Hierarchical Shrinkage in Time-Varying Parameter Models

Miguel A. G. Belmonte
University of Strathclyde

Gary Koop
University of Strathclyde

Dimitris Korobilis
Université Catholique de Louvain

June 22, 2011

Abstract

In this paper, we forecast EU-area inflation with many predictors using time-varying parameter models. The facts that time-varying parameter models are parameter-rich and the time span of our data is relatively short motivate a desire for shrinkage. In constant coefficient regression models, the Bayesian Lasso is gaining increasing popularity as an effective tool for achieving such shrinkage. In this paper, we develop econometric methods for using the Bayesian Lasso with time-varying parameter models. Our approach allows for the coefficient on each predictor to be: i) time varying, ii) constant over time or iii) shrunk to zero. The econometric methodology decides automatically which category each coefficient belongs in. Our empirical results indicate the benefits of such an approach.

Keywords: Forecasting; hierarchical prior; time-varying parameters; Bayesian Lasso

JEL Classification: C11, C52, E37, E47

1 Introduction

The goal of this paper is to forecast EU area inflation using many predictors. Our application (and many similar applications in macroeconomics) has several characteristics that require the development of statistical methods that depart from standard regression-based methods. To explain these departures and motivate our statistical methods, we begin with a generalized Phillips curve specification where inflation, π_t, depends on lags of inflation and other predictors (x_t). In this case, the generalized Phillips curve suitable for forecasting h periods ahead (using the direct method of forecasting) is:

$$\pi_{t+h} = \alpha + \sum_{j=0}^{p-1} \varphi_j \pi_{t-j} + \gamma x_t + \varepsilon_{t+h}. \tag{1}$$

On its own, such a model may be inappropriate for a couple of reasons. First of all the number of parameters to estimate may be large relative to the number of observations in the data set. That is, x_t may contain many predictors. In macroeconomic forecasting, we typically having hundreds of variables to choose from. For instance, De Mol, Giannone and Reichlin (2008) forecast with a regression involving over 100 variables. Banbura, Giannone and Reichlin (2010) forecast using a vector autoregression (VAR) with over 100 variables. Estimation of such models, where the number of parameters is large relative to the number of observations, can lead to imprecise estimation and over-fitting (i.e. the model can fit the noise in the data, rather than finding the pattern useful for forecasting). Both of these can lead to poor forecast performance. This has led many papers (including De Mol, Giannone and Reichlin 2008, and Banbura, Giannone and Reichlin 2010) to use Bayesian methods which use shrinkage to reduce over-fitting problems and improve forecast performance. Closely related to the idea of shrinkage is the idea of variable selection (which can be thought of as shrinking the coefficient on a predictor to zero). The challenge faced by the researcher is often that there are many potential predictors, most are likely to be unimportant but the researcher does not know, a priori, which ones are unimportant. Sequential hypothesis testing procedures run into serious pre-testing problems, which has led researchers to adopt various variable selection measures (see, e.g., George and McCulloch 1997 or Chipman, George and McCulloch, 2001). In the present paper, we draw on one promising approach to shrinkage and variable selection, the Lasso[1] (Park and Casella, 2008).

A second drawback of (1) is that it assumes parameters are constant over time. There is a plethora of evidence of structural breaks and other kinds of parameter change in macroeconomic variables (see, among many others, Stock and Watson, 1996, Cogley and Sargent, 2001, 2005, Primiceri 2005, Sims and Zha, 2006, and D'Agostino, Gambetti and Giannone, 2009). The negative consequences for forecasting of ignoring such parameter change has been stressed by, among many others, Clements and Hendry (1998, 1999) and Pesaran, Pettenuzzo and Timmerman (2006). In this paper, we use a time-varying parameter (TVP) regression model to model parameter change. Constant coefficient models such as (1) can already be over-parameterized. Adding time variation in parameters may exacerbate this problem, suggesting that shrinkage may be useful with TVP models. However, there have been relatively few papers which attempt to ensure shrinkage in TVP models (exceptions include Koop and Korobilis, 2009 and Koop, Leon-Gonzalez and Strachan, 2009).

The purpose of the present paper is to develop an econometric methodology which surmounts these two drawbacks and use it to forecast EU inflation. In particular we develop an econometric methodology which falls in the class of TVP regression models. However, it uses Bayesian shrinkage methods (based on the Lasso) to automatically classify coefficients into three categories: i) those which are time-varying, ii) those which are constant over time and iii) those which are zero (and, thus, the associated predictor does not appear in the model at all).

We extend ideas from the Bayesian Lasso literature (see Park and Casella, 2008) to the case of TVP regression models. TVP regression models are state space models, in order to extend Bayesian Lasso methods, we draw on and extend ideas relating to model selection in state space models developed in Frühwirth-Schnatter and Wagner (2010). The paper is

[1] Lasso is an abbreviation for "least absolute shrinkage and selection operator".

organized as follows: section 2 of the paper discusses our econometric methods. Section 3 uses these methods in an empirical exercise which forecasts EU inflation. We compare our TVP regression methods, involving hierarchical shrinkage, to a range of other common forecast procedures and find Lasso shrinkage to be particularly important on the time-varying coefficients. Section 4 concludes.

2 Econometric Methods

2.1 Overview

The TVP version of the generalized Phillips curve given in (1) can be written as a state space model:

$$\begin{aligned} \pi_{t+h} &= \theta_t^* z_t + \varepsilon_{t+h} \\ \theta_t^* &= \theta_{t-1}^* + \eta_t \end{aligned} \tag{2}$$

where the variable of interest is h-step ahead inflation defined as $\pi_{t+h} = (\log(P_{t+h}) - \log(P_t))$, $z_t = [1, \Delta\log(P_t), .., \Delta\log(P_{t-p+1}), x_t]$, x_t is a $q \times 1$ vector of exogenous predictors, and $\theta_t^* = \left(\alpha_t', \varphi_{t,0}', ..., \varphi_{t,p}', \gamma_t'\right)'$.[2] For the errors we assume $\varepsilon_t \sim N\left(0, \sigma_t^2\right)$ and $\eta_t \sim N(0, \Omega)$. The errors are assumed to be independent of each other and independent at all leads and lags. Note that Ω is of dimension $k \times k$ with $k = 1 + p + q$ which can be large relative to the number of observations. To keep the model relatively parsimonious, we assume Ω is a diagonal matrix, $\Omega = diag\left(\omega_1^2, ..., \omega_k^2\right)$. It is through Ω that we introduce shrinkage in the time-variation in coefficients (i.e. if ω_i^2 is zero then the i^{th} coefficient is constant over time, but larger values of ω_i^2 allow for more variation). Note that we are allowing for heteroskedasticity in the measurement equation. In particular, we will assume a standard stochastic volatility specification for σ_t^2.

[2]Note that, when forecasting with $h = 12$, this means our dependent variable is an annual inflation rate, but our explanatory variables are lags of monthly inflation. Such a timing convention has been found useful in many recent forecasting papers such as Stock and Watson (2011).

It proves convenient to write (2) in an equivalent way, separating out the initial condition, as:

$$\begin{aligned}\pi_{t+h} &= \theta z_t + \theta_t z_t + \varepsilon_{t+h} \\ \theta_t &= \theta_{t-1} + \eta_t \\ \theta_0 &= 0,\end{aligned} \tag{3}$$

where $\theta = \theta_0^*$ and $\theta_t = \theta_t^* - \theta$. This is the well-known result that, in TVP regression models, the initial condition for the states plays the role of a regression effect. Thus, (3) breaks the coefficients into a constant part (i.e. θ) and a time-varying part.

In order to incorporate shrinkage priors, in the TVP regression, we use one more transformation. We use notation where θ_i is the i^{th} constant coefficient, $\theta_{i,t}$ is the i^{th} state and adopt a similar subscript i,t notational convention with other variables and let $\widetilde{\theta}_{i,t} = \frac{\theta_{i,t}}{\omega_i}$. With these conventions, we can write the TVP regression model as:

$$\begin{aligned}\pi_{t+h} &= \sum_{i=1}^{k} \theta_i z_{i,t} + \sum_{i=1}^{k} \omega_i \widetilde{\theta}_{i,t} z_{i,t} + \varepsilon_{t+h} \\ \widetilde{\theta}_{i,t} &= \widetilde{\theta}_{i,t-1} + u_{i,t} \\ \widetilde{\theta}_{i,0} &= 0\end{aligned} \tag{4}$$

where $u_{i,t} \sim N(0,1)$ for $i = 1,..,k$. Fruhwirth-Schnatter and Wagner (2010) refer to this as a non-centered parameterization in their analysis of the dynamic linear trend model and argue for the advantages of this parameterization. Traditionally, Bayesian researchers have used inverted Gamma priors on error variances in state equations such as ω_i^2. Fruhwirth-Schnatter and Wagner (2010) argue (and present strong evidence) in favor of using a normal prior on ω_i. In this paper, we follow this approach and use a hierarchical normal prior for ω_i motivated by the Bayesian Lasso of Park and Casella (2008). We will also adopt a Lasso prior for θ_i (where θ_i for $i = 1,..,k$ are the constant coefficients on the predictors). We will explain these priors shortly, but note first that the Lasso provides shrinkage and, in terms of (4), we will have a model with the properties:

1. If ω_i is shrunk to 0, but θ_i is **not** shrunk to 0, then we have a model with a constant parameter on predictor i.

2. If ω_i is shrunk to 0, and θ_i is shrunk to 0, then predictor i is irrelevant for forecasting inflation.

3. If ω_i is **not** shrunk to 0, but θ_i is shrunk to 0, then we have a small time-varying coefficient on predictor i (since $\theta_{i,0} = 0$ the coefficient is restricted to start at a value of zero).

4. If ω_i is **not** shrunk to 0, and θ_i is **not** shrunk to 0, then we have an unrestricted time-varying coefficient on predictor i.

Thus, we have a methodology which decides, in an automatic fashion whether any predictor is important for forecasting inflation and, if so, whether it has a coefficient which is constant over time or time-varying.

2.2 The Prior

The model given by (4) is parameterized in terms of $\theta = (\theta_1, .., \theta_k)'$, $\widetilde{\theta} = \left(\widetilde{\theta}_{1,t}, .., \widetilde{\theta}_{k,t}\right)'$, $\omega = (\omega_1, .., \omega_k)'$ and σ_t^2. For σ_t^2 we adopt a standard stochastic volatility specification (see the Technical Appendix for details). Our innovation lies in the use of Lasso-type shrinkage priors for the remaining parameters. Such priors have been used with constant coefficient regressions in many places. For instance, Park and Casella (2008) is an important statistical exposition and De Mol, Giannone and Reichlin (2008) is an influential econometric treatment. To explain the basic ideas underlying the Lasso, consider the familiar normal linear regression model:

$$y = X\beta + \varepsilon,$$

where X is a $T \times k$ matrix of regressors, $\beta = (\beta_1, .., \beta_k)'$ and ε is $N\left(0, \sigma^2 I\right)$. Lasso estimates of β are penalized least squares estimates where β is chosen to minimize:

$$(y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^{k} |\beta_j|$$

where λ is a shrinkage parameter. Bayesian treatment of the Lasso arises by noting that Lasso estimates of β are equivalent to Bayesian posterior modes if independent Laplace priors are placed on the elements of β. Additional insight (and the MCMC algorithm used for Bayesian analysis) is obtained by noting that the Laplace distribution can be written as a scale mixture of normals with an exponential mixture density. Thus, Lasso shrinkage can be obtained by using a normal hierarchical prior for β. In this section, we describe how to extend this approach to our TVP regression model.

For the constant coefficients, θ, we use a hierarchical mixtures of normal prior inspired by the traditional Lasso. In particular, θ_i for $i = 1, .., k$ are assumed to be, a priori, independent with

$$\theta_i | \tau_i^2 \sim N\left(0, \tau_i^2\right)$$

and exponential mixing density:

$$\tau_i^2 | \lambda \sim Exp\left(\frac{\lambda^2}{2}\right).$$

This prior is almost the same as the traditional Lasso and has similar shrinkage properties.[3]

The state equation gives us a prior for $\widetilde{\theta}_t$ (for $t = 1, .., T$) of the form:

$$\widetilde{\theta}_t | \widetilde{\theta}_{t-1} \sim N\left(\widetilde{\theta}_{t-1}, I_k\right)$$

where $\widetilde{\theta}_0 = 0$.

[3]The one difference arises since the traditional Lasso has a prior variance of $\sigma^2 \tau_i^2$ instead of our τ_i^2. Since we allow for stochastic volatility and, thus, σ^2 is not constant over time, we cannot adopt the traditional approach.

We extend the Lasso approach to the time-varying coefficients by using a hierarchical prior for ω, each element of which is, a priori, conditionally independent with:

$$\omega_i|\xi_i^2 \sim N\left(0, \xi_i^2\right),$$

also with exponential mixing density:

$$\xi_i^2|\kappa \sim Exp\left(\frac{\kappa^2}{2}\right).$$

Note that, following Fruhwirth-Schnatter and Wagner (2010), we have a normal prior for ω_i. However, the hierarchical nature of the prior gives us the Lasso-type shrinkage of the elements of ω (thus, ensuring shrinkage on the time-varying coefficients).

The shrinkage parameters λ and κ lie at the bottom of the hierarchy and require priors of their own. For these we assume:

$$\lambda^2 \sim Gamma\left(a_1, a_2\right)$$

and

$$\kappa^2 \sim Gamma\left(b_1, b_2\right).$$

Note that the only prior hyperparameters which must be elicited are a_1, a_2, b_1, b_2 and the priors for the parameters in the stochastic volatility specification for σ_t.[4] The Technical Appendix discusses their elicitation.

2.3 Posterior Computation (MCMC algorithm)

In the constant coefficient regression model, an advantage of the Lasso prior is that, conditional on τ_i^2, we have a normal linear regression model with normal prior and standard textbook results can be used to derive the posterior, conditional on τ_i^2. An algorithm for

[4]The formulae in this paper parameterize the Gamma distribution so that its mean is $\frac{a_1}{a_2}$.

drawing τ_i^2 and λ is all that is required to complete an MCMC algorithm and this is provided by Park and Casella (2008). Our MCMC algorithm draws on this strategy to provide blocks for drawing θ, τ_i^2 and λ (conditional on the states and the other parameters in the model).

Similar intuition can be used to develop an algorithm for drawing $\widetilde{\theta}_t$ (for $t = 1, .., T$) conditional on ω (and other parameters). That is, conditional on these other parameters, the model becomes a normal linear state space model and standard methods for posterior simulation from such models can be used to draw $\widetilde{\theta}_t$. We use the algorithm of Carter and Kohn (1994). All that is required to complete an MCMC algorithm is a method for drawing ω and κ (conditional on all the other model parameters). However, these have simple forms.

The precise steps in our MCMC algorithm are given by.

1. Draw θ from the normal conditional posterior:

$$N\left(\left(\widetilde{z}'\widetilde{z}+V_\theta^{-1}\right)^{-1}\widetilde{z}'\widetilde{y},\left(\widetilde{z}'\widetilde{z}+V_\theta^{-1}\right)^{-1}\right)$$

where $V_\theta^{-1} = \left[diag\left(\tau_1^2, ..., \tau_k^2\right)\right]^{-1}$,$\widetilde{z} = \left[\frac{z_1}{\sigma_1}, ..., \frac{z_T}{\sigma_T}\right]'$, $\widetilde{y} = \left[\widetilde{y}_1, ..., \widetilde{y}_T\right]'$ and $\widetilde{y}_t = \frac{y_t - \omega\widetilde{\theta}_t z_t}{\sigma_t}$.

2. Draw $\widetilde{\theta}$ using the algorithm of Carter and Kohn (1994) for the state space model

$$\begin{aligned}\widehat{y}_t &= \widetilde{\theta}_t x_t + \sigma_t \varepsilon_t \\ \widetilde{\theta}_t &= \widetilde{\theta}_{t-1} + u_{i,t}\end{aligned}$$

where $x_t = \omega z_t$, $\widehat{y}_t = y_t - \theta z_t$, $u_{i,t} \sim N(0,1)$and the initial condition is zero ($\widetilde{\theta}_0 = 0$).

3. Draw ω from the normal conditional posterior

$$\omega \sim N\left(\left(\widehat{z}'\widehat{z}+V_\omega^{-1}\right)^{-1}\widehat{z}'\widehat{y},\left(\widehat{z}'\widehat{z}+V_\omega^{-1}\right)^{-1}\right)$$

where $V_\omega^{-1} = \left[diag\left(\xi_1^2, ..., \xi_k^2\right)\right]^{-1}$, $\widehat{z}_t = \frac{\widetilde{\beta}_t z_t}{\sigma_t}$ and $\widehat{y}_t = \frac{y_t - \beta z_t}{\sigma_t}$.

4. Draw τ^2 using the fact that $\frac{1}{\tau_i^2}$ each have independent inverse-Gaussian[5] conditional posteriors

$$IG\left(\sqrt{\frac{\lambda^2}{\beta_i^2}}, \lambda^2\right), \text{ for } i = 1, ..., k$$

5. Draw ξ^2 using the fact that $\frac{1}{\xi_i^2}$ each have independent inverse-Gaussian conditional posteriors

$$IG\left(\sqrt{\frac{\kappa^2}{\omega_i^2}}, \kappa^2\right), \text{ for } i = 1, ..., k$$

6. Draw λ^2 from the Gamma conditional posterior

$$Gamma\left(k + a_1, \frac{1}{2}\sum\nolimits_{j=1}^{k} \tau_j^2 + a_2\right)$$

7. Draw κ^2 from the Gamma conditional posterior

$$Gamma\left(k + a_1, \frac{1}{2}\sum\nolimits_{j=1}^{k} \xi_j^2 + a_2\right)$$

8. Draw σ_t^2 using the algorithm of Kim, Shephard and Chib (1998) for drawing from stochastic volatility models.

Draws from the predictive density are obtained using simulation methods as described, e.g., in section 2.1 of Cogley, Morozov and Sargent (2005). A nonparametric kernel smoothing algorithm is then used on these draws to obtain an approximation of the predictive density.

[5] If x is an inverse-Gaussian random variable with parameters a and b, then its p.d.f. is given by

$$p(x) = \sqrt{\frac{b}{2\pi}x^{-\frac{3}{2}}} \exp\left(-\frac{b(x-a)^2}{2a^2 x}\right)$$

for $x > 0$.

3 Empirical Results

We investigate the performance of our TVP model with hierarchical shrinkage using a relatively short data set where there are potentially a large number of predictors. We are primarily interested in whether LASSO shrinkage is a useful addition to TVP regression models and focus our empirical results on this issue. Accordingly, in addition to our full model described above (labelled "Lasso on constant and TVPs"), we consider several restricted special cases which are the same as the full model (with the same prior choices) except in certain specified dimensions. To be precise, we produce results for the following restricted versions of our full model:

1. Lasso only on constant coefficients: This model omits the Lasso prior on the time-varying coefficients and uses a relatively noninformative (non-hierarchical) normal prior for ω (see the Technical Appendix for details).

2. Lasso only on TVPs: This model omits the Lasso prior on the constant coefficients and uses a relatively noninformative (non-hierarchical) normal prior for θ (see the Technical Appendix for details).

3. TVP regression model: This model omits the Lasso prior everywhere, using the same priors for ω and θ as used in 1. and 2., respectively.

4. Constant coefficient model: This model nearly removes the TVP part of the model by setting prior hyperparameters $b_1 = 100000, b_2 = 0.001$ which implies an extremely tight prior on ω with prior mass concentrated very close to zero.

We consider variants of all models with stochastic volatility (see the Technical Appendix for details) and without. The latter homoskedastic models use the standard noninformative prior for the error variance. For the prior hyperparameters at the lowest level of the hierarchy, noninformative values of $a_1 = a_2 = b_1 = b_2 = 0.001$ are chosen unless otherwise specified.

3.1 Data

We forecast overall and core inflation using a variety of predictors reflecting a range of theoretical considerations. We use real-time data such that, at all points in time, we are using the data that would have been available to the forecaster at that point in time.[6] We have monthly EU data from February 1994 through November, 2010. Precise definitions of our variables follow.

Inflation is constructed as described after (2) based on the harmonized index of consumer prices (HICP). We use overall inflation as well as core inflation (which excludes energy and unprocessed food). Both measures are of interest to policymakers. Neither measure of inflation is seasonally adjusted.

The following predictors are used:

1. I_1MO: 1-month Euribor (Euro interbank offered rate).
2. I_1YR: 1-year Euribor (Euro interbank offered rate).
3. SENT: Percentage change in economic sentiment indicator.
4. STOCK_1: Percentage change in equity index - Dow Jones, Euro Stoxx, Economic sector index financial.
5. STOCK_2: Percentage change in equity index - Dow Jones Eurostoxx 50 index.
6. EXRATE: Percentage change in ECB real effective exchange rate (CPI deflated, broad group of currencies against euro).
7. IP: Percentage change in industrial production index.
8. LOANS: Percentage change in loans (total maturity, all currencies combined).
9. M3: Annual percentage change in monetary aggregate M3.

[6]The data is obtained from the ECB's Statistical Data warehouse with variables being updated in real time taken from its Real Time Data base. Complete descriptions of all variables can be found on the ECB's website. For some of the predictors complete real time data is only available from January 2001. For these variables we use non-real time data before this time. Our forecast evaluations begin in January 2001.

10. CAR: Registrations of new passenger cars.

11. OIL: Percentage change in oil price (brent crude -1 month forward).

12. ORDER: Change in order-book levels.

13. UNEMP: Standardised unemployment rate (all ages, male & female).

In addition, p lags of the logged first difference of the price index are used as predictors, as described after (2). All of the predictors are standardized to have mean zero and variance one. The dependent variable is standardized to have variance one.[7] Since our inflation variables are not seasonally adjusted, we also include an intercept and monthly dummies (omitting the January dummy). We forecast inflation a month ahead and a year ahead ($h = 1$ and $h = 12$).

The results in the body of the paper always use an intercept, $p = 12$, 11 monthly dummies and the 13 predictors listed above. Thus, we have 37 (possibly time-varying) coefficients to estimate with fewer than 18 years of data. In the Empirical Appendix, we present results using various subsets of the 37 variables. These include the case where the 13 predictors are omitted, leading to TVP-AR models, as well the case where the 13 variables and all lags are excluded from the model. This latter case leads to the unobserved components stochastic volatility (UCSV) model of Stock and Watson (2007).[8] Thus, we are investigate the usefulness of hierarchical shrinkage in the context of several popular classes of forecasting models.

3.2 Full Sample Results

Before comparing the forecast performance of the many models we consider, we present some full sample parameter estimates using our model with Lasso priors on constant and time-varying coefficients. For the sake of brevity, we present only results for core inflation.

[7]The standardization is re-done at each period in our recursive forecasting exercise using information available at the time the forecast is being made.

[8]Although we always include the monthly dummies (with possibly time varying coefficients) so our models are not exactly equivalent to the TVP-AR or UCSV models.

We focus on the parameters of most direct relation to the shrinkage properties of the model: ω^2 and τ^2. To gain a feeling for the magnitude of these parameters, remember that they are the prior variances of the errors in the state equations and on the constant coefficients, respectively and that our variables have all been standardized to have a variance of one. We present results using both the $h = 1$ and $h = 12$ models. Table 1 presents results for the predictors, Table 2 for the lags and Table 3 for the monthly dummies.

Note first that ω^2 tends to be much smaller than τ^2. This is as expected, since ω^2 controls the degree of time-variation in coefficients and even a small amount of monthly time-variation can lead to a large degree of change in coefficients. For instance, a value $\omega^2 = 10^{-4}$ implies a value of $\omega = 0.01$ and a standard deviation of the error in the state equation of this magnitude will allow for a moderate degree of change in coefficients over the course of a year. However, if $\omega^2 = 10^{-6}$, then very little change in coefficients is allowed for. With regards to the constant coefficients, if $\tau_i^2 = 0.01$ then, approximately, 95% of the prior probability for θ_i is allocated to the interval $[-0.2, +0.2]$ and, thus, this value for τ_i^2 ensures a fairly high degree of shrinkage, whereas if $\tau_i^2 = 0.1$, then the prior is much more dispersed and the shrinkage much less.

An examination of the tables indicate a moderate, but not dogmatic, amount of shrinkage in most cases. Furthermore, the coefficients in the $h = 1$ case tend to be shrunk more than those with $h = 12$. However, the degree of shrinkage varies across coefficients. For both forecast horizons, both constant and time-varying coefficients on the intercept tend to be shrunk less than the coefficients on the other variables. When $h = 1$, we find the 12th lag and the December dummy to be shrunk much less than other lags and monthly dummies. This finding is not repeated when $h = 12$.

Among the other predictors, for $h = 12$ the unemployment rate has both constant and time-varying coefficients which are shrunk less than coefficients on most of the other predictors. This indicates that the role of the unemployment rate's coefficient in important and time-varying. In contrast, the constant coefficient on the one-year interest rate has a moderately high degree of shrinkage on its constant coefficient, but has very little shrinkage on its time-varying coefficients. This pattern is consistent with this interest rate having having a smaller role, but a role that is changing over time. These patterns are not repeated for $h = 1$, although here the consumer sentiment, car registration and order book level variables exhibit more time variation in their coefficients than the other predictors.

Table 1: Posterior means of ω^2 and τ^2 for predictors (st.dev. in parentheses)

	$h = 1$		$h = 12$	
Predictor	ω^2	τ^2	ω^2	τ^2
Intercept	6.5×10^{-3}	0.061	3.7×10^{-3}	0.926
	(2.9×10^{-4})	(0.035)	(2.3×10^{-4})	(0.302)
I_1MO	7.6×10^{-5}	0.014	3.9×10^{-6}	0.187
	(2.4×10^{-5})	(0.018)	(4.0×10^{-6})	(0.127)
I_1YR	1.7×10^{-5}	0.022	0.010	0.090
	(1.8×10^{-5})	(0.023)	(7.5×10^{-4})	(0.107)
SENT	2.0×10^{-3}	0.014	5.3×10^{-5}	0.088
	(1.2×10^{-4})	(0.019)	(1.9×10^{-5})	(0.102)
STOCK1	8.2×10^{-5}	0.018	4.2×10^{-5}	0.133
	(1.8×10^{-5})	(0.021)	(8.2×10^{-6})	(0.129)
STOCK2	3.0×10^{-4}	0.022	3.3×10^{-3}	0.089
	(3.1×10^{-5})	(0.023)	(2.1×10^{-4})	(0.109)
EXRATE	1.1×10^{-6}	0.013	1.5×10^{-4}	0.108
	(1.6×10^{-6})	(0.018)	(2.4×10^{-5})	(0.107)
IP	2.0×10^{-3}	0.016	3.0×10^{-4}	0.071
	(2.0×10^{-4})	(0.018)	(4.1×10^{-5})	(0.097)
LOANS	3.4×10^{-4}	0.020	6.2×10^{-4}	0.108
	(2.3×10^{-5})	(0.021)	(3.2×10^{-5})	(0.111)
M3	7.7×10^{-6}	0.014	1.2×10^{-3}	0.098
	(6.0×10^{-6})	(0.019)	(7.6×10^{-5})	(0.109)
CAR	2.3×10^{-3}	0.013	1.8×10^{-3}	0.086
	(1.5×10^{-4})	(0.017)	(1.2×10^{-4})	(0.119)
ORDER	1.1×10^{-3}	0.016	7.0×10^{-6}	0.068
	(4.7×10^{-5})	(0.024)	(3.9×10^{-6})	(0.098)
OIL	2.2×10^{-5}	0.015	1.5×10^{-3}	0.103
	(9.4×10^{-6})	(0.018)	(1.2×10^{-4})	(0.120)
UNEMP	5.9×10^{-5}	0.016	2.2×10^{-3}	0.257
	(2.9×10^{-5})	(0.018)	(1.0×10^{-4})	(0.147)

Table 2: Posterior means of ω^2 and τ^2 for lags (st. dev. in parentheses)

	$h=1$		$h=12$	
Lag	ω^2	τ^2	ω^2	τ^2
1	1.6×10^{-3} (7.6×10^{-5})	0.019 (0.021)	6.4×10^{-4} (3.5×10^{-5})	0.109 (0.115)
2	3.9×10^{-4} (9.5×10^{-5})	0.016 (0.018)	6.2×10^{-3} (2.6×10^{-4})	0.085 (0.116)
3	1.8×10^{-6} (2.0×10^{-6})	0.029 (0.026)	7.5×10^{-4} (4.6×10^{-5})	0.104 (0.103)
4	3.5×10^{-5} (7.1×10^{-6})	0.013 (0.018)	4.0×10^{-4} (5.0×10^{-5})	0.108 (0.112)
5	3.6×10^{-4} (6.7×10^{-5})	0.013 (0.018)	1.5×10^{-4} (2.5×10^{-5})	0.108 (0.114)
6	1.8×10^{-4} (5.3×10^{-5})	0.023 (0.020)	9.2×10^{-5} (2.3×10^{-5})	0.082 (0.100)
7	1.7×10^{-4} (1.8×10^{-5})	0.020 (0.023)	1.7×10^{-3} (3.3×10^{-4})	0.075 (0.115)
8	1.2×10^{-5} (5.8×10^{-6})	0.012 (0.015)	2.2×10^{-4} (5.3×10^{-5})	0.069 (0.107)
9	1.5×10^{-4} (2.9×10^{-5})	0.012 (0.019)	5.6×10^{-4} (1.5×10^{-4})	0.097 (0.114)
10	1.0×10^{-4} (1.0×10^{-5})	0.013 (0.018)	6.6×10^{-5} (3.2×10^{-5})	0.072 (0.102)
11	7.3×10^{-5} (1.8×10^{-5})	0.019 (0.022)	5.3×10^{-3} (4.4×10^{-4})	0.094 (0.106)
12	5.1×10^{-5} (8.5×10^{-6})	0.085 (0.040)	1.4×10^{-4} (3.2×10^{-4})	0.096 (0.114)

Table 3: Posterior means of ω^2 and τ^2 for dummies (st. dev. in parentheses)

	$h=1$		$h=12$	
	ω^2	τ^2	ω^2	τ^2
February	$6.7{\times}10^{-4}$	0.041	$1.2{\times}10^{-3}$	0.169
	$\left(1.1 \times 10^{-4}\right)$	(0.028)	$\left(3.1 \times 10^{-4}\right)$	(0.146)
March	$1.4{\times}10^{-3}$	0.027	$1.4{\times}10^{-3}$	0.088
	$\left(4.5 \times 10^{-4}\right)$	(0.024)	$\left(4.2 \times 10^{-4}\right)$	(0.104)
April	$8.9{\times}10^{-5}$	0.015	$2.8{\times}10^{-4}$	0.136
	$\left(4.1 \times 10^{-5}\right)$	(0.020)	$\left(2.9 \times 10^{-4}\right)$	(0.122)
May	$1.4{\times}10^{-5}$	0.034	$8.5{\times}10^{-6}$	0.071
	$\left(1.5 \times 10^{-5}\right)$	(0.026)	$\left(1.6 \times 10^{-5}\right)$	(0.112)
June	$3.8{\times}10^{-5}$	0.016	$4.3{\times}10^{-5}$	0.073
	$\left(4.5 \times 10^{-5}\right)$	(0.020)	$\left(3.6 \times 10^{-5}\right)$	(0.094)
July	$3.6{\times}10^{-5}$	0.014	$2.2{\times}10^{-5}$	0.101
	$\left(2.9 \times 10^{-5}\right)$	(0.020)	$\left(3.6 \times 10^{-5}\right)$	(0.114)
August	$1.7{\times}10^{-4}$	0.015	$4.5{\times}10^{-4}$	0.078
	$\left(2.2 \times 10^{-4}\right)$	(0.020)	$\left(1.4 \times 10^{-4}\right)$	(0.100)
September	$1.1{\times}10^{-4}$	0.016	$4.3{\times}10^{-5}$	0.072
	$\left(9.8 \times 10^{-5}\right)$	(0.020)	$\left(5.0 \times 10^{-5}\right)$	(0.088)
October	$1.4{\times}10^{-4}$	0.028	$5.1{\times}10^{-4}$	0.145
	$\left(5.9 \times 10^{-5}\right)$	(0.023)	$\left(2.0 \times 10^{-4}\right)$	(0.133)
November	$1.6{\times}10^{-3}$	0.022	$2.4{\times}10^{-4}$	0.081
	$\left(1.4 \times 10^{-4}\right)$	(0.020)	$\left(1.5 \times 10^{-4}\right)$	(0.099)
December	$1.5{\times}10^{-4}$	0.079	$5.1{\times}10^{-4}$	0.118
	$\left(1.4 \times 10^{-4}\right)$	(0.036)	$\left(1.1 \times 10^{-4}\right)$	(0.126)

3.3 Forecasting Results

Most Bayesians prefer to use predictive likelihoods for evaluating forecast performance. Our posterior and predictive simulation algorithm provides us with the predictive density for forecasting $y_{\tau+h}$ given information through time τ, which we denote by $p\left(y_{\tau+h}|Data_{\tau}\right)$. This predictive density is evaluated for $\tau = \tau_0, .., T-h$ where τ_0 is January, 2001 and $h=1$ and 12. If we let $y^o_{\tau+h}$ be the observed value of $y_{i,\tau+h}$, then the predictive likelihood is $p\left(y_{\tau+h} = y^o_{\tau+h}|Data_{\tau}\right)$and we use the mean of the log predictive likelihoods (MLPL) for forecast evaluation:

$$MLPL = \frac{1}{T-h-\tau_0+1}\sum_{\tau=\tau_0}^{T-h}\log\left[p\left(y_{\tau+h} = y^o_{\tau+h}|Data_{\tau}\right)\right].$$

In the tables, MLPL results are presented relative to those produced by our full model (i.e. the model with hierarchical shrinkage on both constant and time-varying coefficients and which has stochastic volatility in the measurement equation). Since MLPL is measured in log units, we take the difference between the MLPL for the full model and the restricted model. Thus, positive numbers indicate our full model is forecasting better than the restricted model.

We also present the mean of the squared forecast errors (MSFE) and the mean of the absolute value of the forecast errors (MAFE). In contract to the predictive likelihoods, which evaluate the performance of the entire predictive distribution, MSFEs and MAFE only evaluate the performance of the point forecast. We use the predictive median as our point forecast. Again, we present results relative to the full model. The tables present MSFEs and MAFEs for a restricted model divided by those for the full model. Thus, a number greater than one indicates our full model is forecasting better, using MSFE or MAFE as a metric.

3.3.1 Results for Core Inflation

Tables 4 and 5 present results for core inflation. Consider first results for $h = 1$ (see Table 4). Regardless of whether we use MLPL, MSFE or MAFE to evaluate forecast performance, we find evidence that the inclusion of stochastic volatility causes forecast performance to deteriorate. The predictive likelihoods are substantially higher in the homoskedastic models suggesting that stochastic volatility may not be present in this data set.[9] It is worth noting that the TVP regression model with stochastic volatility (which could be a popular benchmark for the researcher working in this literature, but not interested in adding shrinkage) produces the worst forecasts. And our full model is the second best forecasting model among the set of models with stochastic volatility. This suggests caution should be taken when forecasting with TVP regression models without shrinkage and suggests the importance of shrinkage such as that provided by the Lasso. But beyond this we will say no more about the models with stochastic volatility and focus on the better-forecasting homoskedas-

[9]Note that the product of predictive likelihoods over the entire sample is the marginal likelihood. Thus, MLPL can be used as a method of model selection and here indicates support for homoskedasticity.

tic models.

Among the homoskedastic models with $h = 1$, our model with Lasso prior on both constant and time-varying coefficients forecasts the best when we use predictive likelihoods to evaluate forecast performance. When we use MAFE or MSFE, the constant coefficient model forecasts best. However, our model is a close second best in terms of the latter forecast metrics. In this data set, at this forecast horizon, it appears that there is little need for time-varying coefficients and, thus, our model is forecasting roughly as well as a constant coefficient model. However, it is important to stress that our approach discovered this fact automatically. That is, it allows the researcher to begin with a very flexible model, allowing for features which might be important for forecasting (such as parameter change), but then the statistical methodology decides which aspect is important and which is not. Here our approach is shrinking the time-varying coefficients so as to mostly “turn off” this part of the model. The standard TVP regression model does not do this and exhibits poorer forecast performance. At the other extreme, a researcher who simply began with a constant coefficient model might have been unknowingly working with a badly mis-specified model. Our approach allows the data to decide whether the coefficients are time-varying and, if so, by how much they vary.

The story at the annual forecast horizon ($h = 12$) is a bit different, but also indicates the benefits of using the Lasso, especially on the TVPs. In terms of predictive likelihoods, our full model (including Lasso prior on constant and time-varying coefficients and stochastic volatility) forecasts best. In terms of MSFE and MAFE, our full model also forecasts well, but it is the homoskedastic version of our model which does even better. In this case, our different measures of forecast performance are telling a somewhat conflicting story (especially with regards to the need for stochastic volatility). This is a point we will return to below.

For $h = 1$, we found the simple constant coefficient model to forecast well. For $h = 12$, this is not so. In this case, allowing to time-variation in coefficients is important in achieving a good forecast performance. However, the TVP regression model does not forecast well.

It is allowing for too much time-variation in coefficients. The Lasso prior is allowing for us to estimate the correct degree of time-variation in coefficients in order to obtain a good forecast performance.

Table 4: Measures of Forecast Performance for Core Inflation ($h = 1$)

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
Lasso on constant and TVPs	-0.42	0.77	0.83	0.00	1.00	1.00
Lasso only on constant coeffs.	-0.29	0.88	0.87	2.00	1.93	1.94
Lasso only on TVPs	-0.37	0.85	0.86	0.30	1.59	1.11
TVP regression model	0.09	1.08	0.95	2.54	3.64	1.69
Constant coeff. model	-0.40	0.66	0.78	0.01	0.91	0.94

Note: All results are relative to the benchmark model (Lasso on constant & TVPs)

Table 5: Measures of Forecast Performance for Core Inflation ($h = 12$)

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
Lasso on constant and TVPs	12.83	0.95	0.95	0.00	1.00	1.00
Lasso only on constant coeffs.	2.80	0.88	0.90	0.89	1.23	1.09
Lasso only on TVPs	11.64	1.18	1.02	0.19	1.31	1.14
TVP regression model	3.45	1.27	1.06	1.07	1.80	1.28
Constant coeff. model	6.33	1.09	0.98	0.14	12.48	10.31

Note: All results are relative to the benchmark model (Lasso on constant & TVPs)

Tables 4 and 5 present evidence on average forecast performance. To investigate whether forecast performance varies over time and to shed light on why the MLPL results conflict with the MAFE and MSFE results in Table 5, we present Figures 1 and 2. These are produced using the model with Lasso prior on both constant and time-varying coefficients with $h = 12$ (but similar patterns are found with the other models). Figure 1 plots the forecast errors squared for the homoskedastic and heteroskedastic versions of the model. Figure 2 plots the logs of the predictive likelihood in the same format. Note first that forecast performance does vary over time with a deterioration of forecast performance at the time of the financial crisis and in 2001.

Looking at Figure 1, it can be seen that the homoskedastic and heteroskedastic versions of our model are forecasting roughly as well as each other, in terms of their point forecasts. A similar pattern holds for the predictive likelihoods (see Figure 2) for most of the time. However, at the time of the financial crisis, the homoskedastic version of the model has much lower predictive likelihoods and it is this time period which drives the inconsistency between MLPL and MSFE noted in Table 5. What is happening is that the homoskedastic version of the model is missing the large increase in volatility which began with the financial crisis. This has little impact on the point forecasts and thus, the forecast errors squared do not differ by much between the homoskedastic and heteroskedastic versions of the model. However, the homoskedastic version of the model has an error variance which is much too small. This makes it appear that the point forecast is far in the tails of the predictive density, leading to a very small predictive likelihood. The heteroskedastic version of this model does not run into this problem. Policymakers are increasingly interested in forecast uncertainty and, hence, want more than just a point forecast. Figure 2 shows how the appropriate modelling of the error variance can be crucial in obtaining reliable inference about forecast uncertainty.

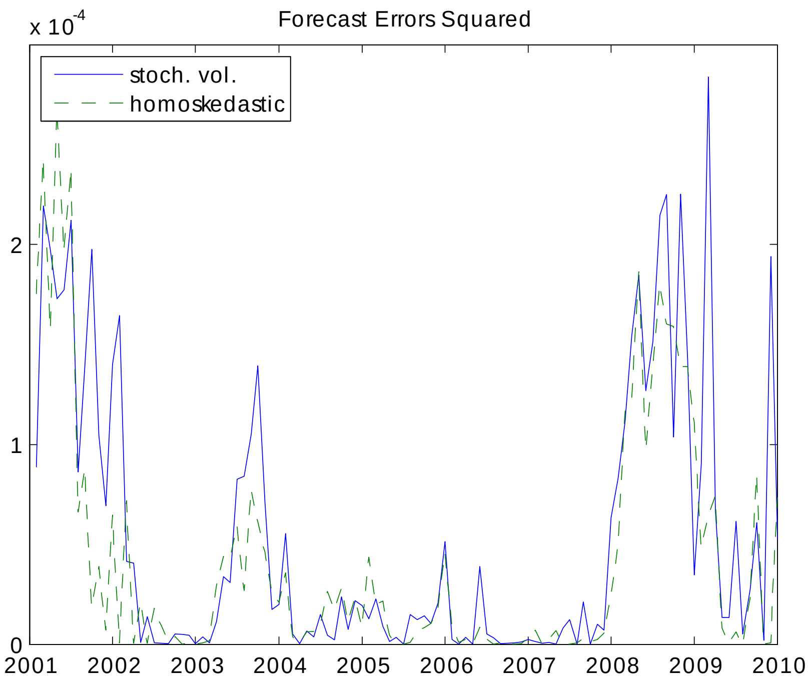

Figure 1: Forecast errors squared for models with Lasso prior on constant and time-varying coefficients, $h = 12$

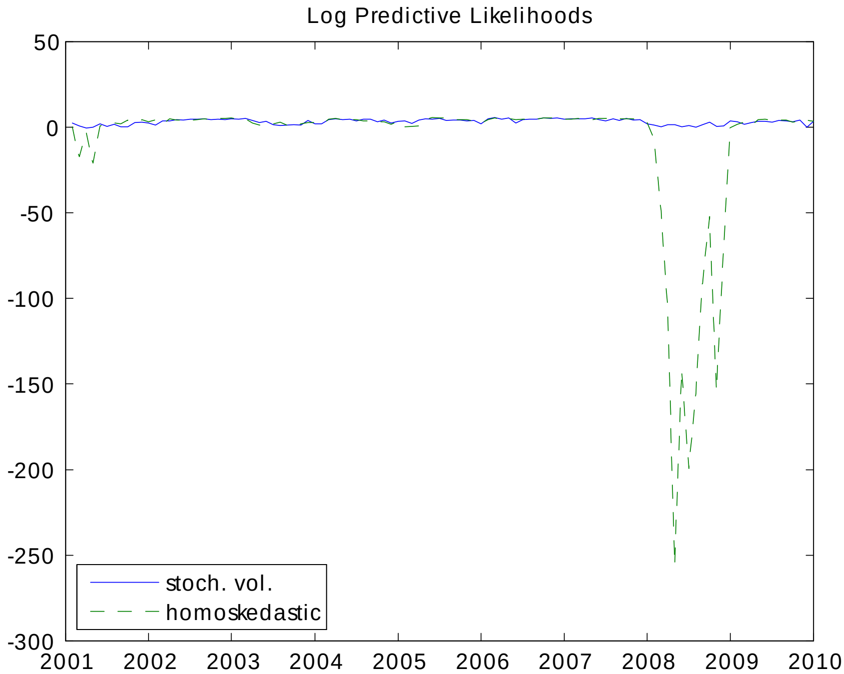

Figure 2: Log predictive likelihoods for models with Lasso prior on constant and time-varying coefficients, $h = 12$

3.3.2 Results for Overall Inflation

Tables 6 and 7 are the same as Tables 4 and 5, except that the former are for overall inflation.

For $h = 1$, we are finding that homoskedastic models tend to do better than those with stochastic volatility (although to a lesser extent than for core inflation). The homoskedastic version of our model with Lasso prior on both time varying and constant coefficients exhibits the best forecast performance regardless of whether one uses MLPL, MSFE or MAFE. Note also that, to a greater extent than with core inflation, our model is forecasting better than two popular benchmarks: the TVP regression model and the constant coefficient model.

For $h = 12$, Table 7 is showing a similar pattern to Table 5. That is, in terms of MLPL, TVP models with stochastic volatility forecast best – but only if a Lasso prior is used on the time-varying coefficients. However, if we look at MSFE or MAFE, models with constant coefficients forecast best. This discrepancy between the MLPL and MSFE results is due to the same reason discussed previously (and illustrated in Figures 1 and 2). However, it is worth noting that, even if we use only MSFE and MAFE for forecast comparison, models with Lasso priors on TVPs forecast much better than the unrestricted TVP regression model.

In summary, in cases where there is time-variation in coefficients, putting a Lasso prior on these coefficients does lead to better forecast performance than unrestricted TVP models. The shrinkage is beneficial in keeping the time-varying coefficients from wandering too widely. In cases where there is little evidence of time-varying coefficients (i.e. where constant coefficient models forecast well), the Lasso prior can automatically discover this lack of time-variation and lead to forecasting results that are almost as good as the constant coefficient model. In these latter cases, unrestricted TVP models can forecast very poorly.

Table 6: Forecast Performance for Overall Inflation ($h = 1$)

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
Lasso on constant and TVPs	-0.24	0.75	0.85	0.00	1.00	1.00
Lasso only on constant coeffs.	-0.14	0.85	0.96	1.99	1.51	1.23
Lasso only on TVPs	-0.18	0.86	0.94	0.04	1.06	1.04
TVP regression model	-0.04	1.00	1.04	1.99	1.56	1.27
Constant coeff. model	0.15	0.85	0.89	0.10	1.03	0.94

Note: All results are relative to the benchmark model (Lasso on constant & TVPs)

Table 7: Forecast Performance for Overall Inflation ($h = 12$)

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
Lasso on constant and TVPs	26.75	0.85	0.95	0.00	1.00	1.00
Lasso only on constant coeffs.	8.43	0.91	0.96	0.81	1.04	1.02
Lasso only on TVPs	16.37	0.97	0.96	-0.08	1.02	1.01
TVP regression model	4.69	1.05	1.02	0.99	1.23	1.08
Constant coeff. model	8.11	0.63	0.80	-0.05	0.76	0.89

Note: All results are relative to the benchmark model (Lasso on constant & TVPs)

3.3.3 Robustness to Different Specifications

Our empirical results are based on large set of explanatory variables. In the Empirical Appendix we investigate the robustness of our results to changes in this set. The reader is referred to the Empirical Appendix for details. We do find our results to be quite robust. We briefly summarize our findings here.

If we omit the 13 predictors listed in Section 3.1, we still have a model with many explanatory variables (an intercept, 11 monthly dummies and 12 lags) and, thus, Lasso-type shrinkage is potentially useful. Table A.1 through A.4 in the Empirical Appendix shows that it is indeed useful in a similar manner to what we found in Tables 4 through 7. This indicates the usefulness of hierarchical shrinkage even in time-varying AR(p) models.

The Empirical Appendix also includes results for models where all predictors and all lags are excluded. Hence, the models include only a (potentially time-varying) intercept and monthly dummies. With the inclusion of stochastic volatility, we have models which are similar to the popular UCSV model of Stock and Watson (2007). For these models, the benefits of hierarchical shrinkage are smaller, but still evident.

4 Conclusions

The macroeconomist often has many variables which can be used in a forecasting exercise. She may also wish to work with a model which allows for the parameter change which is empirically evident in many macroeconomic data sets. These considerations can often lead to models with many parameters, leading to over-fitting and poor forecast performance. In regressions and VARs with constant coefficients, there have been many approaches which try to overcome these problems by shrinking coefficients. However, with TVP models (where we would expect the need for shrinkage to be even greater than in constant coefficient models), few shrinkage approaches have been suggested. In this paper, we have developed a new approach to shrinkage in TVP models based on the Lasso. We have extended Lasso methods, which are popular with constant coefficient models, to TVP models and developed

Bayesian methods for posterior and predictive inference.

To investigate the performance of our approach, we use an EU data set which is relatively short and involves many predictors. Our findings are moderately encouraging in that use of the Lasso on the time-varying coefficients does lead to substantial improvements in forecast performance relative to unrestricted TVP models. Relative to constant coefficient models, a TVP model with Lasso shrinkage in some cases exhibits improved forecast performance. But we find that an advantage of using a TVP model with Lasso shrinkage is that it can automatically produce a model which is similar to a constant coefficient model in the cases where a constant coefficient model is the appropriate forecasting model. Thus, the researcher using the TVP model with Lasso prior can be confident that the risks of misspecification associated with constant coefficient models are avoided, while at the same time avoid the risks of over-parameterization associated with unrestricted TVP models.

References

Banbura, M., Giannone, D. and Reichlin, L. (2010). "Large Bayesian Vector Auto Regressions," *Journal of Applied Econometrics*, 25, 71-92.

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

Chipman, H., George, E. and McCulloch, R. (2001). "The practical implementation of Bayesian model selection," pages 65-134 in Institute of Mathematical Statistics Lecture Notes - Monograph Series, Volume 38, edited by P. Lahiri.

Clements, M. and Hendry, D., 1998, *Forecasting Economic Time Series*. (Cambridge University Press: Cambridge).

Clements, M. and Hendry, D., 1999, *Forecasting Non-stationary Economic Time Series*. (The MIT Press: Cambridge).

Cogley, T., Morozov, S. and Sargent, T. (2005). "Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system," *Journal of Economic Dynamics and Control*, 29, 1893-1925.

Cogley, T. and Sargent, T. (2001). "Evolving post World War II inflation dynamics," *NBER Macroeconomics Annual*, 16, 331-373.

Cogley, T. and Sargent, T. (2005). "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S.," *Review of Economic Dynamics*, 8, 262-302.

D'Agostino, A., Gambetti, L. and Giannone, D. (2009). "Macroeconomic forecasting and structural change," ECARES working paper 2009-020.

De Mol, C., Giannone, D. and Reichlin, L. (2008). "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" *Journal of Econometrics*, 146, 318-328.

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.

George, E. and McCulloch, R. (1997). ”Approaches for Bayesian variable selection,” *Statistica Sinica, 7, 339-373.*

Kim, S., Shephard, N. and Chib, S. (1998). “Stochastic volatility: likelihood inference and comparison with ARCH models,” *Review of Economic Studies*, 65, 361-93.

Koop, G. and Korobilis, D. (2009). “Forecasting inflation using dynamic model averaging,” RCFEA WP 09-34, Rimini Center for Economic Analysis.

Koop, G., Leon-Gonzalez, R., Strachan, R. (2009). “On the evolution of the monetary policy transmission mechanism,” *Journal of Economic Dynamics and Control* 33 (2009), 997-1017.

Park, T. and Casella, G. (2008). “The Bayesian Lasso,” *Journal of the American Statistical Association* 103, 681-686.

Pesaran, M.H., Pettenuzzo, D. and Timmerman, A. (2006). “Forecasting time series subject to multiple structural breaks,” *Review of Economic Studies*, 73, 1057–1084.

Primiceri. G. (2005). “Time varying structural vector autoregressions and monetary policy,” *Review of Economic Studies*, 72, 821-852.

Sims, C. and Zha, T. (2006). “Were there regime switches in macroeconomic policy?” *American Economic Review*, 96, 54-81.

Stock, J. and Watson, M. (1996). “Evidence on structural instability in macroeconomic time series relations, ” *Journal of Business and Economic Statistics*, 14, 11-30.

Stock, J. and Watson, M. (2006). “Forecasting using many predictors,” pp. 515-554 in *Handbook of Economic Forecasting*, Volume 1, edited by G. Elliott, C. Granger and A. Timmerman, Amsterdam: North Holland.

Stock, J. and Watson, M. (2007). “Why has U.S. inflation become harder to forecast?” *Journal of Monetary Credit and Banking* 39, 3-33.

Stock, J. and Watson, M. (2011). “Generalized shrinkage methods for forecasting using many predictors,” manuscript available at http://www.princeton.edu/∼mwatson/papers/stock_watson_generalized_shrinkage_February_2011.pdf.

Technical Appendix

Stochastic volatility

We use a standard stochastic volatility specification for the error variance in the measurement equation. In particular, if $h_t = \ln(\sigma_t)$, then:

$$h_{t+1} = h_t + u_t,$$

where u_t is $N(0, W)$ and is independent over t and of ε_t and η_t. We use the algorithm of Kim, Shephard and Chib (1998) to draw the volatilities. The prior for the initial volatility is:

$$h_0 \sim N(-0.5, 0.5).$$

Since the dependent variable is standardized to have a variance of one, this is only very weakly informative, but is centered over a plausible value for h_0. The prior for W^{-1} is Gamma with prior mean of 10^4 and two prior degrees of freedom.

Priors for λ^2 and κ^2

Unless otherwise specified, our empirical results set $a_1 = a_2 = b_1 = b_2 = 0.001$ which implies proper but very noninformative priors (i.e. the prior mean of these priors is one, but the prior variance is 1000). One of our models almost totally removes the TVP part of the model altogether by setting $b_1 = 100000$. This value for b_1 implies prior mean of 100 for κ (a value which ensures shrinkage of ω to very near zero) and the prior variance is 0.1 ensuring a tight prior around this value.

Priors when Lasso is not used

For models without the Lasso prior on the TVP coefficients, ξ_i^2 and κ do not appear in the model and we use a non-hierarchical prior for ω of the form:

$$\omega \sim N(0, I).$$

For models without the Lasso prior on the constant coefficients, τ_i^2 and λ do not appear in the model and we use a non-hierarchical prior for θ of the form:

$$\theta \sim N\left(0, 9 \times I\right).$$

Empirical Appendix

In this sub-section, we present results for a few additional specifications to show that the results in the body of the paper are robust. Results are presented in the same format as in Tables 4 through 7. That is forecast performance is measured relative to the full model with the same explanatory variables. Tables A.1 and A.2 present results without any of the predictors listed in Section 3.1 (i.e. simply using an intercept, the monthly dummies and $p = 12$). It can be seen that the same patterns noted in Tables 4 through 7 hold. For instance, Table A.1 shows that using the Lasso on the time-varying coefficients leads to substantive forecast improvements over unrestricted TVP models for $h = 1$. It is worth noting, however, that the MLPLs are roughly the same in Tables A.1 and A.2 as in Tables 4 through 7 and the MSFEs and MAFEs are in many cases, somewhat lower in the former tables. This suggests that, in this application, the predictors are adding little. This re-emphasizes the importance of shrinkage methods such as those introduced in this paper. That is, if the researcher is working with a data set with many predictors, our TVP shrinkage methods can, in an automatic fashion, uncover the fact that the predictors are adding little. This may be preferable to a model selection strategy where the researcher seeks to find a single parsimonious forecasting model.

Tables A.3 and A.4 repeat the analysis for models with only an intercept and monthly dummies. Similar patterns to those noted previously are found, although in this relatively parsimonious model, the benefits of Lasso-type shrinkage are smaller. Note, however, that with core inflation, the constant coefficient model does quite poorly (especially when $h = 1$) which contrasts with the results we found with more parameter-rich models.

Table A.1: Forecast Performance for Core Inflation: No Predictors

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
$h = 1$						
Lasso on constant and TVPs	-0.46	0.67	0.77	0.00	1.00	1.00
Lasso only on constant coeffs.	-0.38	0.73	0.81	2.19	1.79	1.31
Lasso only on TVPs	-0.43	0.66	0.79	0.01	0.99	0.94
TVP regression model	-0.31	0.82	0.85	2.14	1.87	133
Constant coeff. model	-0.46	0.56	0.74	-0.16	0.70	0.82
$h = 12$						
Lasso on constant and TVPs	50.29	1.15	1.10	0.00	1.00	1.00
Lasso only on constant coeffs.	31.93	1.16	1.09	1.07	1.51	1.22
Lasso only on TVPs	53.72	1.30	1.18	0.10	1.02	1.05
TVP regression model	28.70	1.25	1.14	1.09	1.59	1.26
Constant coeff. model	-0.44	0.72	0.87	0.05	0.86	0.95

Note: All results are relative to the benchmark model (Lasso on constant & TVPs) for each forecast horizon h

Table A.2: Forecast Performance for Overall Inflation: No Predictors

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
$h = 1$						
Lasso on constant and TVPs	-0.50	0.66	0.79	0.00	1.00	1.00
Lasso only on constant coeffs.	-0.47	0.67	0.82	1.64	1.13	1.03
Lasso only on TVPs	-0.52	0.63	0.79	-0.06	0.86	0.88
TVP regression model	-0.42	0.72	0.86	1.60	1.48	1.22
Constant coeff. model	-0.28	0.69	0.80	-0.18	0.72	0.81
$h = 12$						
Lasso on constant and TVPs	50.14	1.02	0.96	0.00	1.00	1.00
Lasso only on constant coeffs.	41.84	1.09	0.98	1.13	1.47	1.18
Lasso only on TVPs	58.14	1.06	1.01	0.16	0.91	1.05
TVP regression model	30.75	1.25	1.07	1.19	1.49	1.26
Constant coeff. model	0.26	0.58	0.76	-0.07	0.63	0.81

Note: All results are relative to the benchmark model (Lasso on constant & TVPs) for each forecast horizon h

Table A.3: Forecast Performance for Core Inflation: No Predictors nor Lags

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
$h = 1$						
Lasso on constant and TVPs	-0.36	0.77	0.85	0.00	1.00	1.00
Lasso only on constant coeffs.	-0.36	0.74	0.85	0.89	1.09	1.06
Lasso only on TVPs	-0.22	0.78	0.87	0.13	1.13	1.03
TVP regression model	-0.15	0.78	0.86	0.93	1.05	1.01
Constant coeff. model	0.65	2.70	1.55	0.71	3.02	1.67
$h = 12$						
Lasso on constant and TVPs	56.70	0.81	0.89	0.00	1.00	1.00
Lasso only on constant coeffs.	35.07	0.81	0.89	-0.71	0.81	0.88
Lasso only on TVPs	58.98	0.81	0.89	-0.13	0.87	0.91
TVP regression model	39.31	0.81	0.89	-0.75	0.83	0.90
Constant coeff. model	-1.15	0.74	0.88	1.46	1.08	1.11

Note: All results are relative to the benchmark model (Lasso on constant & TVPs) for each forecast horizon h

Table A.4: Forecast Performance for Overall Inflation: No Predictors nor Lags

	Constant Variance			Stochastic Volatility		
	MLPL	MSFE	MAFE	MLPL	MSFE	MAFE
$h = 1$						
Lasso on constant and TVPs	-0.32	0.78	0.90	0.00	1.00	1.00
Lasso only on constant coeffs.	-0.27	0.85	0.90	0.40	1.33	1.16
Lasso only on TVPs	-0.30	0.77	0.88	0.05	1.08	1.05
TVP regression model	-0.23	0.89	0.92	0.49	1.19	1.09
Constant coeff. model	-0.03	1.02	1.03	0.28	1.06	1.02
$h = 12$						
Lasso on constant and TVPs	55.99	1.40	1.14	0.00	1.00	1.00
Lasso only on constant coeffs.	55.23	1.40	1.14	-0.20	1.37	1.13
Lasso only on TVPs	55.53	1.39	1.14	0.25	1.02	0.95
TVP regression model	44.81	1.39	1.14	-0.08	1.38	1.13
Constant coeff. model	-0.16	0.62	0.82	-0.36	0.58	0.76

Note: All results are relative to the benchmark model (Lasso on constant & TVPs) for each forecast horizon h