for change over time controlled by discount factors (e.g. Prado and West, 2010, section 4.3). A lower discount factor allows more substantial changes over time, while a discount factor of 1 corresponds to static coefficients; a brief summary of time t evolution and updating appears in Appendix 1 of the Supplement (Lavine et al., 2020). Importantly, filtering analyses are both analytic and conditionally independent across series j so are done in parallel, while forecasting involves recoupling across series.

At time t for each series j , the conditional (on parental predictors) 1-step ahead forecast distribution $p(y_{j,t}|\mathbf{y}_{pa(j),t}, \mathcal{D}_{t-1})$ is a univariate T-distribution with trivially computed parameters. This yields the joint forecast density function via composition, i.e., $p(\mathbf{y}_t|\mathcal{D}_{t-1}) = \prod_{j=1}^m p(y_{j,t}|\mathbf{y}_{pa(j),t}, \mathcal{D}_{t-1})$. For $k > 1$, forecasting k -steps ahead is based on direct simulation, exploiting the recursive structure of DDNMs. This enables computationally trivial simulation of the path of the multivariate time series over the next k time points. Technically, this simply propagates samples of the paths of states and volatilities $(\boldsymbol{\theta}_{j,*}, \lambda_{j,*})$ for each series j , coupled with sampling from the conditionally normal DLMS to generate the \mathbf{y}_* . Standing at time t , for example, this evaluates the full path forecast distribution by generating Monte Carlo samples, a.k.a. “synthetic

futures”, from $p(\mathbf{y}_{t+1}, \dots, \mathbf{y}_{t+k} | \mathcal{D}_t) = \prod_{h=1}^k p(\mathbf{y}_{t+h} | \mathbf{y}_{t+1}, \dots, \mathbf{y}_{t+h-1}, \mathcal{D}_t)$. All practical forecasting interests over the coming k periods can then be addressed with relevant Monte Carlo summaries (e.g., expected or median paths, prediction of turning points, maxima or minima, value-at-risk, expected utility functions, etc.).

2.3 Model Uncertainty: Predictive Variable Specification

The central model structure uncertainty question in DDNMs is specification of predictor variables in both exogenous/lagged $\mathbf{x}_{j,t}$ terms and parental sets $pa(j)$. Write \mathcal{M}_j for a set of candidate models \mathcal{M}_j^r for series j , indexed by $r \in \{1, \dots, |\mathcal{M}_j|\}$. Mathematically, $\mathcal{M}_j^r \in \{0, 1\}^p$ is a p -dimensional vector selecting predictor variables. An important point is that we will be expanding the framework so that model spaces are effectively time dependent, i.e., $\mathcal{M}_j \rightarrow \mathcal{M}_{j,t}$, but for now maintain the simpler notation. Denote by \mathcal{M} any single multivariate model for \mathbf{y}_t defined by a selection of one model from each of the m sets \mathcal{M}_j .

Consider first the discount learning extension of standard Bayesian model probability analysis, and BMA as a special case. At any time $t - 1$, denote the current model probabilities by $p(\mathcal{M} | \mathcal{D}_{t-1})$. Given a model space discount factor α such that $0 < \alpha \leq 1$, the discount modified model probability at time t is defined by

$$p(\mathcal{M} | \mathcal{D}_t) \propto p(\mathcal{M} | \mathcal{D}_{t-1})^\alpha p(\mathbf{y}_t | \mathcal{M}, \mathcal{D}_{t-1}) \propto p(\mathcal{M} | \mathcal{D}_0)^{\alpha^t} \prod_{h=1}^t p(\mathbf{y}_h | \mathcal{M}, \mathcal{D}_{h-1})^{\alpha^{t-h}},$$

where the earlier notation for 1-step forecast p.d.f.s has been extended to be explicit that it depends on the specific model structure \mathcal{M} . A discount $\alpha < 1$ acts to reduce the impact of historical information in model comparisons, with data from n time points in the past discounted by α^n in the cumulation of model scores. Evidently, standard Bayesian analysis sets $\alpha = 1$. As t increases, $\alpha < 1$ means that model weights will not degenerate; they adapt over time and respond to varying 1-step predictive abilities across the sets of models (West and Harrison, 1989a, p. 445; Raftery et al., 2010; Xie, 2012; Koop and Korobilis, 2013; Zhao et al., 2016). A major potential benefit is that of adapting more rapidly to reweight models based on more “local” behavior in the series, and down-weight models that were historically more favored but are locally of lower predictive value. This often yields improved predictive performance as illustrated in multiple examples in the above references.

A major benefit of DDNMs is that model uncertainty is addressed across series j independently, as dynamic variable selection problems in each of the univariate DLNs. This implies $\sum_{j=1}^m |\mathcal{M}_j|$ possible models \mathcal{M} , whereas a direct multivariate analysis would involve a much more substantial set of $\prod_{j=1}^m |\mathcal{M}_j|$ models. That is, as earlier noted, $p(\mathbf{y}_t | \mathcal{M}, \mathcal{D}_{t-1}) \propto \prod_{j=1}^m p(y_{j,t} | \mathcal{M}_j, \mathbf{y}_{pa(j),t}, \mathcal{D}_{t-1})$ so the contributions to model scores given by the set of m 1-step forecast p.d.f. values are decoupled across series.

Traditional Bayesian analysis—perhaps with the practically motivated but otherwise subjective intervention-based discount model probability variant—proceeds using model scores defined above.

3 Time-Adaptive Variable Selection

3.1 Model Structure Uncertainty and Practical Forecasting

Following discussion and motivation in Section 1, we develop analysis consistent with the following perspectives.

- Performance in prediction with respect to specific, defined forecasting goals should arbitrate model comparisons, combination and selection. Evaluation of alternative models, and the definition of models scores to use in weighting models for aggregation in prediction and selection of future “optimal” models, should consider specific forecast and/or decision goals.
- At each time t , it is desirable to have a single chosen model for communication and use in forecasting, and changes to the chosen model over time justified based on improvements in forecast accuracy modulo specific forecasting goals.
- Consideration of banks of models to assess any “current” model, and combination of selected sets of models for forecasting purposes, should be entertained at any times that forecast accuracy under that chosen model might be questioned. This can be done routinely at each time point, or at selective time points based on model monitoring and assessment of predictive accuracy modulo the specific forecasting goals (West and Harrison, 1986; West, 1986; West and Harrison, 1989b; Gruber and West, 2016, 2017).

The methodological contributions of this paper include a strategy for time-adaptive variable selection that address these desiderata. The resulting adaptive variable selection (AVS) strategy is composed of: (1) so-called Gibbs model probabilities, tying model evaluations with defined forecasting objectives; (2) a local search strategy over model spaces to dynamically explore potential models relative to a “current” selection; (3) a choice of a representative model at each time point for communication, interpretability and as a basis to evolve forward in time; and (4) the use of (1–3) adaptively over time.

3.2 Gibbs Model Probabilities

Our approach relates to the growing interest in Bayesian decision-guided inference with loss or utility functions used to define mechanisms to update subjective probabilities over models (or, more generally, over uncertain states and parameters). We use the term “Gibbs model probabilities”, contributing to the growing literature concerned with so-called generalized belief updating in which data-based evidence is represented in likelihood functions constructed based on defined loss or utility functions (Jiang and Tanner, 2008; Bissiri et al., 2016). Previous work has used purely statistical loss functions, and established that such an approach can provide superior risk performance to Bayesian updating under model misspecification (Zhang, 2006a,b; Jiang and Tanner, 2008). Beyond expanding the ideas to dynamic model structure uncertainty and developing a sequential, adaptive approach, a key focus here is to exploit the approach using loss or

utility functions specific to the main prediction problems of interest, also linked to prior work on explicitly recognizing model selection as a decision (e.g. Hahn and Carvalho, 2015).

Consider series j with (time 0) baseline model probabilities $p(\mathcal{M}_j^r|\mathcal{D}_0)$ over selected models $\mathcal{M}_j^r \in \mathcal{M}_j$. Gibbs model probabilities based on data \mathcal{D}_t (from all m series) observed up to time t are defined by

$$p_j(\mathcal{M}_j^r|\mathcal{D}_t) \propto p(\mathcal{M}_j^r|\mathcal{D}_0)e^{\tau s_{j,t}(\mathcal{M}_j^r)}, \quad (2)$$

where $\tau > 0$ and $s_{j,t}(\mathcal{M}_j^r)$ is a *model score*. A higher model score indicates more support for the model \mathcal{M}_j^r and reflects historical performance in a specific forecasting or decision problem. The scores are defined by choosing a utility function relevant to the specific goals. Examples include simple point forecast metrics or full (log) predictive densities for functions of the outcome time series as used in our examples below. With scores on a known or standardized scale, the parameter τ balances information from the past data with that in the prior. Questions of how to calibrate τ are discussed in Bissiri et al. (2016) and in our settings in Sections 4 and 5 where scores are based on out-of-sample predictive densities. The general setting allows scores to be of different forms across series j , or to be based on a common function but with series j -specific weights, or to involve the same score for each component series. Examples below use the latter, while the generality is important to note for future applications. Gibbs model probabilities are used for model averaging just as in standard model uncertainty analysis; note that the latter arises as a special case when $\tau = 1$ and scores are simply the logs of 1-step ahead predictive densities. More generally, the Gibbs likelihoods can be interpreted as extensions of model likelihood from the 1-step ahead predictive focus to that based on a desired forecast or decision goal.

One of the major benefits of DDNMs is that, as discussed in Section 2.3, marginal likelihoods for a multivariate model are simply the products of likelihoods from each of the m univariate models. This carries over to Gibbs model probabilities assuming that the scores are unrelated and that baseline models are independent across series. Then the overall probabilities on the multivariate model \mathcal{M} defined by the set of m chosen models $\mathcal{M}_j^{r_j}$ is simply the product of terms in (2),

$$p(\mathcal{M}|\mathcal{D}_t) \propto p(\mathcal{M}|\mathcal{D}_0)e^{\tau s_t(\mathcal{M})}, \quad s_t(\mathcal{M}) = \sum_{j=1}^m s_{j,t}(\mathcal{M}_j^{r_j}), \quad (3)$$

where $p(\mathcal{M}|\mathcal{D}_0)$ is the product of the $p(\mathcal{M}_j^r|\mathcal{D}_0)$. The overall probabilities $p(\mathcal{M}|\mathcal{D}_t)$ are used for model averaging for prediction and decisions, and then model selection. That is, model evaluation is decoupled to the levels of the univariate series, then recoupled to assess the overall multivariate model.

3.3 AVS Strategy and Representative Model Selection

The overall strategy of adaptive variable selection (AVS) is summarized in Algorithm 1 below. At each time t , we find a set of models using Shotgun Stochastic Search (SSS),

a strategy to explore regions of strong models (Jones et al., 2005; Hans et al., 2007a,b; Scott and Carvalho, 2008; Wang, 2015). Models are evaluated using Gibbs model probabilities, and averaged together for forecasting and decisions before moving on to the next time point. The AVS strategy is run in parallel over series $j = 1, \dots, m$ to find candidate models and evaluate their Gibbs probabilities. Forecasting from multivariate DDNMs is then trivial via sequential simulation.

Algorithm 1 Adaptive Variable Selection.

```

1: for time  $t$  in 1:T do
2:   for series  $j$  in 1:m do
3:     Find a set of candidate models  $\mathcal{M}_j$  with SSS, seeded by representative model
        $\mathcal{M}_{0,t-1}$ 
4:     Calculate Gibbs probabilities  $p_j(\mathcal{M}_j^r | \mathcal{D}_{t-1})$ 
5:     Forecast with model averaging, where the probability of DDNM  $\mathcal{M}$  is the prod-
       uct of its univariate model probabilities.
6:     Observe  $y_t$ , and for each series  $j$ , update Gibbs Probabilities  $p_j(\mathcal{M}_j^r | \mathcal{D}_t)$ 
7:     Choose a new representative model  $\mathcal{M}_{0,t}$ 

```

As discussed in Section 1 and the desiderata of Section 3.1, it is often desirable for interpretation and communication to operate using a single selected model unless or until changes are suggested based on a breakdown in model performance or external considerations. Thus selecting one model as a representative of the probability-weighted set is of interest. Denote by $\mathcal{M}_{0,t}$ a DDNM chosen as the *representative model* at time t . A natural choice is the modal model with respect to Gibbs model probabilities, i.e., modulo the baseline probabilities that model maximizing the overall score $s_t(\mathcal{M})$. Alternatives would choose $\mathcal{M}_{0,t}$ as a Bayesian decision with respect to the mixture over models. A natural approach would choose the representative model to best approximate (e.g., using Kullback-Leibler divergence) a specific predictive distribution that averages over the full set of models under consideration. This has theoretical and practical appeal, but is computationally expensive relative to selection of the modal model.

Analysis proceeds through the DDNM evolution to time $t + 1$ using the single model $\mathcal{M}_{0,t}$. Then observing \mathbf{y}_{t+1} we face the question of identifying classes of models \mathcal{M}_j and computing Gibbs model probabilities. The theoretical indication that we continuously update scores on all possible models is simply not practicable in realistic settings. Then, as time evolves different models become of interest relative to those that had scored well in the past. Further, interventions at certain times may change the class of models under consideration (e.g. by adding new potential predictors not so far considered). Hence the interest is (a) to identify sets of models at time $t + 1$ that appear competitive with $\mathcal{M}_{0,t}$ in terms of the specific forecasting goals, i.e., in terms of the defined score function, while (b) to do so computationally efficiently as this will be repeated at each time point. Our AVS implementation utilizes an extension of SSS to address these goals, as detailed in Section 3.4 below.

A practical modification is to use the model search and weighting via AVS only at selected time points. That is, at time $t + 1$ and over a number of further time

points, we may simply use the single model $\mathcal{M}_{0,t}$ for evolution and forecasting. At some point, however, consideration of other models will become important, and then the AVS strategy of finding and weighing sets of models will come into play.

3.4 Finding Models: Shotgun Stochastic Search

Originally developed for graphical models and regression, shotgun stochastic search is designed to quickly identify and explore interesting regions in large, discrete model spaces (Jones et al., 2005; Hans et al., 2007a,b; Scott and Carvalho, 2008; Wang, 2015). The proven ability to rapidly transit model spaces based on “local changes” to existing models makes SSS perfectly suited to the AVS context in DDNMs with larger numbers of potential predictor variables per series. At time $t+1$, the current representative model $\mathcal{M}_{0,t}$ serves as an initial “seed model”. Based on this, SSS proceeds as follows:

1. Identify a neighborhood of the seed model, typically the set of models $\{\mathcal{M}_{0,t}^+, \mathcal{M}_{0,t}^o, \mathcal{M}_{0,t}^-\}$, where
 - (a) $\mathcal{M}_{0,t}^+$ is all models with 1 predictor added,
 - (b) $\mathcal{M}_{0,t}^o$ is all models with 1 predictor swapped,
 - (c) $\mathcal{M}_{0,t}^-$ is all models with 1 predictor subtracted.

In DDNMs this applies separately to each of the m decoupled DLMs for univariate series.

2. Evaluate all such models in the neighborhood. This can use posterior model probabilities or Gibbs model probabilities, or any other scoring method desired (e.g., scores from specific decision problems—e.g. West, 2020, Section 2).
3. Record this set of models and scores in a running list.
4. Sample a new seed model from this neighborhood, and repeat. Sampling will be done using model probabilities or the decision-guided Gibbs extensions.

When the seed model is highly scoring, then the set of neighboring models will typically include many other interesting models in terms of the score. SSS therefore fully exploits local modes in model space to swiftly move between individual high probability models to reach varied parts of the model space. Neighboring models can be evaluated in parallel, which is a clear advantage over sequential search methods. This makes SSS particularly suited for situations where full exploration of the model space is computationally impossible, either because the set of models is large, or calculating scores is slow. Importantly, the goal is to identify subsets of highly scoring models to underlie forecasting and evolution to the next time point; the goal is explicitly quite different to that of MCMC-based model search strategies, i.e. of “structure learning”.

Within the SSS search at each time, each model identified requires fitting over a period of past data—possibly all data from $t = 0$ or perhaps over a restricted recent period—

to evaluate model scores based on the historical forecasting record. That DDNMs admit fast, analytic computation is critical here, enabling evaluation of even very large sets of candidate models at each time point; again, these computations are inherently decoupled hence parallelizable within each time point.

4 Synthetic Time Series Example

A simple but relevant and illuminating example with synthetic data illustrates AVS compared to standard model averaging, and demonstrates how AVS selects predictors with stable effects for long-term forecasting. This is clearest in the simple case where $m = 1$ series so the DDNM reduces to a single DLM at $j = m = 1$, with data $y_t \equiv y_{1,t}$. The data are simulated from a model exhibiting both steady and more rapidly changing dynamics. We generate $y_t = c + \theta_{1,t}x_{1,t} + \theta_{2,t}x_{2,t}$, with simulated $\theta_{1,t}$ and $\theta_{2,t}$ displayed in Figure 1. Note that $\theta_{1,t}$ is rapidly changing, while $\theta_{2,t}$ is relatively steady. Predictors $x_{1,t}$ and $x_{2,t}$ are randomly set at 1 or -1 with probability $1/2$.

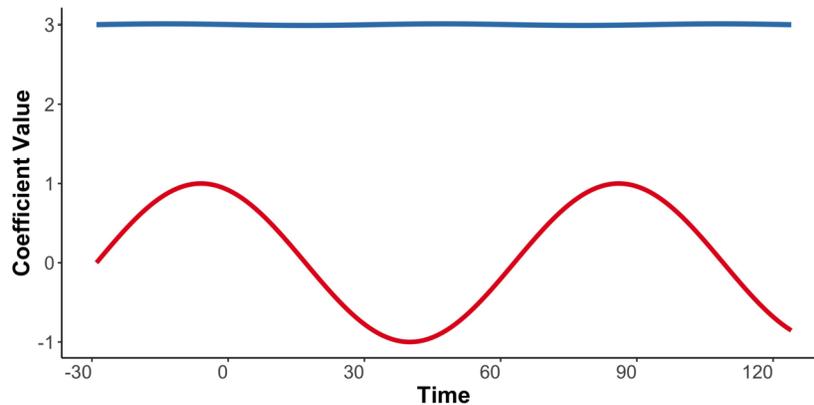


Figure 1: Time-varying coefficients $\theta_{1,t}$ (red) and $\theta_{2,t}$ (blue) underlying synthetic data.

Each model \mathcal{M} is a univariate DLM defined by a choice of predictor variables. This is illustrated for the Gibbs posterior modal model in Figure 2. All models include an intercept so there are 4 possible combinations of the variables $x_{1,t}$ and $x_{2,t}$ for inclusion, defining 4 candidate models at each time. In each DLM, the state vector and volatility processes follow standard random walk evolutions as earlier discussed, with discount factors $\delta = \beta = 0.98$; see also Appendix 1 of the Supplement. Gibbs model probabilities use $\tau = 1$ as scores are log forecast densities so the resulting probabilities are on the same scale as standard Bayesian model probabilities. The baseline priors at $t = 0$ give equal weight to each model, and conjugate normal/inverse gamma priors for the state vector and volatility in each model are based on informal analysis of data from an additional training period of 30 time steps before formal model scoring and AVS analysis begins over $t = 100$ time points.

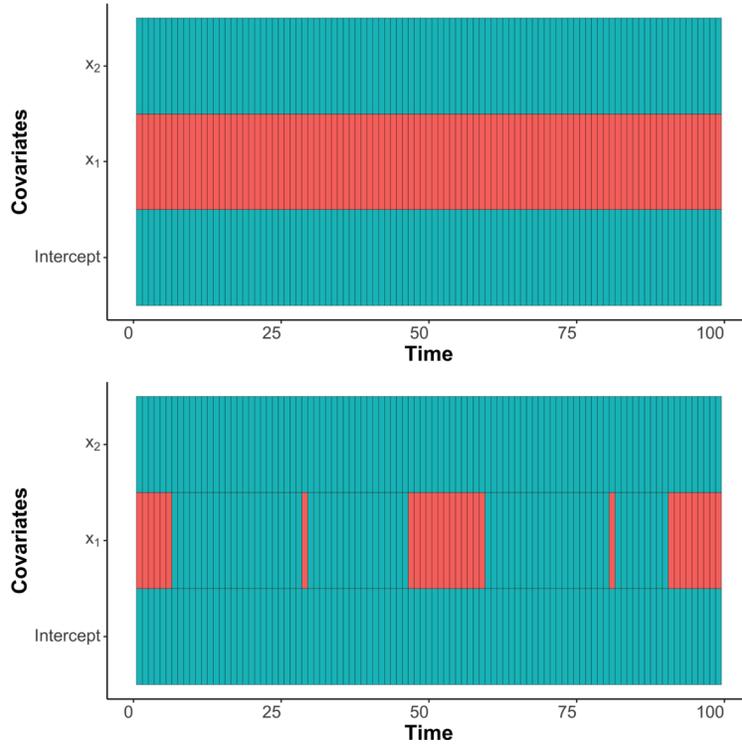


Figure 2: Synthetic data example: Indicators of inclusion (*green*) of $x_{1,t}$ and $x_{2,t}$ in the posterior modal models under AVS (*upper*) and BMA (*lower*).

Figure 3 shows that AVS forecasting dominates BMA in terms of model-averaged predictive density. With $k = 25$ to drive AVS, the k -step ahead based predictive density score naturally improves over the myopic BMA. Smaller differences occur at periods when the BMA drops $x_{1,t}$ or, by chance, $\theta_{1,t} \approx \theta_{1,t+k}$. More deeply, using the same AVS analysis with $k = 25$ in fact improves forecast accuracy over all horizons, as exhibited by the marginal root mean squared forecast error (rMSFE) for each horizon 1 – 25 steps ahead in the figure. This occurs even though the model is weighted by $k = 25$ -step ahead scores only, and the figure highlights the fact that standard BMA will tend to perform well only at short horizons.

Reflecting central interests in multi-step ahead forecasting arising in many applications (e.g. Nakajima and West, 2013a), Gibbs model probabilities at each time are based on model scores of marginal k -step ahead forecast accuracy with $k = 25$ as an example. The model score function $s_t(\mathcal{M}) \equiv s_{1,t}(\mathcal{M}_1)$ on each model \mathcal{M} is simply

$$s_t(\mathcal{M}) = \sum_{h=0}^{t-k} \alpha^{t-k-h} \log(p(y_{h+k}|\mathcal{M}, \mathcal{D}_h)),$$

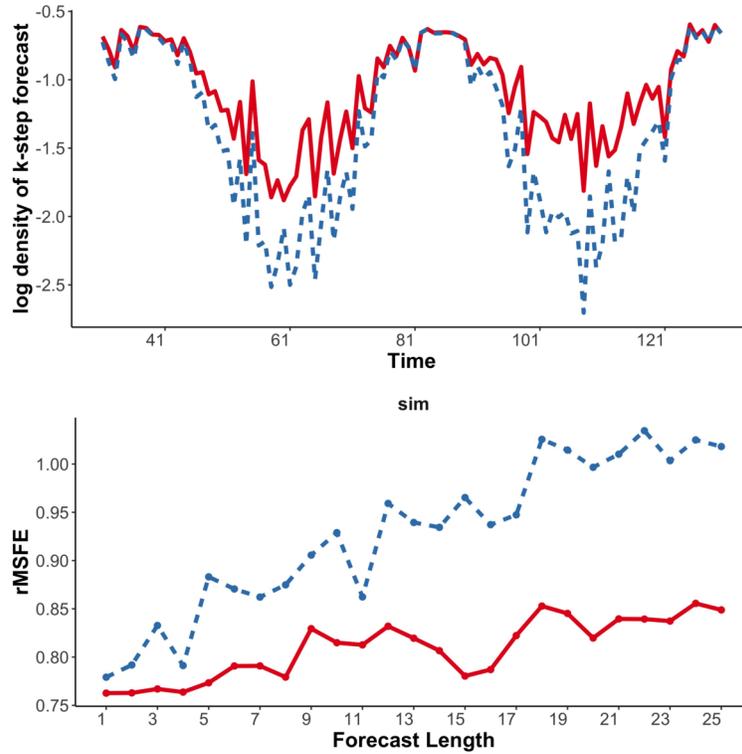


Figure 3: Synthetic data example: Multi-step ahead forecast performance using AVS (red, solid) (with $k = 25$) and BMA (blue, dashed). Log forecast density scores $\log(p(y_{t+25}|\mathcal{D}_t))$ over time t (upper), and root mean squared forecast errors as a function of forecast horizon (lower).

for some model discount factor $\alpha \in (0, 1]$. Here $\alpha < 1$ down-weights more distant past outcomes as in the discount Bayesian model uncertainty analysis of Section 2.3 that arises as the special case when $k = 1$; standard BMA is given with $\alpha = k = 1$. Our example here sets $\alpha = 0.98$ for both Gibbs and Bayesian model probabilities. Previous studies have identified $0.95 < \alpha < 1$ as an appropriate range, with little benefit in considering multiple α values (Raftery et al., 2010; Koop and Korobilis, 2013; Zhao et al., 2016).

The behavior of adaptive variable selection is best illustrated through the identification of the representative model, taken here as the Gibbs posterior modal model at each time point; see Figure 2. Predictor $x_{1,t}$ is uniformly excluded; inclusion of a variable with rapidly changing and unpredictable dynamics generally degrades long-term predictions. In contrast, the posterior modal model from BMA almost always includes $x_{1,t}$, except when the coefficient $\theta_{1,t}$ drops near to 0.

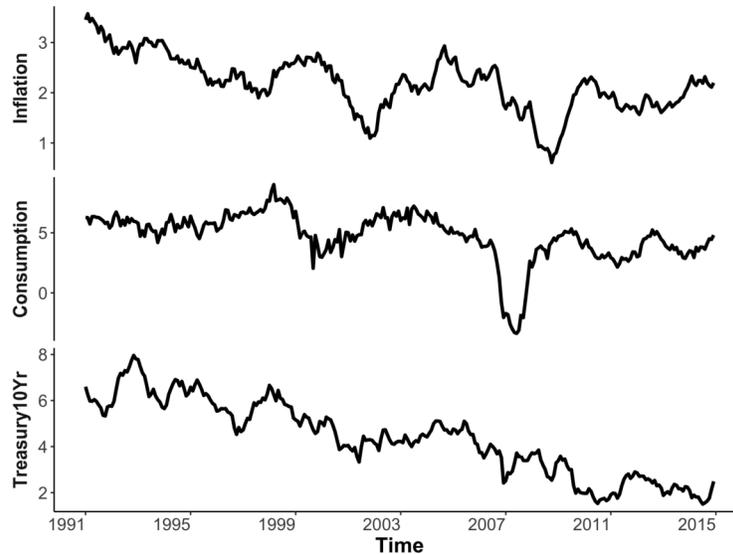


Figure 4: Macroeconomic example: Monthly time series of US Inflation, Consumption, and yield on 10-year Treasury bonds (Tr10Yr) over 24 years up to the end of 2016.

5 Macroeconomic Case Study

5.1 Forecasting Context and Data

We address monthly forecasting of three key US macroeconomic series: year-over-year Inflation, Consumption, and the 10-year yield on Treasury bonds (Tr10Yr). Data over 1991–2016 from the St. Louis Federal Reserve are shown in Figure 4. The sharp drop in both Inflation and Consumption during recessions is clear in 2001 and 2008, while Inflation and Tr10Yr show slight long-term downward trends. Improved forecasting of these and related series is a central concern in national monetary policy, and forecasting more than a few months ahead is notoriously challenging (e.g. Primiceri, 2005). While particular interests lie in forecasting 12–24 months ahead at each time point, central bank concerns spread across forecast horizons. It is becoming increasingly clear that customizing models to the forecast horizon of interest can improve forecast accuracy and potentially generate economic insights into dynamic relationships among series over time (Nakajima and West, 2013a; McAlinn and West, 2019; McAlinn et al., 2020).

Potential predictors include all 1–12 month lags of each series. The DDNM orders series as Inflation-Consumption-Tr10Yr. Hence Consumption and Tr10Yr are potential parents of Inflation, Tr10Yr is a potential parent for Consumption, while Tr10Yr has no parents. Including a possible intercept, the total predictor space has 39 potential predictors for Inflation, 38 for Consumption, and 37 for Tr10Yr. We summarize forecasting results from analyses as follows. Earlier data from 1986–1990 was used informally to choose informative priors at the start of 1991 for all states and volatilities. Analyses

were run for a training period of 5 years, and then full forecast evaluations were made over the 252 month period 1996–2016 inclusive.

5.2 Horizon-Specific Multi-step Forecasting

Initial analysis considers marginal forecasts for $k = 24$ months ahead. The score is the discounted log predictive density at the chosen horizon as in the univariate example in Section 4 but now for the multivariate series; that is,

$$s_t(\mathcal{M}) = \sum_{h=0}^{t-k} \alpha^{t-k-h} \log p(\mathbf{y}_{h+k} | \mathcal{M}, \mathcal{D}_h), \quad (4)$$

where \mathbf{y}_{h+k} is the multivariate observation at time $h+k$, and both AVS and BMA use $\alpha = 0.98$ for this study. Note that the computational cost of AVS scales *linearly* in the number of series because we are able to take advantage of the DDNM structure. Model scores are evaluated for each series j independently, and in parallel, by conditioning on the past values of other series. Forecasts are evaluated using the joint 24-step ahead forecast density. Figure 5 shows that AVS dominates BMA with respect to the long-term forecasting objective function defined by the usual model-averaged predictive density while, as expected, BMA analysis is more accurate in the shorter-term predictions.

Differences between model weightings and selection under AVS and BMA can be visualized in terms of variables included in the modal DDNMs and how these variable sets change over time. For BMA this is simply the model with maximum posterior probability at each time point, while for AVS it is the representative modal model at each time. The DDNM component models for the Inflation series are highlighted in Figure 6. AVS focuses on higher lags of predictor variables, particularly lag-12 Inflation and lag-12 Consumption. Models featuring these higher lags produce more stable and accurate longer-term forecasts, although they are less accurate for 1-step forecasting than models which include the lag-1 variable that are more favored under BMA.

5.3 Multi-step Path Forecasting

A major interest lies in improved *path forecasting*. In this applied context, the main focus is on how the macroeconomy is predicted to evolve over a coming period of months, and how the series are predicted to interact over that time. Such goals are naturally addressed in the Bayesian framework by exploring full joint predictive distributions over multiple months. This is contrasted with the usual horizon-specific, or marginal forecasting analysis, of Section 5.2. With a focus on the path over the next k time points, we are therefore interested in the ($k \times m$ -dimensional) *path forecast density* $p_t(\mathbf{y}_{t+1}, \dots, \mathbf{y}_{t+k} | \mathcal{D}_t)$, and refer to understanding the underlying distribution as path forecasting. The suffix t makes explicit that this is the joint forecast over the next k time points made at time t . In addition to potentially extracting distributional summaries, one key use of models is simulation: generating “synthetic future paths” of the economy that can be explored subjectively and used to interrogate predictions on arbitrary functions of the economic variables (e.g., defined downturns, etc). In terms of model structure assessment, the explicit aim is to find models that balance short and longer-term forecasting, rather than focus on one or more specific horizon.

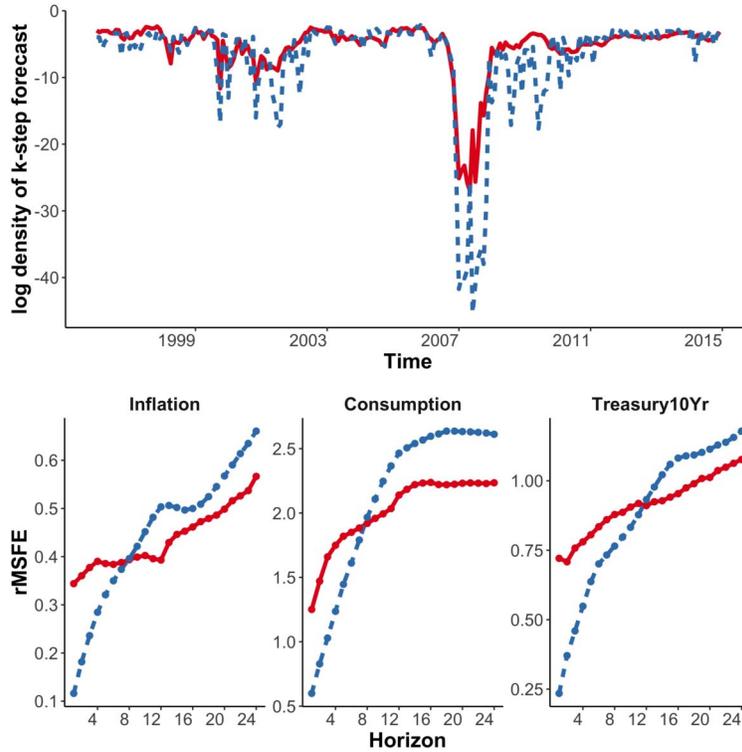


Figure 5: Macroeconomic example: Marginal $k = 24$ -steps ahead forecasting comparisons between AVS (red, solid) and BMA (blue, dashed). Log forecast densities $\log(p(\mathbf{y}_{t+24}|\mathcal{D}_t))$ over time t (upper), and marginal root mean squared forecast errors over 1 to 24 month forecast horizons (lower).

Define the corresponding log path forecast density (LPFD) score at any time t via

$$s_t(\mathcal{M}) = \sum_{h=0}^{t-k} \alpha^{t-k-h} \log(p_h(\mathbf{y}_{h+1}, \dots, \mathbf{y}_{h+k}|\mathcal{M}, \mathcal{D}_h)), \tag{5}$$

with discount $\alpha \in (0, 1]$. As above, the example sets $\alpha = 0.98$ for both AVS and BMA. As the loss function is based on a k -dimensional joint density, the natural setting for the scale parameter τ in Gibbs model probabilities is $\tau = 1/k$; this puts the model score on the same scale as in standard Bayesian updating of model probabilities. As in Section 5.2, the computation scales linearly in the number of series because the LPFD score can be factored into independent terms for each univariate series.

In any chosen DDNM, forecast evaluation of path scores is via Monte Carlo. This involves simple, direct/forward simulation of state vectors and volatilities in the usual recursive form within each time point, and then sequentially over the next k time points.

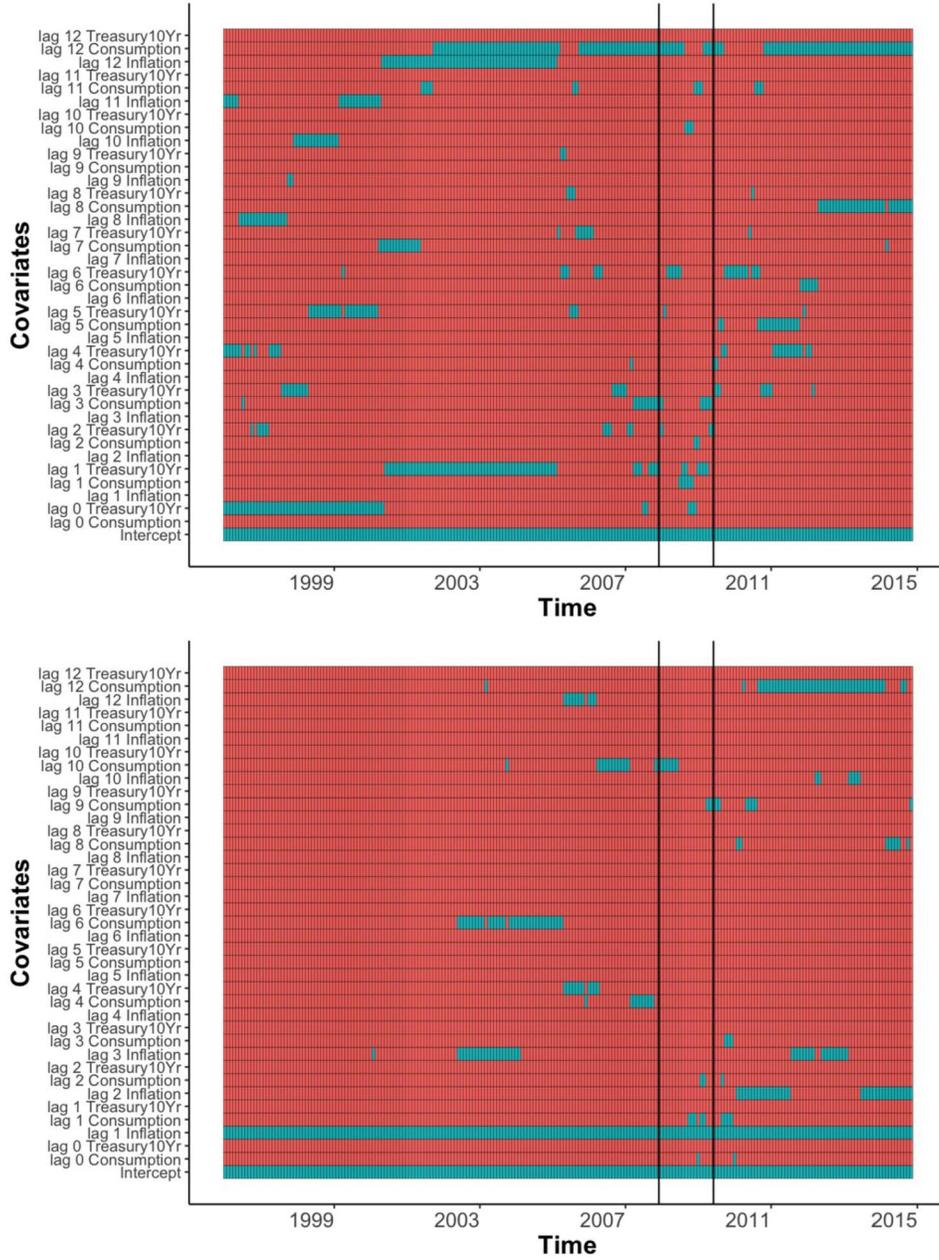


Figure 6: Macroeconomic example: Marginal $k = 24$ -steps ahead forecasting comparisons. Dynamic variable inclusion (*green*) for DDNM series $j = 1$, Inflation, in the posterior modal models: under marginal 24-step AVS (*upper*) and BMA (*lower*). Vertical bars indicate the start and end of the great recession, during which AVS rapidly adapts to changing economic conditions.

This generates samples from the predictive distributions of these latent parameter processes, each of which defines a full set of conditional multivariate normal distributions for the outcome path $\mathbf{y}_{h+1}, \dots, \mathbf{y}_{h+k}$. Monte Carlo averages of these normals evaluated at the eventual outcome data provide Monte Carlo evaluations of path densities.

5.4 Path Forecasting Results

The expectation is that path forecast-guided AVS will improve longer-term (up to horizon k) forecast accuracy while still favoring models with realistic short-term forecasts. This is borne out. Figure 7 shows summary information for 1- and 12-month ahead Inflation forecasts. Related plots for Consumption and Tr10Yr are in Appendix 2 of the Supplement. Improved forecasting accuracy under AVS can be seen in the marginal rMSFEs. The trend is for the short-term forecast accuracy to be very similar to BMA, with AVS offering increasing improvements over BMA at longer forecast horizons.

Trajectories over time of indicators of variables included in modal models provide insights into differences between AVS and BMA in this path forecasting context; this is illustrated in Figure 8 in DDNM model components for predicting series $j = 3$, Tr10Yr. It is typical and to be expected that the lag-1 value of a given series is a dominant predictor of that series, especially with data at monthly levels. This holds true in the models selected by both BMA and AVS for all 3 series, exemplified in this figure for Tr10Yr. However, with the LPFD score using $k = 24$, AVS does better in capturing longer-term dynamics; the figure highlights the involvement of higher lags of all series in the AVS analysis relative to that using standard BMA.

5.5 Higher Dimensional Results

As described in Sections 5.2 and 5.3, the computational load of AVS analysis in DDNMs scales *linearly* in the number of series, enabled by evaluating model scores independently across series in the DDNM hierarchy. A further study of a higher-dimensional series confirms the ability to scale computations while also reinforcing the role of the AVS strategy in goal-focused forecasting.

We expand the previous analysis to 7-dimensions by adding time series of year-over-year Wage Growth, M2 Money Stock, Moody's BAA Corporate Bond, and Gold prices. These variables are important indicators of the labor market, monetary policy, and corporate activity. We consider as potential predictors 1, 3, 6, and 12 month lags of each series. The model score is the marginal $k = 24$ month ahead log predictive density, as defined in Section 5.2. As in the previous example, we set $\alpha = 0.98$ for both AVS and BMA, and $\tau = 1$ to put the model score on the same scale as standard Bayesian model probabilities.

Figures 9 and 10 compare the performance of AVS and BMA. The higher-dimensional DDNM provides a greater diversity of economic signals to choose from as predictors, leading to larger differences between AVS and BMA. As in the 3-dimensional example, AVS consistently dominates BMA with respect to the 24-month ahead log forecast density, especially during periods of economic upheaval. The marginal rMSFE shows that

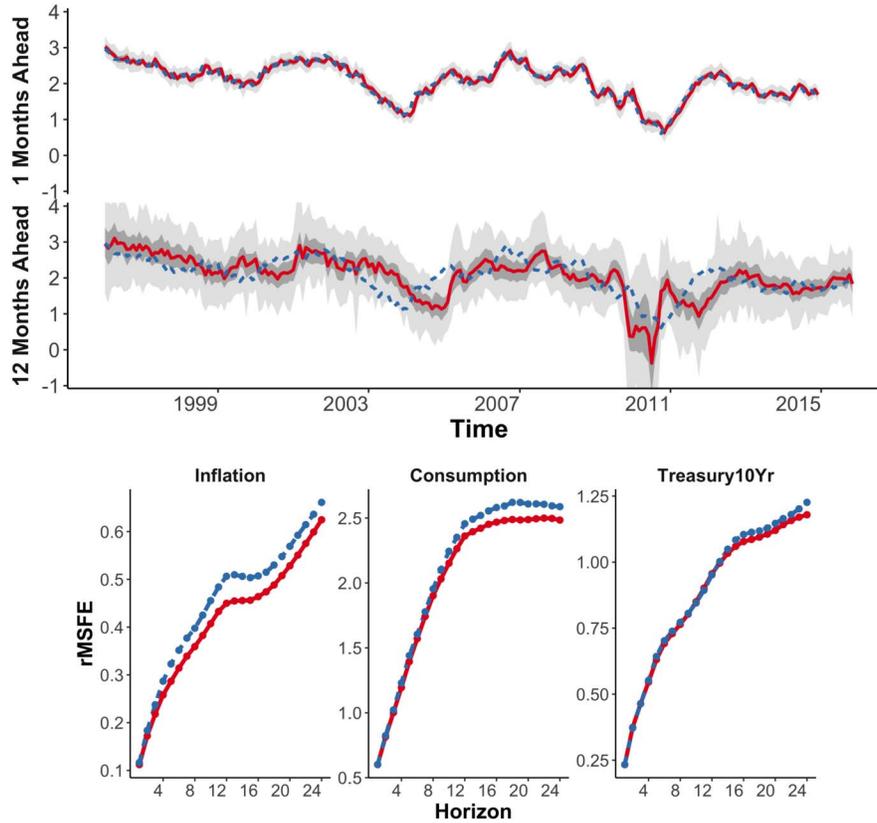


Figure 7: Macroeconomic example: Path forecasting comparisons. The upper frame shows 1-month and 12-months ahead forecasts of Inflation using AVS: data (*blue, dashed*), forecast means (*red, solid*), 50% prediction intervals (*dark gray bands*) and 95% prediction intervals (*light gray bands*). The lower frames show marginal rMSFE measures for Inflation, Consumption, and Tr10Yr over 1 to 24 month forecast horizons for AVS (*red, solid*) and BMA (*blue, dashed*).

BMA performs better at short-term forecasting, while AVS produces superior results over longer horizons.

6 Additional Comments

6.1 Conclusions

AVS builds on concepts in the recent Bayesian literature to define goal-oriented model structure uncertainty analysis that avoids the shortcomings of standard approaches. We do this in sequential, dynamic time series contexts by adapting goal-focused Gibbs

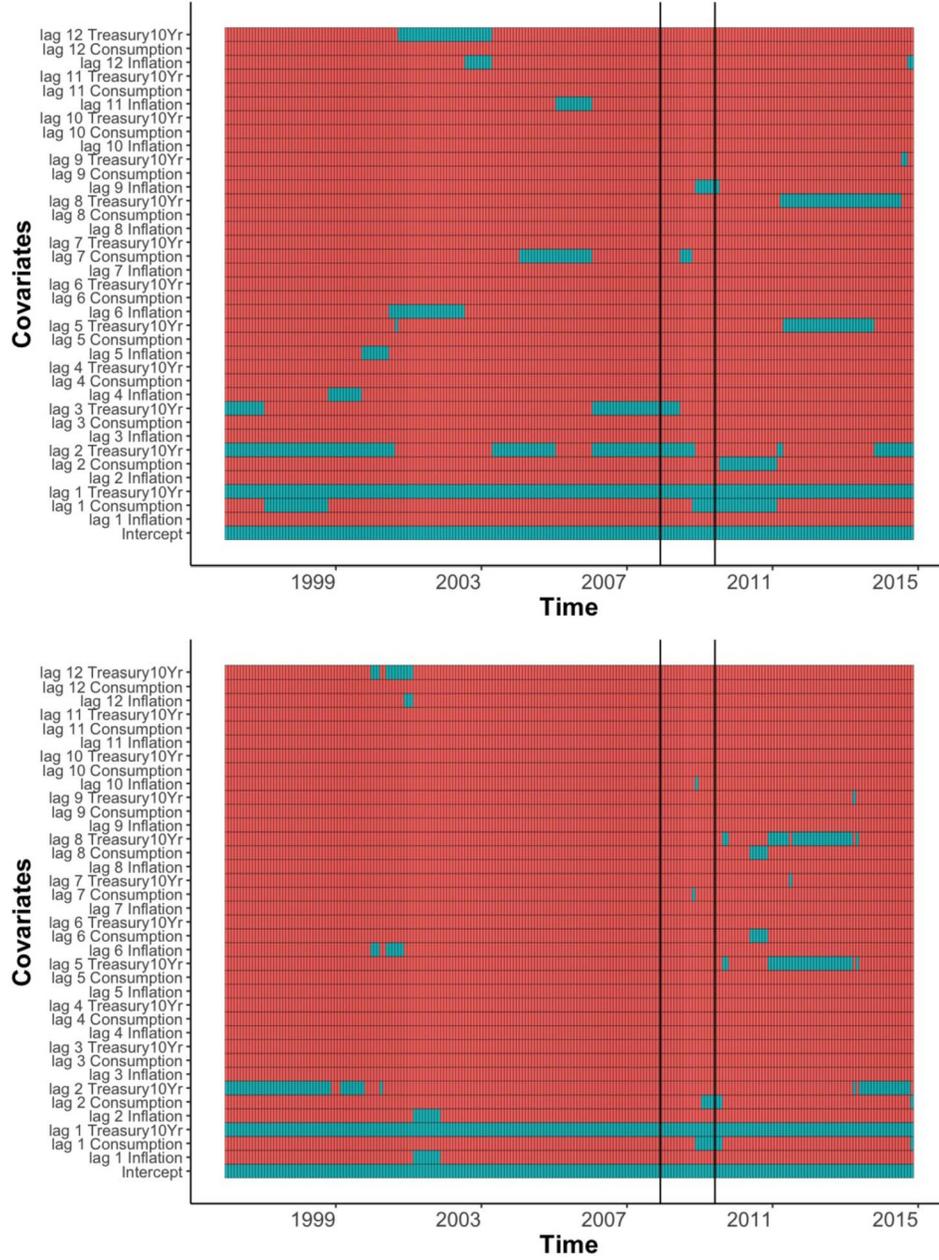


Figure 8: Macroeconomic example: Path forecasting comparisons. Dynamic variable inclusion (*green*) for DDNM series $j = 3$, Tr10Yr, in the posterior modal models: under 24-step path forecasting AVS (*upper*) and BMA (*lower*). Vertical bars indicate the start and end of the great recession, during which AVS focuses on shorter term predictors to adapt to changing economic conditions.

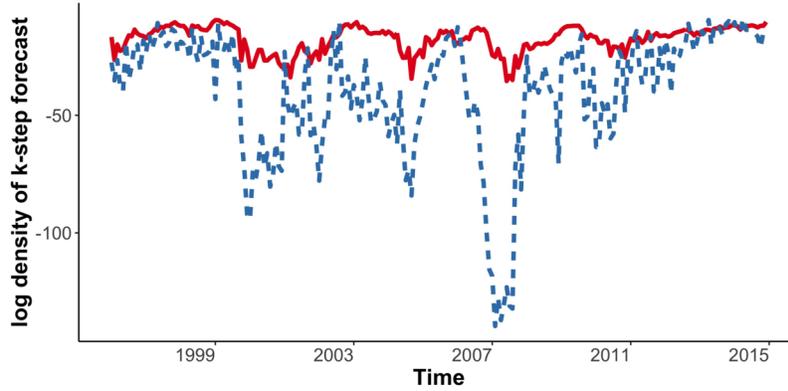


Figure 9: 7-Dimensional macroeconomic example: Log forecast densities $\log(p(\mathbf{y}_{t+24}|\mathcal{D}_t))$ over time t for AVS (*red, solid*) and BMA (*blue, dashed*).

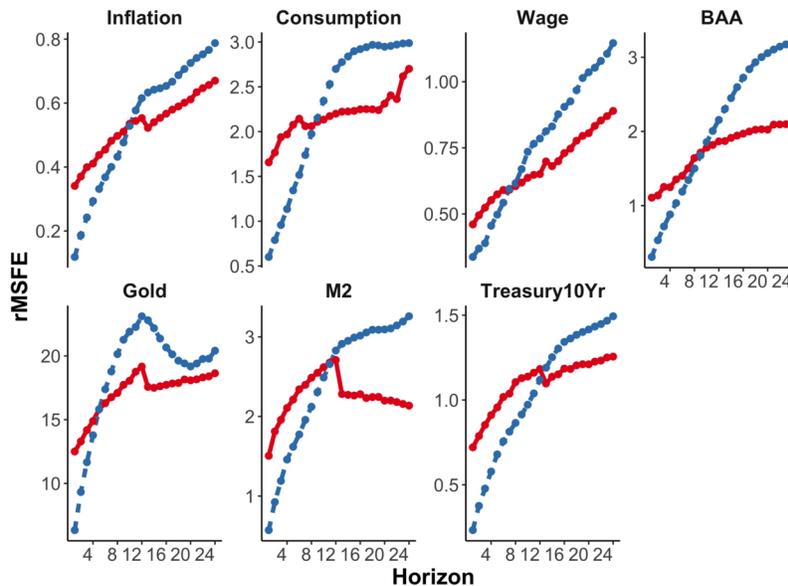


Figure 10: 7-Dimensional macroeconomic example: Marginal rMSFE measures for all series over 1 to 24 month forecast horizons for AVS (*red, solid*) and BMA (*blue, dashed*).

model probabilities coupled with efficient shotgun stochastic search over spaces of model structures, overlaid on standard Bayesian analysis in DDNMs. The resulting methodology maximally exploits analytic Bayesian computations within DDNMs, and is open to partial parallelization of both analytic and direct simulation-based computations for forecasting for increasingly high-dimensional series.

The examples with simulated and real economic data show the ability of goal-focused AVS to achieve superior results to standard Bayesian model structure learning. AVS improves longer-term forecast accuracy by identifying, weighting and averaging over models whose structure is different to that identified by standard Bayesian analysis; time series models more heavily weighted for longer-term forecasting naturally involve longer-lagged predictors. The important context of path forecasting emphasizes the benefits of AVS while also highlighting the relevance of standard Bayesian model probability analysis in connection with shorter-term forecasting.

6.2 Potential Extensions and Related Research

The paper uses DDNMs as context and examples. The related class of Simultaneous Graphical Dynamic Linear Models (SGDLMs) relax restrictions on the selection of parental predictors (Gruber and West, 2016, 2017), with a trade-off in terms of additional computational needs. AVS is directly extensible to SGDLMs (and, to other multivariate dynamic model frameworks, at least in principle) and further development in that direction can be anticipated.

The presentation and examples in the paper focus on the use of model scores based on a specific, defined forecasting or decision goal. In other settings, there may be several—possibly competing—goals. For example, in multi-step forecasting we may consider marginal forecasts at each horizon $h = 1, 2, \dots, k$ to be of explicit interest. Fitting separate models and AVS analyses for each horizon would be consistent with “models for goals” as in Bayesian predictive synthesis approaches to model combination (McAlinn et al., 2018; McAlinn and West, 2019; McAlinn et al., 2020), and with the over-arching motivation for AVS. This obviously raises questions of computational demands as well as of how to balance and potentially combine AVS analyses across goals. An alternative view is to use some form of aggregate score that balances interest across the several goals, consistent with practice in multi-objective Bayesian decision analysis.

Supplementary Material

Adaptive Variable Selection for Sequential Prediction in Multivariate Dynamic Models – Supplementary Material – (DOI: [10.1214/20-BA1245SUPP](https://doi.org/10.1214/20-BA1245SUPP); .pdf).

References

- Amisano, G. and Geweke, J. (2017). “Prediction using several macroeconomic models.” *The Review of Economics and Statistics*, 5: 912–925. [1060](#)
- Berry, L. R., Helman, P., and West, M. (2020). “Probabilistic forecasting of heterogeneous consumer transaction-sales time series.” *International Journal of Forecasting*, 36: 552–569. Published online Nov 25 2019. [1060](#)

- Berry, L. R. and West, M. (2020). “Bayesian forecasting of many count-valued time series.” *Journal of Business and Economic Statistics*, 38: 872–887. Published online: 25 Jun 2019. [1060](#)
- Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). “A general framework for updating belief distributions.” *Journal of the Royal Statistical Society (Series B: Statistical Methodology)*, 5: 1103–1130. [MR3557191](#). doi: <https://doi.org/10.1111/rssb.12158>. [1064](#), [1065](#)
- Clyde, M. and Iversen, E. S. (2013). “Bayesian model averaging in the \mathcal{M} -open framework.” In Damien, P., Dellaportes, P., Polson, N. G., and Stephens, D. A. (eds.), *Bayesian Theory and Applications*, 483–498. Clarendon: Oxford University Press. [MR3221178](#). doi: <https://doi.org/10.1093/acprof:oso/9780199695607.003.0024>. [1060](#)
- Crespo Cuaresma, J., Doppelhofer, G., Feldkircher, M., and Huber, F. (2019). “Spillovers from US monetary policy: evidence from a time varying parameter global vector auto-regressive model.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(3): 831–861. [MR3955500](#). doi: <https://doi.org/10.1111/rssa.12439>. [1062](#)
- Eickmeier, S., Lemke, W., and Marcellino, M. (2015). “Classical time varying factor-augmented vector auto-regressive models—estimation, forecasting and structural analysis.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3): 493–533. [MR3348347](#). doi: <https://doi.org/10.1111/rssa.12068>. [1062](#)
- George, E. I. (2010). “Dilution priors: Compensating for model space redundancy.” In Berger, J. O., Cai, T. T., and Johnstone, I. M. (eds.), *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown*, 158–165. Beachwood, Ohio, USA: Institute of Mathematical Statistics. [MR2798517](#). [1060](#)
- Giannone, D., Lenza, M., and Primiceri, G. E. (2018). “Economic predictions with big data: The illusion of sparsity.” *Federal Reserve Bank of New York: Staff Report, No. 847*. [1060](#)
- Gruber, L. F. and West, M. (2016). “GPU-accelerated Bayesian learning in simultaneous graphical dynamic linear models.” *Bayesian Analysis*, 11: 125–149. [MR3447094](#). doi: <https://doi.org/10.1214/15-BA946>. [1060](#), [1064](#), [1079](#)
- Gruber, L. F. and West, M. (2017). “Bayesian forecasting and scalable multivariate volatility analysis using simultaneous graphical dynamic linear models.” *Econometrics and Statistics*, 3: 3–22. [MR3666239](#). doi: <https://doi.org/10.1016/j.ecosta.2017.03.003>. [1060](#), [1064](#), [1079](#)
- Hahn, P. R. and Carvalho, C. M. (2015). “Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective.” *Journal of the American Statistical Association*, 435–448. [MR3338514](#). doi: <https://doi.org/10.1080/01621459.2014.993077>. [1065](#)

- Hans, C., Dobra, A., and West, M. (2007a). “Shotgun stochastic search in regression with many predictors.” *Journal of the American Statistical Association*, 102: 507–516. MR2370849. doi: <https://doi.org/10.1198/016214507000000121>. 1060, 1066, 1067
- Hans, C., Wang, Q., Dobra, A., and West, M. (2007b). “SSS: High-dimensional Bayesian regression model search.” *Bulletin of the International Society for Bayesian Analysis*, 8–9. 1060, 1066, 1067
- Irie, K. and West, M. (2019). “Bayesian emulation for multi-step optimization in decision problems.” *Bayesian Analysis*, 14: 137–160. MR3910041. doi: <https://doi.org/10.1214/18-BA1105>. 1061
- Jiang, W. and Tanner, M. A. (2008). “Gibbs posterior for variable selection in high-dimensional classification and data mining.” *The Annals of Statistics*, 5: 2207–2231. MR2458185. doi: <https://doi.org/10.1214/07-AOS547>. 1064
- Jones, B., Dobra, A., Carvalho, C. M., Hans, C., Carter, C., and West, M. (2005). “Experiments in stochastic computation for high-dimensional graphical models.” *Statistical Science*, 388–400. MR2210226. doi: <https://doi.org/10.1214/088342305000000304>. 1066, 1067
- Kapetanios, G., Mitchell, J., Price, S., and Fawcett, N. (2015). “Generalized density forecast combinations.” *Journal of Econometrics*, 188: 150–165. MR3371665. doi: <https://doi.org/10.1016/j.jeconom.2015.02.047>. 1060
- Koop, G. and Korobilis, D. (2013). “Large time-varying parameter VARs.” *Journal of Econometrics*, 177: 185–198. MR3118555. doi: <https://doi.org/10.1016/j.jeconom.2013.04.007>. 1063, 1070
- Lavine, I., Lindon, M., and West, M. (2020). “Supplementary Material of “Adaptive variable selection for sequential prediction in multivariate dynamic models”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/20-BA1245SUPP>. 1062
- Lopes, H. F., McCulloch, R. E., and Tsay, R. S. (2018). “Parsimony inducing priors for large scale state-space models.” Technical report, Booth School of Business, University of Chicago. 1061
- McAlinn, K., Aastveit, K. A., and West, M. (2018). “Bayesian predictive synthesis—discussion of: Using stacking to average Bayesian predictive distributions, by Y. Yao *et al.*” *Bayesian Analysis*, 13: 971–973. MR3853125. doi: <https://doi.org/10.1214/17-BA1091>. 1060, 1079
- McAlinn, K. and West, M. (2019). “Dynamic Bayesian predictive synthesis in time series forecasting.” *Journal of Econometrics*, 210: 155–169. MR3944768. doi: <https://doi.org/10.1016/j.jeconom.2018.11.010>. 1060, 1071, 1079

- McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2020). “Multivariate Bayesian predictive synthesis in macroeconomic forecasting.” *Journal of the American Statistical Association*, 115: 1092–1110 . Published online: Oct 9 2019. [MR4143452](#). doi: <https://doi.org/10.1080/01621459.2019.1660171>. 1060, 1071, 1079
- Nakajima, J. and West, M. (2013a). “Bayesian analysis of latent threshold dynamic models.” *Journal of Business and Economic Statistics*, 31: 151–164. [MR3055329](#). doi: <https://doi.org/10.1080/07350015.2012.747847>. 1060, 1061, 1069, 1071
- Nakajima, J. and West, M. (2013b). “Bayesian dynamic factor models: Latent threshold approach.” *Journal of Financial Econometrics*, 11: 116–153. 1061, 1062
- Nakajima, J. and West, M. (2015). “Dynamic network signal processing using latent threshold models.” *Digital Signal Processing*, 47: 6–15. [MR3425313](#). doi: <https://doi.org/10.1016/j.dsp.2015.04.008>. 1061, 1062
- Nakajima, J. and West, M. (2017). “Dynamics and sparsity in latent threshold factor models: A study in multivariate EEG signal processing.” *Brazilian Journal of Probability and Statistics*, 31: 701–731. [MR3738175](#). doi: <https://doi.org/10.1214/17-BJPS364>. 1061
- Prado, R. and West, M. (2010). *Time Series: Modeling, Computation & Inference*. Chapman & Hall/CRC Press. [MR2519461](#). 1061, 1062
- Primiceri, G. E. (2005). “Time varying structural vector autoregressions and monetary policy.” *Review of Economic Studies*, 72: 821–852. [MR2148143](#). doi: <https://doi.org/10.1111/j.1467-937X.2005.00353.x>. 1061, 1071
- Raftery, A. E., Kárný, M., and Ettler, P. (2010). “Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill.” *Technometrics*, 52: 52–66. [MR2654989](#). doi: <https://doi.org/10.1198/TECH.2009.08104>. 1063, 1070
- Scott, J. G. and Carvalho, C. M. (2008). “Feature-inclusion stochastic search for Gaussian graphical models.” *Journal of Computational and Graphical Statistics*, 17: 790–808. [MR2649067](#). doi: <https://doi.org/10.1198/106186008X382683>. 1066, 1067
- Shirota, S., Omori, Y., Lopes, H. F., and Piao, H. (2017). “Cholesky realized stochastic volatility model.” *Econometrics and Statistics*, 3: 34–59. 1061
- Wang, H. (2015). “Scaling it up: Stochastic search structure learning in graphical models.” *Bayesian Analysis*, 10: 351–377. [MR3420886](#). doi: <https://doi.org/10.1214/14-BA916>. 1066, 1067
- West, M. (1986). “Bayesian model monitoring.” *Journal of the Royal Statistical Society (Series B: Statistical Methodology)*, 48: 70–78. [MR0848052](#). 1064
- West, M. (2020). “Bayesian forecasting of multivariate time series: Scalability, structure uncertainty and decisions (with discussion).” *Annals of the Institute of Statistical Mathematics*, 72: 1–44. Published online Dec 9 2019. [MR4052649](#). doi: <https://doi.org/10.1007/s10463-019-00744-0>. 1059, 1060, 1067

- West, M. and Harrison, P. J. (1986). “Monitoring and adaptation in Bayesian forecasting models.” *Journal of the American Statistical Association*, 81: 741–750. [1064](#)
- West, M. and Harrison, P. J. (1989a). *Bayesian Forecasting & Dynamic Models*. Springer, 1st edition. [1063](#)
- West, M. and Harrison, P. J. (1989b). “Subjective intervention in formal models.” *Journal of Forecasting*, 8: 33–53. [1064](#)
- West, M. and Harrison, P. J. (1997). *Bayesian Forecasting & Dynamic Models*. Springer, 2nd edition. [MR1482232](#). [1061](#)
- Xie, M. (2012). “Discount-weighted Bayesian model averaging for portfolio decisions in matrix variate dynamic linear models.” Unpublished undergraduate thesis, Department of Statistical Science, Duke University. [1063](#)
- Yao, Y., Vehtari, A., Simpson, D., and Gelman, A. (2018). “Using stacking to average Bayesian predictive distributions.” *Bayesian Analysis*, 3: 917–1003. [MR3853125](#). doi: <https://doi.org/10.1214/17-BA1091>. [1060](#)
- Zhang, T. (2006a). “From ϵ -entropy to KL-entropy: Analysis of minimum information complexity density estimation.” *The Annals of Statistics*, 5: 2180–2210. [MR2291497](#). doi: <https://doi.org/10.1214/009053606000000704>. [1064](#)
- Zhang, T. (2006b). “Information theoretic upper and lower bounds for statistical estimation.” *IEEE Transactions on Information Theory*, 4: 1307–1321. [MR2241190](#). doi: <https://doi.org/10.1109/TIT.2005.864439>. [1064](#)
- Zhao, Z. Y., Xie, M., and West, M. (2016). “Dynamic dependence networks: Financial time series forecasting and portfolio decisions (with discussion).” *Applied Stochastic Models in Business and Industry*, 32: 311–339. [MR3518020](#). doi: <https://doi.org/10.1002/asmb.2161>. [1061](#), [1062](#), [1063](#), [1070](#)
- Zhou, X., Nakajima, J., and West, M. (2014). “Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models.” *International Journal of Forecasting*, 963–980. [1061](#)

Acknowledgments

The research reported here was developed while Isaac Lavine and Michael Lindon were PhD students in Statistical Science at Duke University. Aspects of the research developed in the working group on *Bayesian Analysis & Decisions* of the 2016–2017 program on *Optimization* of the Statistical and Applied Mathematical Sciences Institute (SAMSI). Isaac Lavine and Michael Lindon were the co-recipients of the 2016–17 BEST Award for Student Research at Duke University, and acknowledge the BEST Foundation for partial financial support in early stages of the research reported here.