



CREaTE

Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

<http://create.canterbury.ac.uk>

Please cite this publication as follows:

Kalli, M. and Griffin, J. (2014) Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, 178 (2). pp. 779-793. ISSN 0304-4076.

Link to official URL (if available):

<http://dx.doi.org/10.1016/j.jeconom.2013.10.012>

This version is made available in accordance with publishers' policies. All material made available by CReaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: create.library@canterbury.ac.uk



Time-varying sparsity in dynamic regression models

Maria Kalli and Jim E. Griffin *

March 27, 2013

Abstract

We propose a novel Bayesian method for dynamic regression models where both the values of the regression coefficients and the importance of the variables are allowed to change over time. The parsimony of the model is important for good forecasting performance and we develop a prior which allows the shrinkage of the regression coefficients to suitably change over time. An efficient MCMC method for computation is described. The new method is then applied to two forecasting problems in econometrics: equity premium prediction and inflation forecasting. The results show that this method outperforms current competing Bayesian methods.

Keywords: Time-Varying Regression; Shrinkage priors; Normal-Gamma priors; Markov chain Monte Carlo; Equity Premium; Inflation.

*Maria Kalli, Canterbury Christ Church University Business School, Canterbury Christ Church University, Canterbury, UK (email: maria.kalli@canterbury.ac.uk), Jim Griffin, School of Mathematics, Statistics & Actuarial Science, University of Kent, Canterbury, UK (email: J.E.Griffin-28@kent.ac.uk).

1 Introduction

Forecasting, the estimation of a future value of a variable, plays an important role in both decision making and strategic planning and has been extensively studied in econometrics. For example, forecasts of inflation affect the decisions of monetary and fiscal policymakers, investors who wish to hedge against the risk of nominal assets, trade unions and management when they negotiate wage contracts, to name a few. Similarly, forecasts of equity premiums plays an important role for investors who wish to diversify their equity portfolios to hedge against adverse market movements. The quality of the forecast depends on: the time scale involved (how far into the future we are trying to predict), the time period of the empirical sample, and the model used.

Regression models are a popular choice of technique for forecasting since the value of other variables can be used to inform predictions. However, their use with observations made over time is complicated by several problems. Firstly, it has been found that these models can produce poor out-of-sample forecasts when the predictors' effects are assumed constant over time. This is generally taken as evidence that the effect of variables are time-varying. Sims (1980), Stock and Watson (1996), Cogley and Sargent (2001, 2005), Primiceri (2005), Paye and Timmermann (2006), Ang and Bekaert (2007), Canova (2007), and Lettau and Van Nieuwerburgh (2008) are some studies providing evidence of time varying regressor coefficients in inflation and equity premium forecasting. Secondly, the increasing availability of large economic datasets has lead to interest in using regression models with many regressors. It is well-known that the estimation of regression models becomes more complicated when a large number of predictors is used due to the increased potential for over-fitting which can lead to poor out-of-sample forecasts or predictions. The problem of over-fitting can be alleviated by looking for sparse regression estimates where many regression coefficients are set to zero or values close to zero. This is usually achieved using regularisation of the

regression coefficients or variable selection.

The problem of time-varying predictor effects can be addressed using dynamic regression models (DR), which is a form of time-varying parameter model, where the regression coefficients are assumed to evolve according to some stochastic process. This defines a dynamic linear model (DLM) (West and Harrison, 1999), or a state-space model.

The problem of a large number of variables has been addressed in several ways. Initial work concentrated on models which assume a global measure of the importance of a variable. Groen et al. (2009) introduced a latent variable which indicates whether a variable is included in or excluded from the model. The approach is restricted so that the decision to include or exclude a predictor is irreversible. Belmonte et al. (2011) combined the Bayesian Lasso of Park and Casella (2008) with the model selection methods of Frühwirth-Schnatter and Wagner (2010) in order to have shrinkage in a DR setting. This approach allows some regression coefficients to be shrunk very close to zero for the whole time series and so effectively achieve variable selection. These methods have the potentially important limitation that the importance of variables cannot change over time. For example, in some problems certain predictors could be useful for forecasting at particular times but not at others. In the Bayesian literature, this problem has been approached by allowing variables to enter and exit the model over time. Koop and Korobilis (2012) used the dynamic model averaging (DMA) method of (Raftery et al., 2010) to select a suitable time-varying parameter model. However, the dynamics on model space are only implicitly defined by their approach. Alternatively, Chan et al. (2012) constructed the class of time-varying dimension (TVD) models which uses an explicitly constructed stochastic process for the subset of variables to include in the model. This leads to a dynamic mixture model for which efficient posterior computational methods can be developed using the approach of Gerlach et al. (2000). However, both approaches are limited to the number

of models that they can consider. The DMA approach uses full enumeration of posterior probabilities and the number of models with p regressors is 2^p precluding large values of p . Posterior computation in the TVD model also potentially involves all 2^p models but the authors suggest using a much restricted set of possible models.

The DMA and TVD approaches build on Bayesian variable selection techniques which explicitly consider all possible regression models. An alternative class of methods is Bayesian regularisation methods which use absolutely continuous priors and encourage small regression effects to be aggressively shrunk towards zero under the posterior (see Carvalho et al. (2010), and Polson and Scott (2011)). These authors have shown that these methods can lead to posteriors which place substantial mass on combinations of regression coefficients which are sparse (that is most of the regression coefficients have values very close to zero). Belmonte et al. (2011) have already extended one such prior, the Bayesian Lasso, to the dynamic regression setting. Our methodological contribution differs from their work in two main respects. Firstly, our prior for the time-varying regression coefficients extends the more general Normal-Gamma (NG) prior (see Caron and Doucet (2008) and Griffin and Brown (2010)) to DR models and, secondly, our prior accounts for both time-varying regression coefficients and time-varying sparsity.

The paper is organized as follows: Section 2 introduces the Normal-Gamma AutoRegressive (NGAR) prior, considers some of its properties and describes the full Bayesian model for dynamic regression with time-varying sparsity. Section 3 describes the required MCMC method for fitting a dynamic regression model with an NGAR prior. Section 4 applies the NGAR model to simulated data, while Section 5 considers empirical studies in equity premium and inflation forecasting. Section 6 summarises our findings and conclusions.

2 A Bayesian Dynamic Regression Model with Time-Varying Sparsity

A dynamic regression (DR) model links a response y_t to regressors $x_{1,t}, \dots, x_{m,t}$ (which have both been observed at time t) by

$$y_t = \sum_{i=0}^m x_{i,t} \beta_{i,t} + \epsilon_t, \quad t = 1, \dots, T \quad (1)$$

where $x_{0,t} = 1$ for all t (and so allows an intercept), $\beta_{i,t}$ is a vector of unknown coefficients for the i -th regressor at time t , ϵ_t is the innovation term at time t , and T is the time of the final observation. The regressors may include both lags of the response and exogenous variables. The model is usually completed by assuming that $\beta_{1,t}, \dots, \beta_{m,t}$ follow a linear stochastic process (such as a random walk or vector autoregression) and that the variance of the innovations is time-varying so that $\epsilon_t \sim N(0, \sigma_t^2)$.

In regression models with a large number of regressors, it is common to assume that only a subset of the regressors is important for prediction. In DR models, this assumption is most naturally extended to allow the subset of important regressors to change over time. Therefore, we want to construct a stochastic process which allows regressors to be either removed from the model, or equivalently some of $\beta_{1,t}, \dots, \beta_{m,t}$ to take values close (or equal) to zero. We will refer to the proportion of time that $\beta_{i,t}$ is close to zero (for $t = 1, \dots, T$) as the sparsity of the vector $\beta_i = (\beta_{i,1}, \dots, \beta_{i,T})$. This allows us to have time-varying sparsity since the i -th regression coefficient is effectively removed from the regression model at time t if $\beta_{i,t}$ is very close to zero (and that proportion is controlled by parameters of the stochastic process). We achieve this by assuming that β_1, \dots, β_m follow independent stochastic processes and using a new form of process, the Normal-Gamma Autoregressive (NGAR) process, for β_i , which is defined below. Let $x \sim \text{Ga}(a, b)$ mean that x follows a Gamma distribution with shape

parameter a and mean a/b and $x \sim \text{Pn}(\mu)$ mean that x follows a Poisson distribution with mean μ .

Definition 1 *The Normal-Gamma Autoregressive (NGAR) process for β_i is defined by*

$$\kappa_{i,s-1}|\psi_{i,s-1} \sim \text{Pn}\left(\frac{\rho_i \frac{\lambda_i}{\mu_i} \psi_{i,s-1}}{1 - \rho_i}\right), \quad \psi_{i,s}|\kappa_{i,s-1} \sim \text{Ga}\left(\lambda_i + \kappa_{i,s-1}, \frac{\lambda_i}{\mu_i(1 - \rho_i)}\right)$$

and

$$\beta_{i,s} = \sqrt{\frac{\psi_{i,s}}{\psi_{i,s-1}}} \varphi_i \beta_{i,s-1} + \eta_{i,s}, \quad \eta_{i,s}|\psi_{i,s} \sim N(0, (1 - \varphi_i^2)\psi_{i,s}) \quad s = 2, \dots, T$$

where

$$\psi_{i,1} \sim \text{Ga}(\lambda_i, \lambda_i/\mu_i), \quad \beta_{i,1}|\psi_{i,1} \sim N(0, \psi_{i,1}).$$

This process will be written $\beta_i \sim \text{NGAR}(\lambda_i, \mu_i, \varphi_i, \rho_i)$.

The NGAR prior can also be represented as the product of two independent stochastic processes: $\psi_i = (\psi_{i,1}, \dots, \psi_{i,T})$ and $\phi_i = (\phi_{i,1}, \dots, \phi_{i,T})$. Under this representation, $\beta_{i,t} = \sqrt{\psi_{i,t}} \phi_{i,t}$ where $\phi_i = (\phi_{i,1}, \dots, \phi_{i,T})$ is generated from an AR(1) process with autocorrelation parameter φ_i such that ϕ_i has the standard Normal as its stationary distribution (i.e. $\phi_i = \varphi_i \phi_{i-1} + \varsigma_i$ where $\varsigma_i \sim N(0, 1 - \varphi_i^2)$). We generate $\psi_i = (\psi_{i,1}, \dots, \psi_{i,T})$ from a non-Gaussian AR(1) process with a $\text{Ga}(\lambda_i, \frac{\lambda_i}{\mu_i})$ as its marginal distribution, using the method described in (Pitt et al., 2002; Pitt and Walker, 2005) and later, independently, developed as the autoregressive gamma process by Gouriéroux and Jasiak (2006). The conditional density of $\psi_{i,t}$ given $\psi_{i,t-1}$ is equal to

$$\sum_{\kappa=0}^{\infty} w_{\kappa, \psi_{i,t-1}} \text{Ga}\left(\psi_{i,t} \mid \lambda_i + \kappa, \frac{\lambda_i}{\mu_i(1 - \rho_i)}\right) \quad (2)$$

which is a mixture of Gamma distributions with parameters $\lambda_i + \kappa$, and $\frac{\lambda_i}{\mu_i(1 - \rho_i)}$, and

Poisson weights such that

$$w_{\kappa, \psi_{i,t-1}} = \frac{\exp \left\{ -\frac{\rho_i \lambda_i}{\mu_i (1 - \rho_i)} \psi_{i,t-1} \right\} \left(\frac{\rho_i \lambda_i}{\mu_i (1 - \rho_i)} \psi_{i,t-1} \right)^\kappa}{\kappa!}. \quad (3)$$

The mean of $\psi_{i,t}$ given $\psi_{i,t-1}$ is

$$E[\psi_{i,t} | \psi_{i,t-1}] = \mu_i (1 - \rho_i) + \rho_i \psi_{i,t-1},$$

which has an autoregressive structure, and its conditional variance is

$$\text{Var}[\psi_{i,t} | \psi_{i,t-1}] = \frac{\mu_i^2 (1 - \rho_i)^2}{\lambda_i} + \frac{2\rho_i \mu_i (1 - \rho_i) \psi_{i,t-1}}{\lambda_i}.$$

It is clear from the second representation of the NGAR that $E[\beta_{i,t} | \psi_{i,t}] = 0$ and $\text{Var}[\beta_{i,t} | \psi_{i,t}] = \psi_{i,t}$. Therefore the value of $\psi_{i,t}$ plays a pivotal role in the NGAR set-up. It affects the mass of the conditional prior distribution of $\beta_{i,t}$ given $\psi_{i,t}$ around zero. As $\psi_{i,t}$ decreases more mass is placed close to zero. This leads to the posterior mean of $\beta_{i,t}$ conditional on $\psi_{i,t}$ being increasingly shrunk to zero as $\psi_{i,t}$ decreases. We can thus describe $\psi_{i,t}$ as the relevance of the i -th regressor at time t . A smaller value of $\psi_{i,t}$ implies smaller relevance and so the posterior mean is increasingly shrunk to zero. This interpretation of the variance of a normal prior distribution for a regression coefficient dates back to Tipping (2000) and Bishop and Tipping (2000). Since $\psi_{i,t}$ and $\phi_{i,t}$ are independent and stationary, the process β_i is also stationary and has a Normal-Gamma stationary distribution. The unconditional variance of $\beta_{i,t}$ is $\text{Var}(\beta_{i,t}) = \mu_i$ and the excess kurtosis is $\varkappa(\beta_{i,t}) = 3/\lambda_i$. Thus for a fixed prior mean, μ_i , as the value of λ_i decreases more mass is placed at zero and as it increases less mass is placed at zero, affecting the impact of the i -th regressor. Griffin and Brown (2010) study the use of a Normal-Gamma (NG) prior for regression problems. They conclude that the proportion of prior mass close to zero is important for obtaining sparse posterior estimates (where

many regression coefficients are shrunk close to zero) and show that it is controlled by the shape parameter λ_i , which we will refer to as the “sparsity parameter” of the prior. Smaller values of λ_i imply larger levels of sparsity.

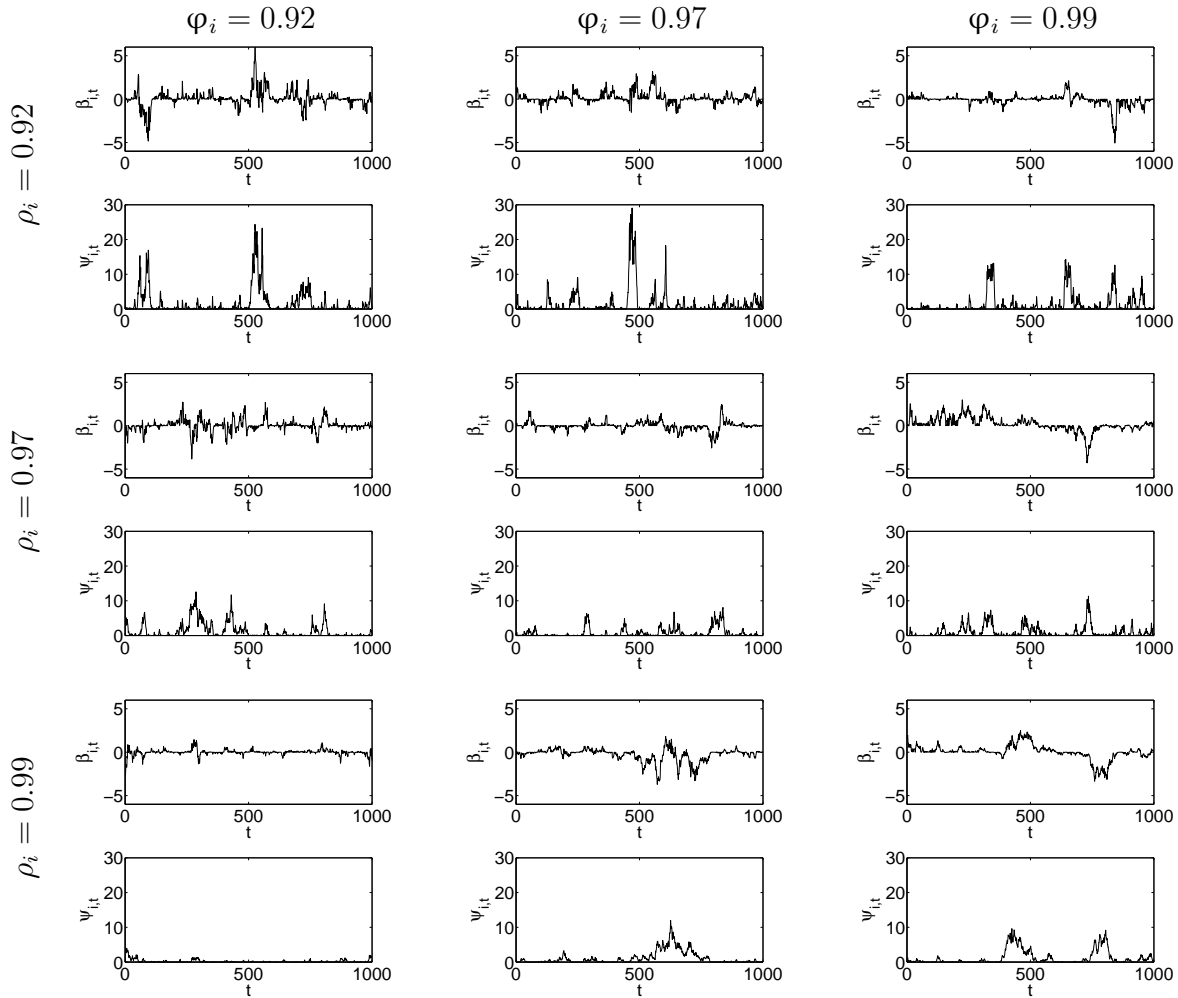


Figure 1: Simulated paths of β_t and ψ_t with different values of ρ and φ with $\lambda = 0.2$ and $\text{Var}[\beta_{i,t}] = \mu_i = 1$.

Figures 1 and 2 display simulated paths of the NGAR prior for both $\psi_{i,t}$ and $\beta_{i,t}$ with different combinations of λ_i , φ_i and ρ_i . These illustrate the ability of the prior to generate periods where the regression coefficients are close to zero and periods where the regression coefficients are away from zero.

The sparsity parameter λ_i clearly controls the proportion of time that the regres-

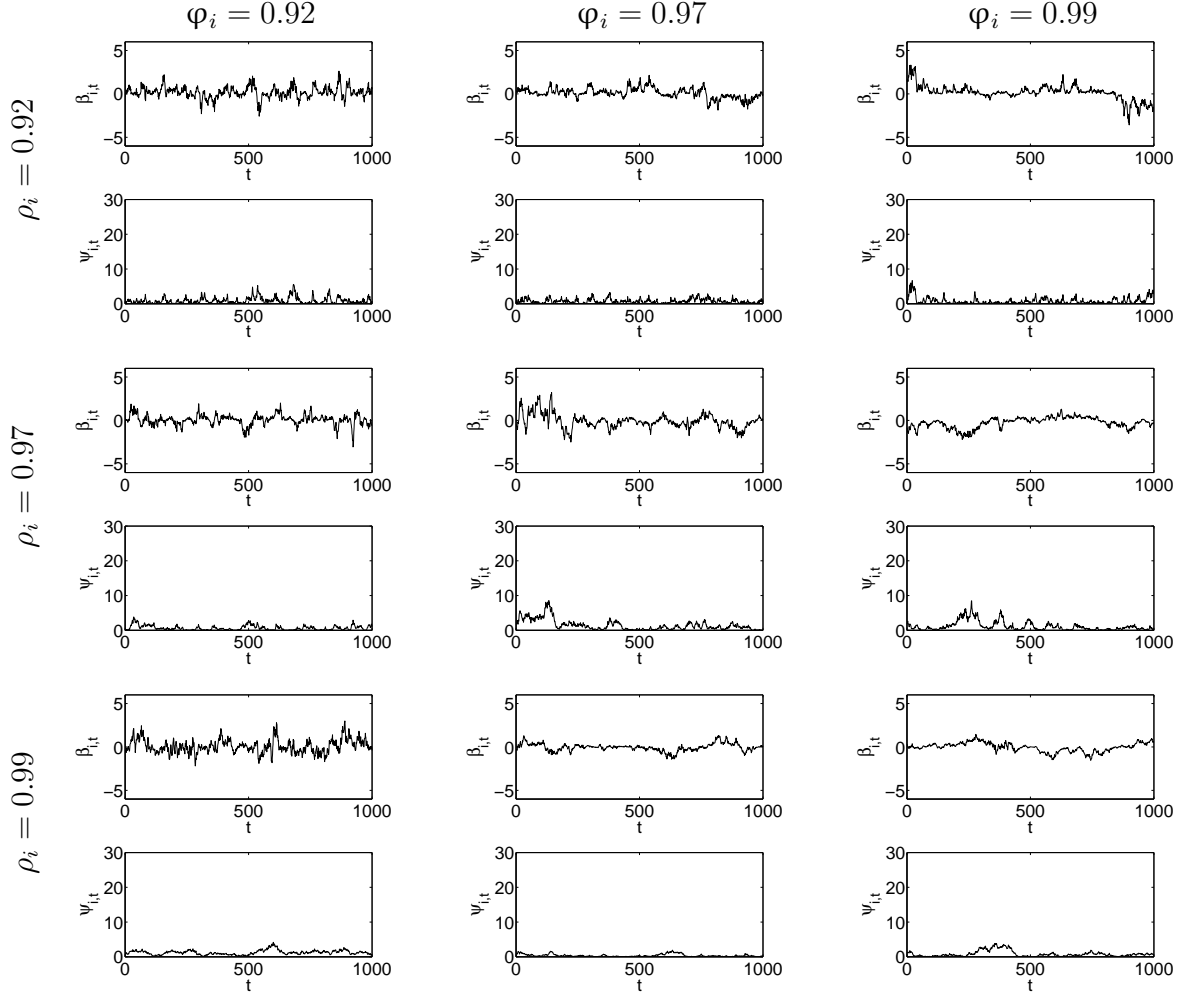


Figure 2: Simulated paths of $\beta_{i,t}$ and $\psi_{i,t}$ with different values of ρ_i and φ_i with $\lambda_i = 1$ and $\text{Var}[\beta_{i,t}] = \mu_i = 1$.

sion coefficient spends close to zero. This proportion becomes larger as λ_i decreases which is illustrated in Figures 1 and 2 where $\lambda_i = 0.2$ and $\lambda_i = 1$ respectively. Smaller values of λ_i lead to “spikier” processes for $\psi_{i,t}$ and $\beta_{i,t}$ which favours increasingly rapid changes from small to large values. The autocorrelation parameter ρ_i controls the dependence between $\psi_{i,t-1}$ and $\psi_{i,t}$. Larger values of ρ_i leads to a larger autocorrelation and which favours processes which spend longer periods close to zero or away from zero. Decreasing the value of ρ_i allows the regressors to increasingly jump in and out of the DR model. The autocorrelation parameter φ_i controls the dependence between

$\beta_{i,t}$ and $\beta_{i,t-1}$ conditional on the $\psi_i = (\psi_{i,1}, \dots, \psi_{i,T})$ process.

We complete our model by assuming a stochastic volatility process for the observational errors, $\sigma_1^2, \dots, \sigma_T^2$ in equation (1). These will be given the gamma autoregressive process uses in the construction of the NGAR (Pitt et al., 2002; Pitt and Walker, 2005; Gouriéroux and Jasiak, 2006) which results in a process specified by

$$\kappa_{t-1}^\sigma | \psi_{t-1}^\sigma \sim \text{Pn}(\lambda^\sigma \rho^\sigma \sigma_{t-1}^2 / ((1 - \rho^\sigma) \mu^\sigma)) \text{ and } \sigma_t^2 \sim \text{Ga}(\lambda^\sigma + \kappa_{t-1}^\sigma, \lambda^\sigma / (\mu^\sigma (1 - \rho^\sigma))),$$

for $t = 2, \dots, T$ with $\sigma_1^2 \sim \text{Ga}(\lambda^\sigma, \lambda^\sigma / \mu^\sigma)$.

We will make Bayesian inference in the DR with independent NGAR processes for the regression coefficients. The prior that we described allow us to control the sparsity of the posterior distribution of the regression coefficients (*i.e.* the proportion of regression coefficients with mass close to zero at a time t), and assumes that regressors should rarely jump in and out of the model. Both assumption are important to avoid over-fitting of the dynamic regression model.

The parameter μ_i acts as an overall relevance parameter for the i -th regression coefficient since it controls the marginal variance of $\beta_{i,t}$. In particular, $\beta_{i,t}$ will be close to zero for all t if μ_i is small. Therefore, a hierarchical prior is specified for μ_1, \dots, μ_m where

$$\mu_i \sim \text{Ga}(\lambda^*, \lambda^* / \mu^*), \quad i = 0, \dots, m$$

with

$$\lambda^* \sim \text{Ex}(1/s^*), \quad p(\mu^*) \propto (\mu^* + 2b^*)^{-3}$$

where $\text{Ex}(\gamma)$ represents an exponential distribution with mean $1/\gamma$. This introduces a second level of sparsity (at the level of the regressors rather than the time-varying regression coefficients). This is particularly important in problem with many regressors where some regressors have no regression effect across all observations. The hyperpa-

parameter s^* is the prior mean of λ^* and so gives an initial idea of the level of sparsity using the ideas described in Griffin and Brown (2010). The parameter μ^* is given a heavy-tailed prior with b^* being the prior mean which is given a value suitable for the spread of the regression coefficients in the particular application.

The prior now has two sparsity parameters. The parameter λ^* is the sparsity parameter for the whole series of each regression coefficient. Smaller values of λ^* indicate that more μ_i 's are close to zero, and so $\beta_{i,t}$ is close to zero at all time t for more regressors. In contrast, λ_i controls the sparsity within the time series of the i -th regression coefficient and a small value of λ_i would indicate that the regression coefficient is close to zero for a large proportion of observations. The shape parameter λ_i is given the prior

$$p(\lambda_i) \propto \lambda_i(0.5 + \lambda_i)^{-4}$$

which is a heavy-tailed prior giving values around 1. This centres the prior over the Lasso cases (which arises when $\lambda_i = 1$).

The dependence parameters φ_i and ρ_i play a key role in our model. We make the assumption that the processes for the regression coefficients and the relevances are strongly autocorrelated. We therefore choose informative priors. The flexibility of the NGAR can lead to over fitting when the values of φ_i and ρ_i are small. The problem of over fitting is particularly acute in DR models since we have m regression coefficients at each time point. The realisations in Figures 1 and 2 confirm that even a value of ρ_i close to 0.9 allows regressors to quickly be excluded from the DR model. Therefore informative priors effectively exclude models which allow the regression coefficients to rapidly change over time (and so will lead to overfitting). The priors used were

$$\varphi_i \sim \text{Be}(77.6, 2.4), \quad \rho_i \sim \text{Be}(77.6, 2.4), \quad i = 0, \dots, m,$$

which gives a prior mean of 0.97 with most mass over 0.9.

The priors for the parameters of the volatility process σ_t^2 are chosen as

$$\lambda^\sigma \sim \text{Ga}(3, 1), \quad p(\mu^\sigma) \propto (1 + \mu^\sigma)^{-3/2}, \quad \rho^\sigma \sim \text{Be}(38, 2).$$

$$\sigma_1^2 \sim \text{Ga}(\lambda^\sigma, \lambda^\sigma / \mu^\sigma),$$

The choice for λ^σ signifies that the volatility process will have stationary distribution which is less heavy tailed than a Laplace distribution. The mean μ^σ is given a very heavy tailed prior to allow for a wide-range of possible values. The dependence parameter ρ^σ is given an informative prior that enforces stationarity and which places most of its mass on values greater than 0.85. This seems reasonable given the value usually associated with stochastic volatility models.

3 Computation

MCMC methods to fit the dynamic regression model in (1) with the NGAR will be described in this section. The MCMC sampler exploits the observation that the regression model conditional on ψ is a Gaussian state-space model and so the marginal likelihood $p(y|X, \psi, \sigma^2, \rho, \varphi)$, where $\psi = (\psi_1, \dots, \psi_m)$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_T^2)$, $\rho = (\rho_1, \dots, \rho_m)$ and $\varphi = (\varphi_1, \dots, \varphi_m)$, can be efficiently calculated using the Kalman filter. Therefore, the posterior is sampled integrating over β_1, \dots, β_m , realisations of these parameters can be generated using standard forward-filtering backwards-sampling (Frühwirth-Schnatter, 1994; Carter and Kohn, 1994) in the Gibbs sampler. The steps of the MCMC sampler are as follows.

Updating ψ

The full conditional distribution of $\psi_{i,1}$ is proportional to

$$\psi_{i,1}^{\lambda_i + \kappa_{i,1} - 3/2} \left(\frac{\lambda_i \rho_i}{(1 - \rho_i) \mu_i} \right)^{\kappa_{i,1}} \exp \left\{ -\frac{\psi_{i,1} \lambda_i}{\mu_i (1 - \rho_i)} \right\} \exp \left\{ -\frac{1}{2} \left[\frac{\beta_{i,1}^2}{\psi_{i,1}} + \frac{(\beta_{i,2} - \varphi_i \sqrt{\frac{\psi_{i,2}}{\psi_{i,1}}} \beta_{i,1})^2}{\psi_{i,2} (1 - \varphi_i^2)} \right] \right\},$$

the full conditional distribution of $\psi_{i,t}$ is proportional to

$$\begin{aligned} & \psi_{i,t}^{\lambda_i + \kappa_{i,t-1} + \kappa_{i,t} - 3/2} \left(\frac{\rho_i \lambda_i}{(1 - \rho_i) \mu_i} \right)^{\kappa_{i,t}} (1 - \varphi^2)^{1/2} \exp \left\{ -\frac{\lambda_i \psi_{i,t}}{\mu_i} \left(1 + \frac{2\rho_i \lambda_i}{(1 - \rho_i) \mu_i} \right) \right\} \\ & \times \exp \left\{ -\frac{1}{2} \left[\frac{(\beta_{i,t} - \varphi_i \sqrt{\frac{\psi_{i,t}}{\psi_{i,t-1}}} \beta_{i,t-1})^2}{\psi_{i,t} (1 - \varphi_i^2)} + \frac{(\beta_{i,t+1} - \varphi_i \sqrt{\frac{\psi_{i,t+1}}{\psi_{i,t}}} \beta_{i,t})^2}{\psi_{i,t+1} (1 - \varphi_i^2)} \right] \right\} \end{aligned}$$

if $1 < t < T$ and the full conditional distribution of $\psi_{i,T}$ is proportional to

$$\psi_{i,T}^{\lambda_i + \kappa_{i,T-1} - 3/2} (1 - \varphi_i^2)^{-1/2} \exp \left\{ -\frac{\lambda_i \psi_{i,T}}{(1 - \rho_i) \mu_i} \right\} \exp \left\{ -\frac{1}{2} \left[\frac{(\beta_{i,T} - \varphi_i \sqrt{\frac{\psi_{i,T}}{\psi_{i,T-1}}} \beta_{i,T-1})^2}{\psi_{i,T} (1 - \varphi_i^2)} \right] \right\}.$$

The parameter can be updated using a Metropolis-Hastings random walk step where a new value of $\psi_{i,t}$ is proposed from a Normal distribution with the current value of $\psi_{i,t}$ as the mean and variance $\tau_{i,t}$. The variance $\tau_{i,t}$ is tuned adaptively using the method of Atchadé and Rosenthal (2005) which updates $\tau_{i,t}$ to $\tau_{i,t} + \frac{1}{\gamma \eta} (\alpha_\gamma - 0.3)$ where γ is the iteration number of the MCMC sampler, α_γ is the acceptance probability of the Metropolis-Hastings step at the γ -th iteration and $0.5 < \eta < 1$. This choice of η guarantees that the average acceptance rate converges to 0.3. We use the value $\eta = 0.6$ in our examples.

Updating $\kappa_{i,t}$

The full conditional distribution of $\kappa_{i,t}$ is proportional to

$$\left(\frac{\lambda_i}{(1-\rho_i)\mu_i} \right)^{\lambda_i + \kappa_{i,t}} \left(\frac{\psi_{i,t}\lambda_i\rho_i}{(1-\rho_i)\mu_i} \right)^{\kappa_{i,t}} \psi_{i,t+1}^{\lambda_i + \kappa_{i,t} - 1} \frac{1}{\kappa_{i,t}! \Gamma(\lambda_i + \kappa_{i,t})}$$

for $1 \leq t \leq T-1$. We update this parameter using a Metropolis-Hastings step with the proposed value $\kappa'_{i,t}$ where $\kappa'_{i,t} = \kappa_{i,t} + 1$ with probability 1/2 and $\kappa'_{i,t} = \kappa_{i,t} - 1$ with probability 1/2. The move is rejected if $\kappa'_{i,t} < 0$.

Updating λ_i , μ_i and ρ_i

These parameters are updated using a Metropolis-Hastings sampler. Let $\theta_i = (\lambda_i, \mu_i, \rho_i)$. Updating from the full conditional distribution of θ_i can lead to slow mixing. This problem can be addressed by jointly proposing a new value θ'_i from a transition kernel $q(\theta_i, \theta'_i)$ with new values of ψ_i and κ_i conditional on θ'_i using a retrospective method (Papaspiliopoulos et al., 2007). These new values of ψ_i and κ_i will be denoted ψ'_i and κ'_i and are proposed in the following way

$$\psi'_{1,i} = \begin{cases} \frac{\lambda_i \mu'_i}{\lambda'_i \mu_i} \psi_{i,1} + \text{Ga}(\lambda'_i - \lambda_i, \lambda'_i / \mu'_i) & \text{if } \lambda'_i > \lambda_i \\ \frac{\lambda_i \mu'_i}{\lambda'_i \mu_i} \psi_{i,1} \text{Be}(\lambda'_i, \lambda_i - \lambda'_i) & \text{if } \lambda'_i < \lambda_i \end{cases}$$

$$\kappa'_{i,t} = \begin{cases} \kappa_{i,t} + \text{Pn} \left(\frac{\rho_{j'}'}{1-\rho_{j'}} \frac{\psi'_{i,t}\lambda'_i}{\mu'_i} - \frac{\rho_j}{1-\rho_j} \frac{\psi_{i,t}\lambda_i}{\mu_i} \right) & \text{if } \frac{\rho_{i'}'}{1-\rho_{i'}} \frac{\psi'_{i,t}\lambda'_i}{\mu'_i} > \frac{\rho_i}{1-\rho_i} \frac{\psi_{i,t}\lambda_i}{\mu_i} \\ \text{Bi} \left(\kappa_{i,t}, \frac{\rho_{i'}'(1-\rho_i)}{\rho_i(1-\rho_{i'}')} \frac{\mu_i \psi'_{i,t}\lambda'_i}{\mu'_i \psi_{i,t}\lambda_i} \right) & \text{if } \frac{\rho_{i'}'}{1-\rho_{i'}} \frac{\psi'_{i,t}\lambda'_i}{\mu'_i} < \frac{\rho_i}{1-\rho_i} \frac{\psi_{i,t}\lambda_i}{\mu_i} \end{cases}$$

for $t = 1, \dots, T-1$ and

$$\psi'_{i,t} = \begin{cases} \frac{\lambda_i \mu'_i (1-\rho_{i'})}{\lambda'_i \mu_i (1-\rho_i)} \psi_{i,t} + \text{Ga} \left(\lambda'_i + \kappa'_{i,t-1} - \lambda_i - \kappa_{i,t-1}, \frac{\lambda'_i}{\mu'_i (1-\rho_{i'})} \right) & \text{if } \lambda'_i + \kappa'_{i,t-1} > \lambda_i + \kappa_{i,t-1} \\ \frac{\lambda_i \mu'_i (1-\rho_{i'})}{\lambda'_i \mu_i (1-\rho_i)} \psi_{i,t} \text{Be}(\lambda'_i + \kappa'_{i,t-1}, \lambda_i + \kappa_{i,t-1} - \lambda'_i - \kappa'_{i,t-1}) & \text{if } \lambda'_i + \kappa'_{i,t-1} < \lambda_i + \kappa_{i,t-1} \end{cases}$$

for $t = 2, \dots, T$. The acceptance ratio for the Metropolis-Hastings algorithm is

$$\min \left\{ 1, \frac{p(y|X, \psi', \sigma^2)p(\theta)q(\theta', \theta)}{p(y|X, \psi, \sigma^2)p(\theta)q(\theta, \theta')} \right\}.$$

It is possible to update λ_i , μ_i and ρ_i jointly but we have updated each parameter one-at-a-time using a random walk proposal. These proposals are: $\log \lambda_i' \sim N(\log \lambda_i, \tau_i^\lambda)$, $\log \mu_i' \sim N(\log \mu_i, \tau_i^\mu)$ and

$$\log \rho_i' - \log (1 - \rho_i') \sim N(\log \rho_i - \log (1 - \rho_i), \tau_i^\rho).$$

The values of τ_i^λ , τ_i^μ and τ_i^ρ are tuned automatically using the method described in the section on updating ψ .

Updating φ_i

The full conditional distribution of φ_i is proportional to

$$p(y|X, \psi, \sigma^2, \varphi) \rho^{\varphi_i 77.6} (1 - \varphi_i)^{2.4}.$$

The parameter can be updated using a Metropolis-Hastings random walk step where the new value φ_i' is simulated by

$$\log \varphi_i' - \log (1 - \varphi_i') \sim N(\log \varphi_i - \log (1 - \varphi_i), \tau_i^\varphi).$$

The variance τ_i^φ is updated automatically using the method described in the updating of ψ .

Updating σ_t^2

The full conditional distribution of σ_t^2 follows a generalised inverse Gaussian distribution which has density

$$\frac{(c/d)^{h/2}}{2K_h(\sqrt{cd})} (\sigma_t^2)^{h-1} \exp \left\{ -\frac{1}{2} \left(c\sigma_t^2 + \frac{d}{\sigma_t^2} \right) \right\},$$

where K_h is a modified Bessel function of the second kind. The parameter values of the full distribution for σ_t^2 are

$$d = \left(y_t - \sum_{i=0}^m \beta_{i,t} x_{i,t} \right)^2, \quad t = 1, \dots, T,$$

$$c = \begin{cases} \frac{2\lambda^\sigma}{\mu^\sigma(1-\rho^\sigma)}, & t = 1, T \\ 2\frac{\lambda^\sigma + \rho^\sigma \lambda^\sigma}{\mu^\sigma(1-\rho^\sigma)}, & 1 < t < T \end{cases}$$

and

$$h = \begin{cases} \kappa_t^\sigma + \lambda^\sigma - 0.5, & t = 1 \\ \kappa_t^\sigma + \kappa_{t-1}^\sigma + \lambda^\sigma - 0.5, & 1 < t < T \\ \kappa_{t-1}^\sigma + \lambda^\sigma - 0.5, & t = T \end{cases}.$$

Updating κ_t^σ

The full conditional distribution of κ_t^σ is proportional to

$$\left(\frac{\lambda^\sigma}{(1-\rho^\sigma)\mu^\sigma} \right)^{\lambda^\sigma + \kappa_t^\sigma} \left(\frac{\sigma_t^2 \lambda^\sigma \rho^\sigma}{(1-\rho^\sigma)\mu^\sigma} \right)^{\kappa_t^\sigma} (\sigma_{t+1}^2)^{\lambda^\sigma + \kappa_t^\sigma - 1} \frac{1}{\kappa_t^\sigma! \Gamma(\lambda^\sigma + \kappa_t^\sigma)}$$

for $1 \leq t \leq T - 1$. We update this parameter using a Metropolis-Hastings step with the proposed value $\kappa_t^{\sigma'}$ where $\kappa_t^{\sigma'} = \kappa_t^\sigma + 1$ with probability 1/2 and $\kappa_t^{\sigma'} = \kappa_t^\sigma - 1$ with probability 1/2. The moves is rejected if $\kappa_t^{\sigma'} < 0$.

Updating μ^*

The full conditional distribution of μ^* is proportional to

$$(\mu^* + 2\beta^*)^{-3} \left(\frac{\lambda^*}{\mu^*}\right)^{m\lambda^*} \exp\left\{-\frac{\lambda^*}{\mu^*} \sum_{i=1}^m \mu_i\right\},$$

where m is the number of regressors and b^* a positive constant. The parameter can be updated using a Metropolis-Hastings random walk step where the proposed value μ^{*l} is simulated according to $\log \mu^{*l} \sim N(\log \mu^*, \tau^{\mu^*})$. The variance τ^{μ^*} is tuned automatically using the method described in the updating of ψ . In both the simulated and empirical examples $\beta^* = 0.05$.

Updating λ^*

The full conditional distribution of λ^* is proportional to

$$\exp\{-\lambda^*/s^*\} \left(\frac{\lambda^{*\lambda^*}}{\mu^{*\lambda^*}\Gamma(\lambda^*)}\right)^m \exp\left\{-\frac{\lambda^*}{\mu^*} \sum_{i=1}^m \mu_i\right\} \prod_{i=1}^m \mu_i^{\lambda^*}.$$

The parameter can be updated using a Metropolis-Hastings random walk step where the proposed value λ^{*l} is simulated according to $\log \lambda^{*l} \sim N(\log \lambda^*, \tau^{\lambda^*})$. The variance τ^{λ^*} is again tuned automatically using the method described in the updating of ψ . In both simulated and empirical examples $s^* = 10$.

We ran the MCMC sampler for both the simulated and empirical examples using 12000 iterations and discarded the initial 2000 as burn in, and stored every 2nd draw. An iMac with a 2.66 GHz Intel Core i5 processor, and memory 4 GB 1067 MHz DDR3 was used, and the computing time was between 5 to 10 hours for the equity premium example and between 12 to 19 for the inflation examples.

4 Simulated Example

The following simulated example illustrates the ability of the NGAR prior to allow time-varying sparsity in dynamic regression. We generated the data from equation (1) with $m = 5$, $x_{i,t} \sim N(0, I_5)$ and $x_{i,1}, \dots, x_{i,T}$ independent. We introduced five regression coefficients: $\beta_{1,t}$ followed an AR(1) process with AR parameter 0.97 and a Normal marginal distribution with mean 2 and variance 0.25, $\beta_{2,t}$ followed an AR(1) process with AR parameter 0.97 and a Normal marginal distribution with mean 0 and variance 0.25 for $t < 100$ and $\beta_{2,t} = 0$ for $t > 100$ with $\beta_{2,1} \sim N(2, 0.25)$,

$$\beta_{3,t} = \begin{cases} 0 & \text{if } t \leq 20, 51 \leq t \leq 120, \text{ and } 151 \leq t \leq 200 \\ -2 & \text{if } 21 \leq t \leq 50, \text{ and } 121 \leq t \leq 150 \end{cases},$$

and $\beta_{0,t}$ (the intercept), $\beta_{4,t}$ and $\beta_{5,t}$ were zero for all times. The innovation variance σ_t^2 was generated using an AR(1) process on the log scale

$$\log \sigma_t^2 = \log(0.01) + 0.97(\log \sigma_{t-1}^2 - \log(0.01)) + \sqrt{\frac{0.01}{1 - 0.97^2}} \nu_t$$

where $\nu_t \sim N(0, 1)$. The initial value of each parameter was drawn from its stationary distribution. The generated values for the regression coefficients are shown in the first row of Figure 3 where $\beta_{1,t}$ is always important, $\beta_{4,t}$ and $\beta_{5,t}$ are never important, the importance of $\beta_{2,t}$ tends to decrease until $t = 100$ after which the value of $\beta_{2,t}$ is zero, and $\beta_{3,t}$ enters and exits the model abruptly on two occasions.

The second row of Figure 3 shows the estimated regression coefficients which follow the true values closely. The posterior median and 95% credible interval of $\beta_{0,t}$, $\beta_{4,t}$ and $\beta_{5,t}$ are very close to zero. The NGAR prior is also able to adapt to the changing importance of $\beta_{2,t}$ and the abrupt behaviour of $\beta_{3,t}$ in the model. The third row of Figure 3 shows the posterior inference on the regressor time-varying relevance factor $\sqrt{\psi_{i,t}}$. The values for the intercept, x_1 , x_4 and x_5 are fairly constant and close to zero.

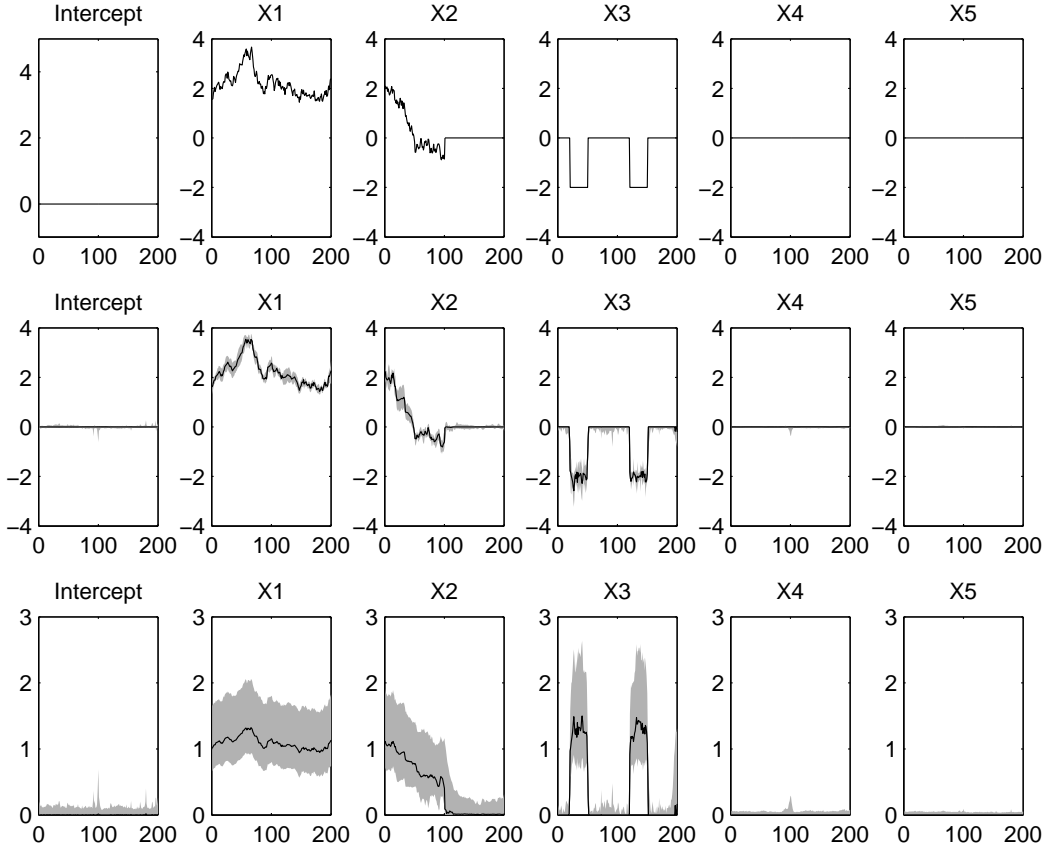


Figure 3: Simulated: First row - true regression coefficients $\beta_{i,t}$, Second row - the posterior median (solid line) and 95% credible interval (grey shading) of $\beta_{i,t}$, and Third row - the posterior median (solid line) and 95% credible interval (grey shading) of $\sqrt{\psi_{i,t}}$.

The posterior median of the relevance factor for x_2 is decreasing until about $t = 100$ and then takes a value close to zero, whereas the posterior median of the relevance factor for x_3 correctly replicates the abrupt entries and exits of the regressor from the model. We can therefore conclude that the results of this simulation illustrate the ability of the NGAR prior to shrink values close to zero when the data supports.

5 Empirical Examples

In this section we apply the dynamic regression model with an NGAR prior for the regression coefficients to both equity premium and inflation datasets. Our aim is to

provide evidence that the NGAR prior adequately accounts for the time-varying effect of the regression coefficients, and produces good out-of-sample forecasts by identifying those variables that lead to the best predictive performance. This latter point also relates to the identification of variables that not only lead to better prediction but also the time period for which they are most relevant.

In Section 2, we discussed the NGAR prior for the time-varying regression coefficient, $\beta_{i,t}$, and the time-varying relevance of the i^{th} regressor, $\psi_{i,t}$. Recall that a smaller value of $\psi_{i,t}$ implies that the i -th predictor is less important at time t . We present two plots for each data set. The first plot displays the posterior median of $\sqrt{\psi_{i,t}}$ as it changes over time and shows the importance of each predictor over time (including periods where it has most impact). The second plot displays the posterior median of $\beta_{i,t}$ over time, which evaluates the effect of each relevant predictor. It is natural to expect that when a predictor is not relevant (when the posterior median of $\sqrt{\psi_{i,t}}$ is zero), then the value of $\beta_{i,t}$ should be very close to zero. We also plot the time-varying innovation variance, σ_t^2 , to identify the periods when σ_t^2 changes. All plots display the 95% credible intervals (CI).

5.1 Equity Premium Prediction

The set of variables relevant to equity premium forecasting is large. It ranges from variables relating to dividends and earnings such as dividend yield and price earnings ratio to interest rates, bond yields, and inflation. For our empirical study we use the same data set as Goyal and Welch (2008). The response variable is the value weighted monthly return of the S&P 500 obtained from the CRSP database. For our illustration we considered all the twelve predictors (see Appendix A for the complete list), including cross sectional beta premium (CSP) (see Roll and Ross, 1994). For this reason the sample period is restricted from May 1937 to December 2002, as it is the period where values of CSP are available.

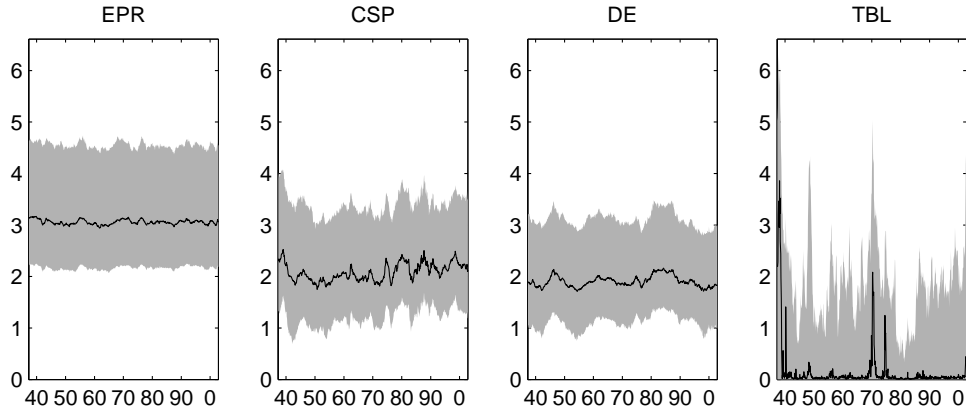


Figure 4: Equity Premium: Posterior medians (solid line) of the time-varying regression relevance, $\sqrt{\psi_{i,t}}$. (y_{axis} = value of $\sqrt{\psi_{i,t}}$, x_{axis} = time) with 95% credible interval (grey shading)

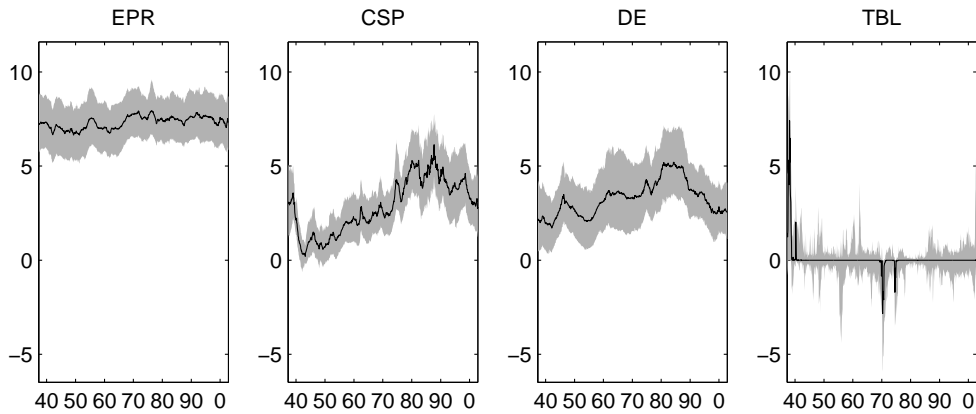


Figure 5: Equity Premium: Posterior medians (solid line) of the time-varying regression coefficients, $\beta_{i,t}$. (y_{axis} = value of $\beta_{i,t}$, x_{axis} = time) with 95% credible interval (grey shading)

The plots of the posterior medians and 95% credible intervals (CI's) of $\sqrt{\psi_{i,t}}$ for various predictors, and the posterior medians and 95% CI's of $\beta_{i,t}$ for those predictors are displayed in Figures 4 and 5 respectively. From the original list of twelve predictors, eight had posterior medians and 95% CI's for both $\sqrt{\psi_{i,t}}$ and $\beta_{i,t}$ that were very close to zero, and therefore are not displayed in our plots. These excluded variables were: B/M, LTY, NTIS, INFL, LTR, D/Y, and DFY (see Appendix A for full details). The four most relevant predictors of the equity premium, which are displayed in Figures 4

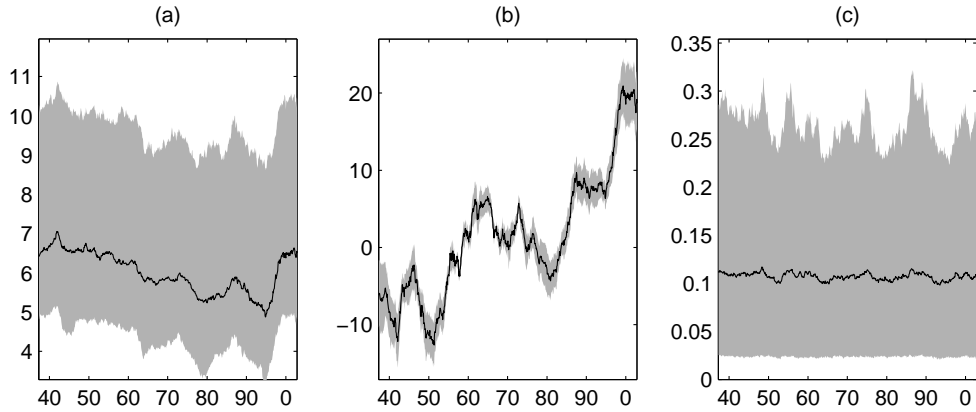


Figure 6: Equity Premium: (a) the posterior median (solid line) and 95% CI (grey shading) of the relevances of the intercept $\sqrt{\psi_{0,t}}$, (b) the posterior median (solid line) and 95% CI (grey shading) of the intercept $\beta_{0,t}$, and (c) posterior median (solid line) and 95% CI (grey shading) of the time-varying innovation volatility, σ_t^2 . (y_{axis} = values of $\sqrt{\psi_{0,t}}$, $\beta_{0,t}$ and σ_t^2 respectively, x_{axis} = time)

and 5, are: EPR (earnings price ratio), CSP (relative valuation of high and low beta stocks), DE (dividend payout ratio), and TBL (3m T-bill rate). The relevance of EPR is relatively constant over time. The same is true for its regression coefficient, which has a posterior median around 7 for the whole period. EPR has a positive effect on the equity premium for the whole period which is expected as it signals a firm's profitability. The regression coefficients for CSP and DE show more fluctuation in their relevance over time. However, their relevance is still relatively constant over time. The coefficient of CSP is almost always positive. It increases from the mid 1950's up to the mid 1980's and then it decreases. In addition we can also observe an oscillating pattern within this gradual increase and decrease of the CSP effect. One possible explanation is that within each decade there are years of high economic growth followed by years of slow growth. The beta of the firm is a measure of the firm risk that is attributed to the market and cannot be diversified. The beta will be high in times of recession and will affect equity premiums more than during periods of high growth. The coefficient of DE is also positive for all time periods, with its effect being largest during

the 1980's, the period when the Reagan administration began the deregulation of US financial markets. The effect decreases from the 1990's to 2010's and this could be due to the shift of emphasis in investment decisions from DE to firm growth prospects. The final important predictor is TBL. Its importance clearly changes from period to period. It is a very important predictor during the 1930's, 1950's and 1970's. During these periods its coefficient is positive except in the 1970's when it is negative. This could be attributed to the 1973 oil shock which led the US (and the rest of the world) into recession. During this period the interest rates soared to the double digits thus negatively affecting equity premiums. Goyal and Welch (2008) do not provide estimates for the regression coefficients but look at the importance of predictors by running simple regressions for different periods within the sample.

Figure 6 displays the time-varying relevance and the effect of the intercept (the first two plots) and the behaviour of the innovation variance, σ_t^2 over time. The importance of the intercept is fairly constant. Its effect is increasing over time and is positive with the exception of the period of the second World War and the beginning of the 1950's which was a period of reorganisation following the War. The innovation variance is fairly constant over time, around 0.11.

5.2 Inflation Forecasting

Forecasts of inflation are usually classified according to the type of explanatory variables used. The size of the set of potential variables is huge and is usually split into four subsets: past inflation forecasts, where the explanatory variables are previous lags of inflation; Phillips curve forecasts, which involve activity variables, such as economic growth rate or output gap, unemployment rate, and lagged inflation; forecasts based on variables which are themselves forecasts of asset prices (combination indices), term structures of nominal debt, and consumer surveys; and forecasts based on other exogenous variables such as government investment, the number of new private houses

built, etc.

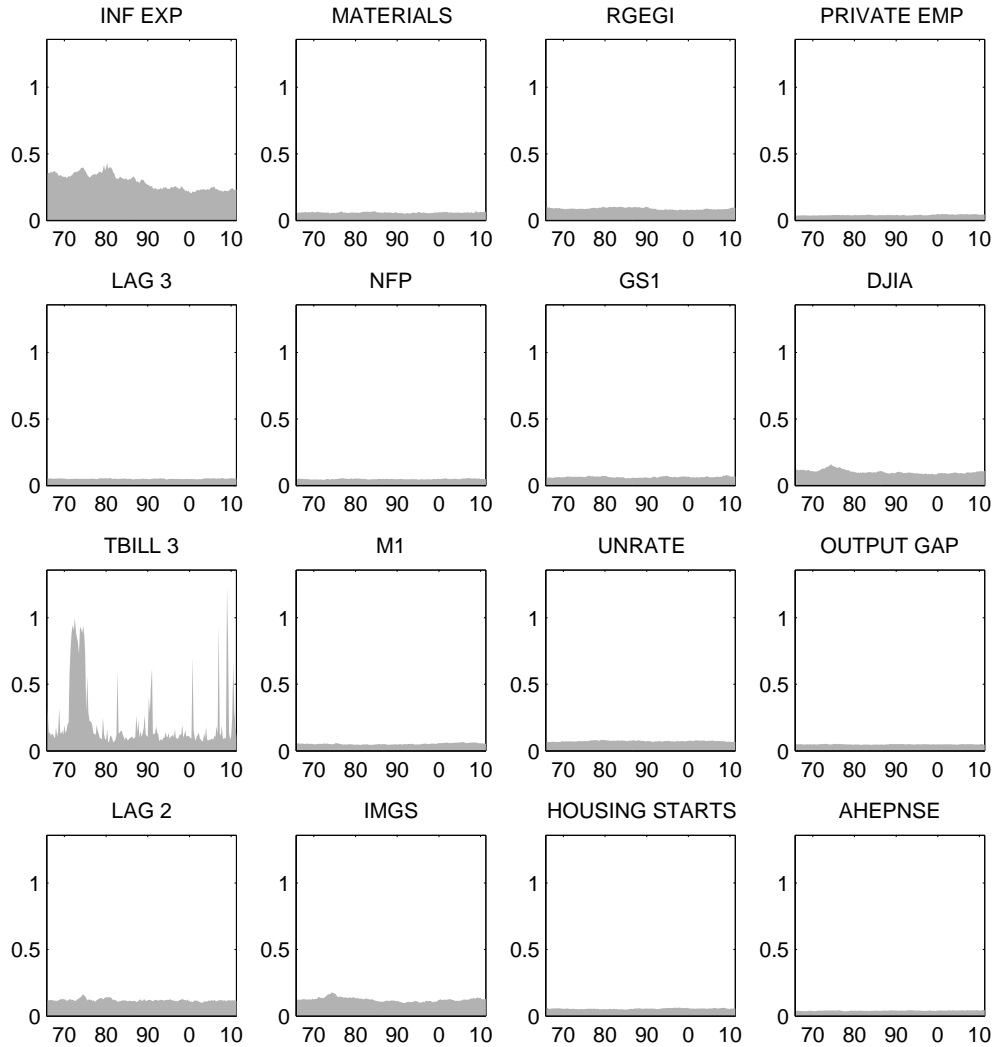


Figure 7: PCE deflator: Posterior medians (solid line) of the time-varying regression relevance, $\sqrt{\psi_{i,t}}$. (y_{axis} = value of $\sqrt{\psi_{i,t}}$, x_{axis} = time) with 95% CI (grey shading)

For our inflation forecasting study we constructed a data set using data series obtained from: FRED, the economic database of the Federal Reserve Bank of St Louis, the consumer survey database of the University of Michigan, the Federal Reserve Bank of Philadelphia, and the Institute of Supply Management. We use two different quarterly measures of US inflation as the response variable, the personal consumption expenditure (PCE) deflator and the gross domestic product (GDP) deflator. We therefore fit two separate models. The sample period for both is from the second quarter of 1965

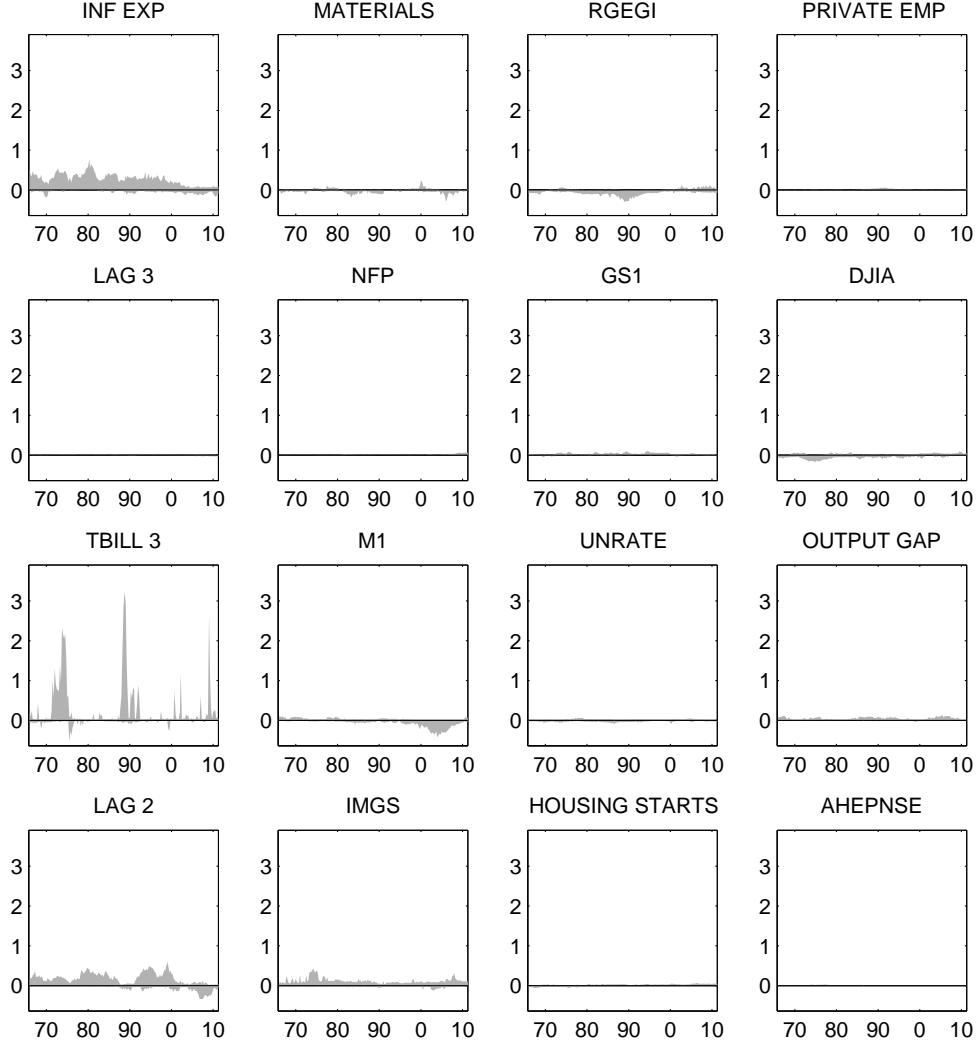


Figure 8: PCE deflator: Posterior medians (solid line) of the time-varying regression coefficients, $\beta_{i,t}$. (y_{axis} = value of $\beta_{i,t}$, x_{axis} = time) with 95% CI (grey shading)

to first quarter of 2011. Our data set includes 31 predictors, from activity and term structure variables to survey forecasts and previous lags. The full list with details of each is included in Appendix A.

We first discuss the results based on the PCE deflator. The plots of the posterior median and 95% CI of $\sqrt{\psi_{i,t}}$ for each of the predictors are displayed in Figure 7. The posterior median and 95% CI of $\beta_{i,t}$ for each predictor are displayed in Figure 8. From the thirty one predictors eight were noticeably relevant in forecasting the PCE deflator and are displayed in the plots of Figures 7 and 8. These are: TBill 3m rate, INF EXP (inflation expectation), Lag 4, IMGS (import of goods and services) growth, Lag2,

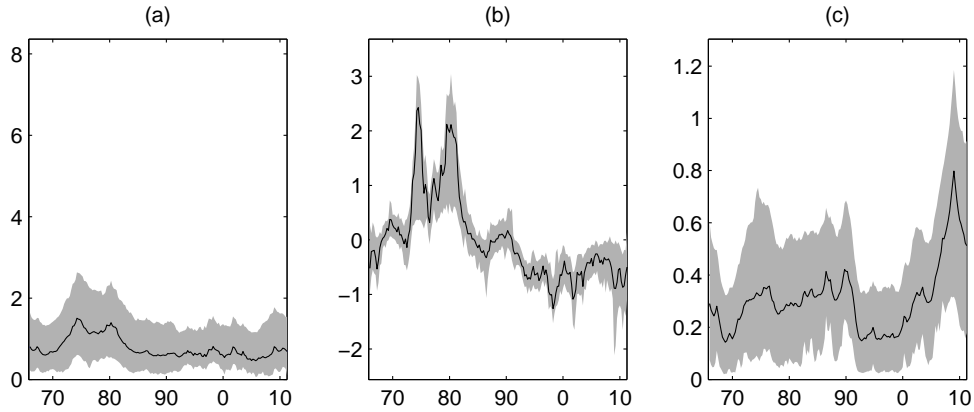


Figure 9: PCE deflator: (a) the posterior median (solid line) and 95% CI (grey shading) of the relevances of the intercept $\sqrt{\psi_{0,t}}$, (b) the posterior median (solid line) and 95% CI (grey shading) of the intercept $\beta_{0,t}$, and (c) posterior median (solid line) and 95% CI (grey shading) of the time-varying innovation volatility, σ_t^2 . (y_{axis} = values of $\sqrt{\psi_{0,t}}$, $\beta_{0,t}$ and σ_t^2 respectively, x_{axis} = time)

DJIA (S&P 500 returns), RGEGI (real government consumption expenditure and gross investment) growth, and M1 (narrow commercial bank money). Unlike equity premiums, the posterior median of $\beta_{i,t}$ is rarely far from zero. Therefore, we interpret the 95% credible interval as a set of plausible values for the regression coefficients. This allows us to identify times when a large absolute value of the regression coefficient is implausible and also variables for which it is implausible that the regression coefficient takes a particular sign (either positive or negative).

The 3m Tbill rate is the most important predictor of PCE. It appears that its importance is more obvious during periods of economic slowdown, financial crises and recessions. These periods are the 1970s the late 1980's and early and late 2000's. Its coefficient also reflects this. It is highly positive during all the aforementioned periods, so the PCE deflator increases when interest rates increase. The importance of INF EXP is decreasing over time. It is more important up to the early 1980's and then it becomes less important. Its coefficient is positive and it is decreasing over time. Its effect is more obvious from the 1970's up to the 1990's. From then on its coefficient approaches zero. The importance of the remaining six predictors is fairly constant over

time. Their coefficients are also fairly constant, however they provide more information as to which periods their effect was more obvious. The coefficient of IMGS growth is positive during the mid 1970's (the period of the oil shock) showing its impact on the PCE deflator. From then on its effect is small to almost insignificant. The coefficient of RGEGI growth is small for almost all periods except the late 1980's when it is negative. The coefficient of M1 is almost zero up to the start of the 2000's and then for most of the decade it is negative. Though empirical evidence suggests that rapid increases in the money supply lead to rapid increases in inflation, when we looked at the time plots of the changes in M1 and the changes in PCE deflator we found that during the late 2000's, the start of the current financial crisis, the rapid increase in the supply of money had a negative impact on inflation. It is therefore reasonable to observe this negative coefficient for M1. The coefficient of DJIA is very close to zero except during the period of the oil shock when it had a negative effect on inflation. Finally, of the two lags of PCE deflator it is the coefficient of the second lag that exhibits the most changes over time. It is positive up to the 1970's approaching zero during the end of this period, it picks up again during the 1980's approaching zero in the early 1990's and it then turns negative up to the late 2000's.

The time-varying importance and effect of the intercept are shown in the first two plots of Figure 9, whereas the third plot displays the behaviour of the innovation variance σ_t^2 . The intercept is more important during periods of slow economic growth and recession (1970's, late 1980's and 2000's). Its coefficient is positive in the 1970's and 1980's and turns negative in the 2000's. The innovation variance oscillates over time and reaches its peak around 2008, the start of the recent financial crisis.

The posterior median and 95% CI plot of $\sqrt{\psi_{i,t}}$ for each of the predictors of the GDP deflator are displayed in the plots of Figure 10. The posterior median and 95% CI plot of $\beta_{i,t}$ for each predictor are displayed Figure 11. Based on our NGAR model we identified sixteen important predictors for the GDP deflator, double the number of

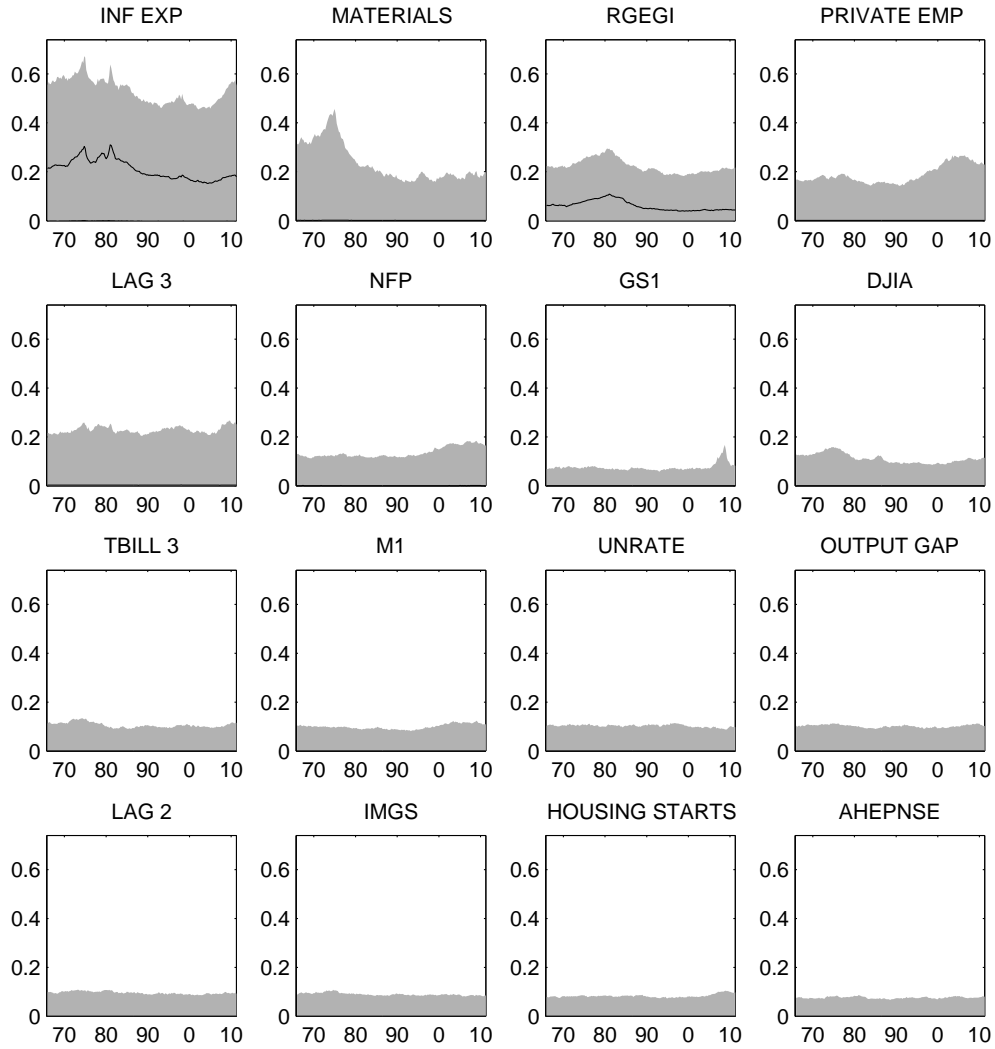


Figure 10: GDP deflator, $h = 1$: Posterior medians (solid line) of the time-varying regression relevance, $\sqrt{\psi_{i,t}}$. (y_{axis} = values of $\sqrt{\psi_{i,t}}$, x_{axis} = time) with 95% CI (grey shading)

those of the PCE deflator. This is reasonable as GDP reflects the value of all finished goods and services produced within the country whereas the PCE reflects personal consumption of goods and services. The five common predictors are: INF EXP, DJIA, TBILL 3m, RGEGI growth, and IMGS growth. INF EXP is clearly an important predictor of the GDP deflator. It is more important in the mid 1970's and mid 1980's. Its coefficient is positive from the start of our sample period up to the start of the 2000s, when it starts to approach zero. The lower band of its 95% CI suggests that it may also have a nega-

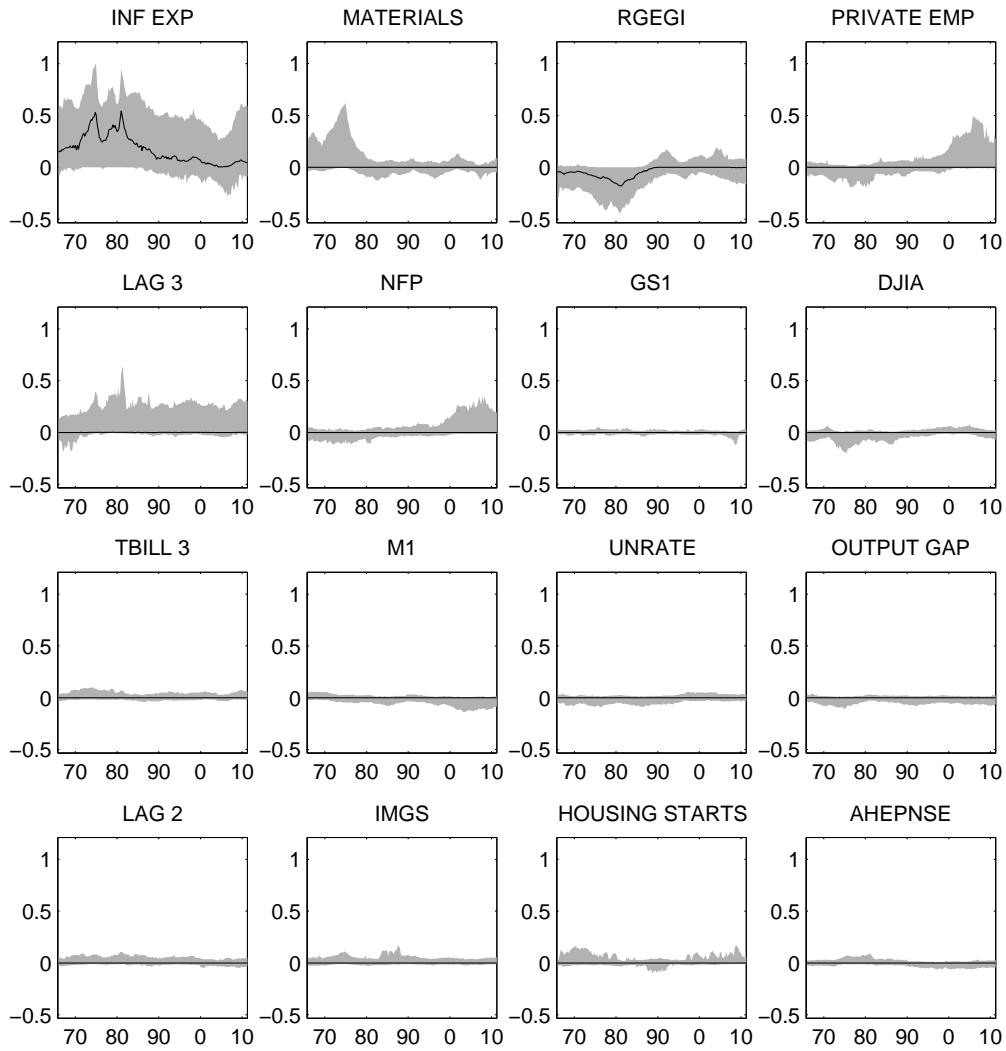


Figure 11: GDP deflator: Posterior medians (solid line) of the time-varying regression coefficients, $\beta_{i,t}$. (y_{axis} = values of $\beta_{i,t}$, x_{axis} = time) with 95% CI (grey shading)

tive effect on the GDP deflator. The TBILL 3m rate has on the other hand a small fairly constant effect on the GDP deflator as its coefficient is fairly constant at zero for all time periods. The TBILL's effect is clearly different to that on the PCE deflator as personal consumption expenditure is more directly affected by changes in interest rates. The coefficient of RGEGI growth is negative in the 1980's as was the case with the PCE deflator, however its effect on the GDP deflator is clearly more obvious. Unlike the case of the PCE deflator, the coefficient of M1 is fairly constant over time and very close to zero, whereas the coefficient of DJIA is very close to zero just as we observed for

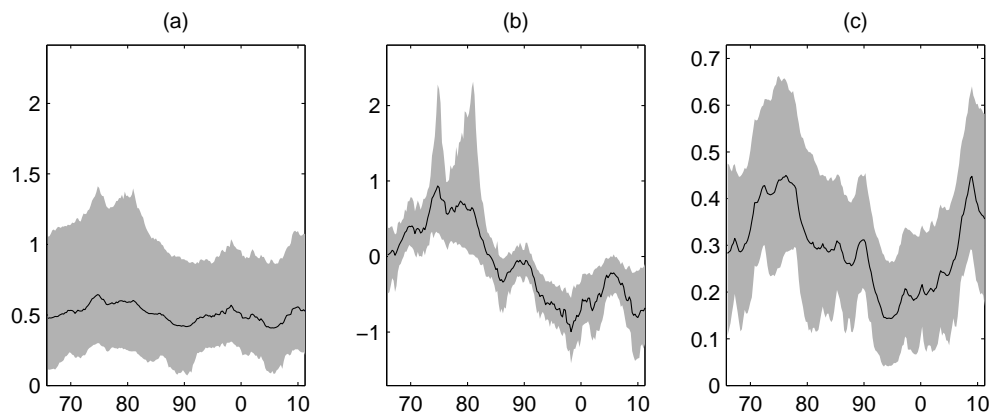


Figure 12: GDP deflator: (a) the posterior median (solid line) and 95% CI (grey shading) of the relevances of the intercept $\sqrt{\psi_{0,t}}$, (b) the posterior median (solid line) and 95% CI (grey shading) of the intercept $\beta_{0,t}$, and (c) posterior median (solid line) and 95% CI (grey shading) of the time-varying innovation volatility, σ_t . (y_{axis} = values of $\sqrt{\psi_{0,t}}$, $\beta_{0,t}$ and σ_t^2 respectively, x_{axis} = time)

the PCE deflator. In terms of the remaining eleven predictors, the coefficients of GS1, UNRATE, Output Gap, Lag 2 and AHEPNSE are fairly constant for all time periods. In the case of AHEPNSE (average hourly earning of private non managerial employees) we can observe a possible switch in the sign of its coefficient. During the 1980's it changes from positive to negative. The coefficients of the other six predictors exhibit more time-varying behaviour. For MATERIALS the coefficient is positive between the 1970's and 1980's possibly due the oil shock which led to increases in prices and thus to higher inflation. From the 1980's onwards this effect becomes smaller and perhaps at some time points turning negative. The coefficient of private employment shows an inverse pattern to that of materials. It starts fairly small (possibly negative) and it then becomes positive (from the 1990's to late 2000's). Periods of economic growth always provide more employment opportunities in the private sector. The coefficient of NFP (non farm payroll) also has the same effect on the GDP deflator. Finally the coefficient of the third lag of GDP deflator is almost always positive as is the case with that of IMGS growth. The time-varying importance and effect of the intercept are shown in the first two plots of Figure 12, whereas the third plot displays the behaviour of the in-

novation variance σ_t^2 . The intercept is more important during periods of slow economic growth and recession (1970's , late 1980's and 2000's). Its coefficient is positive in the 1970's and 1980's and turns negative in the 2000's. The innovation variance oscillates over time and reaches its peak in the mid 1970's and around 2008, two recessionary periods.

5.3 Comparison to other methods

We compare the predictive performance of the DR model with our NGAR prior to other Bayesian variable selection and regularisation methods that have recently been proposed for DR models with a large number of potential predictors. These methods are: Time Varying Dimension (TVD) models (Chan et al., 2012), the dynamic model average (DMA) approach (Koop and Korobilis, 2012), and the hierarchical shrinkage (HierShrink) prior of Belmonte et al. (2011). In case of the former two methods we use the priors suggested in the related papers and in case of the DMA we set the “forgetting” parameters $\lambda = \alpha = 0.99$, as the paper suggested. We also fit a rolling window Bayesian Model Averaging (BMA) using a g -prior for prediction. We use the default choices of Fernandez et al. (2001) for the g -prior with the previous k observations, i.e. y_{t-k}, \dots, y_{t-1} , to predict at each time point. Finally, we use the random walk model of Atkeson and Ohanian (2001) as a benchmark model. We focus on one step ahead forecasts and our comparison metric is the root mean square error (RMSE) using the posterior mean as our estimated calculated on the second half of the data

$$\sqrt{\frac{1}{T-s} \sum_{t=s+1}^T (y_t - \mathbb{E}[y_t | y_1, \dots, y_{t-1}, x_1, \dots, x_t])^2}$$

where $x_t = (x_{0,t}, x_{1,t}, \dots, x_{m,t})$ and $s = \lfloor \frac{T}{2} \rfloor$ (i.e. the largest integer less than or equal to $T/2$). The posterior predictive means of y_t for $t = s + 1, \dots, T$ were calculated using

particle filtering methods and includes uncertainty in all parameters.

	Equity Premium	PCE Inflation	GDP Inflation
RW	1.100	0.635	0.373
NGAR	0.977	0.611	0.410
DMA	1.01	0.660	0.422
TVD1	2.193	2.688	2.688
TVD2	0.986	0.623	0.481
TVD3	0.992	0.628	0.500
HierShrink	1.547	1.131	2.556
gprior ¹	2.822	0.796	0.660
gprior ²	1.648	0.712	0.588
gprior ³	1.282	0.681	0.516

Table 1: RMSE of out-of-sample prediction with different priors for the three data sets. The smallest RMSE for each data set is written in bold.

Table 1 displays the RMSE for all three data sets under the different models. There are three versions of TVD which make different assumptions about the evolution of the regression coefficients and which are fully described in (Chan et al., 2012). The window lengths for the three g-priors were 100 (gprior¹), 200 (gprior²) and 300 (gprior³) for the equity premium data and 50 (gprior¹), 70 (gprior²) and 90 (gprior³) for the inflation data. The choice of the window is controlled by the number of regressors included (which must be less than the window length) and the number of observations in the sample. The NGAR is the best performing approach for two data sets (equity premium and PCE inflation) and the second best performing for the GDP inflation data (with only the random walk giving better predictions). The TVD2 and TVD3 model and DMA also perform well across the three data sets. In general, the models which are the complexity of the model to change over time (NGAR, TV and DMA) outperform the other approaches (HierShrink and rolling window g-prior). This illustrates the importance of allowing time-variation in the relevance of regression coefficients. The poor performance of the HierShrink prior suggests that the double exponential prior may be unsuitable with these data and imply too little sparsity. This illustrates the importance of allowing for time varying sparsity in these data.

6 Discussion

This paper introduces a new approach to time-varying sparsity in dynamic regression models. The time-varying regression coefficients follow a stochastic process with a Normal-Gamma marginal distribution and smaller values of the shape parameter imply that the process will spend more time at values close to zero. This allows us to identify periods when regression coefficients are very close to zero and so are effectively removed from the model. A Normal-Gamma prior on the variance of the marginal distribution of $\beta_{i,t}$ encourages shrinkage of the whole path of $\beta_{i,t}$ close to zero. The empirical examples illustrate that the method leads to smaller out-of-sample predictive RMSE than several recently proposed approach to dynamic regression models with many regressors.

References

- Ang, A. and Bekaert, G. (2007). Stock return predictability: is it there?, *Review of Financial Studies* 20(3): 657–707.
- Atchadé, Y. F. and Rosenthal, J. S. (2005). On Adaptive Markov chain Monte Carlo algorithms, *Bernoulli* 11: 815–828.
- Atkeson, A. and Ohanian, L. (2001). Are Philips curves useful for forecasting inflation?, *Quarterly review*, Federal Reserve Bank of Minneapolis.
- Belmonte, M., Koop, G. and Korobilis, D. (2011). Hierarchical shrinkage in time varying parameter models, *CORE Discussion Papers 2011036*, Universite Catholique de Louvain.
- Bishop, C. M. and Tipping, M. E. (2000). Variational relevance vector machines, *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 46–53.

- Canova, F. (2007). G-7 inflation forecasts: random walk, Phillips curve, or what else?, *Macroeconomic Dynamics* 11: 1–30.
- Caron, F. and Doucet, A. (2008). Sparse Bayesian nonparametric regression, *Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland*.
- Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika* 81: 541–553.
- Carvalho, C., Polson, N. and Scott, J. (2010). The horseshoe estimator for sparse signals, *Biometrika* 97: 465–480.
- Chan, J., Koop, G., Leon-Gonzales, R. and Strachan, R. (2012). Time Varying Dimension Models, *Journal of Business and Economic Statistics* 30: 358–367.
- Cogley, T. and Sargent, T. (2001). Evolving post World War II US inflation dynamics, *NBER Macroeconomics Annual*, MIT Press/NBER, pp. 331–373.
- Cogley, T. and Sargent, T. (2005). Drifts and volatilities: monetary policies and outcomes in the post World War US, *Review of Economic Dynamics* 8(2): 262–302.
- Fernandez, C., Ley, E. and Steel, M. (2001). Model uncertainty in cross-country growth regressions, *Journal of Applied Econometrics* 16: 563–576.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of Time Series Analysis* 15: 183–202.
- Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models, *Journal of Econometrics* 154: 85–100.
- Gerlach, R., Carter, C. and Kohn, R. (2000). Efficient Bayesian Inference in Dynamic Mixture Models, *Journal of the American Statistical Association* 95: 819–828.

- Gourieroux, C. and Jasiak, J. (2006). Autoregressive gamma processes, *Journal of Forecasting* **25**: 129–152.
- Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**: 1455–1508.
- Griffin, J. and Brown, P. (2010). Inference with normal-gamma prior distributions in regression models., *Bayesian Analysis* **5**: 171–188.
- Groen, J., Paap, R. and Ravazzolo, F. (2009). Real time inflation forecasting in a changing world, *Technical Report 19*, Economic Institute, Erasmus University, Rotterdam.
- Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging, *International Economic Review* **53**: 867–886.
- Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence, *Review of Financial Studies* **21**: 1607–1652.
- Papaspiliopoulos, O., Roberts, G. O. and Skold, M. (2007). A general framework for the parameterization of hierarchical models, *Statistical Science* **22**(1): 59–73.
- Park, T. and Casella, G. (2008). The Bayesian Lasso, *Journal of the American Statistical Association* **103**: 672–680.
- Paye, B. and Timmermann, A. (2006). Instability of return prediction models, *Journal of Empirical Finance* **13**: 274–315.
- Pitt, M., Chatfield, C. and Walker, S. (2002). Constructing first order stationary autoregressive models via latent processes, *Scandinavian Journal of Statistics* **29**(4): 657–663.
- Pitt, M. and Walker, S. (2005). Constructing stationary time series models using auxiliary variables with applications, *Journal of the American Statistical Association* **100**(470): 554–564.

- Polson, N. G. and Scott, J. G. (2011). Shrink globally, act locally: sparse Bayesian regularization and prediction, *Bayesian Statistics 9*, Clarendon Press, pp. 501–538.
- Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**(821–852).
- Raftery, A., Karny, M. and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill, *Technometrics* **52**(52-66).
- Roll, R. and Ross, S. (1994). On the cross sectional relation between expected returns and betas, *The Journal of Finance* **49**: 101–121.
- Sims, C. (1980). Macroeconomics and reality, *Econometrica* **48**(1): 1–48.
- Stock, J. and Watson, M. (1996). Evidence on structural instability in macroeconomic time series relations, *Journal of Business and Economic Statistics* **14**: 11–30.
- Tipping, M. E. (2000). The Relevance Vector Machine, *Advances in Neural Information Processing Systems 12*, Cambridge, U. S., pp. 652–658.
- West, M. and Harrison, J. (1999). *Bayesian forecasting and dynamic models*, Springer Series in Statistics, 2nd edn, Springer.

A Data Appendix

Name	Description
B/M	ratio of book to market value for the Dow Jones Industrial Average
TBL	3m Treasury Bill: Secondary Market Rate
LTY	Difference between the long term yield on government bonds and treasury bill
NTIS	ratio of 12m moving sums of net issues by NYSE listed stocks to total year end market cap
INFL	Consumer Price Index
LTR	Long term government bond yield
SVAR	Sum of squared daily returns of S&P500
CSP	Cross-sectional beta premium (relative valuation of high and low beta firms)
D/Y	Dividend yield: difference between the log of dividends and the log of lagged prices (S&P500)
EPR	Earnings price ratio: difference between the log of earnings and the log of prices (S&P500)
DE	Dividend payout ratio: difference between the log of dividends and the log of earnings (S&P 500)
DFY	Default yield spread: difference between BAA and AAA-rated corporate bond yields

Table 2: Equity Return Data. Source: Goyal and Welch (2008)

Name	Description
GDP	Difference in logs of real gross domestic product
PCE	Difference in logs of real personal consumption expenditure
GPI	Difference in logs of real gross private investment
RGEGI	Difference in logs of real government consumption expenditure and gross investment
IMGS	Difference in logs of imports of goods and services
NFP	Difference in logs non-farm payroll
M2	Difference in logs M2 (commercial bank money)
ENERGY	Difference in logs of oil price index
FOOD	Difference in logs of food price index
MATERIALS	Difference in logs of producer price index (PPI) industrial commodities
OUTPUT GAP	Difference in logs of potential GDP level
GS10	Difference in logs of 10yr Treasury constant maturity rate
GS5	Difference in logs of 5yr Treasury constant maturity rate
GS3	Difference in logs 3yr Treasury constant maturity rate
GS1	Difference in logs 1yr Treasury constant maturity rate
PRIVATE EMPLOYMENT	Log difference in total private employment
PMI MANU	Log difference in PMI-manufacturing index
AHEPNSE	Log difference in average hourly earnings of private non management employees
DJIA	Log difference in Dow Jones Industrial Average Returns
M1	Log difference in M1 (narrow-commercial bank money)
ISM SDI	Institute for Supply Management (ISM) Supplier Deliveries Inventory
CONSUMER	University of Michigan consumer sentiment (level)
UNRATE	Log of the unemployment rate
TBILL3	3m Treasury bill rate
TBILL SPREAD	Difference between GS10 and TBILL3
HOUSING STARTS	Private housing (in thousands of units)
INF EXP	University of Michigan inflation expectations (level)
LAG1, LAG2, LAG3, LAG4	The first, second, third and fourth lags of the response variable

Table 3: Inflation Data. Sources: FRED database, Federal Reserve Bank of St.Louis, University of Michigan Consumer Survey database, Federal Reserve Bank of Philadelphia, and Institute of Supply Management.