Bayesian model selection approaches to MIDAS regression.

Abel Rodriguez

University of California, Santa Cruz, USA

Gavino Puggioni

University of North Carolina, Chapel Hill, USA

Summary. We describe Bayesian models for economic and financial time series that use regressors sampled at finer frequencies than the outcome of interest. The models are developed within the framework of dynamic linear models, which provide a great level of flexibility and direct interpretation of results. The problem of collinearity of intraperiod observations is solved using model selection and model averaging approaches which, within a Bayesian framework, automatically adjust for multiple comparisons and allows us to accurately account for all uncertainty when predicting future observations. We also introduce novel formulations for the prior distribution on model space that allow us to include additional information in a flexible manner. We illustrate our approach by predicting the gross domestic product of United Stated using the term structure of interest rates.

Keywords: Mixed Frequency Data, Model Selection, Model Averaging, Interest Rates, Gross National Product

1. Introduction

Some classes of economic data are typically available only at low frequencies; examples include gross domestic product (GDP), gross national product (GNP), consumption index, inflation rate and other macroeconomic indicators. However, many of these series can be effectively explained by predictors that are observed at higher frequencies. For instance, GNP, which is sampled quarterly, has been shown to be explained by the behavior of interest rates, which is available on a monthly basis (see Estrella & Hardouvelis (1991) and references therein), or even daily (Ghysels & Wright, 2006). A similar situation arises with weekly and monthly stock market volatility, which can be explained by high frequency market data, often tick-by-tick (Ghysels et al., 2006).

Simultaneously using all low frequency frequency predictors available is typically unfeasible, as collinearity might lead to overfitting. Traditional methods solve the frequency disparity by aggregating the variables at the highest frequency available. In

many cases this implies a loss of valuable information that might potentially enhance the predictive power of the model.

Borrowing from existing literature about distributed lag models, Ghysels et al. (2002) and Ghysels et al. (2007) recently proposed a class of models, called MIDAS (Mixed Data Sampling) regression, which allows dependent and independent variables to be sampled at heterogeneous frequencies. In this approach lagged explanatory variables are weighted by coefficients that come from deterministic specifications such as Almon lags and Beta polynomials. As a consequence the problem of parameter proliferation is parsimoniously solved without sacrificing a certain degree of flexibility.

MIDAS regression is particularly attractive when the number of regressors is large and fast computation is required. For instance, financial data are often available at 5-minutes intervals and other methods would be cumbersome to apply in such cases. However, in many circumstances a deterministic and time-static structure as the one used in MIDAS, is not flexible enough. On one hand, enforcing a deterministic structure on the coefficients of the model might potentially induce artifacts in the analysis. For example, periodic patterns, with relevant regressors separated by a fixed number of periods, cannot be captured by a typical MIDAS specifications. On the other hand, MIDAS regression is static in nature, but the influence of the predictors can easily change in time as other (unobserved) concomitant variables evolve.

The main purpose of this paper is to show how Bayesian model selection and model averaging can be used to model mixed frequency data while preserving a parsimonious model structure and automatically adjust for multiple comparisons. Our approach is developed within the context of dynamic linear models (DLMs) (West & Harrison, 1998), which have become a standard tool in Bayesian time series analysis, especially in finance and econometrics (Quintana & West, 1987; Quintana, 1992; Putnam & Quintana, 1995; Aguilar & West, 2000; Carvalho & West, 2007). DLMs are an extremely flexible class of model that allow us to easily combine dynamic regression with trends, periodicities and autoregressive components. Specifically, this paper discusses how model selection can be introduced in dynamic regression models and how prior distributions on the space of models can be used to enforce "soft" restrictions on the paramters, similar to those implemented in MIDAS regression.

Model comparison approaches that use frequentist hypothesis tests have been previously used in the context of mixed frequency data to deal with the problem of multicollinearity (see for example Birchenhall et al. (1999) and Campos et al. (2003)). The Bayesian approach we advocate in this paper has a number of theoretical and practical advantages. First, model selection (in contrast to the structural constraints used in MIDAS models) can provide interesting insights into the economic process being modeled. Second, Bayesian methods for model comparison automatically adjust for multiple comparisons and allow us to efficiently explore the full model space. Third, model averaging (as opposed to model selection) is optimal for prediction purposes,

in the sense of minimizing predictive squared error loss. Finally, predictions obtained by Bayesian model averaging (in contrast to frequentist prediction) correctly account for model and estimation uncertainty.

The price to pay for these advantages is slightly more complicated and time-consuming computational algorithms. Since the posterior distributions arising from the models discussed here are too complex to be analytically tractable, we resort to simulation algorithms. In particular, we focus on Markov Chain Monte Carlo (MCMC) samplers (see Smith & Roberts (1993) and Robert & Casella (2004) for an introduction). As the name indicates, MCMC algorithms construct a Markov chain whose stationary distribution corresponds to the posterior distribution of interest. This is done by iteratively sampling from the full conditional distribution of blocks of parameters. The result is a sequence of *dependent* samples from the posterior distribution of the model; inference on functionals of parameters can be easily performed using ergodic averages. This paper shows that, in the context of dynamic linear model, it is possible to implement MCMC algorithms that move across models of varying dimension without without resorting to cumbersome transdimensional algorithms like reversible jump MCMC samplers (Green, 1995).

As an illustration of our methodology, we discuss models that explain quarterly GNP growth as a function of the monthly spreads in interest rates, calculated as the difference between 3 months and 10 years yields of treasury bonds. The models also include trends and autoregressive components, providing a powerful prediction tool as well as interesting insight into monetary policy.

The paper is organized as follows: In Section 2 we describe the basic MIDAS model of Ghysels et al. (2002) and provide a more detailed motivation for our approach. In Section 3 we review Bayesian model selection and model averaging. In Section 4 we review dynamic linear models and describe how model selection can be performed in this setting. Section 5 discusses the role of prior distributions in model selection and presents priors that can be used to mimic the weighted structure behind MIDAS models while providing additional flexibility. In Section 6 we develop a model to predict GNP growth that incorporates the slope of the term structure as predictors. Finally, Section 7 provides a brief discussion and some future directions for research.

2. Motivation: Mixed frequency data and the MIDAS regression model

As an initial motivation, consider a multiple regression model

$$y_t = \beta_0 + \sum_{j=1}^{p} \beta_j x_{t-j/3} + \epsilon_t \qquad \epsilon_t \sim \mathsf{N}(0, \sigma^2) \tag{1}$$

where y_t and x_t represent, respectively, the rate of GNP growth and the slope of the term structure of the interest rate at period $t = 1, \ldots, T$. The fractional notation in

the subindexes simply recognizes the fact that interest rates are reported monthly, while GNP is reported quarterly.

Since the term structure of interest rates tends to evolve slowly in time, the set of predictors in (1) tends to be highly correlated, possibly leading to overfitting. In MIDAS regression, this problem is solved by imposing a parametric structure on the coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. For example, Beta polynomials can be used, in which case,

$$\beta_i = b\frac{w(i/p, \lambda_1, \lambda_2)}{\sum_{s=1}^{p} w(i/p, \lambda_1, \lambda_2)} \qquad w_i(z, \lambda_1, \lambda_2) = z^{\lambda_1 - 1}(1-z)^{\lambda_2 - 1} \qquad \lambda_1, \lambda_2 \geq 1 \qquad (2)$$

Note that $\sum_{j=1}^{p} \beta_j = b$. Therefore, we can think of the weights as spreading the predictive power of the interest rates across the different lags of the term structure. This is done in such a way that coefficients change smoothly with the lag. Since the parametric form depends only on three parameters (b, λ_1 and λ_2), parsimony is preserved no matter how many lags are included. However, the use of a parametric representation for the coefficients of the model also implies strong restrictions on what lags of the interest rate significantly affect GDP. In particular, the beta polynomials in (2) imply that the distribution of weights is unimodal, with the weights growing until they reach a maximum around $z = (\lambda_1 - 1)/(\lambda_1 + \lambda_2 - 2)$ and then decay smoothly. If there is a periodic pattern in the effect of interest rates (for example, if it is the slope of the term structure at the end of each quarter that affects GDP), Beta polynomials are ill suited to capture it. Choosing more flexible parametric forms (including mixtures of Beta polynomials) can help alleviate the problem, but might unnecessarily increase the number of parameters.

The main insight behind MIDAS models is that if a given lag of the term structure is significant to explain GDP, then lags that are close to it should also be significant. In other words, only a small number of adjacent lags are necessary for prediction, and the colinearity problem can be fixed by forcing the coefficients of the other lags to be close to zero. This motivates an alternative approach to mixed frequency regression, in which variable selection approaches are used to identify important lags. Rather than constraining the value of the coefficients themselves as is done in MIDAS, we can use prior distributions to favor clustered patterns in the coefficients to be included in the model.

Although conceptually straightforward, the use of variable selection in the context of mixed frequency data implies a number of practical challenges. First of all, the number of models to be compared grows exponentially with the number of lags under consideration. This means that any procedure needs to control family-wise error rate, which will be in general larger than the error rate associated with every individual pairwise test. It also means that we need to design algorithms that efficiently explore the model space if the models are to be implemented in settings where the number of

lags is large. In second place, predictions generated by the model need to account for the uncertainty involved in the selection of the set of relevant lags in order to avoid underestimation of prediction errors. Finally, it is important that the methodology developed can be incorporated into more complex hierarchical models; in particular, it is important that we can allow the coefficients relating the term structure of interest rates to GDP to evolve smoothly in time. For example, in the analysis of historical data over long periods of time, institutional and other changes might imply that dependence among variables structure will change in time. In the following sections, we develop models that tackle these three issues.

3. Bayesian model selection and prediction

Since different lags of the term structure provide essentially equivalent information, a natural alternative to a strong parametric form on the coefficients of the model is to select just a subset of the variables to be included in the model. In that case, we are left with a collection of models $\mathcal{M} = \{M_1, \ldots, M_K\}$, corresponding to each one of the possible $K = 2^p$ possible subset of regressors. It is usually convenient to represent these models using indicator variables $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)$ such that $\xi_i = 1$ if variable i in included in the model, and 0 otherwise. Model M_k is associated with a likelihood $p_k(\mathbf{y}|\boldsymbol{\beta}_k)$ and a prior distribution $p(\boldsymbol{\beta}_k)$, where $\boldsymbol{\beta}_k$ is the subset of coefficients $\boldsymbol{\beta}$ corresponding to the variables included in M_k.

Given prior probabilities $\mathsf{Pr}(M_1), \ldots, \mathsf{Pr}(M_K)$ for each one of the models in $\mathcal{M}$, we can use Bayes theorem to obtain updated posterior probabilities,

$$\mathsf{Pr}(M_k|\mathbf{y}) = \frac{m_k(\mathbf{y})\mathsf{Pr}(M_k)}{\sum_{s=1}^{K} m_s(\mathbf{y})\mathsf{Pr}(M_s)}$$

where

$$m_k(\mathbf{y}) = \int p_k(\mathbf{y}|\boldsymbol{\beta}_k)p_k(\boldsymbol{\beta}_k)d\boldsymbol{\beta}_k \tag{3}$$

Posterior model probabilities measure the relative strength of evidence in favor of each model. When a single model needs to be picked (for example, if we are interested in which lags of the term structure affect the GNP) the optimal choice under a 0-1 loss function is the highest posterior probability model, M_{k^*}, where

$$k^* = \arg\max_k \{\mathsf{Pr}(M_k|\mathbf{y})\}$$

Note that choosing a specific model might be particularly relevant from the point of view of policy design. For example, choosing the relevant lags of the term structure

in the previous model allows us to determine how long it will take for FED intervention to affect the real economy, or if it will have any effect at all.

Bayesian procedures for model selection like the one just described automatically adjust for multiple comparisons, as long as a hierarchical structure is used to specify $\{\mathsf{Pr}(M_k)\}_{k=1}^K$ (Jefferys & Berger, 1992; Scott & Berger, 2003); some examples of hierarchical specifications relevant for mixed frequency data will be discussed in Section 5.1. This means that no ad-hoc adjustment (Bonferroni, FDR, etc) is necessary to control family-wise error rates.

Maximum probability models can be very helpful in understanding the underlying economic process generating he data. However, since they are derived from a 0-1 loss function, maximum probability models are not necessarily optimal for prediction (Draper, 1995; Hoeting et al., 1999). Instead, the optimal prediction rule for a new observation y^* under squared prediction error loss function is obtained as a weighted average the predictions provided by each model under consideration,

$$\hat{y}^* = \sum_{k=1}^{K} \mathsf{E}(y^*|M_k, \mathbf{y})\mathsf{Pr}(M_k|\mathbf{y})$$

Even if for any reason model average prediction is not viable and a single model is to be used for forecasting, the median probability model is typically a better option than the maximum probability model (Barbieri & Berger, 2004). The median probability model is defined as the model that includes all variables whose marginal inclusion probability q_i is greater than 1/2, where $q_i = \sum_{M_k \in \mathcal{M}_i} \mathsf{Pr}(M_k|\mathbf{y})$ and $\mathcal{M}_i$ is the subset of $\mathcal{M}$ containing all models that include variable i. A note of caution on the use of marginal inclusion probabilities is in order. Unlike, frequentist approaches where p-values and confidence bands are equivalent, Bayesian posterior credible intervals and posterior model probabilities might yield contradicting results (see Jefferys (1990) for an example). This might happen when data is unusual under both hypothesis, but happens to be more unusual under the complex model, and can be seen as a consequence of Lindley's paradox (Lindley, 1957; Shafer, 1982). Our application in Section 6 provides an enlightening example.

Computation of Bayes factors can be a complex task, as it requires the calculation of the multidimensional integrals in (3). When these integrals are available in closed form (as in Gaussian linear models) and the number of model is moderate (typically, $p \leq 20$) we can design simulation algorithms that directly sample over the full space $\mathcal{M}$ at each iteration and, conditionally on the model chosen, generate samples from the model parameters. A simple description of the algorithm is,

(a) Sample the model index $k \in \{1, \ldots, K\}$ according to their posterior probabilities $\{\mathsf{Pr}(M_k|\mathbf{y})\}_{k=1}^K$

(b) Conditional on k, sample $\boldsymbol{\beta}$ by setting the coefficients corresponding to variables not included in k to zero and by sampling the rest (contained in $\boldsymbol{\beta}_k$) from

$$p_k(\boldsymbol{\beta}_k|\mathbf{y}) \propto p_k(\mathbf{y}|\boldsymbol{\beta}_k)p_k(\boldsymbol{\beta}_k)$$

If (3) is available in closed form but the parameter space is too large, step (a) above can be replaced by a Metropolis step that allows for *local* moves on the space of models (see George & McCulloch (1997) for an excellent review). It is typical to use a symmetric proposal distribution

$$q(\boldsymbol{\xi}^{(p)}|\boldsymbol{\xi}^{(c)}) = q_d \text{ if } \sum_{i=1}^{p} |\xi_i^{(p)} - \xi_i^{(c)}| = d$$

where $\boldsymbol{\xi}^{(c)}$ and $\boldsymbol{\xi}^{(p)}$ correspond to the current and the proposed model respectively, and q_d is the probability of proposing a change that implies adding/deleting d variables to the model (see George & McCulloch (1997) for details). This is the approach we adopt in the following sections.

The most complex case when (3) is not available in closed form requires the use of transdimensional MCMC algorithms that allow us to move across spaces of different dimension, for example, the reversible jump algorithm (Green, 1995). In the following section we show how these ideas just described in the context of linear regression can be extended to dynamic linear models.

4. Dynamic linear models and dynamic regression

Using the notation in West & Harrison (1998), let y_t for $t = 1, \ldots, T$ be a univariate time series. Given the quadruplets

$$\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}_{t=1}^{T}$$

where $\mathbf{F}_t$ is a known n-dimensional row vector, $\mathbf{G}_t$ is a $n \times n$ matrix, V_t is a positive scalar, and $\mathbf{W}_t$ is a $n \times n$ symmetric positive-semidefinite matrix, a dynamic linear model (DLM) is defined by two groups of equations. The observational equations take the form

$$y_t = \mathbf{F}_t\boldsymbol{\theta}_t + \nu_t \qquad \boldsymbol{\nu}_t \sim \mathsf{N}(\mathbf{0}, V_t) \tag{4}$$

and describe how observations are linked to the sequence of (unobserved) state vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T$, while the state equations

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim \mathsf{N}(\mathbf{0}, \mathbf{W}_t) \tag{5}$$

describe how the state vectors evolve in time. Both noise processes are assumed to be independent, and an initial condition $\theta_0 \sim \mathsf{N}(\mathbf{m}_0, \mathbf{C}_0)$ is assumed. DLMs are extremely flexible and are able to accommodate a number of different behaviors. For example, a simple alternative to capture trends is to use a local constant polynomial, in which case $n = 1$, $G_t = 1$ and $F_t = 1$, i.e., the evolution equation describes a simple unidimensional random walk. Autoregressive components of order q can be accommodated by taking

$$\mathbf{F}_t' = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{G}_t = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{q-1} & \phi_q \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & & 0 \cdots & 1 & 0 \end{pmatrix} \quad \mathbf{W}_t = \begin{pmatrix} U & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

where ϕ_i stands for the autoregressive coefficient associated with the i-th lag. Finally if a set p predictors $x_{1t}, \ldots, x_{pt}$ are available at time t a dynamic regression can be accommodated by taking $\mathbf{F}_t = (x_{1t}, \ldots, x_{pt})$ and $\mathbf{G}_t = \mathbf{I}_p$, where $\mathbf{I}_p$ denotes the identity matrix of dimension p. Note that, taking $\mathbf{W}_t = \mathbf{0}$ leads directly to the static multiple regression model we used as motivation in Section 2; hence, the presence of evolution noise allows the regression coefficients to adapt in time generating a model that is a linear regression only *locally*. See West & Harrison (1998) for a complete description.

When the quadruplets $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}_{t=1}^T$ are assumed to be known, the posterior and marginal predictive distributions can be obtained in closed form. However, this is rarely a realistic setting, as structural terms like autocorrelation coefficients or evolution are rarely known a priori. A simple mechanism to model the evolution variance $\mathbf{W}_t$ is information discounting (see West & Harrison (1998), Chapters 2.4 and 6.4). In the simplest case of a single discount factor $0 < \delta \leq 1$, the evolution variance is specified as $\mathbf{W}_t = (1-\delta)\mathsf{Var}(\boldsymbol{\theta}_{t-1}|y_{t-1}, \ldots, y_1)/\delta$. Note that $\delta = 1$ implies $\mathbf{W}_t = 0$, and no evolution is allowed. In most application, $\delta \in [0.8, 1]$.

A fully hierarchical specification of the DLM typically requires the use of MCMC methods for inference. In this case, the state parameters are sampled conditionally on the quadruplets using a Forward-Filtering-Backward-Sampling (FFBS) algorithm (Carter & Kohn, 1994; Frühwirth-Schnatter, 1994), while the structural parameters are in turn sampled conditionally on the current imputed values for the state parameters. In the forward step of the FFBS algorithm, the posterior distributions are sequentially updated by exploiting conjugacy. Starting with $p(\boldsymbol{\theta}_0)$, we can obtain $p(\boldsymbol{\theta}_t|y_t, \ldots, y_1)$ and $p(y_t|y_{t-1}, \ldots, y_1)$ from $p(\boldsymbol{\theta}_{t-1}|y_{t-1}, \ldots, y_1)$. In the backward step, the conditional distributions $p(\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_t, y_T, \ldots, y_1)$ are constructed and used to sequentially sample from the joint distribution $p(\boldsymbol{\theta}_T, \ldots, \boldsymbol{\theta}_1|y_T, \ldots, y_1)$. Incidentally,

note that the FFBS algorithm also provides an efficient mechanism to integrate out the unknown parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T$ and compute $p(y_T, \ldots, y_1)$ since

$$p(y_T, \ldots, y_1) = \prod_{t=1}^{T} p(y_t | y_{t-1}, \ldots, y_1).$$

In the context of variable selection in dynamic regression, this allows us to implement the simple MCMC schemes described at the end of Section 2 and avoid transdimensional algorithms.

5. Prior distributions for mixed frequency data regressions

5.1. Priors on model space

When the model space is finite, uniform distributions are frequently used as a default option. However, uniform priors are often restrictive and, not being hierarchical, do not automatically adjust for multiple comparisons. In this section, we discuss three types of hierarchical priors on model space, Bernoulli-type priors, order-inducing priors and truncated geometric priors. Bernoulli-type priors are standard in Bayesian model selection, especially in stochastic search variable selection schemes (George & McCulloch, 1993). Given $0 < \eta < 1$, a prior in this class is obtained by setting $p(\xi_i) = \eta^{\xi_i}(1-\eta)^{1-\xi_i}$ independent for all i. This implies that

$$p(M_k) = \eta^{|M_k|}(1-\eta)^{p-|M_k|}$$

where $|M_k| = \xi^{(k)'}\xi^{(k)}$ is the number of variables in model M_k (dimension of the model). In other words, $|M_k|$ is the number of regressors that is included in M_k. The hyperparameter η, representing the prior probability that any single variable is included in the model, controls the sparsity induced by the prior. Therefore, values of η close to zero favor models that include few regressors, and viceversa. Since sparsity depends so crucially on η, we employ a Beta hyperprior, $\eta \sim \mathsf{Be}(a_0, b_0)$, where $a_0/(a_0+b_0)$ corresponds to the proportion of the variables that are expected to be significant a priori.

Bernoulli priors are easy to elicit and understand, but treat all variables equally. However, one of the most interesting insights from MIDAS models is that high frequency predictors are more likely to be relevant if the adjacent ones are. Positional information can be introduced in the prior using Markov-switching structures. In particular, consider a prior such that $p(\xi_1) = \eta^{\xi_1}(1-\eta)^{1-\xi_1}$ and for $i > 1$

$$p(\xi_i | \xi_{i-1}) = \begin{cases} (1-\pi_0)^{\xi_i}\pi_0^{1-\xi_i} & \xi_{i-1} = 0 \\ \pi_1^{\xi_i}(1-\pi_1)^{1-\xi_i} & \xi_{i-1} = 1 \end{cases}$$

In words, the most recent lag of the predictor variable is included in the regression with probability η, and from then on lag i is included in the model with probability π_1 if lag $i-1$ is also included, and with probability $1-\pi_0$ otherwise. Therefore, the sequence $\xi_1, \ldots, \xi_p$ defining the model follows a Markov chain with stationary distribution,

$$e_0 = \frac{1-\pi_0}{2-\pi_0-\pi_1} \qquad\qquad e_1 = 1 - e_0$$

The values of π_0 and π_1 close to 1 imply high persistence rates, therefore favoring models that include consecutive lags. Again, since the properties of the prior crucially depend on π_0 and π_1, we estimate them using hyperpriors $\pi_0 \sim \mathsf{Be}(a_1, b_1)$ and $\pi_1 \sim \mathsf{Be}(a_2, b_2)$ and choose $a_1/(a_1+b_1) > 1/2$ and $a_2/(a_2+b_2) > 1/2$. Rather than modeling η separately, we set $\eta = e_1 = (1-\pi_1)/(2-\pi_0-\pi_1)$. Therefore, η can also be interpreted as the marginal probability that any variable is included in the model, just as in the Bernoulli prior.

Note that the Markov switching prior includes the Bernoulli prior as a special case when $\pi_1 = 1-\pi_0$. In addition, it allows us to induce ordering in the lags being included in the model, in a similar way to the Almond lags and Beta weights specifications used for MIDAS regression. However, the constraints introduced by our model, while also preserving parsimony, are much less restrictive. On one hand, it allows any single regression coefficients to take any value independently of other coefficients, while preserving parsimony by excluding redundant variables from the model.

Truncated geometric distributions are an alternative to order inducing priors that are appropriate when interest focuses in models where a regressor is included *only* if all its predecessors are also included. Principal component regression is an excellent example where this type of specification is natural. Specifically, a truncated geometric prior takes the form

$$\mathsf{Pr}(M_k) = \frac{\nu(1-\nu)^k}{1-(1-\nu)^{p+1}}$$

where M_k corresponds to the model such that $\xi_i = 1$ for $i \leq k$ and $\xi_i = 0$ otherwise. Note that the support of this prior is relatively small, as only $p+1$ models have non-zero probabilities. As before, ν controls the complexity of the models; each additional regressor decreases the probability of the model by a factor of $(1-\nu)$. Also, taking the limit $\nu \to 0$ yields a uniform prior on model space. In the sequel, we estimate the hyperparameter ν by using a Beta hyperprior, $\nu \sim \mathsf{Be}(a_\nu, b_\nu)$. Note that truncated geometric priors are reminiscent of MIDAS regression models with step functions that are covered, for instance, in (Ghysels et al., 2007).

5.2. Priors on parameter space

Although Bayesian model comparison methods that use proper priors are consistent under mild regularity condition, improper priors can be dangerous in Bayesian model selection and model average (Kass & Raftery, 1995). This is unlike estimation problems, where reference or other objective but improper priors are typically adequate.

In nested models like the mixed data regression model we have been discussing, improper priors are typically adequate *only* for parameters that are common to *all* models under consideration (a formal argument for this practice is given in Berger et al. (1998)). When used for parameters under selection, standard improper priors (which are determined only up to a multiplicative constant) typically lead to indeterminate posterior probabilities. The same happens for proper but diffuse priors (Natarajan & McCulloch, 1998), where the problem can be hard to capture as lack of mixing in the sampler might be hard to detect.

There is ample literature in objective priors for model selection in general (Berger & Pericchi, 2001), and linear regression (Berger & Pericchi, 1996b,a) and dynamic linear models (Rodriguez & Pericchi, 2000) in particular. However these objective priors are hard to adapt to multilevel hierarchical specifications like the ones we describe in the following section. Instead of trying to elicit objective priors, we note that in mixed frequency regression, enough prior information is typically available to elicit reasonable prior distributions for the parameters in the model. In the context of our motivating example, expectations on the maximum and minimum value for GNP and interest rates spread can be used to decide on reasonable ranges for the priors on the regression coefficients. For example, it is reasonable to assume that annualized GNP growth rates over 8% (in absolute value) are highly unlikely in the US; similarly, we do not expect to see spreads over 10% in either direction. Therefore, we would use a prior distribution for each coefficient β_i that places most of its probability on the interval $(-0.8, 0.8)$. We use this type of specification along with sensitivity analysis in our empirical studies.

6. Understanding the relationship between GNP growth and interest rates in the US

6.1. The data

In this section we construct a model for the quarterly growth of GNP between January 1955 and April 2005. The dataset is seasonally adjusted. Netting in the initial values, a total of $T = 201$ observations are available for the dependent variable

$$y_t = 400 \log\left(\frac{GNP_t}{GNP_{t-1}}\right),$$

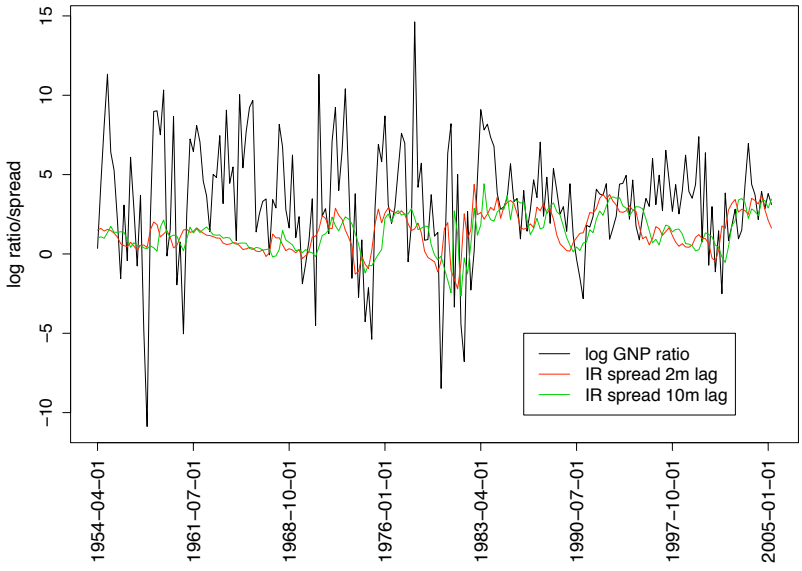

Fig. 1. Annualized GNP growth rate and lagged interest rate spreads in the US economy between 1954 and 2005

where GNP_t denotes the value of GNP at quarter t (hence, y_1 corresponds to the annualized growth rate during the first quarter of 1955). Among other components, the model includes information on the monthly behavior of interest rates during the previous 2 years as an explanatory variable in the model. This information is summarized through the slope of the term structure, namely the spread between the long term (10 year) and the short term (3 month) interest rates. The GNP series together with the lagged values of the interest rates spread are depicted in Figure 1.

6.2. Model formulation

6.2.1. Bayesian mixed frequency regression models for macroeconomic data

The model is built by the superimposition of three distinct blocks: a local polynomial trend μ_t capturing the average level of the series, a latent autoregressive process z_t capturing the inertia in the behavior of the GNP, and a dynamic regression component

with coefficients $\boldsymbol{\beta}_t = (\beta_{t1}, \ldots, \beta_{tp})'$ capturing the influence of interest rates levels during the previous year. Specifically, the observational equation for the model takes the form

$$y_t = z_t + \mu_t + \mathbf{X}_t^k \boldsymbol{\beta}_t^k + \nu_t \qquad \nu_t \sim \mathsf{N}(0, V) \tag{6}$$

For the three components model, we structure the regression vector as

$$\mathbf{X}_t = (x_{t-2/3}, x_{t-1}, \ldots, x_{t-23/3})$$

where x_l denotes the spread on quarter l. Again, we use fractional subindexes to denote the infraperiods on which the predictor is observed (in this case, months). We include a total of $p = 22$ months of information on interest rates, starting with the first month of the quarter and going back two years, but other schemes are possible. The superscript $k \in \{1, \ldots, 2^{22}\}$ denotes the regression model, corresponding to one of the subsets of the p variables being entertained. The inclusion of an autoregressive term follows standard practice in macroeconomics modeling (for example, see Shen (1996), Clements & Galvao (2008) and references therein). However, as pointed out by Ghysels et al. (2002) and Clements & Galvao (2008), standard AR-MIDAS models can exhibit a seasonal response of y to some of the x's, irrespective of whether the x's displays a seasonal pattern. Clements & Galvao (2008) solve this by introducing the autoregressive dynamics as a common factor, in the style of Hendry & Mizon (1978). Including the autoregressive component as a latent process, as we do in this paper, is an alternative route to circumvent this issue.

The parameters of the model are allowed to evolve according to the state equations

$$z_t = \phi z_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathsf{N}(0, U) \tag{7}$$

$$\begin{pmatrix} \mu_t \\ \boldsymbol{\beta}_t^k \end{pmatrix} = \begin{pmatrix} \mu_{t-1} \\ \boldsymbol{\beta}_{t-1}^k \end{pmatrix} + \boldsymbol{\varpi}_t \qquad \boldsymbol{\varpi}_t \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma}_t^k) \tag{8}$$

where ϕ is the autoregressive coefficient of the process. Therefore, we assume that the evolution noise for the autoregressive component of the model has a constant (but unknown) variance U. On the other hand, the joint evolution of the variance for the trend and dynamic regression coefficients is specified through two discount factors, δ_μ and δ_β, such that

$$\boldsymbol{\Sigma}_t^k = \boldsymbol{\Delta} \mathsf{Var}(\mu_{t-1}, \boldsymbol{\beta}_{t-1}^k | y_1, \ldots, y_{t-1}) \boldsymbol{\Delta}' \quad \boldsymbol{\Delta} = \mathrm{diag}\left\{ \sqrt{\frac{1-\delta_\mu}{\delta_\mu}}, \sqrt{\frac{1-\delta_\beta}{\delta_\beta}} \mathbf{1}_p' \right\} \tag{9}$$

where $\mathbf{1}_p$ denotes a column vector of dimension p with 1's in all entries (see West & Harrison (1998), Chapters 2.4 and 6.4). Therefore δ_μ controls the information

discounting for the trend, while δ_β controls the level of information discounting for the regression coefficients.

Modeling the evolution of the autoregressive component separately from the trend and regression simplifies prior elicitation and interpretation of the model. For any given model k, we can rewrite (5), (6) and (7) in terms of the DLM formulation in (3) and (4) as

$$y_t = \mathbf{F}_t^k \boldsymbol{\theta}_t^k + \nu_t \qquad \nu_t \sim \mathsf{N}(0, V) \tag{10}$$

$$\boldsymbol{\theta}_t^k = \mathbf{G}^k \boldsymbol{\theta}_{t-1}^k + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim \mathsf{N}(\mathbf{0}, \mathbf{W}_t) \tag{11}$$

where $\boldsymbol{\theta}_t^k = (z_t, \mu_t, \boldsymbol{\beta}_t^k)'$ is the $|M_k| + 2$ vector of state variables at time t for model k, $\mathbf{G}^k = \text{diag}\{\phi, 1, 1, \ldots, 1\}$ is the $(|M_k| + 2) \times (|M_k| + 2)$ evolution matrix for model k, $\boldsymbol{\omega}_t = (\epsilon_t, \boldsymbol{\varpi}_t')'$ is the evolution $|M_k| + 2$ vector of describing the evolution noise on the state variables, $\mathbf{W}_t^k = \text{diag}\{U, \boldsymbol{\Sigma}_t^k\}$ is $(|M_k| + 2) \times (|M_k| + 2)$ the block diagonal evolution covariance matrix for model k at time t, and $\mathbf{F}_t^k = (1, 1, \mathbf{X}_t^k)$ is the $|M_k| + 2$ design vector at time t for model k.

The model is completed by specifying priors on the unknown parameters, for which we exploit subjective information. For the prior on model space, we employ the order-inducing priors described in Section 5.1, and set mildly informative priors $\pi_0 \sim \mathsf{Be}(3, 1)$ and $\pi_1 \sim \mathsf{Be}(3, 1)$. Conditional on being included in the model, the dynamic regression coefficients are assigned independent normal priors at time 0, $\beta_{0i} \sim \mathsf{N}(0, 0.25^2)$ if $\xi_i = 1$. Note that, following the discussion in Section 5.2, the variance has been chosen to provide high support to the interval $(-0.75, 0.75)$. Sensitivity analysis was performed, with values of the standard deviation between 0.1 and 1 yielding similar results. Priors on the variance components are set as conditionally conjugate $U \sim \mathsf{IGam}(2, 2)$ and $V \sim \mathsf{IGam}(10, 0.5)$, which implies that $\mathsf{E}(U) = 1$ and $\mathsf{E}(V) = 0.05$ a priori. This pair of priors favor small observational variances compared to the evolution variance of the autoregressive process. The discount factors δ_μ and δ_β are assigned discrete priors over the set $\{0.85, 0.90, 0.95, 0.99, 0.999\}$ in order to simplify computation. Finally, the initial states for the trend and autoregressive term are assigned conditionally conjugate normal priors $z_0 \sim \mathsf{N}(0, 3^2)$ and $\mu_0 \sim \mathsf{N}(0, 3^2)$, again in line with our expectations about rates of growth.

6.2.2. Bayesian principal component regression for macroeconomic data

We also discuss a similar model as the one described above that replaces the lagged slopes of the term structure by their principal components, ordered by the proportion of the variability they explain. Principal components analysis has been repeatedly used in the literature to deal with multicollinearity problems. Indeed, principal components are orthogonal by construction, and typically a handful of them is enough to explain most of the variability in the predictors.

In this application, the principal components have a very interesting interpretation; for example, the first principal component (explaining 66% of the variability in the data) is approximately the average slope of the term structure during the last two years, while the second principal component (explaining 18% of the variability) contrasts the average variability during the first half of the period with the variability during the second half. Similar interpretations can be attached to the other principal components. A more detailed discussion, along with part of the loadings matrix, can be seen in Appendix B. Instead of an order inducing prior, we employ a truncated geometric prior, which allows us to decide how many of the most relevant principal components should be included in the model. As discussed in Section 5.1, a prior of this type greatly simplifies computation by dramatically reducing the number of models to be considered, but for the same reason it reduces the flexibility of the models. The hyperparameter ν, controlling the sparsity in the model, is given a prior $\nu \sim \mathsf{Be}(1, 9)$

6.3. Results

We employ the MCMC sampler detailed in Appendix A to fit the model described above. The algorithm was run for 40,000 iterations following a 5,000 iteration burn-in period. Examination of diagnostic plots showed adequate mixing and no evidence of lack of convergence. In order to corroborate this observation, we used the Gelman-Rubin convergence test (Gelman & Rubin, 1992), which compares the variability within and between multiple runs of the sampler with overdispersed starting values; we monitored the marginal probabilities of inclusion of each of the 22 lags under consideration, the autocorrelation coefficient and the observational and structural variances as the parameters of interest. In every case, confidence intervals for the convergence statistic R contained the reference value 1, as expected for non-divergent chains.

Figure 2 shows a decomposition of the original GNP series into the three structural components described in the previous section. The decomposition is almost identical for both the regression and the principal component models. It reveals a very subtle trend, with the average growth rate increasing from 1.9% to 1.97% in the 50 years under consideration The posterior mode for the discount factor of the trend is δ_μ is 0.99. Most of the variability in the sample can be explained by the autoregressive process. Our estimate of the autocorrelation coefficient is 0.24 (posterior mean), with a 90% symmetric credible interval (0.08 , 0.36). The variance of the observational noise is negligible, with the posterior expectation of V being around 0.0011. It is interesting to note that the data favors a static regression model. Indeed, the posterior distribution for the discount factor δ_β concentrates on 0.999, leading to almost constant estimates of the regression coefficients.

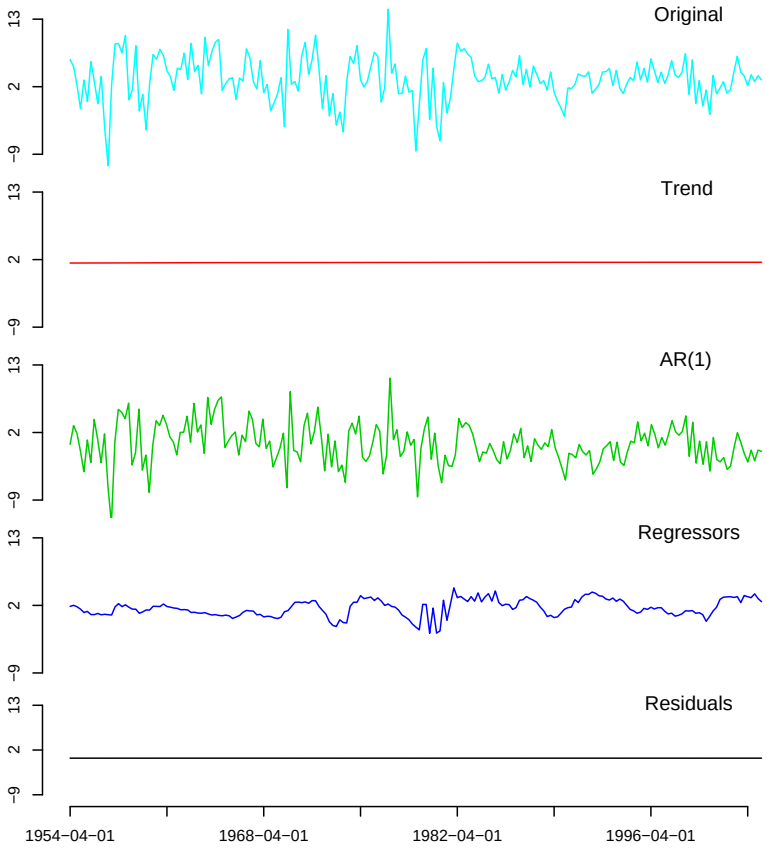

Fig. 2. Decomposition of the US GNP time series into three structural terms. Residuals are shown, reveling no apparent patterns and an extremely good fit.

Table 1. Marginal probability of inclusion for the principal components of the lagged slope of the term structure.

Prin. comp.	1	2	3	4	5	6	7	8	9	10	11
Marginal Pr.	0.98	0.85	0.54	0.48	0.46	0.45	0.45	0.39	0.36	0.34	0.32
Prin. comp.	12	13	14	15	16	17	18	19	20	21	22
Marginal Pr.	0.31	0.30	0.26	0.22	0.18	0.15	0.12	0.08	0.06	0.03	0.02

Next, we explore in more detail the influence of the term structure on growth. For the Bayesian model averaging specification, most of the posterior probability (0.67) is assigned to models that contain between 7 and 13 lags. Posterior probability for models containing two or less lags is less than 10^{-3}, and the posterior probability of the model without any lag included is less than 10^{-5}. However, the maximum probability model provides very little information; it includes lags 4, 6, 8, 12, 15, 18, 20 and 22, but its estimated posterior probability is only 10^{-4}. On the other hand, the second most likely model has a similar posterior probability but includes a different set of lags.

Since individual models are of little help in understanding how the different lags affect growth, we turn our attention to the individual probabilities of inclusion for the coefficients, which are presented in Table 2. Note that, marginally, only the 4, 8 and 12-month lags of the term structure have a significant impact on GNP, although a number of other lags are close to significance. This shows that information on consecutive lags is decidedly redundant, indicating that the standard assumptions underlying MIDAS models might not be appropriate in this problem. Compatible with these observations, we note that $\mathsf{E}(\pi_0) = 0.47$ and $\mathsf{E}(\pi_1) = 0.41$, with 90% symmetric credible probability intervals $(0.24, 0.71)$ and $(0.22, 0.61)$ respectively, indicating low persistence. The differences between using credible intervals and posterior probabilities for model selection is demonstrated in this problem by the coefficient corresponding to lag 22. Its marginally probability of inclusion is only 0.36, but the probability of being negative given that it is different from zero is 0.95. Had we used credible bands to decide which coefficients should be included (which is approximately equivalent to using a p-value based on a normal approximation), our conclusion would have been the opposite.

For the principal component regression model, Table 1 shows the marginal inclusion probabilities for each of the principal components. The results are enlightening; note that most of the predictive information contained in the slope of interest rates is provided by the average slope over the previous two years (PC1), the biannual rate of change in the slope (PC2), and the average annual rate of change (PC3). This suggests that the common practice of aggregating rates is reasonable, but falls short of incorporating all the important information.

Table 2. Marginal probability of inclusion for different lags of the slope of the term structure of interest rates.

Lag	2	3	4	5	6	7	8	9	10	11	12
Marginal Pr.	0.32	0.46	0.63	0.48	0.47	0.35	0.62	0.44	0.44	0.38	0.50
Lag	13	14	15	16	17	18	19	20	21	22	23
Marginal Pr.	0.45	0.36	0.48	0.38	0.37	0.37	0.41	0.35	0.36	0.36	0.42

Finally, we discuss the predictive performance of the model. Table 3 presents one-step-ahead, out-of-sample predictions for the growth rate between the first quarter of 2003 and the first quarter of 2005, as well as the associated mean squared prediction error, for seven alternative models: MA, which corresponds to the model average Bayesian regression prediction, MM corresponds to the median regression model, MPM is the maximum probability regression model, PC is the model average prediction from the principal component regression model, LAR is a latent autoregressive model without regressors, MIDAS-AR corresponds to a static MIDAS regression model using Beta weights (see the description in Section 2) and incorporating an AR component (Clements & Galvao, 2008), and AR, a purely autoregressive process. The MIDAS model was fitted by adapting the code designed by Arthur Sinko and available at `http://www.unc.edu/~sinko/MATLAB.html`, while the parameters of the AR process were estimated using maximum likelihood methods.

Note that the best performance is produced by model averaging over the original space of predictors, closely followed by MIDAS-AR and the principal component (PC) regression model. The predictions from the AR and the LAR model are very similar; this is consistent with our low estimate of the observational noise V (indeed, taking $V = 0$ reduces the LAR to an AR model). Also, in line with our discussion in Section 3, both the median probability model and maximum probability model have a larger mean squared predictive error than both MA and PC. However, it is surprising that the predictive error of the MPM model is larger than the predictive error form the AR and LAR models.

6.4. Discussion

The results presented above suggest that: 1) the AR component plays a key model in the performance of predictive models for economic output, and 2) including high-frequency term structure information can also improve prediction. This is in line with results previously reported in the literature (see Clements & Galvao (2008) and references therein). However, the results for the discount factors and the MM and MIDAS-AR predictive errors seem to indicate that, at least for this data set, a DLM specification is unnecessary and a regular static regression is a reasonable model.

The results also suggest that introducing soft constraints through prior distribu-

Table 3. Mean-squared out-of-sample prediction error associated with eight alternative prediction models. MA corresponds to the model average prediction, MPM as the maximum probability model, MM to the median model, PC is the model incorporating principal components, LAR is a Bayesian latent AR model without predictors, MIDAS-AR to a static MIDAS regression model incorporating an AR component, and AR to the purely autoregressive process. The true column corresponds to the values actually observed.

Quarter	MA	MPM	MM	PC	LAR	MIDAS-AR	AR	True
2003-I	4.06	4.40	4.29	4.36	2.68	3.11	2.71	4.38
2003-II	4.81	4.70	4.66	4.72	3.57	3.70	3.54	6.96
2003-III	4.52	5.13	4.83	4.90	3.97	4.49	3.99	4.38
2003-IV	5.34	5.38	5.51	5.40	3.08	4.16	3.06	3.74
2004-I	4.62	5.16	4.43	4.45	3.98	4.27	3.98	2.17
2004-II	4.11	4.36	4.50	4.53	2.62	3.82	2.62	3.94
2004-III	5.23	5.37	5.54	5.42	3.97	4.51	3.99	2.86
2004-IV	3.95	4.26	3.50	3.98	3.01	3.13	2.99	3.80
2005-I	3.97	4.15	4.01	3.90	3.97	3.34	3.94	3.11
MSE	2.19	2.79	2.46	2.31	2.51	2.23	2.52	*

tions produces modest improvements over the hard constrains used by the AR-MIDAS models, at least for this data set. However, the choice of model to be used for prediction seems to be critical when soft constrains are used. Maximum probability models can produce very poor predictive results, while principal component models (with truncated geometric priors) can generate competitive results, albeit not as good as those produced by MIDAS-AR or model average over the full space of regression models.

7. Conclusions

This paper describes a Bayesian approach to mixed frequency data regression. In particular, we discussed a dynamic model for quarterly GNP growth prediction that incorporates high frequency (monthly) information on the slope of the term structure of interest rates a predictor. In order to deal with the strong multicollinearity in the predictor series, our approach uses Bayesian model selection and model averaging tools along with structured priors on the model space. Besides greater flexibility and improved forecasting power, the models we discuss provide structural information about the underlying economic process that is unavailable from aggregation or MIDAS models. For example, MIDAS models do not provide a way to test for significance of individual lags on GNP.

One of the main contributions of this paper is to introduce hierarchical distributions on the space of models that automatically adjust for multiple comparisons

and allow us to include relevant prior information. In particular, the order-inducing priors based on a simple Markov process can be used to mimic the specification of MIDAS models, while allowing for greater flexibility. In addition, we introduce priors for dynamic principal component regression that enforce components explaining large proportions of the variability in the predictors to be included in the model.

Although we have focused on models that are conditionally homoscedastic and Gaussian, our approach can be easily extended to include more general error specification. For example, heteroskedasticity in the observational equation can be accommodated using the stochastic volatility model discussed in Uhlig (1997). Computation of the marginal predictive distributions as well as posterior sampling can be done through a small adaptation of the FFBS algorithm discussed in this paper, for details see (West & Harrison, 1998, Chapter 10). Similarly, non-Gaussian error structures can be obtained through mixing. A number of widely used distributions, including the Student t and the power exponential family can be obtained through scale mixtures of normals. Also, finite mixtures of normals can be used to approximate any absolutely continuous distribution, as long as enough component are used. Models of this type can be fitted by introducing a sequence of independent and identically distributed auxiliary variables $\zeta_1, \ldots, \zeta_T$ associated with each observation. Conditionally on the auxiliary variables, the error distribution is Gaussian and computation proceeds using the same algorithms discussed here, while each ζ_t has to be sampled from its associated full conditional distribution.

¿From a methodological perspective, our examples illustrates the difference between prediction (for which model averaging is optimal), and the validation of a scientific theory (for which model comparison/variable selection approaches are necessary). Although both approaches use essentially the same inputs, the insights obtained are in general very different.

The models we describe in this paper can be easily extended to include multiple variables sampled at various frequencies. Similarly, they can accommodate more complex error structures through the use of hierarchical specifications. In particular heavy-tailed error distributions and stochastic volatility models can be incorporated through mixing over the parameters of the normal errors. Along this lines, we are currently working on applications of this class of models to multiresolution models for volatility prediction, with applications to option pricing.

The model selection approach discussed in this paper assumes that significant coefficients remain so as time evolve. A more realistic approach would allow a different set of regressors at each point in time. We are currently working on this type of approach, which presents some additional difficulties. In first place, we need to design dynamic priors on the space of models that allow for simple computation while at the same time providing enough flexibility. Second, we need to deal with possible identifiability issues arising from allowing both coefficients and models to evolve in

time.

8. Acknowledgements

We would like to thank Mike West for his kind support in the initial stages of this project. We would also like to thank the Editor Gael Maritn and three anonymous referees for suggestions that helped to greatly improve this manuscript.

A. Markov chain Monte Carlo sampler

Given initial values for all unknown parameters in the model, the algorithm updates their values iteratively by sampling from the following full conditional distributions:

Step 1 Conditional on V, δ^{β}, δ^{μ}, ϕ, and U sample the model k and the state parameters $\{(\mu_t, z_t, \boldsymbol{\beta}_t)\}_{t=1}^{T}$ using a FFBS algorithm. To do this, let $\boldsymbol{\theta}_t^k = (z_t, \mu_t, \boldsymbol{\beta}_t^k)$ be the $|M_k|+2$ vector of state variables at time t for model k, $\mathbf{G}^k = \text{diag}(\phi, 1, 1, \ldots, 1)$ the $(|M_k|+2) \times (|M_k|+2)$ evolution matrix, and $\mathbf{W}_t^k = \text{diag}(U, \boldsymbol{\Sigma}_t^k)$ the block diagonal $(|M_k|+2) \times (|M_k|+2)$ evolution covariance matrix and $\mathbf{F}_t^k = (1, 1, \mathbf{X}_t^k)$ be the $|M_k|+2$ observational design vector.

- Forward Filtering step. For $k = 1, \ldots, 2^p$ and $t = 1, \ldots, T$ compute $\boldsymbol{\theta}_t^k | D_t, \cdots \sim \mathsf{N}(\mathbf{m}_t^k, \mathbf{C}_t^k)$, where:

$$
\begin{aligned}
\mathbf{m}_t^k &= \mathbf{a}_t^k + \mathbf{A}_t^k e_t^k \\
\mathbf{C}_t^k &= \mathbf{R}_t^k - \mathbf{A}_t^k Q_t^k (\mathbf{A}_t^k)' \\
\mathbf{A}_t^k &= \mathbf{R}_t^k (\mathbf{F}_t^k)' (Q_T^k)^{-1} \\
e_t^k &= y_t - f_t^k \\
\mathbf{a}_t^k &= \mathbf{G}^k \mathbf{m}_{t-1} \\
\mathbf{R}_t^k &= \mathbf{G}^k \mathbf{C}_t^k (\mathbf{G}^k)' + \mathbf{W}_t^k \\
f_t^k &= \mathbf{F}_t^k \mathbf{a}_t^k \\
Q_t^k &= \mathbf{F}_t^k \mathbf{R}_t^k (\mathbf{F}_t^k)' + V
\end{aligned}
$$

- Sample k from a discrete distribution with probabilities proportional to:

$$
p_k(y_1, \ldots, y_T | \cdots) = \prod_{t=1}^{T} p_k(y_t | D_t, \cdots) = \prod_{t=1}^{T} \mathsf{N}(y_t | f_t^k, Q_t^k)
$$

– Backward Sampling. Conditional of k, and for $t = T-1, \ldots, 0$ set the the entries of $\boldsymbol{\beta}$ not included in $\boldsymbol{\beta}^k$ equal to zero and sample

$$(\boldsymbol{\theta}_t^k | \boldsymbol{\theta}_{t+1}^k) \sim \mathsf{N}(\mathbf{h}_t^k, \mathbf{H}_t^k)$$

where:

$$\begin{aligned}
\mathbf{h}_t^k &= \mathbf{m}_t^k + \mathbf{B}_t^k(\boldsymbol{\theta}_t^k - \mathbf{a}_t^k) \\
\mathbf{H}_t^k &= \mathbf{C}_t^k - \mathbf{B}_t^k \mathbf{R}_{t+1}^k \mathbf{B}_t^{k'} \\
\mathbf{B}_t^k &= \mathbf{C}_t^k \mathbf{G}^{k'} (\mathbf{R}_{t+1}^k)^{-1}
\end{aligned}$$

Step 2 Conditional on the the state parameters and the data, sample V from

$$V^{-1} | \cdots \sim \mathsf{Gam}\left(a_\nu + \frac{T}{2}, b_\nu + \frac{\sum_{t=1}^T \hat{\nu}_t^2}{2}\right)$$

where $\hat{\nu}_t = y_t - z_t - \mu_t - \mathbf{F}_t^k \boldsymbol{\beta}_t^k$ and a prior $V^{-1} \sim \mathsf{Gam}\,(a_\nu, b_\nu)$

Step 3 Conditional on the the state parameters and U, sample ϕ from $\phi | \cdots \sim \mathsf{N}(\hat{m}, \hat{C})$ where

$$\hat{m} = \hat{C}\left(\frac{m_\phi}{C_\phi} + \frac{(\sum_{t=1}^T z_{t-1}^2)(\sum_{t=1}^T z_{t-1} z_t)}{U}\right) \quad \hat{C} = \left(\frac{1}{C_\phi} + \frac{(\sum_{t=1}^T z_{t-1}^2)}{U}\right)^{-1}$$

and a priori $\phi \sim \mathsf{N}(m_\phi, C_\phi)$.

Step 4 Conditional on the the state parameters and ϕ sample U from

$$U^{-1} | \cdots \sim \mathsf{Gam}\left(a_\epsilon + \frac{T}{2}, b_\epsilon + \frac{\sum_{t=1}^T \hat{\epsilon}_t^2}{2}\right)$$

where $\hat{\epsilon}_t = z_t - \phi z_{t-1}$ and a priori $U^{-1} \sim \mathsf{Gam}\,(a_\epsilon, b_\epsilon)$

Step 5 Conditional on all other parameters, sample δ^β and δ^μ from a discrete distribution with probabilities proportional to:

$$\prod_{t=1}^T \mathsf{N}(\boldsymbol{\theta}_t | \mathbf{G}\boldsymbol{\theta}_{t-1}, \mathbf{W}_t(\delta_\mu, \delta_\beta))$$

Table 4. Loadings matrix for the principal components analysis of the slope of the term structure of interest rates.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Lag	65.9%	18.2%	5.4%	3.2%	2.3%	1.8%	1.0%	0.4%
2	-0.1735	0.2868	-0.2642	-0.2408	-0.1279	0.2821	-0.5250	0.4433
3	-0.1890	0.2821	-0.2711	-0.2238	-0.1936	0.1433	-0.0552	-0.1570
4	-0.1950	0.2875	-0.2601	-0.1896	-0.1734	-0.0084	0.2678	-0.4022
5	-0.2075	0.2660	-0.2120	-0.0479	-0.0322	-0.0883	0.3100	-0.1563
6	-0.2177	0.2356	-0.1271	0.0725	0.1837	-0.2757	0.0926	0.3030
7	-0.2203	0.2178	-0.0338	0.1888	0.3406	-0.3428	0.0190	0.2782
8	-0.2283	0.1803	0.0556	0.2655	0.2492	-0.2092	-0.0176	-0.1057
9	-0.2350	0.1406	0.1197	0.2740	0.1372	0.1917	-0.1765	-0.2199
10	-0.2337	0.1029	0.2257	0.3090	0.0521	0.3724	-0.1381	-0.1837
11	-0.2365	0.0541	0.2881	0.2189	-0.0527	0.2465	-0.0234	-0.0255
12	-0.2408	0.0188	0.2454	0.0592	-0.3079	0.0313	0.1759	0.1119
13	-0.2353	-0.0273	0.2850	-0.0622	-0.4235	-0.0976	0.2001	0.1987
14	-0.2326	-0.0827	0.2777	-0.1544	-0.2614	-0.1114	0.0577	0.1578
15	-0.2340	-0.1059	0.1927	-0.2476	0.0376	-0.2235	-0.2329	0.0198
16	-0.2254	-0.1502	0.1501	-0.3585	0.2012	-0.2183	-0.2789	-0.1644
17	-0.2176	-0.2012	0.0710	-0.3121	0.2064	-0.0766	-0.0501	-0.3212
18	-0.2179	-0.2118	-0.0432	-0.1688	0.2198	0.2599	0.1372	-0.0231
19	-0.2058	-0.2466	-0.1455	-0.0911	0.2427	0.3654	0.2682	0.1951
20	-0.1921	-0.2831	-0.1972	0.0185	0.1582	0.1772	0.1909	0.2217
21	-0.1906	-0.2780	-0.2310	0.1633	-0.1052	-0.1003	0.1007	0.0761
22	-0.1751	-0.2879	-0.2932	0.2703	-0.2331	-0.1984	-0.0707	-0.1107
23	-0.1571	-0.2956	-0.3008	0.2611	-0.2323	-0.0989	-0.3768	-0.1255

B. Principal component analysis for the term structure of interest rates

Table 4 provides the first eight columns of the loadings matrix of the loadings matrix for the principal components analysis of the slope of the term structure of interest rates. Components are ordered according to the percentage of variability explained, computed as the variance of the component divided by the total variance in the data. We can roughly interpret the first principal component which explains 66% of the variability in the interest rates, as the (negative) average of the slope during the last two years. Similarly, the second principal component contrasts the very recent and the very old interest rates, roughly how much the slope has changed over the last two years, while the third principal component compares the rates around the center of the series with the rates at the beginning and the end of the two years period under analysis. Surprisingly enough, this decomposition is very similar to the one that would have been obtained from a wavelet analysis of the series using a Haar basis system.

References

Aguilar, O. & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics* **18**, 338–357.

Barbieri, M. M. & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics* **32**, 870–897.

Berger, J. & Pericchi, L. (2001). Objective Bayesian methods for model selection: introduction and comparison (with discussion). In *Model Selection*, Ed. P. Lahiri, volume 38 of *Monograph Series*. Institute of Mathematical Statistics Lecture Notes.

Berger, J., Pericchi, L. & Varshavsky, J. (1998). Bayes factors and marginal distributions in invariant situations. *Sankhya A* **60**, 307–321.

Berger, J. O. & Pericchi, L. (1996a). The intrinsic Bayes factor for linear models. In *Bayesian Statistics 5*, Eds. A. P. D. J. M. Bernardo, J. O. Berger & A. F. M. Smith, pp. 23–42. Oxford: Oxford University Press.

Berger, J. O. & Pericchi, L. (1996b). Posterior odds ratios for selected regression hyotheses. In *Bayesian Statistics: Proceedings of the First International Meeting held in Valencia (Spain)*, Eds. D. V. L. J. M. Bernardo, M. H. DeGroot & A. F. M. Smith, pp. 585–603. Valencia: University Press.

Birchenhall, C. R., Jessen, H. & Osborn, P., D. R. Simpson (1999). Predicting u.s. business cycle regimes. *Journal of Business and Economic Statistics* **17**, 313–323.

Campos, J., Hendry, D. f. & Krolzig, H.-M. (2003). Consistent model selection by an automatic gets approach. *Oxford Bulletin of Economics and Statistics* **65**, 803–819.

Carter, C. K. & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541–553.

Carvalho, C. M. & West, M. (2007). Dynamic matrix-variate graphical models. *Bayesian Analysis* **2**, 69–98.

Clements, M. P. & Galvao, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business and Economic Statistics* **26**, 546–554.

Draper, D. (1995). Assesment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* **57**, 45–97.

Estrella, A. & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *The Journal of Finance* **46**, 555–576.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* **15**, 183–202.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7**, 457–472.

George, E. I. & McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of American Statistical Association* **88**, 881–889.

George, E. I. & McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–373.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2002). The MIDAS touch: Mixed data sampling regression models. Technical report, University of North Carolina, Chapel Hill and University of California, Los Angeles.

Ghysels, E., Santa-Clara, P. & Valkanov, R. (2006). Predicting volatility: Getting the most out of return data sampled at different frequencies. *Journal of Econometrics* **131**, 59–95.

Ghysels, E., Sinko, A. & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews* **26**, 53–90.

Ghysels, E. & Wright, J. H. (2006). Forecasting professional forecasters. Technical report, Federal Reserve Working Papers.

Green, P. J. (1995). Reversible jump MCMC computation and bayesian model determination. *Biometrika* **82**, 711–732.

Hendry, D. F. & Mizon, G. E. (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal* **88**, 549–563.

Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with discussion). *Statistical Science* **14**, 382–417.

Jefferys, W. H. (1990). Bayesian analysis of random event generator data. *Journal of Scientific Exploration* **4**, 153–169.

Jefferys, W. H. & Berger, J. O. (1992). Ockham's razor and bayesian analysis. *American Scientist* **80**, 64–72.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.

Lindley, D. V. (1957). A statistical paradox. *Biometrika* **44**, 187–192.

Natarajan, R. & McCulloch, R. E. (1998). Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference? *Journal of Computational and Graphical Statistics* **7**, 267–277.

Putnam, B. H. & Quintana, J. M. (1995). The evolution of bayesian forecasting models. In *Asset Allocation: Applying Quantitative Discipline to Asset Allocation*, Ed. B. H. Putnam, pp. 139–146. London: Global Investor, Euromoney Publications.

Quintana, J. M. (1992). Optimal portfolios of forward currency contracts. In *Bayesian Statistics 4*, Eds. A. D. J.O. Berger, J.M. Bernardo & A. Smith. Oxford University Press.

Quintana, J. M. & West, M. (1987). An analysis of international exchange rates using multivariate DLMs. *The Statisticians* **36**, 275–281.

Robert, C. P. & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, second edition edition.

Rodriguez, A. & Pericchi, L. (2000). Intrinsic Bayes factors for dynamic linear models. In *Proceedings of the International Society for Bayesian Analysis World Meeting*. Greece.

Scott, J. & Berger, J. O. (2003). An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference* **137**, 2144–2162.

Shafer, G. (1982). Lindley's paradox. *Journal of the American Statistical Association* **77**, 325–334.

Shen, C.-H. (1996). Forescasting macroeconomic variables using data of different periodicities. *International Journal of Forecasting* **12**, 269–282.

Smith, A. F. M. & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **55**, 3–23.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica* **65**, 59–74.

West, M. & Harrison, J. (1998). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, second edition edition.