A new approach to forecasting in the presence of in and out of sample breaks

Jiawen Xu[*] **Pierre Perron**[†]

Boston University **Boston University**

November 4, 2012

Abstract

We present a novel frequentist-based approach to forecast time series in the presence of in-sample and out-of-sample breaks in the parameters of the forecasting model. We first model the parameters as following a random level shift process, with the occurrence of a shift governed by a Bernoulli process. In order to have a structure so that changes in the parameters be forecastable, we introduce two modifications. The first models the probability of shifts according to some covariates that can be forecasted. The second incorporates a built-in mean reversion mechanism to the time path of the parameters. Similar modifications can also be made to model changes in the variance of the error process. Our full model can be cast into a non-linear non-Gaussian state space framework. To estimate it, we use particle filtering and a Monte Carlo expectation maximization algorithm. Simulation results show that the algorithm delivers accurate in-sample estimates, in particular the filtered estimates of the time path of the parameters follow closely their true variations. We provide a number of empirical applications and compare the forecasting performance of our model with a variety of alternative methods. These show that substantial gains in forecasting accuracy are obtained.

Keywords: instabilities; structural change; forecasting; random level shifts; particle filter.

JEL Classification: C22, C53

[*]Department of Economics, Boston University, 270 Bay State Rd., Boston MA 02215 (jwxu@bu.edu).

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston MA 02215 (perron@bu.edu).

1 Introduction

Forecasting is obviously of paramount importance in time series analyses. The theory of constructing and evaluating forecasting models is well established in the case of stable relationships. However, there is growing evidence that forecasting models are subject to instabilities, leading to imprecise and unreliable forecasts. This is so in a variety of fields including macroeconomics and finance. Indeed, Stock and Watson (1996) documented widespread prevalence of instabilities in macroeconomic time series relationships. A prominent example is forecasting inflation; see, e.g., Stock and Watson (2007). This problem is also prevalent in finance. Pastor and Stambaugh (2001) document structural breaks in the conditional mean of the equity premium using long time return series. Paye and Timmermann (2006) examined model instability in the coefficients of ex post predictable components of stock returns. See also Pesaran and Timmermann (2002), Rapach and Wohar (2006) and Pettenuzzo and Timmermann (2011).

There is a vast literature on testing for and estimating structural changes within a given sample of data; see, e.g., Andrews (1993), Bai and Perron (1998, 2003) and Perron (2006) for a survey. Much of the literature does not model the breaks as being stochastic. Hence, the scope for improving forecasts is limited. There can be improvements by relying on the estimates of the last regime (or at least putting more weights on them) but even then such improvements are possible if there are no out-of-sample breaks. In the presence of out-of-sample breaks the limitation imposed by treating the breaks as deterministic mitigates the forecasting ability of models corrected for in-sample breaks. This renders forecasting in the presence of structural breaks quite a challenge; see, e.g., Clements and Hendry (2006).

Some Bayesian models have been proposed to address this problem; see, e.g., Pesaran et al. (2006), Koop and Porter (2007), Maheu and Gordon (2008), Maheu and McCurdy (2009) and Hauwe et al. (2011). The advantage of the Bayesian approach steams from the fact that it treats the parameters as random and by imposing a prior (or meta-prior) distribution one can model the breaks and allow them to occur out-of-sample with some probability. Such methods can, however, be sensitive to the exact prior distributions used.

We propose a frequentist-type approach with a forecasting model in which the changes in the parameters have a probabilistic structure so that the estimates can help forecast future out-of-sample breaks. Our approach is best suited to the case for which breaks occur both in and out-of-sample, which in particular avoids the problematic use of a trimming window assumed to have a stable structure. The method will work best indeed if there are many in-sample breaks, so that a long span of data is beneficial. This is unavoidable since good out-of-sample forecasts of breaks require in-sample information about the process generating such breaks, the more so the more efficient the forecasts will be. The same applies to previously proposed Bayesian methods, though the use of tight priors can partially substitute for the lack of precise in-sample information. Having said that, our method still yields considerable improvements even if relatively few breaks are present in-sample.

Our approach is similar in spirit to unobserved components models in which the parameters are modeled as random walk processes. There are, however, important departures. Most importantly, a shift need not occur every period. It does so with some probability dictated by a Bernoulli process for the occurrence of shifts and a normal random variable for its magnitude. This leads to a specification in which the parameters evolve according to a random level shift process. Some or all of the parameters of the model can be allowed to change and the latent variables that dictate the changes can be common or different for each parameters. Also, the variance of the errors may change in a similar manner.

The basic random level shift model has been used previously to model changes in the mean of a time series, whether stationary or long-memory, in particular to try to assess whether a seemingly long-memory model is actually a random level shift process or a genuine long-memory one; see Ray and Tsay (2002), Perron and Qu (2010), Lu and Perron (2010), Qu and Perron (2012), Varneskov and Perron (2012), Li and Perron (2012) and Xu and Perron (2012). It has been shown to provide improved forecasts over commonly used short or long-memory models. Our basic framework is a generalization in which any or all parameters of a forecasting model are modeled as random level shift processes.

To improve the forecasting performance we augment the model in two directions. First, we model the probability of shifts as a function of some covariates which can be forecasted. Second, we allow a mean-reversion mechanism such that the parameters tend to revert back to the pre-forecast average. This last feature is especially influential in providing improvements in forecasting performance at long horizons. Functional forms for these two modifications are suggested for which the parameters can be estimated and incorporated in the forecast scheme to model the future path of the parameters.

Our model can be cast into a non-linear non-Gaussian state space framework for which standard Kalman filter type algorithms cannot be used. The state space representation of our model is actually a linear dynamic mixture models in the sense that it is linear and Gaussian conditional on some latent random variables. See Giordani et al. (2007) who discuss the advantages of the class of conditionally linear and Gaussian state space models. To provide a computationally efficient method of estimation, we rely on recent developments on particle filtering methods. The predictive distribution of the state is approximated by a weighted sum of particles. The key to particle filtering is turning integrals into sums via discrete approximations. The EM algorithm is used to obtain the maximum likelihood estimates of the parameters. This allows treating the latent state variables as missing data (see Bilmes, 1998) and using a complete or data-augmented likelihood function which is easier to evaluate than the original likelihood. Since the missing information is random, the complete-data likelihood function is a random variable and we end up maximizing the expectation of the complete-data log-likelihood with respect to the missing data. Wei and Tanner (1990) introduced the Monte Carlo EM algorithm where the evaluation step is executed by Monte Carlo methods. Random samples from the conditional distribution of the missing data (state variables) can be obtained via a particle smoothing algorithm. For an application of the use of such methods to the estimation of stochastic volatility models, see Kim (2006). The forecasting procedure is then relatively simple and can be carried out in a straightforward fashion once the model has been estimated. Simulations show that the estimation method provides very reliable results in finite samples. The parameters are estimated precisely and the filtered estimates of the time path of the parameters follow closely the true process.

We apply our forecasting model to a variety of series which have been the object of considerable attention from a forecasting point of view. These include the equity premium, inflation, exchange rates and the Treasury bill interest rates. In each case, we compare the forecast accuracy of our model relative to the most important forecasting methods applicable for each variable. We also consider different forecasting sub-samples or periods. The results show clear gains in forecasting accuracy, sometimes by a very wide margin; e.g., over 80% reduction in mean squared forecast error for the equity premium over all popular contenders.

Finally, note that given the availability of the proper code for estimation and forecasting, the method is very flexible and easy to implement. For a given forecasting model, all that is required by the users are: 1) which parameters (including the variance of the errors if desired) are subject to change; 2) whether the same or different latent Bernoulli processes dictates the timing of the changes in each parameters; 3) which covariates are potential explanatory variables to model the probability of shifts.

The rest of the paper is organized as follows. Section 2 describes the basic model with random level shifts in the parameters. Section 3 discusses the modifications introduced to improve forecasting: the modeling of the probability of shifts and the allowance for a mean-reverting mechanism. Section 4 presents the estimation methodology: the particle filtering algorithm in Section 4.1, the particle smoothing algorithm in Section 4.2, the Monte Carlo Expectation Maximization method to evaluate the likelihood function in Section 4.3, and in Section 4.4 issues related to initialization and the construction of the standard errors of the estimates. Section 5 presents results pertaining to the accuracy of the estimation method in finite samples. Section 6 contains the various applications and comparisons with other forecasting methods. Section 7 offers brief concluding remarks.

2 The basic model

We consider a basic forecasting model specified by

$$y_t = X_t \beta_t + e_t \tag{1}$$

where y_t is a scalar variable to be forecasted, X_t is a k-vector of covariates and, in the base case $e_t \sim i.i.d.\ N(0, \sigma_e^2)$. It is assumed that the some or all of the parameters are time-varying and exhibit structural changes at some unknown time. The specification adopted for the time-variation in the parameters is the following:

$$\beta_t = \beta_{t-1} + K_t^{\beta} \eta_t$$

where $K_t^{\beta} = diag(K_{1,t}^{\beta}, ..., K_{k,t}^{\beta})$ and $\eta_t = (\eta_{1,t}, ..., \eta_{k,t})' \sim i.i.d.\ N(0, \Sigma)$. The latent variables $K_{j,t}^{\beta} \sim Ber(p^{(j)})$ and are independent across j. Hence, the parameter evolves according to a Random Level Shift (RLS) process such that the shifts are dictated by the outcomes of the Bernoulli random variables $K_{j,t}^{\beta}$. When $K_{j,t}^{\beta} = 1$, a shift $\eta_{j,t}$ occurs drawn from a $N(0, \sigma_{\eta,j}^2)$ distribution, otherwise when $K_{j,t}^{\beta} = 0$, the parameter does not change. The shifts can be rare (small values of $p^{(j)}$) or frequent (larger values of $p^{(j)}$).

This specification is ideally suited to model changes in the parameters occurring at unknown dates. Many specifications are possible depending on the assumptions imposed on K_t^{β} and Σ. First, when $K_{1,t}^{\beta} = ... = K_{k,t}^{\beta}$, we can interpret the model as one in which all parameters are subject to change at the same times, akin to the pure structural change model of Bai and Perron (1998). A partial structural model, can be obtained by setting $p^{(j)} = 0$ for the parameters not allowed to change, or equivalently by setting the corresponding rows and columns of Σ to 0. The case with $K_{1,t}^{\beta} = ... = K_{k,t}^{\beta}$ is arguably the most interesting for a variety of applications. However, it is also possible not to impose equality for the different $K_{j,t}^{\beta}$. This allows the timing of the changes in the different parameters to be governed by different independent latent processes. This may be desirable in some cases. For instance, it is reasonable to expect changes in the constant to be related to low frequency variations of the random level shifts type, while changes in the coefficients associated with random regressors to be related to business-cycle type variations. In such cases, it would therefore be desirable to allow the timing of the changes to be different for the constant and the other parameters. Of course, many different specifications are possible, and the exact structure needs to be tailored to the specific application under study.

The assumption that the latent Bernoulli processes $K^{\beta}_{j,t}$ are independent across j may seem strong. It implies that the timing of the changes are independent across parameters. As stated above, this can be relaxed by imposing a perfect correlation, i.e., setting some latent variables to be the same. Ideally, one may wish to have a more flexible structure that would allow imperfect though non-zero correlation. This generalization is not feasible in our framework. In many cases, it may also be sensible to impose that Σ is a diagonal matrix. This implies that the magnitudes of the changes in the various parameters are independent. In our applications, we follow this approach as it appears the most relevant case in practice and also considerably reduces the complexity of the estimation algorithm to be discussed in the next section. Hence, for the j^{th} parameter β_j $(j = 1, ..., k)$, we have

$$\beta_{j,t} = \beta_{j,t-1} + K^{\beta}_{j,t}\eta_{j,t} \tag{2}$$

where $\eta_{j,t} \sim N(0, \sigma^2_{\eta,j})$ and $K^{\beta}_{j,t} \sim Ber(p^{(j)})$.

In some cases, it may also be of interest to allow for changes in the variance of the errors. The specification for the distribution is then $e_t = \sigma_{\varepsilon,t}\varepsilon_t$ with

$$ln\sigma^2_{\varepsilon,t} = \phi ln\sigma^2_{\varepsilon,t-1} + K^{\sigma}_t v_{\varepsilon,t} \tag{3}$$

where $\varepsilon_t \sim N(0, 1)$, $K^{\sigma}_t \sim Ber(p^{\sigma})$ and $v_{\varepsilon,t} \sim N(0, \sigma^2_v)$.

Remark 1 *When $p^{(j)} = p^{\sigma} = 0$ for all j, the model reduces to the classic regression model with time invariant parameters. When $p^{(j)} = 1$ for all j and $p^{\sigma} = 0$, it becomes the standard time varying parameter model; e.g., Rosenberg (1973), Chow (1984), Nicholls and Pagan (1985) and Harvey (2006).*

Remark 2 *The model can be extended to a multiple regressions framework such as VAR models. However, the focus here will be on a single equation model.*

3 Modifications useful for forecasting improvements

The framework laid out in the previous section is well tailored to model in-sample breaks in the parameters. However, as such it does not allow future breaks to play a role in forecasting. In order to be able to do so, we incorporate some modifications. Two features that are likely to improve the fit and the forecasting performance is to allow for changes in the probability of shifts and model explicitly a mean-reverting mechanism for the level shift component. In the first step, we specify the jump probability to be

$$p_t^{(j)} = f(p, w_t)$$

where p is a constant, w_t are covariates that would allow to better predict the probability of shifts and f is a function that ensure that $p_t \in [0, 1]$. Note that w_t needs to be in the information set at time t in order for the model to be useful for forecasting. We shall adopt a linear specification with the standard normal cumulative distribution function $\Phi(\cdot)$, so that $K_{j,t}^{\beta} \sim Ber(p_t)$ with $p_t = \Phi(r_0 + r_1 w_t)$. As similar specification can be made for the probability of the Bernoulli random variable K_t^{σ} affecting the shifts in the variance of the errors.

The second step involves building a mean reverting mechanism to the level shift model. The motivation for doing so is that we observe evidence that parameters does not jump arbitrarily and that large upward movement tend to be followed by a decrease. This feature can be beneficial to improve the forecasting performance if explicitly modeled. The specification we adopt is the following:

$$\begin{aligned} \eta_{j,t} &\sim N(\mu_{\eta,j,t}, \sigma_{\eta,j}^2) \\ \mu_{\eta,j,t} &= \rho(\beta_{j,t-1} - \bar{\beta}_j^{(t-1)}) \end{aligned}$$

where $\beta_{j,t-1}$ is the filtered estimate of the parameter subject to change at time $t-1$ and $\bar{\beta}_j^{(t-1)}$ is the mean of all the filtered estimates of the jump component from the beginning of the sample up to time $t-1$. This implies a mean-reverting mechanism provided $\rho < 0$. The magnitude of ρ then dictates the speed of reversion. Note that the specification involves using data only up to time $t-1$ in order to be useful for forecasting purposes. Also, it will have an impact on forecasts since being in a high (low) values state implies that in future periods the values will be lower (higher), and more so as the forecasting horizon increases. Hence, this specification has an effect on the forecasts of both the sign and size of future jumps in the parameters. Similar specifications can be made to p^{σ} and $v_{\varepsilon,t}$ for the changes in the variance of the errors.

The out-of-sample forecasts are then constructed in two steps. The first involves forecasting the covariates w_t using a preliminary model; e.g., using an $AR(p)$. The h-step ahead forecast of the jump probability is then $p_{t+h|t} = \Phi(r_0 + r_1 w_{t+h|t})$ where $w_{t+h|t}$ is the h-step ahead forecast of w_{t+h} at time t. Note that one can also forecast the regressors X_t to obtain predicted values denoted by $X_{t+h|t}$. The second step is to forecast $\beta_{j,t}$ as a weighted sum of two different possible outcomes: one with structural breaks and one without. For example, the one-step-ahead forecast at time t is

$$\beta_{j,t+1|t} = E(\beta_{j,t+1}|\beta_{j,t}) = (1 - p^{(j)}_{t+1|t})\beta_{j,t} + p^{(j)}_{t+1|t}(\beta_{j,t} + \mu_{\eta,j,t+1|t}),$$

where $\mu_{\eta,j,t+1|t} = \rho(\beta_{j,t|t} - \bar{\beta}^{(t)}_j)$. Longer horizon forecasts are based on forward iterations to compute future conditional means. Therefore, the h-step ahead forecast is

$$\begin{aligned}\beta_{j,t+h|t} &= E(\beta_{j,t+h}|\beta_{j,t}) = E(\beta_{j,t} + \sum_{k=1}^{h} p^{(j)}_{t+k|t}\mu_{\eta,j,t+k|t}|\beta_{j,t}) \\ &= \beta_{j,t} + \sum_{k=1}^{h} p_{t+k|t}\mu_{\eta,j,t+k|t} = \beta_{j,t} + \sum_{k=1}^{h} p^{(j)}_{t+k|t}\rho(\beta_{j,t+k-1|t} - \bar{\beta}^{(t+k-1)}_j)\end{aligned}$$

As the forecast horizon increases, the probability of future structural changes also increases. This feature is also present in some Bayesian-type forecasting methods for out-of-sample structural breaks; see, e.g., Hauwe et al. (2011).

4 Estimation Methodology

The model described is within the class of non-linear non-Gaussian State Space models of the form

$$\begin{aligned}y_t &= H_t(x_t, v_t) \\ x_t &= F_t(x_{t-1}, u_t)\end{aligned}$$

where y_t is the variable to be forecasted and x_t are the latent processes. The measurement equation is (1) and the transition equations are (2) and (3). This implies that standard Kalman filter type algorithms are not appropriate and that an extended estimation method is needed. The one adopted is discussed in this section.

4.1 Particle filtering

As an alternative to simulation-based algorithms like Markov Chain Monte Carlo (MCMC) methods, particle filters and smoothers are sequential Monte Carlo methods used to evaluate the probability distribution of some variable x, which is hard to compute directly as in cases for which the analytic solutions are not available. They approximate the continuous distribution of x by a discrete distribution involving a set of weights and particles $\{w^{(i)}, x^{(i)}\}_{i=1}^{M}$. We can view particle filters and smoothers as generalizations of the Kalman filters and smoothers for general state space models. Since our model setup includes mixtures of normal errors and stochastic volatility, standard Kalman filtering and smoothing techniques are not applicable. Of particular interest is the fact that sequential Monte Carlo filtering or particle filtering enables us to approximate the conditional density $f(x_t|y^{(t)})$ as a weighted sum of particles, where $y^{(t)} = (y_1, ..., y_t)$ denotes the history of the data up to time t.

The filtered distribution of x_{t+1} conditional on information up to time $t+1$ is

$$p(x_{t+1}|y^{(t+1)}) \propto p(y_{t+1}|x_{t+1})p(x_{t+1}|y^{(t)}) \tag{4}$$

The likelihood $p(y_{t+1}|x_{t+1})$ is usually known, and the predicting density $p(x_{t+1}|y^{(t)})$ is given by:

$$p(x_{t+1}|y^{(t)}) = \int p(x_{t+1}|x_t)p(x_t|y^{(t)})dx_t \tag{5}$$

Particle filtering methods approach the filtering problem through simulations and a discrete approximation of the optimal filtering distribution. More precisely, particle methods approximate $p(x_t|y^{(t)})$ by

$$p^M(x_t|y^{(t)}) = \sum_{i=1}^{M} \omega_t^{(i)} \delta_{x_t^{(i)}}$$

where δ is the Dirac delta function and $\{\omega_t^{(i)}, x_t^{(i)}\}_{i=1}^{M}$ denote a set of weights and particles. Here M is the number of particles. See Johannes and Polson (2006) for a brief introduction to particle filtering, and Doucet et al. (2001) and Ristic et al. (2004) for a textbook discussion. For applications to stochastic volatility models using particle filtering, see Kim et al. (1998), Chib et al. (2006) and Malik and Pitt (2009). Pitt (2005) applies particle filtering to maximum likelihood estimation, while Fernandez and Rubio (2005) apply it to dynamic macroeconomic models. See also Creal (2012) for a survey of applications of sequential Monte Carlo methods in economics and finance.

Via resampling $\{w_t^{(i)}, x_t^{(i)}\}_{i=1}^{M}$ can yield an equally weighted random sample from $p(x_t|y^{(t)})$. Therefore, we can discretely approximate $p(x_t|y^{(t)})$ by

$$p^M(x_t|y^{(t)}) = \frac{1}{M}\sum_{i=1}^{M} p(y_t|x_t)p(x_t|x_{t-1}^{(i)}).$$

Hence, using (5) we can update $p(x_t|y^{(t)})$ to $p(x_{t+1}|y^{(t)})$, and using (4) we can obtain sample particles from $p(x_{t+1}|y^{(t+1)})$. There are different sampling strategies developed in the literature, such as sampling/importance resampling (SIR), sequential importance sampling (SIS), exact particle filtering and auxiliary particle filtering algorithms. We adopt the sequential importance sampling with resampling (SISR) algorithm to get particles from p$(x_{t+1}|y^{(t+1)})$. This algorithm was introduced by Gordon et al. (1993) to add a resampling step within the SIS algorithm that can mitigate the weight degeneracy problem. A sequential importance density $q(x_{t+1}|y^{(t+1)})$ is introduced, which is easier to sample from than $p(x_{t+1}|y^{(t+1)})$. By the change of measure formula

$$E[f(x_{t+1})|y^{(t+1)}] = \frac{E_q[w_{t+1}f(x_{t+1})|y^{(t+1)}]}{E_q[w_{t+1}|y^{(t+1)}]},$$

with $w_t = p(x_t|y^{(t)})/q(x_t|y^{(t)})$. Given samples from the importance density,

$$E^M[f(x_{t+1})|y^{(t+1)}] \propto \sum_{i=1}^{M} w_{t+1}^{(i)} f(x_{t+1}^{(i)})$$

Given the recursive specifications for $p(x_{t+1}|y^{(t+1)})$ and $q(x_{t+1}|y^{(t+1)})$, we have

$$p(x_{t+1}|y^{(t+1)}) \propto p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)p(x_t|y^{(t)})$$

and

$$q(x_{t+1}|y^{(t+1)}) \propto q(x_{t+1}|x_t, y^{(t+1)})q(x_t|y^{(t)})$$

The weights are also recursive, so that:

$$\begin{aligned} w_{t+1}^{(i)} &= \frac{p(x_{t+1}^{(i)}|y^{(t+1)})}{q(x_{t+1}^{(i)}|y^{(t+1)})} = \frac{p(y_{t+1}|x_{t+1}^{(i)})p(x_{t+1}^{(i)}|x_t^{(i)})}{q(x_{t+1}^{(i)}|x_t^{(i)}, y^{(t+1)})} \frac{p(x_t^{(i)}|y^{(t)})}{q(x_t^{(i)}|y^{(t)})} \\ &= \frac{p(y_{t+1}|x_{t+1}^{(i)})p(x_{t+1}^{(i)}|x_t^{(i)})}{q(x_{t+1}^{(i)}|x_t^{(i)}, y^{(t+1)})} w_t^{(i)} \end{aligned}$$

The first equation follows from the definition of the change of probability measures, and the second from the recursive representation for $p(x_{t+1}|y^{(t+1)})$ and $q(x_{t+1}|y^{(t+1)})$. We summarize the steps involved in implementing the SISR algorithm in the context of our forecasting model with one parameter subject to change and no stochastic volatility, in which case the state or latent variables are $\{K_t^\beta\}$ and $\beta_{j,t}$.

- **Particle filtering algorithm based on SISR**: First, for $i = 1, \ldots, M$: generate $K_0^{\beta(i)} \sim Ber(p_0)$, then $\beta_0^{(i)} \sim K_0^{\beta(i)} * N(\mu_{\eta_0}, \sigma_{\eta_0}^2)$. Set the initial weights to $w_0^{(i)} = (1/M)$. Second, for $t = 1, \ldots, T$: generate $K_t^{\beta(i)} \sim Ber(p_t)$ and $\beta_t^{(i)} = \beta_{t-1}^{(i)} + K_t^{\beta(i)} * N(\mu_{\eta,t}, \sigma_\eta^2)$, compute
$$w_t^{(i)} \propto p(y_t|x_t^{(i)}) w_{t-1}^{(i)} \propto \frac{1}{\sqrt{2\pi\sigma_e^2}} exp(-\frac{(y_t - X_t\beta_t^{(i)})^2}{2\sigma_e^2}),$$
for $i = 1, \ldots, M$, and normalize the weights to get $\hat{w}_t^{(i)} = w_t^{(i)} / \sum_{i=1}^M w_t^{(i)}$. Then, resample $\{\beta_t^{(i)}, K_t^{\beta(i)}\}_{i=1}^M$ with probability $\hat{w}_t^{(i)}$, and set $w_t^{(i)} = (1/M)$. Repeat the steps above increasing from $t + 1$ until T.

Remark 3 *Resampling the particles $\{\beta_t^{(i)}, K_t^{\beta(i)}\}_{i=1}^M$ implies replicating a new population of particles from the existing population in proportion to their normalized importance weights. In Gordon et al. (1993), resampling is carried out every time period, and M random variables are drawn with replacement from a multinomial distribution with probabilities $\{\hat{w}_t^{(i)}\}_{i=1}^M$. After resampling, we set the weights of the particles to a constant $(1/M)$. This resampling scheme allows solving the weight degeneracy problem of the SIS algorithm.*

Remark 4 *Liu and Chen (1995) suggest to resample only when the importance weights are unstable to decrease the effect of Monte Carlo variation impacted to the estimator. The effective sample size (ESS) as a measure of the weight instability is defined as:*
$$ESS = \frac{1}{\sum_{i=1}^M (\hat{w}_t^{(i)})^2}$$
At each time period, ESS is calculated and compared to a user chosen threshold. If ESS drops below the threshold, then resampling is performed. Usually the threshold is picked as a percentage of the number of particles, e.g., in the range 0.5 to 0.75. In our applications, we use 0.5.

4.2 Particle Smoothing

The particle smoothing algorithm is designed to obtain particle smothers $\{s_t^{(i)}\}_{i=1}^M$ with certain weights $\{w_t^{(i)}\}_{i=1}^M$ from $p(x_t|y^{(T)})$. Godsill et al. (2004) provide a forward-filtering and backward-simulation smoothing procedure. It allows drawing random samples from the joint density $p(x_0, x_1, \ldots x_T|y^{(T)})$, not only the individual marginal smoothing densities $p(x_t|y^{(T)})$. The smoothing algorithm relies on a pre-filtering procedure and previously obtained set of particles $\{w_t^{(i)}, x_t^{(i)}\}_{i=1}^M$ for each time period. The main ingredients behind the smoothing algorithm are the relations:

$$p(x_1, x_2, \ldots, x_T|y^{(T)}) = p(x_T|y^{(T)}) \prod_{t=1}^{T-1} p(x_t|x_{t+1}, \ldots, x_T, y^{(T)})$$

and

$$\begin{aligned} p(x_t|x_{t+1}, \ldots, x_T, y^{(T)}) &= p(x_t|x_{t+1}, y^{(t)}) \\ &= \frac{p(x_t|y^{(t)})p(x_{t+1}|x_t)}{p(x_{t+1}|y^{(t)})} \propto p(x_t|y^{(t)})p(x_{t+1}|x_t) \end{aligned}$$

The first equality follows from the Markov property of the model and the second from Bayes' rule. Since random samples $\{x_t^{(i)}\}_{i=1}^M$ from $p(x_t|y^{(t)})$ can be obtained from the particle filtering algorithm, $p(x_t|x_{t+1}, \ldots, x_T, y^{(T)})$ can be approximated as $\sum_{i=1}^M w_{t|t+1}^{(i)} \delta_{x_t^{(i)}}(x_t)$ with modified weights

$$w_{t|t+1}^{(i)} = \frac{w_t^{(i)} p(x_{t+1}|x_t^{(i)})}{\sum_{i=1}^M w_t^{(i)} p(x_{t+1}|x_t^{(i)})}.$$

This procedure is performed in a reverse-time direction conditioning on future states. Given a random sample $\{s_{t+1}, \ldots, s_T\}$ drawn from $p(x_{t+1}, \ldots, x_T|y^{(T)})$, we take one step back and sample s_t from $p(x_t|s_{t+1}, \ldots, s_T, y^{(T)})$. The smoothing algorithm is summarized as follows in the context of the simple version of our model.

- **Particle smoothing algorithm**: Consider the weighted particles obtained from the filtering algorithm $\{w_t^{(i)}, \beta_t^{(i)}, K_t^{\beta(i)}\}_{i=1}^M$ for $i = 1, \ldots, M$, and $t = 1, \ldots, T$. Let $\{s_{\beta,t}^{(j)}, s_{K_1,t}^{(j)}\}_{j=1}^M$ be a set of particle smoothers. First set $s_{\beta,T}^{(j)} = \beta_T^{(i)}$ and $s_{K_1,T}^{(j)} = K_T^{\beta(i)}$ with probability $(1/M)$. Then, for $t = T-1, T-2, \ldots, 1$, compute

  $$w_{t|t+1}^{(i)} \propto w_t^{(i)} p(s_{\beta,t+1}^{(j)}|\beta_t^{(i)}) \propto \{p_{t+1} exp(-\frac{(s_{\beta,t+1}^{(j)} - \beta_t^{(i)} - \mu_{\eta,t})^2}{2\sigma_\eta^2})\}^{s_{K_1,t+1}^{(j)}} \{1 - p_{t+1}\}^{1 - s_{K_1,t+1}^{(j)}}$$

  for $i = 1, \ldots, M$, and let $s_{\beta,t}^{(j)} = \beta_t^{(i)}$ and $s_{K_1,t+1}^{(j)} = K_t^{\beta(i)}$ with probability $w_{t|t+1}^{(i)}$. Repeat the steps above decreasing from $t-1$ until 1 to obtain $\{s_{\beta,t}^{(j)}, s_{K_t^\beta,t+1}^{(j)}\}$ as approximations to $p(\beta_t, K_t^\beta|y^{(T)})$, for $j = 1, \ldots, M$.

4.3 MCEM algorithm

Frequentist likelihood-based parameter estimation of non-linear and non-Gaussian state space models using particle filters and smoothers is not straightforward. The gradient-based optimizer suffers from a discontinuity problem caused by the resampling. Here, we follow the Monte Carlo Expectation Maximization (MCEM) method proposed by Olsson et al. (2008). The Basic EM algorithm is a general method to obtain the maximum-likelihood estimates of the parameters of an underlying distribution from a given data set with missing values. Suppose the complete data set is $Z = (Y, X)$, in which Y is observed but X is unobserved. For the joint density $p(z|\Theta) = p(y, x|\Theta) = p(y|\Theta)p(x|y, \Theta)$, we define the complete-data likelihood function by $L(\Theta|Y, X) = p(Y, X|\Theta)$. The original likelihood $L(\Theta|Y)$ is the incomplete-data likelihood. Since X is unobserved and may be generated from an underlying distribution, e.g., the transition equation in a state space model, $L(\Theta|Y, X)$ is indeed a random variable. Therefore, we maximize the expectation of $logL(\Theta|Y, X)$ with respect to X, with the expectation defined by:

$$Q(\Theta, \Theta^{(k-1)}) = E[logL(\Theta|Y, X)|Y, \Theta^{(k-1)}] = \int logp(Y, x|\Theta)p(x|Y, \Theta^{(k-1)})dx$$

The difference between Monte Carlo EM algorithm and the basic EM algorithm is that when evaluating $Q(\Theta, \Theta^{(k-1)})$, the MCEM uses a Monte-Carlo based sample average to approximate the expectation. The Monte Carlo Expectation or E-step is:

$$Q^*(\Theta, \Theta^{(k-1)}) = \frac{1}{M}\sum_{i=1}^{M} logp(Y, x^{(i)}|\Theta)$$

where $\{x^{(i)}\}_{i=1}^{M}$ are random samples from $p(x|Y, \Theta^{(k-1)})$. Given current parameter estimates, random samples from $p(x|Y, \Theta^{(k-1)})$ are simply the particle smoothers $\{s_t^{(i)}\}_{i=1}^{M}$ obtained as described above. The Maximization or M-step is:

$$\Theta^{(k)} = \arg\max_{\Theta} Q(\Theta, \Theta^{(k-1)})$$

These two steps are repeated until $\Theta^{(k)}$ converges. The rate of convergence has been studied by many researchers; e.g., Dempster et al. (1977), Wu (1983) and Xu and Jordan (1996). In the context of the simple version of our model, the specifics of the algorithm are as follows. For the E-step, the complete likelihood of $\{\beta_1, \ldots, \beta_T, K_1^\beta, \ldots, K_T^\beta, y_1, \ldots, y_T\}$ is

$$f(\beta, K_1, Y) = \prod_{t=1}^{T} f(\beta_t|\beta_{t-1}, K_t^\beta)\prod_{t=1}^{T} f(K_t^\beta)\prod_{t=1}^{T} f(y_t|\beta_t, K_t^\beta)$$

$$= \{\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma_\eta^2}}exp(-\frac{(\beta_t - \beta_{t-1} - \mu_{\eta,t})^2}{2\sigma_\eta^2})\}^{K_t^\beta}\prod_{t=1}^{T} p_t^{K_t^\beta}(1-p_t)^{1-K_t^\beta}\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma_e^2}}exp(-\frac{(y_t - X_t\beta_t)^2}{2\sigma_e^2})$$

The log-likelihood function is:

$$
\begin{aligned}
-2logf(\beta, K^\beta, Y) &= \sum_{t=1}^{T} K_t^\beta [log(\sigma_\eta^2) + \frac{(\beta_t - \beta_{t-1} - \mu_{\eta,t})^2}{\sigma_\eta^2}] \\
&\quad -2\sum_{t=1}^{T}[K_t^\beta log(p_t) + (1-K_t^\beta)log(1-p_t)] \\
&\quad + \sum_{t=1}^{T}[log(\sigma_e^2) + \frac{(y_t - X_t\beta_t)^2}{\sigma_e^2}]
\end{aligned}
$$

The expectation of the complete log-likelihood function with respect to the unknown state variables β, K^β given Y and current parameter estimates $\Theta^{(k-1)}$ is the objective function to be maximized (or minimized if using the negative of the log-likelihood function). For the Monte Carlo EM algorithm, we approximate the expectation by Monte Carlo sample average with random samples drawn from $p(\beta_t, K_t^\beta | y^T)$ obtained using the particle smoothing algorithm. Then,

$$
\begin{aligned}
Q(\Theta, \Theta^{(k-1)}) &= E[-2logf(\beta, K^\beta, Y)|Y, \Theta^{(k-1)}] \\
&= \frac{1}{M}\sum_{i=1}^{M}\{\sum_{t=1}^{T} K_t^{\beta(i)}[log(\sigma_\eta^2) + \frac{(\beta_t^{(i)} - \beta_{t-1}^{(i)} - \mu_{\eta,t}^{(i)})^2}{\sigma_\eta^2}] \\
&\quad -2\sum_{t=1}^{T}[K_t^{\beta(i)} log(p_t) + (1-K_t^{\beta(i)})log(1-p_t)] \\
&\quad + \sum_{t=1}^{T}[log(\sigma_e^2) + \frac{(y_t - X_t\beta_t^{(i)})^2}{\sigma_e^2}]\}
\end{aligned}
$$

For the M-step, note that conditional on K^β, the model is a linear Gaussian state space model. Hence, standard maximum likelihood estimates are obtained by solving the first order condition.

Remark 5 *For the full model with stochastic volatility, the estimation methodology is the same. The difference is that instead of having two state variables, we now have four state variables* $\{\beta_t, K_t^\beta, ln\sigma_{\varepsilon,t}^2, K_t^\sigma\}$. *Similarly, if different parameters are allowed to vary independently, we simply add the additional latent variables* $(\beta_{jt}, K_{jt}^\beta)$.

4.4 Selection of the initial values and construction of the standard errors

In order to speed up the convergence of the estimation algorithm, we can use information from the data in order to provide better initial parameter values. Consider for example, the simple model

$$\begin{aligned} y_t &= \beta_t + e_t \\ \beta_t &= \beta_{t-1} + K_t^\beta \eta_t \end{aligned}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$, $e_t \sim N(0, \sigma_e^2)$ and $K_t^\beta \sim Ber(p)$. The initial parameter values are set to $p^{(0)}\sigma_\eta^{2(0)} = |var(y - y_{-2}) - var(y - y_{-1})|$ and $\sigma_e^{2(0)} = (var(y - y_{-1}) - p^{(0)}\sigma_\eta^{2(0)})/2$. We set $p^{(0)}$ according to prior judgment about the frequency of the jumps.

To construct the standard errors of the estimates, Louis (1982) provides a way of obtaining the information matrix when using the EM algorithm. It is given by

$$\begin{aligned} I &= \sum_{t=1}^{T} E[B(\chi_t, \hat{\Theta})|\chi_t] - \sum_{t=1}^{T} E[S(\chi_t, \hat{\Theta})S^T(\chi_t, \hat{\Theta})|\chi] \\ &\quad -2\sum_{t<k}^{T} E[S(\chi_t, \hat{\Theta})|\chi]E[S(\chi_k, \hat{\Theta})|\chi]' \end{aligned}$$

where $S(\chi_t, \hat{\Theta})$ and $B(\chi_t, \hat{\Theta})$ are the first and second order derivatives, respectively and χ refers to the complete data set including both observed data and unobserved state variables. However, since simulations are used in the EM algorithm, this may cause discontinuities, in which case this method is unstable and cannot always provide a positive definite covariance matrix. Duan and Fulop (2011) proposed a stable estimator of the information matrix applicable to the EM algorithm. They estimate the variance using the smoothed individual scores. Define $a_t(\Theta) = E[\partial log f(x_t|\chi_{t-1}, \Theta)/\partial\Theta|Y, \Theta]$, then the estimate of the information matrix is

$$\hat{I} = \Omega_0 + \sum_{j=1}^{l} w(l)(\Omega_j + \Omega_j')$$

where $\Omega_j = \sum_{t-1}^{T-j} a_t(\hat{\Theta})a_{t+j}(\hat{\Theta})'$ and $w(j) = 1 - j/(l+1)$. This method is easy to compute and does not require evaluations of the second-order derivatives of the complete data log-likelihood.

5 Simulations

We now present simulation results to assess the adequacy of our estimation method in providing good estimates in finite samples. All simulation results are obtained from $N = 1000$ particles and the sample size is $T = 1000$. The number of replications is 500.

We start with a simple model in which only a constant is included as a regressor and the variance of the errors does not change. The model is then

$$\begin{aligned} y_t &= \beta_t + e_t \\ \beta_t &= \beta_{t-1} + K_t^{\beta} \eta_t \end{aligned} \tag{6}$$

where $e_t \sim i.i.d.\ N(0, \sigma_e^2)$, $\eta_t \sim i.i.d.\ N(0, \sigma_\eta^2)$ and $K_t^{\beta} \sim Ber(p_t)$ with $p_t = \Phi(r_0 + r_1 w_t)$. We start with a case with infrequent shifts with parameters given by $\theta_0 = (r_0, r_1, \sigma_e, \sigma_\eta) = (-1.96, 4, 0.2, 0.2)$. Since we are not concerned about forecasting here, we set $\mu_{\eta,t} = 0$. The covariate w_t is a vector in which every 50 time periods $w_t = 1$, and 0 otherwise. Hence, a shift occurs with probability very close to one every 50 periods, otherwise the probability of a shift is 2.5%. We also consider a case with frequent jumps so that a shift occurs with probability 0.5 every time period. In this case the parameter values are $\theta_0 = (0, 0, 0.2, 0.2)$. In both cases, we use the true parameter values as the initial conditions. Table 1 (panel A) presents the mean and standard errors of the estimates showing that, in both cases, they are very accurate. Figure 1(a,b) presents a plot of the true path of the process β_t along with the filtered estimates $\beta_{t|t}$ obtained for the particle filter algorithm. This is done for a single realization chosen randomly. These reveal that the filtered estimates provide very accurate estimates of the time path of the parameter.

We now present results when using adding a mean reverting component and allowing the variance of the errors to change. Hence, $e_t = \sigma_{\varepsilon,t} \varepsilon_t$ with

$$ln\sigma_{\varepsilon,t}^2 = \phi ln\sigma_{\varepsilon,t-1}^2 + K_t^{\sigma} v_{\varepsilon,t} \tag{7}$$

where $\varepsilon_t \sim i.i.d.\ N(0, 1)$, $K_t^{\sigma} \sim Ber(p^{\sigma})$ and $v_{\varepsilon,t} \sim i.i.d.\ N(0, \sigma_v^2)$. Also,

$$\begin{aligned} \eta_t &\sim N(\mu_{\eta,t}, \sigma_\eta^2) \\ \mu_{\eta,t} &= \rho(\beta_{t-1} - \bar{\beta}^{(t-1)}) \end{aligned}$$

The true parameters are $\theta_0 = (r_0, r_1, p^{\sigma}, \phi, \sigma_v, \sigma_\eta, \rho) = (-1.96, 4, 0.5, 0.95, 0.2, 0.2, -0.1)$. The covariate w_t is as specified before. The mean and standard deviations of the estimates are presented in Panel B of Table 1. Figure 2 presents a graph of the path of the true β_t and $ln\sigma_{\varepsilon,t}^2$ along with their filtered estimates, again for a single realization chosen randomly. The results show that the mean values of the estimates are close to the true values. The filtered estimates of β_t and $ln\sigma_{\varepsilon,t}^2$ follow the general time variations of the true processes, though not as precisely as in the simplified case.

To assess the robustness of our estimation method, we first consider the simplified model (6) with $K_t^\beta \sim Ber(p)$ for two extreme cases involving a constant probability of changes. One has β_t constant (jump probability 0) and the other has β_t changing every period (jump probability 1). Here, the parameter space of interest is $\theta = (p, \sigma_e, \sigma_\eta)$. The simulation results for the means and standard deviations are presented in Table 2 for both cases. For the case with $p = 0$, the estimates are very accurate. When $p = 1$, p is very precisely estimated but the estimates of σ_e and σ_η are slightly biased upward.

The next experiment aims to assess whether it is detrimental to introduce a mean reversion component when none is present. To that effect, we use model (6) with the addition that

$$\begin{aligned} \eta_t &\sim N(\mu_{\eta,t}, \sigma_\eta^2) \\ \mu_{\eta,t} &= \rho(\beta_{t-1} - \bar{\beta}^{(t-1)}) \end{aligned}$$

The true parameter values are $\theta_0 = (r_0, r_1, \sigma_e, \sigma_\eta, \rho) = (-1.96, 4, 0.2, 0.2, 0)$. The results presented in the last panel of Table 2 show that the estimate of all parameters are precise so that no efficiency loss is incurred.

6 Forecasting applications

We consider a variety of forecasting applications pertaining to variables which have been the object of intense attention in the literature: the equity premium, inflation, the treasury bill rate and exchange rates. We compare the forecasting performance of our model relative to popular forecasting methods applicable to the different variables. In all cases, our model provides improved forecasts, in some cases by a considerable margin.

Throughout, the out-of-sample forecasting experiments aim at evaluating the experience of a real-time forecaster by performing all model specifications and estimations using data through date t, making a h-step ahead forecast for date $t + h$, then moving forward to date $t + 1$ and repeating this through the sub-sample used to construct the forecasts. The estimation of each model is recursive, using an increasing data window starting with the same initial observation. The forecasting performance is evaluated using the mean square forecast error (MSFE) criterion defined as

$$MSFE(h) = \frac{1}{T_{out}} \sum_{t=1}^{T_{out}} (\bar{y}_{t,h} - \bar{y}_{t+h|t})^2$$

where T_{out} is the number of forecasts produced, h is the forecasting horizon, $\bar{y}_{t,h} = \sum_{k=1}^{h} y_{t+k}$ and $\bar{y}_{t+h|t} = \sum_{k=1}^{h} y_{t+k|t}$ with y_{t+k} the actual observation at time $t+k$ and $y_{t+k|t}$ its forecast conditional at time t. To ease presentation, the MSFE are reported relative to some benchmark model, usually the most popular forecasting model in the literature. In all cases, we allow mean reversion in the paramters of the model when constructing forecasts using our model.

6.1 Equity premium

Forecasts of excess returns at both short and long-horizons are important for many economic decisions. Much of the existing literature has focused on the conditional return dynamics and studied the implications of structural breaks in regression coefficients including the lagged dividend yield, short interest rate, term spread and the default premium. However, most of the research has focused on modeling the equity premium assuming a certain number of structural breaks in-sample while ignoring potential out-of-sample structural breaks. Recently, Maheu and McCurdy (2009) studied the effect of structural breaks on forecasts of the unconditional distribution of returns, focusing on the long-run unconditional distribution in order to avoid model misspecification problems. Their empirical evidence strongly rejects ignoring structural breaks for out-of-sample forecasting. We consider using our forecasting model with different specifications. One models the unconditional mean of excess returns incorporating random level shifts in mean, with the time varying jump probabilities influenced by the absolute rate of growth in the earning price ratio. We also consider a conditional mean model using the dividend yield as the explanatory variable.

Following Jagannathan et al. (2000), we approximate the equity premium of S&P 500 returns as the difference between stock yield and bond yield. The data were obtained from Robert Shiller's website (http://www.econ.yale.edu/~shiller/data.htm). According to Gordon's valuation model, stock returns are the sum of the dividend yields and the expected future growth rate in stock dividends. We use the average dividend growth rate to proxy for the expected future growth rate. The data is monthly and covers the period from 1871 to 2012.5. High quality monthly data are available after 1927, before 1927 the monthly data are interpolated from lower frequency data. We use the 10-year Treasury constant maturity rate (GS10) as the risk free rate.

We start with a simple random level shift model without explanatory variables given by:

$$\begin{aligned} y_t &= \beta_t + e_t \\ \beta_t &= \beta_{t-1} + K_t^{\beta}\eta_t \end{aligned} \tag{8}$$

where $e_t \sim i.i.d.\ N(0, \sigma_e^2)$, $\eta_t \sim i.i.d.\ N(\mu_{\eta,t}, \sigma_\eta^2)$, $\mu_{\eta,t} = \rho(\beta_{t-1} - \bar{\beta}^{(t-1)})$, $K_t^{\beta} \sim Ber(p_t)$ with $p_t = \Phi(r_0 + r_1 w_t)$. The covariate w_t used to model the time variation in the probability of shifts is the lagged absolute value of the rates of changes in the EP ratio. The rational for doing so that it is expected that large fluctuations in the earning price ratio induce a higher probability that excess stock returns will experience a level shift in the unconditional mean. We also allow a mean reversion component and to assess its effect we also consider a version without it. To implement the forecasts, we use an $AR(p)$ model to forecast w_t for which, here and throughout all applications, the number of lags is selected using the Akaike Information Criterion (AIC) with a maximal value of 4.

We also consider a conditional forecasting model that uses the lagged dividend price ratio as the regressor. The specifications are

$$y_t = \beta_{1t} + \beta_{2t} dp_{t-1} + e_t \tag{9}$$

where, with $\beta_t = (\beta_{1t}, \beta_{2t})$,

$$\beta_t = \beta_{t-1} + K_t^{\beta}\eta_t.$$

Lettau and van Nieuwerburgh (2007) analyzed the implications of structural breaks in the mean of the dividend price ratio for conditional return predictability. Xia (2001) studied model instability using a continuous time model relating excess stock returns to dividend yields. They model the coefficient β_t to follow an Ornstein–Uhlenbeck process and the ensuing estimates of the time varying coefficient β_{2t} revealed instability of the forecasting relationship. Hence, instabilities have been shown to be of concern when using this conditional forecasting model, which motivates the use of our forecasting model. Besides the addition of the lagged dividend price ratio as regressors, the specifications are the same as for the unconditional mean model (8).

We consider various versions depending on which coefficients are allowed to change and if so whether they change at the same time. These are: 1) the unconditional mean model (8) with level shifts, 2) the conditional mean model (9) with the coefficient on the lagged dividend yield allowed to change ($K_{1t}^{\beta} = 0$) , 3) the conditional mean model (9) with the constant allowed to change ($K_{2t}^{\beta} = 0$). We compare our forecasting model with the most popular forecasting models used in the literature. These are: 1) a rolling ten-years average (used as the benchmark model); 2) the historical average; 3) the conditional model with a constant and the lagged dividend price ratio as the regressors without changes in the parameters.

We first consider 1998-2012 as the forecasting period, with forecasting horizons $1, 3, 6, 12, 18, 24, 30, 36, 40$. The results are presented in Table 3.1. The first thing to note is that all three versions involving random level shifts perform very well and are comparable. The best model for horizons up to 6 months is the conditional mean model (9) with the coefficient on the lagged dividend yield allowed to change ($K_{1t}^{\beta} = 0$), though the difference are quite minor. For longer horizons, the the unconditional mean model (8) with level shifts is the best. What is noteworthy is that our model performs much better than any competing forecasting models. This is especially the case at short-horizons, for which the gain in forecasting accuracy translates into a reduction in MSFE of up to 90% when compared to the conditional model with no breaks (and even more so when compared to the rolling 10 year average or the historical average, the latter performing especially badly). At longer horizons, the unconditional mean model (8) with level shifts still perform better than the conditional model with constant coefficient but to a lesser extent. Figure 3 presents a plot of the forecasts obtained from the various methods (without the historical average) for horizons 1, 12, 24 and 36 months. On can see that the forecasts from the random level shift model track the actual data quite well.

To assess the robustness of the results we also consider the forecasting period 1988-1996, given that it offers an historical episode with different features. What is noteworthy is that the conditional mean model with constant parameters now performs very poorly with MSFE more than four times the rolling 10 year average. On the other hand the models with random level shifts continue to perform very well, with MSFE around 10% of the rolling 10 years average at short horizons, and around 20% at longer horizons. All models with random level shifts have comparable performance at short horizons, but the conditional mean model (9) with the constant allowed to change ($K_{2t}^{\beta} = 0$) is best at longer horizons.

In summary, the evidence provides strong evidence that our forecasting model offers marked improvements in forecast accuracy. It does so at all horizons with results that are robust to different forecasting periods.

6.2 Inflation

Stock and Watson (2007) documented the fact that the rate of price inflation in the United States has become easier to forecast in the sense that using standard methods yields a lower MSFE since the mid-1980s (the "Great Moderation"). At the same time, however, they showed that the advantage of using a multivariate forecasting model such as the backward-looking Phillips curve has declined concurrently. Hence, in fact inflation has become harder to forecasts except for the fact that the variance of the shocks is smaller. They argued that the best forecasting model is an unobserved components model with stochastic volatility (UC-SV model) which allows for changing inflation dynamic in both the conditional mean and the variance. They conjectured two reasons for the deterioration. One is due to the changes in the variance of the various activity measures used as predictors and the other is due to changes in coefficients. We consider extending both the UC-SV model and the popular backward-looking Phillips curve model to incorporate random level shifts in the parameters and stochastic volatility.

The data used were collected from the Federal Reserve Bank of St. Louis and the US Bureau of Labor Statistics and are the monthly CPI for all items and the civilian unemployment rates (seasonally adjusted) from 1960-2012. Annual inflation rates are constructed as $\pi_t = 1200 ln(P_t/P_{t-1})$.

The unobserved components model with stochastic volatility for inflation y_t is given by:

$$\begin{aligned} y_t &= \beta_t + e_t \\ \beta_t &= \beta_{t-1} + K_t^{\beta}\eta_t \\ e_t &= \sigma_{\varepsilon,t}\varepsilon_t \\ ln\sigma^2_{\varepsilon,t} &= ln\sigma^2_{\varepsilon,t-1} + v_{\varepsilon,t} \end{aligned}$$

with the various variables as defined in Section 2. In this case the probability of shifts is modelled using the unemployment rate as the covariate.

The other class of models considered are based on the popular backward looking Phillips curve

$$\triangle\pi_{t+1} = \beta_t\triangle\pi_t + \mu + \varphi(L)\triangle\pi_{t-1} + \alpha u_t + \gamma(L)\triangle u_t + \sigma_{\varepsilon,t}\varepsilon_t \tag{10}$$

where only the coefficient of the current value of the first-differences in inflation is allowed to change and $\varphi(B)$, $\gamma(B)$ are polynomials in the lag operator L whose order is selected using AIC. Also, β_t, $\sigma_{\varepsilon,t}$ and ε_t are as specified above, with the mean reversion mechanism incorporated for β_t. The covariate w_t used to model the probability of shifts in β_t is the effective federal funds rate.

We also considered several models used in Stock and Watson (2007) for forecasting comparisons. Those are: 1) the $AR(AIC)$ model, which simply uses an AR(p) model for the first-differences of inflation with the lag order selected using the AIC. 2) AO model as suggested by Atkeson and Ohanian (2001) which is simply a 12 periods backward average so that the h-steps ahead forecast is given by

$$\pi_{t+h|t} = \frac{1}{12}(\pi_t + \pi_{t-1} + \cdots + \pi_{t-11}).$$

Note that in this case multi-steps forecasts are the same for all horizons. Since we are using monthly data, the forecasts of future inflation are based on a rolling but fixed length window of average inflation for the previous 12 months. 3) Backward-looking Phillips curve which is the same as model (10) but without allowing for changes in coefficients and stochastic volatility. We also consider a slight modification that drops the regressor u_t and only include the stationary predictors $\triangle u_t$. This version is labelled as model PC-$\triangle u$. 4) UC-SV (unobserved component with stochastic volatility) model:

$$\begin{aligned} \pi_t &= \tau_t + \eta_t \\ \tau_t &= \tau_{t-1} + \varepsilon_t \end{aligned}$$

where $\varepsilon_t = \sigma_{\varepsilon,t}\zeta_{\varepsilon,t}$, $\eta_t = \sigma_{\eta,t}\zeta_{\eta,t}$, $ln\sigma^2_{\eta,t} = ln\sigma^2_{\eta,t-1} + v_{\eta,t}$, $ln\sigma^2_{\varepsilon,t} = ln\sigma^2_{\varepsilon,t-1} + v_{\varepsilon,t}$ with $\zeta_t = (\zeta_{\eta,t}, \zeta_{\varepsilon,t}) \sim i.i.d.\ N(0, I_2)$, $v_t = (v_{\eta,t}, v_{\varepsilon,t}) \sim i.i.d.\ N(0, \gamma I_2)$. Here, γ is the only parameter in the model. It controls the smoothness of the stochastic volatility process. We follow Stock and Watson (2007) and set $\gamma = 0.2$ as a prior when forecasting. Since this is a one step ahead model, multistep forecasts are calculated using an iterated method.

The results are presented in Table 4 for the forecasting period 1984 to the end of the sample and forecast horizons $h = 1, 4, 12, 24, 36, 48, 60, 72, 84, 96$ months. The UC-SV model is used as the benchmark. Consistent with the results of Stock and Watson (2007), the UC_SV performs best at short horizons up to 12 months with the two models with random level shifts close second and third. At longer forecasting horizons, the UC_SV model with random level shifts performs best with reductions in MSFE of up to 20% at the 5 years horizon. At the longest horizon considered, 96 months, the naive AO forecasts are the best but the methods is, however, very unreliable at short horizons. The forecasts of the basic UC_SV model, the UC_SV model with random level shifts and the Phillips Curve model with random level shifts are presented in Figure 4 for foreacst horizons $h = 4$, 48 and 96 months ahead. The results show how the the UC_SV model with random level shifts tracks actual inflation more closely.

Overall, the UC_SV model with random level shifts provides substantial improvements over the standard UC_SV model at long horizons and is still competitive at short horizons.

6.3 Exchange rates

Since the work by Meese and Rogoff (1983a,b), the prevalent view has been that a simple random walk model has a better forecasting performance than macroeconomic models of exchange rates. However, the forecasting failure of fundamental models may not be due to the lack of correlation between fundamentals and exchange rate fluctuations but may be the outcome of instabilities in the relationships. Such instabilities have been documented by Rossi (2006) and Kilian and Taylor (2003), among others. Using tests robust to parameter instability, they reached the conclusion that exchange rates are not random walks in-sample. Here, we shall directly consider two popular fundamentals-based models of exchange rate allowing for time variations in the parameters and compare their forecasting performance to the simple random walk model.

The first is the Uncovered Interest Rate Parity Model (UIRP) which specifies that the first-differences of the logarithm of the bilateral nominal exchange rate, s_t, is determined by

$$s_{t+1} - s_t = \beta_1 + \beta_2 z_t + \epsilon_{t+1}$$

where $z_t = f_t - s_t$, with f_t the long-run equilibrium level of the nominal exchange rate determined by macroeconomic fundamentals. In the UIRP model, $f_t = (i_t - i_t^*) + s_t$, where $i_t - i_t^*$ is the short-term interest rate difference between the home and foreign countries. The UIRP model with time-varying parameters, can then be written as:

$$s_{t+1} - s_t = \beta_{1t} + \beta_{2t}(i_t - i_t^*) + \epsilon_{t+1} \tag{11}$$

We shall apply this model using monthly data (not seasonally adjusted) of the exchange rates for the Japanese Yen and the Canadian Dollar relative to the US dollar. The exchange rate data were obtained from the Board of Governors of the Federal Reserve System. The monthly interest rates for government securities and bonds for Japan, Canada and the USA were obtained from the IMF's International Financial Statistics database. The sample is from 1971:1 to 2012:4. Data prior to 1984 are used for the in-sample estimation, and we consider forecasts up to 24-months ahead.

The second model is one with the so-called Taylor Rule fundamentals as proposed by Molodtsova and Papell (2007). In this model, the interest rate of the home country is assumed to follow a Taylor rule (Taylor, 1993) given by:

$$i_t = \pi_t + \delta(\pi_t - \pi^T) + \gamma y_t^{gap} + r$$

where π_t is the inflation rate, π^T is the target level of inflation, y_t^{gap} is the output gap and r is the equilibrium level of the real interest rate. If we assume that the coefficients of the Taylor rule in the foreign countries are similar to the those of the home country, we obtain, in first-differences, that:

$$i_t - i_t^* = (1+\delta)(\pi_t - \pi_t^*) + \gamma(y_t^{gap} - y_t^{gap*})$$

Then the exchange rate model with time varying parameters is:

$$s_{t+1} - s_t = \beta_{1t} + \beta_{2t}(\pi_t - \pi_t^*) + \beta_{3t}(y_t^{gap} - y_t^{gap*}) + \epsilon_{t+1} \tag{12}$$

For the inflation series, we use monthly data of the CPI for all items from the Federal banks of St. Louis database and calculate the compounded annual rate of change. The output gap series are constructed based on industrial production indices (seasonally adjusted) as the percentage difference between actual and potential output at time t, where potential output is measured by the linear time trend in output.

The models with time variations in the parameters that we consider are; 1) UIRP_level shift: Model (11) with $K_{2t}^{\beta} = 0$; 2) UIRP_RLS: Model (11) with $K_{1t}^{\beta} = 0$; 3) UIRP_ K_t: Model (11) with $K_{1t}^{\beta} = K_{2t}^{\beta}$ so that both parameters are allowed to change according to the same latent Bernoulli variable; 4) UIRP_ K_{1t}, K_{2t}: Model (11) with $K_{1t}^{\beta} \neq K_{2t}^{\beta}$ so that both parameters are allowed to change according to different latent Bernoulli variables; 5) Taylor-Rule_level shift: Model (12) with $K_{2t}^{\beta} = K_{3t}^{\beta} = 0$; 6) Taylor-Rule_RLS (inflation): Model (12) with $K_{1t}^{\beta} = K_{3t}^{\beta} = 0$; 7) Taylor-rule_RLS (output gap): Model (12) with $K_{1t}^{\beta} = K_{2t}^{\beta} = 0$; 8) Taylor-Rule_constant+inflation_K_t: Model (12) with $K_{1t}^{\beta} = K_{2t}^{\beta}$ and $K_{3t}^{\beta} = 0$; 9) Taylor-Rule_constant+output_K_t: Model (12) with $K_{1t}^{\beta} = K_{3t}^{\beta}$ and $K_{2t}^{\beta} = 0$; 10) Taylor-Rule_constant+inflation_K_{1t}, K_{2t}: Model (12) with $K_{1t}^{\beta} \neq K_{2t}^{\beta}$ and $K_{3t}^{\beta} = 0$; 11) Taylor-Rule_constant+output_K_{1t}, K_{2t}: Model (12) with $K_{1t}^{\beta} \neq K_{3t}^{\beta}$ and $K_{2t}^{\beta} = 0$. We also consider a random walk model with random level shifts. In all cases, mean-reversion in the parameters is allowed and, in all cases with a single latent Bernoulli random variable, the covariate w_t used to model the time-varying probabilities is the change in stock returns, constructed from the logarithm of the monthly S&P stock price index. For models with two Bernoulli random variables, the additional covariate is the change in M2. The popular competing models to which we make comparisons are: 1) the random walk model; 2) the UIRP with constant coefficients; 3) the Taylor-Rule with constant coefficients.

The forecasting period is 1969:1 to the end of the sample and we consider the forecast horizons $h = 1, 4, 8, 12, 16, 20, 24$ months. The results are presented in Table 5.1 for Canada and 5.2 for Japan. For Canada, the best forecasting models are the various versions of the UIRP with random level shift parameters and the random walk model with random level shifts. The model with the smallest MSFE for all horizons is the UIRP_ K_{1t}, K_{2t} for which both parameters are allowed to change according to different latent Bernoulli variables. The gains in forecast accuracy range from 2% to 8%. These reductions are more modest in comparisons with the other series analyzed but still important given the difficulty in forecasting exchange rates. The results are broadly similar for Japan, though in this case the UIRP_RLS version has smallest MSFE at longer horizons.

6.4 Interest rate forecasting

Another variable of interest, which has attracted attention from a forecasting perspective is the U.S. T-bill rates. Various studies have shown that it exhibits structural instability in both mean and variance, see, e.g. Garcia and Perron (1996), Gray (1996), Ang and Bekaert (2002) and Pesaran and Timmermann (2006). We use monthly data on the 3-months Treasury Bill rates from 1947:07-2002:12, obtained from the Federal Bank of St. Louis database. The period prior to 1968:12 is used for in-sample estimation, and we consider forecasting horizons of 12, 24, 36, 48 and 60 months. The basic model adopted is a simple AR(1) process with stochastic volatility given by:

$$\begin{aligned} y_t &= \beta_{1t} + \beta_{2t} y_{t-1} + \sigma_{\varepsilon,t}\varepsilon_t \\ ln\sigma^2_{\varepsilon,t} &= ln\sigma^2_{\varepsilon,t-1} + v_{\varepsilon,t} \end{aligned}$$

In all cases, we allow mean-reversion in the parameters and the covariate w_t used to model the time-varying probabilities of shifts is the growth rate of GDP when a single latent Bernoulli varaible is present. When two are present, the additional covariate is the absolute change in stock returns (S&P 500). We consider five possible specifications: 1) AR_level shift_SV ($K^{\beta}_{2t} = 0$ and $K^{\sigma}_t \neq 0$); 2) AR_RLS_SV ($K^{\beta}_{1t} = 0$ and $K^{\sigma}_t \neq 0$); 3) AR_RLS ($K^{\beta}_{1t} = 0$ and $K^{\sigma}_t = 0$); 4) AR_level shift ($K^{\beta}_{2t} = 0$ and $K^{\sigma}_t = 0$); 5) AR_K^{β}_{1t}_K^{β}_{2t} with K^{β}_{1t} and K^{β}_{2t} allowed to be different latent Bernoulli processes. The performance of the models is assessed relative to four commonly used forecasting methods: 1) a 5 years rolling average (used as the benchmark model); 2) a 10 years rolling average; 3) a recursive OLS based an a first-order autoregression with fixed parameters; 4) a time-varying probability model in which β_{2t} is modelled as a random walk.

The results are presented In Table 6 for various forecast periods and forecast horizons $h = 12, 24, 36, 48, 60$ months. Consider first the results for the longest forecasting period 1968-2002. Here, the best forecasting model for all horizons is the AR_RLS with the coefficient on the lagged dependent variable allowed to follow a random level shift process. The gains in forecast accuracy vary between 6 and 17% and increase as the forecasting horizon increases. The same model with added stochastic volatility is the second best. We then separate the forecasting period into three decades: the 70s, the 80s and the 90. In the 70s, the 5 years rolling average is overall the best, though all models perform about the same with the exception of the TVP and the AR_K_{1t}^{β}_K_{2t}^{β} models whose performances are inferior. For the 80s, the best forecasting models are again the AR_RLS with the coefficient on the lagged dependent variable allowed to follow a random level shift process and its variant that incorporate a stochastic volatility component. The improvements in forecast accuracy are quite impressive, ranging from 14% at short-horizon to 60% at long-horizon relative to the benchmark model. For the 90s, the best performing model is the AR_K_{1t}^{β}_K_{2t}^{β} for which both the mean and autoregressive coefficient are changing according to different latent Bernoulli processes. Note, however, that all models with random level shifts in parameters perform better than the benchmark 5 years rolling average.

Overall, the evidence again indicates that important gains in forecast accuracy can be obtained using our forecasting models and that they are robust in the sense that in no case do hey performs substantially worse than the popular forecasting methods. Overall, for the application, the AR_RLS model with the coefficient on the lagged dependent variable allowed to follow a random level shift process is the best and most robust across the various specifications considered.

7 Conclusion

We proposed a forecasting framework based on modeling the parameters as random level shift processes dictated by a Bernoulli process for the occurrence of shifts and a normal random variable for its magnitude. Some or all of the parameters of the model can be allowed to change and the latent variables that dictate the changes can be common or different for each parameters. Also, the variance of the errors may change in a similar manner. To improve the forecasting performance we augmented the basic model to allow the probability of shifts to be a function of some covariates which can be forecasted and to incorporate a mean-reversion mechanism such that the parameters tend to revert back to the pre-forecast average.

Our model can be cast into a non-linear non-Gaussian state space framework for which standard Kalman filter type algorithms cannot be used. To provide a computationally efficient method of estimation, we rely on recent developments on particle filtering methods. Simulations show that the estimation method provides very reliable results in finite samples. The parameters are estimated precisely and the filtered estimates of the time path of the parameters follow closely the true process.

We apply our forecasting model to a variety of series which have been the object of considerable attention from a forecasting point of view. These include the equity premium, inflation, exchange rates and the Treasury bill interest rates. In each case, we compare the forecast accuracy of our model relative to the most important forecasting methods used applicable for each different variable. We also consider different forecasting sub-samples or periods. The results show clear gains in forecasting accuracy, sometimes by a very wide margin; e.g., over 80% reduction in mean squared forecast error for the equity premium over all popular contenders.

Finally, note that given the availability of the proper code for estimation and forecasting, the method is very flexible and easy to implement. For a given forecasting model, all that is required by the users are: 1) which parameters (including the variance of the errors if desired) are subject to change; 2) whether the same or different latent Bernoulli processes dictates the timing of the changes in each parameters; 3) which covariates are potential explanatory variable to model the probability of shifts.

References

Andrews, D.W.K. (1993). “Tests for parameter instability and structural change with unknown change point,” Econometrica 61: 821-856.

Ang, A., and Bekaert, G. (2002). “Regime switches in interest rates,” Journal of Business and Economic Statistics 20: 163-182.

Atkeson, A. and Ohanian, L.E. (2001). “Are Phillips curve useful for forecasting inflation?” Federal Reserve Bank of Minneapolis Quarterly Review 25: 2-11.

Bai, J. and Perron, P. (1998). “Estimating and testing linear models with multiple structural changes,” Econometrica 66: 47-78.

Bai, J. and Perron, P. (2003). “Computation and Analysis of Multiple Structural Change Models,” Journal of Applied Econometrics 18*,* 1-22.

Bilmes, J.A. (1998). “A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models,” working paper, International Computer Science Institute.

Chow, G.C. (1984). “Random and changing coefficient models,” in Griliches, Z., Intriligator, M. (Eds.), Handbook of Econometrics, vol. 2. North Holland, Amsterdam, 1213-1245.

Chib, S., Nardari, F. and Shephard, N. (2006). “Analysis of high dimensional multivariate stochastic volatility models,” Journal of Econometrics 134: 341-371.

Clements, M. and Hendry, D. (2006). “Forecasting with breaks,” in G. Elliott, C. Granger and A. Timmermann (Eds.), Handbook of Economic Forecasting, Elsevier Science, Amsterdam, 605-657.

Creal, D. (2012). “A survey of sequential Monte Carlo methods for economics and finance,” Econometric Reviews 31: 245-296.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). “Maximum-likelihood from incomplete data via the EM algorithm,” Journal of the Royal Statistical Society. Series B 39: 1-38.

Doucet, A., Freitas, N. de and Gordon, N. (2001). “Sequential Monte Carlo methods in practice,” New York: Springer-Verlag, Series Statistics for Engineering and Information Science.

Duan, J. and Fulop, A. (2011). “A stable estimator of the information matrix under EM for dependent data,” Statistics and Computing 21: 83-91.

Fernandez-Villaverde, J. and Rubio-Ramirez, J. (2005). “Estimating macroeconomic models: a likelihood approach,” working paper, University of Pennsylvania.

Garcia, R. and Perron, P. (1996). "An Analysis of the real interest rate under regime shifts," Review of Economics and Statistics 78: 111-125.

Giordani, P., Kohn, R. and van Dijk, D. (2007). "A unified approach to nonlinearity, structural change, and outliers," Journal of Econometrics 137: 112-133.

Godsill, S.J., Doucet, A. and West, M. (2004). "Monte Carlo smoothing for nonlinear time series," Journal of the American Statistical Association 99: 156-168.

Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," Radar and Signal Processing, IEE Proceedings F 140: 107-113.

Gray, S., (1996). "Modeling the conditional distribution of interest rates as regime-switching process," Journal of Financial Economics 42: 27-62.

Harvey, A. (2006). "Forecasting with unobserved components time series models," in G. Elliott, C. Granger and A. Timmermann (Eds.), Handbook of Economic Forecasting, Elsevier Science, Amsterdam, 327-408.

Hauwe, S., Paap, R. and van Dijk, D. (2011). "An alternative Bayesian approach to structural breaks in time series Models," Tinbergen Institute Discussion Paper.

Jagannathan, R., McGrattan, E.R. and Scherbina, A. (2000). "The declining U.S. equity premium," Federal Reserve Bank of Minneapolis Quarterly Review. 24(4): 3-19.

Johannes, M. and Polson, N. (2006). "Particle filtering", Working Paper, University of Columbia.

Kilian, L., and Taylor, M.P. (2003). "Why is it so difcult to beat the random walk forecast of exchange rates?," Journal of International Economics 60: 85-107.

Kim, J. (2006). "Parameter estimation in stochastic volatility models with missing data using particle methods and the EM algorithm," University of Pittsburgh.

Kim, S., Shephard, N. and Chib, S. (1998). "Stochastic volatility: likelihood inference and comparison with ARCH models," Reviews of Economic Studies 65: 361-393.

Koop, G. and Potter, S.M. (2007). "Estimation and forecasting in models with multiple breaks," Review of Economics Studies 74: 763-789.

Lettau, M. and Van Nieuwerburgh, S. (2008). "Reconciling the return predictability evidence," Review of Financial Studies, 21: 1607-1652.

Li, Y. and Perron, P. (2012). "Modelling exchange rate volatility with random level shifts," Unpublished Manuscript, Department of Economics, Boston University.

Liu, J.S. and Chen, R. (1995). "Sequential Monte Carlo methods for dynamic dystems," Journal of the American Statistical Association 93: 1032-1044.

Louis, T.A. (1982). "Finding the observed information matrix when using the EM algorithm," Journal of the Royal Statistical Society. Series B 44: 226-233.

Lu, Y.K. and Perron, P. (2010). "Modeling and forecasting stock return volatility using a random level shift model," Journal of Empirical Finance 17: 138-156.

Maheu, J.M. and Gordon, S. (2008). "Learning, forecasting and structural breaks," Journal of Applied Econometrics 23: 553-583.

Maheu, J.M. and McCurdy, J.H. (2009). "How useful are historical data for forecasting the long run equity return distribution?" Journal of Business & Economic Statistics 27: 95-112.

Malik, S. and Pitt, M.K. (2009). "Modeling stochastic volatility with leverage and jumps: a simulated maximum likelihood approach via particle filtering," Working Paper, University of Warwick.

Meese, R., and Rogoff, K., (1983a). "Exchange rate models of the seventies. Do they fit out of sample?," The Journal of International Economics 14: 3-24.

Meese, R., and Rogoff, K., (1983b). "The out of sample failure of empirical exchange rate models," in Jacob Frankel (Ed.), Exchange Rates and International Macroeconomics. Chicago: University of Chicago Press for NBER.

Molodtsova, T., and Papell, D.H., (2007). "Out-of-sample exchange rate predictability with Taylor rule fundamentals," manuscript, University of Houston.

Nicholls, D.F. and Pagan, A.R. (1985). "Varying coefficient regression," in Hannan, E.J., Krishnaiah, P.R., Rao, M.M. (Eds.), Handbook of Statistics, vol. 5. North-Holland, Amsterdam, pp. 413-450.

Olsson, J., Cappe, O., Douc, R. and Moulines, E. (2008). "Sequential Monte Carlo smoothing with application to parameter estimation in non-linear state space models," Bernoulli 14: 155–179.

Pastor, L. and Stambaugh, R.F. (2001). "The equity premium and structural breaks," Journal of Finance 4: 1207-1231.

Paye, B.S. and Timmermann, A. (2006). "Instability of return prediction models," Journal of Empirical Finance 13: 274-315.

Perron, P. (2006). "Dealing with structural breaks," in K. Patterson and T. Mills (Eds.), Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory, Palgrave Macmilan, Hampshire, 278-352.

Perron, P. and Qu, Z. (2010). "Long-memory and level shifts in the volatility of stock market return indices," Journal of Business and Economic Statistics 28, 275-290.

Pesaran, M.H., Pettenuzzo, D. and Timmermann, A. (2006). "Forecasting time series subject to multiple structural breaks," Review of Economic Studies 73: 1057-1084.

Pesaran, M.H. and Timmermann, A. (2002). "Market timing and return prediction under model instability," Journal of Empirical Finance 9: 495-510.

Pettenuzzo, D. and Timmermann, A. (2011). "Predictability of stock returns and asset allocation under structural breaks," Journal of Econometrics 164: 60-78.

Pitt, M. (2005). "Smooth particle filters for likelihood evaluation and maximization," working paper, University of Warwick.

Qu, Z. and Perron, P. (2012). "A stochastic volatility model with random level shifts and its application to S&P 500 and NASDAQ return indices," forthcoming in the *Econometrics Journal*.

Rapach, D.E. and Wohar, M.E. (2006). "Structural breaks and predictive regression models of aggregate U.S. stock returns," Journal of Financial Econometrics 4: 238-274.

Ray, B.K. and Tsay, R.S. (2002). "Bayesian methods for change-point detection in long-range dependent processes," Journal of Time Series Analysis 23, 867-705.

Ristic, B., Arulampalam, S. and Gordon, N. (2004). Beyond the Kalman filter: Particle Filters for Tracking Applications. Artech House, Boston.

Rosenberg, B. (1973). "Random coefficient models: the analysis of a cross-section of time series by stochastically convergent parameter regression," Annals of Economic and Social Measurement 2: 399-428.

Rossi, B. (2006). "Are exchange rates really random walks? Some evidence robust to parameter instability," Macroeconomic Dynamics 10: 20-38.

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

Stock, J.H. and Watson, M.W. (2007). "Why has U.S. inflation become harder to forecast?," Journal of Money, Credit and Banking 39: 3-33.

Taylor, J.B. (1993). "Discretion versus policy rules in practice," Carnegie-Rochester Conference Series on Public Policy 39: 195-214.

Varneskov, R.T. and Perron P. (2012). "Combining long memory and level shifts in modeling and forecasting the volatility of asset returns," Unpublished Manuscript, Department of Economics, Boston University.

Wei, G. and Tanner, M.A. (1990). "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," Journal of the American Statistical Association 85: 699–704.

Wu, C.F.J. (1983). “On the convergence properties of the EM algorithm,” The Annals of Statistics 11: 95-103.

Xia, Y. (2001). “Learning about predictability: the effects of parameter uncertainty on dynamic asset allocation,” Journal of Finance 56: 205-246.

Xu, J. and Perron, P. (2012) “ Modeling and foreacsting stock return volatility: level shift model with time varying jump probablity and mean reversion,” Unpublished Manuscript, Department of Economics, Boston University.

Xu, L. and Jordan, M.I. (1996). “On convergence properties of the EM algorithm for Gaussian mixtures,” Neural Computation 8: 129-151.

Table 1: Simulation Results for the Basic and Full Models

Panel A:		Basic Model					
Parameter		**r0**	**r1**	**σ_e**	**σ_η**		
TRUE		**-1.96**	**4**	**0.2**	**0.2**		
Mean		-2.01	4.01	0.20	0.18		
Median		-2.01	4.01	0.20	0.18		
s.e.		(0.015)	(0.092)	(0.004)	(0.008)		
TRUE		**0**	**0**	**0.2**	**0.2**		
Mean		-0.06	-0.01	0.20	0.18		
Median		-0.06	-0.01	0.20	0.18		
s.e.		(0.007)	(0.047)	(0.003)	(0.002)		
Panel B:			**Full Model**				
Parameter	r_0	r_1	p2	θ	σ_v	σ_η	ρ
TRUE	**-1.96**	**4**	**0.5**	**0.95**	**0.2**	**0.2**	**-0.1**
Mean	-1.96	4.01	0.50	0.95	0.20	0.20	-0.10
Median	-1.96	4.01	0.50	0.95	0.20	0.20	-0.10
s.e.	(0.014)	(0.098)	(0.002)	(0.004)	(0.001)	(0.003)	(0.015)

Table 2: Robustness Check

Parameter	p		σ_u		σ_η
TRUE	**0**		**0.2**		**0.2**
Mean	0		0.20		-
Median	0		0.20		-
s.e.	(0)		(0.004)		-
TRUE	**1**		**0.2**		**0.2**
Mean	1		0.32		0.30
Median	1		0.31		0.30
s.e.	(0)		(0.007)		(0.009)
Parameter	r_0	r_1	σ_e	σ_η	ρ
TRUE	**-1.96**	**4**	**0.2**	**0.2**	**0**
Mean	-2.01	4.01	0.20	0.18	-0.02
Median	-2.00	4.01	0.20	0.18	-0.00
s.e.	(0.015)	(0.105)	(0.004)	(0.009)	(0.050)

Table 3.1: Equity Premium Forecasting Comparisons for the Period 1998-2012

				Cumulative	MSFE				
	h=1	h=3	h=6	h=12	h=18	h=24	h=30	h=36	h=40
rolling 10 years	1.40	13.40	57.08	240	551	1019	1667	2545	3300
			Relative	Cumulative	MSFE				
rolling 10 years	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
historical average	7.35	6.80	6.23	5.68	5.36	4.92	4.43	3.92	3.58
dividend_no break	**0.90**	**0.86**	**0.81**	**0.74**	**0.69**	**0.64**	**0.58**	**0.52**	**0.48**
RLS_meanrevert	**0.14**	**0.18**	**0.22**	**0.28***	**0.34***	**0.36***	**0.37***	**0.36***	**0.35***
constant+dividend_RLS_mean	**0.09***	**0.13***	**0.19***	**0.31**	**0.44**	**0.51**	**0.54**	**0.53**	**0.51**
constant_RLS_mean+dividend	**0.11**	**0.16**	**0.21**	**0.29**	**0.37**	**0.42**	**0.45**	**0.48**	**0.49**

Table 3.2: Equity Premium Forecasting Comparisons for the Period 1988-1996

				Cumulative	MSFE		
	h=1	h=5	h=10	h=15	h=20	h=25	h=30
rolling 10 years	1.74	44	179	413	758	1229	1837
			Relative	Cumulative	MSFE		
rolling 10 years	1.00	1.00	1.00	1.00	1.00	1.00	1.00
historical average	13.13	12.95	12.59	12.11	11.54	10.96	10.39
dividend_no break	4.13	4.33	4.48	4.59	4.63	4.62	4.57
RLS_meanrevert	**0.09**	**0.12**	**0.15***	**0.17***	**0.16**	**0.15**	**0.15**
constant+dividend_RLS_mean	**0.10**	**0.14**	**0.19**	**0.23**	**0.25**	**0.27**	**0.29**
constant_RLS_mean+dividend	**0.09**	**0.14**	**0.17**	**0.17**	**0.16***	**0.14***	**0.14***

	Cumulative		MSFE			
	h=35	h=40	h=45	h=50	h=55	h=60
rolling 10 years	2591	3483	4514	5677	6968	8375
		Relative	Cumulative	MSFE		
rolling 10 years	1.00	1.00	1.00	1.00	1.00	1.00
historical average	9.89	9.50	9.21	8.98	8.80	8.67
dividend_no break	4.50	4.47	4.48	4.50	4.55	4.61
RLS_meanrevert	**0.15**	**0.15**	**0.17**	**0.18**	**0.18**	**0.18**
constant+dividend_RLS_mean	**0.31**	**0.35**	**0.39**	**0.44**	**0.48**	**0.51**
constant_RLS_mean+dividend	**0.13***	**0.13***	**0.13***	**0.13***	**0.13***	**0.12***

Note: Bold numbers indicate entries with smaller MSFE than for the benchmark model. Numbers with an asterisk refer to the model with the smallest MSFE amongst all models. RLS_meanrevert: unconditional mean model with level shifts and mean reversion; Constant+dividend_RLS_mean: multivariate model with a constant term and lagged dividend yield as regressors and the coeffcient of the lagged dividend yield follows a level shift process with mean reversion; Constant_RLS_mean+dividend: multivariate regression model with a constant term and lagged dividend yield as regressors and the constant term follows a level shift process with mean reversion.

Table 4: Inflation Forecasting Comparisons

				Cumulative	MSFE					
	h=1	h=4	h=12	h=24	h=36	h=48	h=60	h=72	h=84	h=96
UC_SV	0.08	3.19	75	464	1324	2497	3979	5583	7592	10034
				Relative	MSFE					
UC_SV	1.00*	1.00*	1.00*	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR(AIC)	2.22	1.58	1.41	1.39	1.40	1.41	1.40	1.39	1.36	1.35
AO_12	7.02	3.24	1.59	1.22	1.03	**0.97**	**0.92**	**0.92**	**0.91**	**0.93***
PC_U	1.56	1.14	1.12	1.12	1.14	1.15	1.18	1.18	1.20	1.19
PC_dU	2.04	1.39	1.33	1.28	1.25	1.25	1.27	1.26	1.26	1.24
UC_SV modified	1.33	1.16	1.04	**0.96***	**0.89***	**0.83***	**0.81***	**0.84***	**0.90***	**0.99**
PC_RLS_SV	1.15	1.11	1.04	1.04	1.04	1.04	1.03	1.03	1.04	1.04

Note: Bold numbers indicate entries with smaller MSFE than for the benchmark model. Numbers with an asterisk refer to the model with the smallest MSFE amongst all models. UC_SV refers to the unobserved components stochastic volatility model; AR(AIC) is the autoregression model with lag order determined by AIC; AO_12 is the Atkeson-Ohanian model using the previous 12 months average as the forecast; PC_U refers to the backward-looking Phillip Curve model using the unemployment rate as regressor; PC_dU refers to the backward-looking Phillips Curve model using the first differences of the unemployment rate; UC_SV-modified is the the unobserved components model with a random level shit process; PC_RLS_SV is the Phillips Curve model with coefficient following a random level shift process and with random walk stochastic volatility.

Table 5.1: Exchange Rate Forecasting Comparisons; Canada

	MSFE						
	h=1	h=4	h=8	h=12	h=16	h=20	h=24
Random walk	0.0004	0.0163	0.1283	0.4238	0.9987	1.9178	3.3082
			Relative	MSFE			
	h=1	h=4	h=8	h=12	h=16	h=20	h=24
Random walk	1	1	1	1	1	1	1
UIRP	1.0059	1.0011	1.0012	1.0016	1.0022	1.0029	1.0036
Taylor	1.0168	1.0122	1.0125	1.0125	1.0119	1.0110	1.0100
Random walk+levleshift	**0.9462**	**0.9795**	**0.9932**	**0.9971**	**0.9995**	1.0008	1.0016
UIRP+level shift	**0.9417**	**0.9717**	**0.9881**	**0.9922**	**0.9949**	**0.9960**	**0.9968**
UIRP_RLS	**0.9491**	**0.9760**	**0.9965**	1.0009	1.0028	1.0034	1.0039
UIRP_Kt	**0.9657**	**0.9391**	**0.9693**	**0.9780**	**0.9859**	**0.9897**	**0.9915**
UIRP_K1t,K2t	**0.9352***	**0.9269***	**0.9666***	**0.9771***	**0.9822***	**0.9842***	**0.9853***
Taylor+level shift	**0.9531**	**0.9845**	**0.9995**	1.0031	1.0044	1.0040	1.0031
Taylor_RLS(inflation)	1.0221	1.0135	1.0129	1.0126	1.0119	1.0109	1.0099
Taylor_RLS(output gap)	**0.9734**	**0.9736**	**0.9997**	1.0045	1.0057	1.0056	1.0048
Taylor_constant+inflation_Kt	**0.9927**	**0.9647**	**0.9883**	**0.9957**	1.0002	1.0016	1.0016
Taylor_constant+output_Kt	**0.9762**	**0.9389**	**0.9781**	**0.9931**	**0.9994**	1.0012	1.0013
Taylor_constant+inflation_K1t,K2t	**0.9710**	**0.9963**	1.0085	1.0108	1.0111	1.0102	1.0088
Taylor_constant+output_K1t,K2t	1.0139	**0.9483**	**0.9676**	**0.9776**	**0.9885**	**0.9950**	**0.9973**

Table 5.2: Exchange Rate Forecasting Comparisons; Japan

	MSFE						
	h=1	h=4	h=8	h=12	h=16	h=20	h=24
Random walk	0.0008	0.0350	0.2496	0.8194	2.0560	4.1894	7.4684
			Relative	MSFE			
	h=1	h=4	h=8	h=12	h=16	h=20	h=24
Random walk	1	1	1	1	1	1	1
UIRP	**0.9919**	**0.9880**	**0.9858**	**0.9857**	**0.9872**	**0.9886**	**0.9897**
Taylor	1.0037	1.0015	**0.9999**	**0.9974**	**0.9974**	**0.9975**	**0.9973**
Random walk+levleshift	**0.9628**	**0.9788**	**0.9800**	**0.9733**	**0.9758**	**0.9781**	**0.9797**
UIRP+level shift	**0.9557**	**0.9787**	**0.9829**	**0.9794**	**0.9832**	**0.9866**	**0.9892**
UIRP_RLS	**0.9956**	**0.9671**	**0.9707**	**0.9555***	**0.9575***	**0.9600***	**0.9623***
UIRP_Kt	**0.9281***	**0.9450**	**0.9685**	**0.9689**	**0.9723**	**0.9752**	**0.9769**
UIRP_K1t,K2t	**0.9741**	**0.9441***	**0.9630***	**0.9590**	**0.9645**	**0.9688**	**0.9717**
Taylor+level shift	**0.9593**	**0.9890**	**0.9972**	**0.9969**	**0.9977**	**0.9983**	**0.9986**
Taylor_RLS(inflation)	**0.9865**	**0.9980**	**0.9997**	**0.9995**	**0.9996**	**0.9997**	**0.9998**
Taylor_RLS(output gap)	**0.9584**	**0.9874**	**0.9978**	**0.9962**	**0.9957**	**0.9953**	**0.9952**
Taylor_constant+inflation_Kt	1.2174	1.0056	**0.9831**	**0.9628**	**0.9642**	**0.9673**	**0.9700**
Taylor_constant+output_Kt	**0.9741**	**0.9816**	1.0098	1.0079	1.0087	1.0094	1.0093
Taylor_constant+inflation_K1t,K2t	1.0567	**0.9897**	**0.9874**	**0.9852**	**0.9856**	**0.9868**	**0.9879**
Taylor_constant+output_K1t,K2t	1.0625	1.0034	1.0052	**0.9827**	**0.9787**	**0.9807**	**0.9820**

Note: Bold numbers indicate entries with smaller MSFE than for the benchmark model. Numbers with an asterisk refer to the model with the smallest MSFE amongst all models.

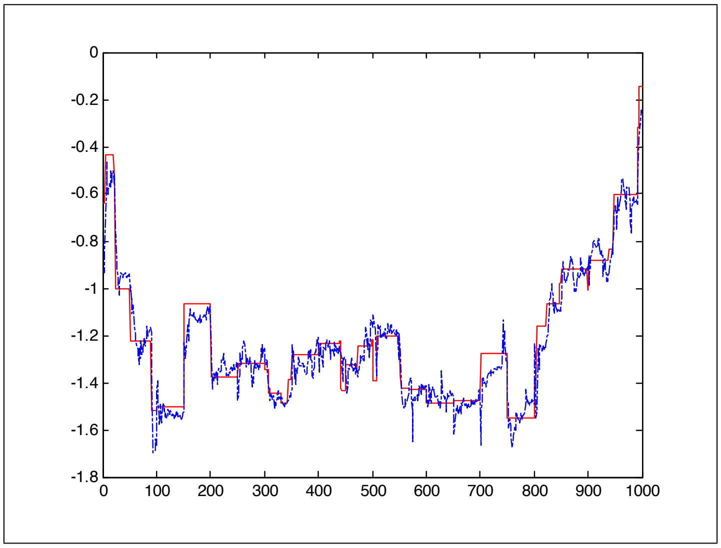
Table 6: Treasury Bill Rate Forecasting Comparisons

Full out-of-sample forecasting relative MSFE	***1968-2002***				
	h=12	h=24	h=36	h=48	h=60
Rolling 5 years	1.00	1.00	1.00	1.00	1.00
Rolling 10 years	1.00	1.00	**0.96**	**0.96**	**0.96**
Recursice OLS	**0.99**	**0.97**	**0.92**	**0.93**	**0.91**
TVP	1.46	2.17	2.47	2.28	2.19
AR_RLS_mean	**0.94***	**0.90***	**0.84***	**0.85***	**0.83***
AR_level shift	1.00	1.07	1.10	1.15	1.21
AR_RLS_SV	**0.96**	**0.91**	**0.86**	**0.87**	**0.86**
AR_level shift_SV	1.04	1.07	1.09	1.14	1.20
AR_K1t,K2t	1.34	1.18	1.02	**0.97**	**0.93**
Sub out-of-sample forecast	*relative*	*MSFE*	***1970's***		
	h=12	h=24	h=36	h=48	h=60
Rolling 5 years	1.00	1.00	1.00	1.00	1.00
Rolling 10 years	1.04	1.06	1.02	**0.99***	1.01
Recursice OLS	1.05	1.09	1.08	1.10	1.09
TVP	1.31	1.55	1.66	1.75	1.66
AR_RLS_mean	1.06	1.12	1.10	1.13	1.18
AR_level shift	**0.99***	1.07	1.11	1.15	1.16
AR_RLS_SV	1.08	1.15	1.14	1.18	1.25
AR_level shift_SV	1.06	1.10	1.11	1.15	1.16
AR_K1t,K2t	1.28	1.34	1.30	1.31	1.36
			1980's		
Rolling 5 years	1.00	1.00	1.00	1.00	1.00
Rolling 10 years	**0.92**	**0.85**	**0.87**	**0.94**	**0.95**
Recursice OLS	**0.97**	**0.88**	**0.81**	**0.84**	**0.83**
TVP	1.65	3.06	3.55	2.98	2.72
AR_RLS_mean	**0.86***	**0.63**	**0.51**	**0.50**	**0.40**
AR_level shift	1.01	1.10	1.16	1.26	1.37
AR_RLS_SV	**0.87**	**0.62***	**0.48***	**0.48***	**0.36***
AR_level shift_SV	1.03	1.07	1.13	1.24	1.37
AR_K1t,K2t	1.44	1.14	**0.72**	**0.59**	**0.45**
			1990's		
Rolling 5 years	1.00	1.00	1.00	1.00	1.00
Rolling 10 years	1.24	1.12	1.00	**0.95**	**0.87**
Recursice OLS	**0.92**	**0.79**	**0.72**	**0.64**	**0.53**
TVP	**0.99**	**0.87**	**0.88**	**0.82**	**0.82**
AR_RLS_mean	**0.92**	**0.80**	**0.74**	**0.66**	**0.57**
AR_level shift	**0.88**	**0.79**	**0.65**	**0.56**	**0.58**
AR_RLS_SV	**0.95**	**0.80**	**0.62**	**0.49***	**0.53**
AR_level shift_SV	**0.91**	**0.77**	**0.63**	**0.50**	**0.50***
AR_K1t,K2t	**0.78***	**0.65***	**0.56***	**0.54**	**0.55**

Note: Bold numbers indicate entries with smaller MSFE than for the benchmark model. Numbers with an asterisk refer to the model with the smallest MSFE amongst all models. Recursive OLS refers to the OLS method with an expanding estimation window; Rolling 5 years and 10 years refer to OLS models with window lengths set at 5 and 10 years; TVP stands for time varying parameter model; AR_RLS_mean is the AR(1) model with AR cofficient following a level shift process with mean reversion; AR_level shift is the AR(1) model allowing for level shifts in the constant term; AR_levelshift_SV and AR_RLS_SV also incorporate stochastic volatility into the error term. AR_K_{1t}, K_{2t} allows for both the constant term and the AR coefficient to follow a level shift process with two different latent variables and mean reversion.

Figure 1: Particle Filtered Estimates and Trues Values of the Parameter Process

Panel A: Less frequent breaks



Panel B: More frequent breaks

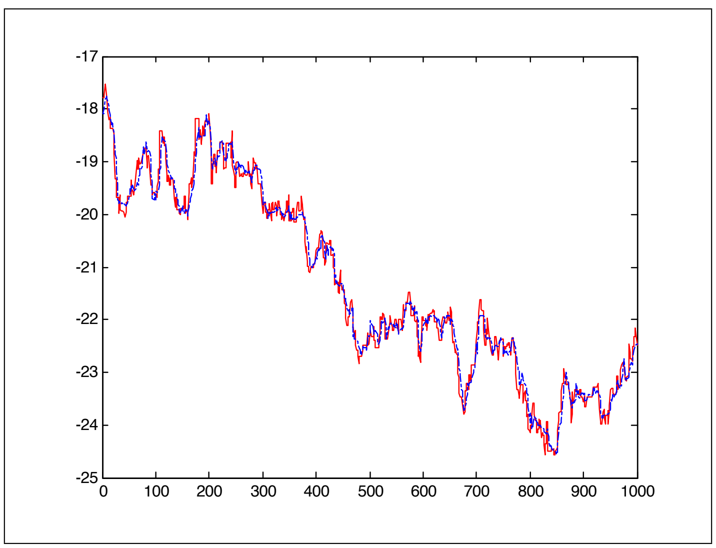
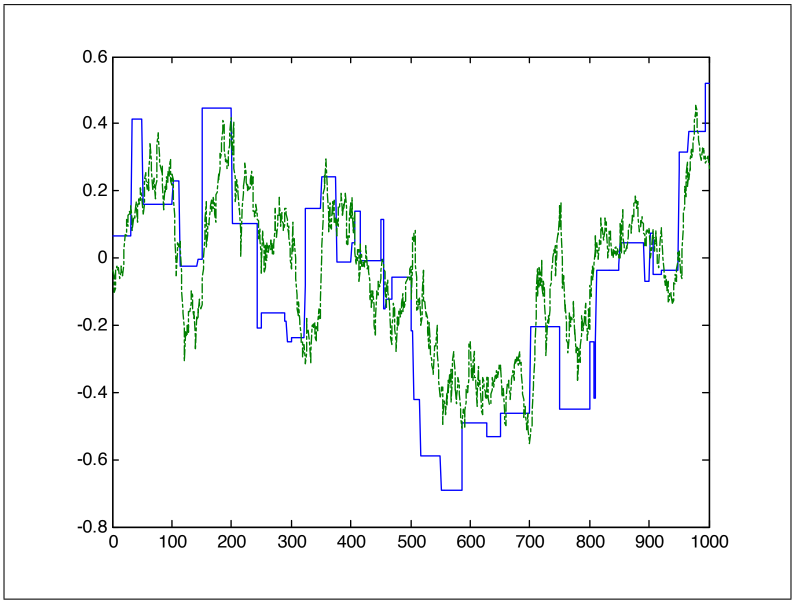


Figure 2: Particle Filtered Estimates and True Values of the Parameter and Stochastic Volatility Processes

Panel A: Particle Filtered Estimates of the Parameter Process



Panel B: Particle Filtered Estimates of the Stochastic Volatility

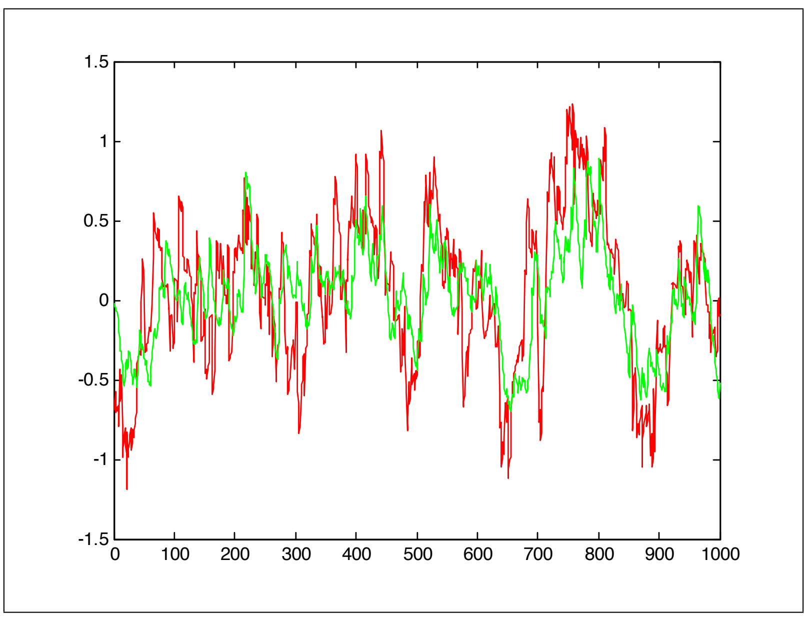


Figure 3: Out-of-Sample Forecasts of the Equity Premium

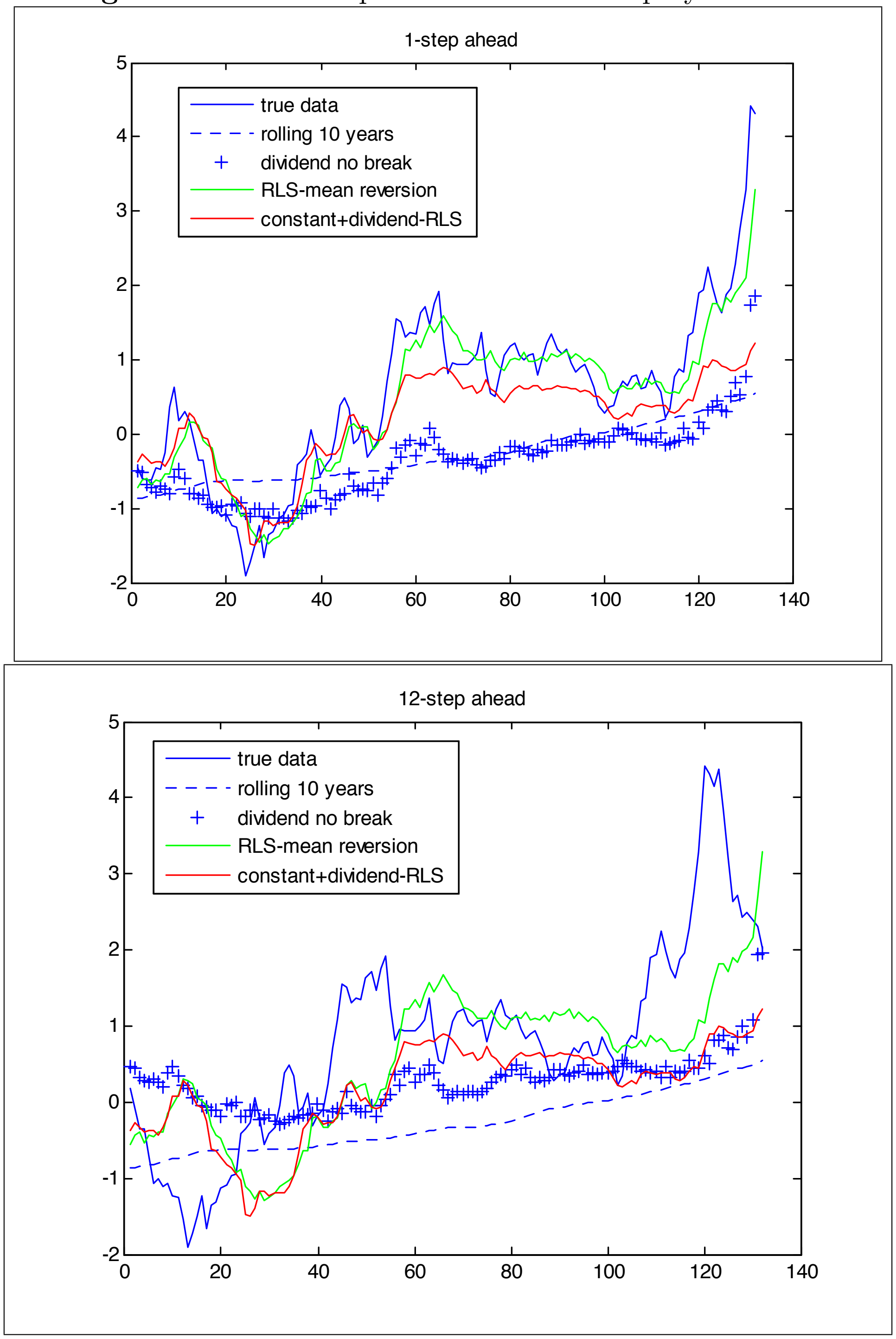
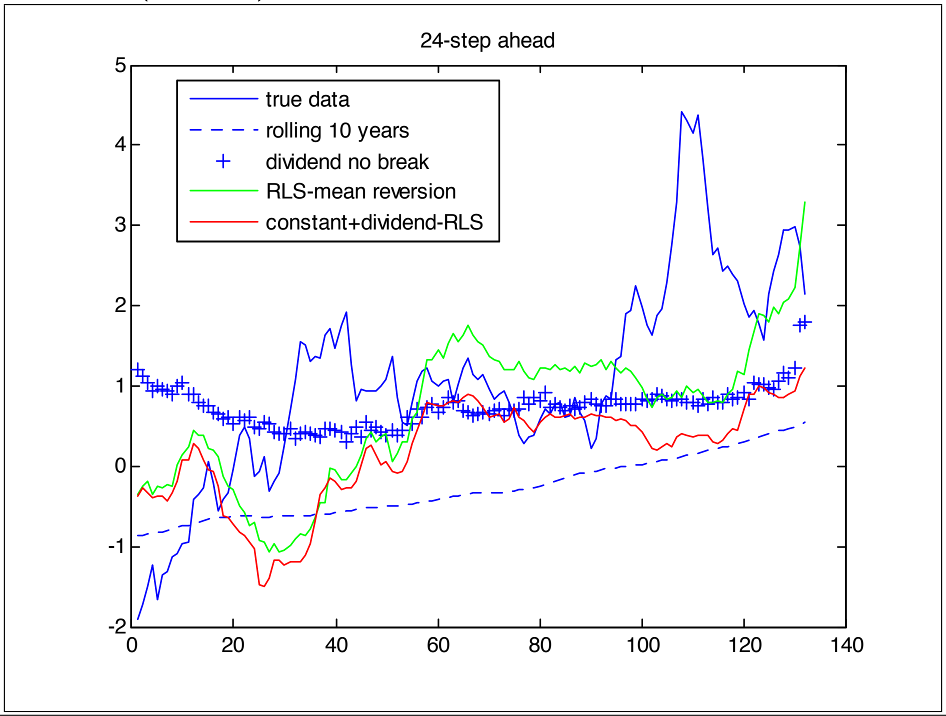


Figure 3 (Cont'd): Out-of-Sample Forecasts of the Equity Premium



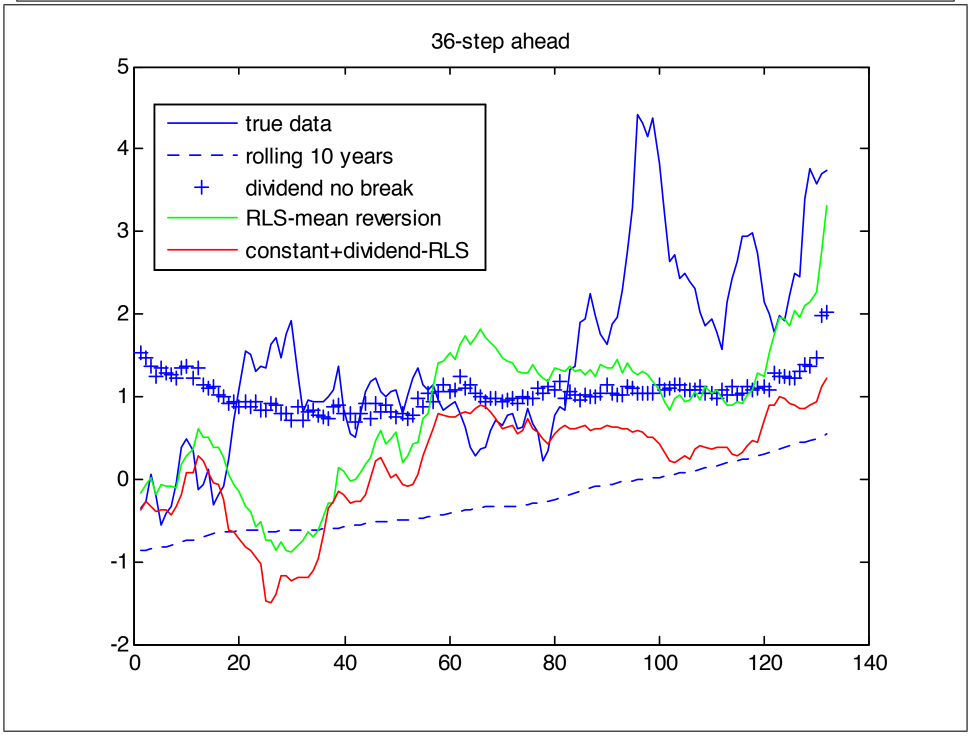


Figure 4: Out-of-Sample Forecasts of Inflation