University of Kent

School of Economics Discussion Papers

Forecasting with the Standardized Self-Perturbed Kalman Filter

Stefano Grassi, Nima Nonejad and Paolo Santucci de Magistris

February 2014

KDPE 1405

Forecasting with the Standardized Self-Perturbed Kalman Filter *

Stefano Grassi[†]
University of Kent and CREATES

Nima Nonejad [‡]
Aarhus University and CREATES

Paolo Santucci de Magistris[§]
Aarhus University and CREATES.

February 24, 2014

Abstract

A modification of the self-perturbed Kalman filter of Park and Jun (1992) is proposed for the on-line estimation of models subject to parameter instability. The perturbation term in the updating equation of the state covariance matrix is weighted by the measurement error variance, thus avoiding the calibration of a design parameter. The standardization leads to a better tracking of the dynamics of the parameters compared to other on-line methods, especially as the level of noise increases. The proposed estimation method, coupled with dynamic model averaging and selection, is adopted to forecast $S\&P500$ realized volatility series with a time-varying parameters HAR model with exogenous variables.

Keywords: TVP models, Self-Perturbed Kalman Filter, Dynamic Model Averaging, Dynamic Model Selection, Forecasting, Realized Variance.

JEL Classification: C10, C11, C22, C80

*The authors acknowledge the research support of CREATES (funded by the Danish National Research Foundation).

[†]**Corresponding Author**: School of Economics, Canterbury, Kent, CT2 7NZ, England; phone: +44 (0) 1227 824715; email address: S.Grassi@kent.ac.uk

[‡]Department of Economics and Business , Fuglesangs Alle 4; DK-8210 Aarhus V Denmark; phone: +45 8716 5325; email address: nnonejad@creates.au.dk.

[§]Department of Economics and Business, Fuglesangs Alle 4; DK-8210 Aarhus V, Denmark; phone: +45 8716 5319; email address: psantucci@creates.au.dk.

1 Introduction

Over the past two decades, time-varying parameter (TVP) models have attracted increasing interest in econometrics as tools for understanding and predicting the presence of structural breaks in macroeconomic and financial time series. In particular, TVP models are attractive since they allow for empirical insights which are not available with the traditional, constant coefficient models. Recently, TVP models are proved to be successful in macroeconomics by Primiceri (2005), Cogley and Sargent (2005) and Koop and Strachan (2009), among others. For example, Primiceri (2005) and Cogley and Sargent (2005) use time-varying VAR models to study the dynamic effects of alternative monetary policies on the real outcomes. Alternatively, Watson and Stock (2007), Cogley et al. (2010) and Grassi and Proietti (2010) focus on the US inflation series. They all find a strong evidence for a reduction of the variance in the last 25 years, a well known phenomenon called *Great Moderation*. Moreover, the coefficients on the predictors of inflation are also found to vary over time being subject to structural breaks. This phenomenon is called *time-varying* Phillips curve. In finance, the successful class of ARCH-GARCH models of Engle (1982) and stochastic volatility models could be tough as alternative ways to generate time-varying standard deviations of returns. Dangl and Halling (2012) propose a time-varying predictive regression, finding that a non-negligible portion of the out-of-sample total return variance can be predicted when the coefficients are allowed to change over time. Unfortunately, although TVP models proved to be successful in describing the changing behaviour of the US economics and of the stock returns, the estimation methods employed are extremely computationally intensive, as they generally require simulation based algorithms as MCMC.

A solution to this problem has been proposed by Raftery et al. (2010) and Koop and Korobilis (2012). Following Fagin (1964) and Jazwinsky (1970), they suggest to estimate the TVP models with a modified Kalman filter algorithm by an approximation of the updating step of the latent states covariance matrix. The updating equation of the states covariance matrix is assumed to follow a decay function that depends on an additional parameter, the *forgetting factor*. This simplifying assumption avoids to resort on computationally intensive algorithms based on simulations. The main drawback of this methodology is related to the nature of the forgetting factor, that is calibrated and assumed fixed over time. Recently Koop and Korobils

(2013) allow the forgetting factor to be time varying and dynamically selected.

The contribution of this paper is threefold. First, a new method for the on-line estimation of the TVP models is proposed. The new estimation procedure is an extension of the self-perturbing Kalman filter of Park and Jun (1992). Differently from the method based on the forgetting factor, the method of Park and Jun (1992) induces dynamics in the parameters by means of a perturbation term that is a function of the squared prediction errors. A modification of the perturbation function is introduced, as the squared prediction errors are standardized by their variance, thus avoiding the calibration of a design parameter. In other words, the new updating function dynamically calibrates the perturbation mechanism since the contribution of the squared prediction errors is weighted by the measurement error variance, which is allowed to vary according to an exponential weighted moving average.

Second, the proposed estimation procedure is compared to other methods by means of Monte Carlo simulations. The main advantage of the new methodology over other on-line methods is that it does not require to specify the forgetting factor, i.e. the decay rate of the covariance matrix of the latent states. The variation in the coefficients is instead endogenously determined by the standardized squared residuals. It emerges that the standardized self-perturbing Kalman filter has the best performance in tracking the variability of the parameters when the series at hand is noisy. This makes the new method particularly appealing in forecasting financial time series, which are typically characterized by an high noise-to-signal ratio.

Third, the TVP-HAR model with explanatory variables is proposed to forecast the realized volatility series. As noted by Koop and Korobilis (2012), the main advantage of the on-line estimation algorithms is the possibility of selecting between a large number of potential predictors at each point in time. The optimal set of predictors of realized volatility is selected at each point in time by means of the predictive likelihood as in Koop and Korobilis (2012) and Koop and Korobils (2013). The on-line nature of the self-perturbing Kalman filter allows to dynamically select between a large number of potential models. The use of predictive measures of fit, as explained in Eklund and Karlsson (2007), offers protection against in-sample over-fitting and improves the forecast performance. We find that the proposed perturbation method, combined with a dynamic model selection technique, provides superior forecasting performances, in terms of the predicted likelihood, root mean square forecast error (RMFE) and continuous ranked probability score (CRPS), see Groen et al. (2013). The good performance is mainly due to the fact that only few predictors are chosen at each point in time, as a consequence of the precision in tracking the dynamics of the parameters, thus reducing the uncertainty on the selection of the relevant variables.

The paper is organized as follows. Section 2 introduces the general model and discusses the old and new estimation methods. Section 2.4 presents a Monte Carlo exercise that shows the usefulness of the proposed estimation strategy. The technique for an optimal dynamic forecast combination is discussed in Section 3. The empirical application to the $S\&P500$ realized volatility series is presented in Section 4. Finally Section 5 draws some conclusions.

2 Online Methods for TVP models

The state-space representation of a generic TVP model is:

$$\begin{aligned} y_t &= \mathrm{Z}_t\theta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, \mathrm{H}_t), \\ \theta_t &= \theta_{t-1} + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \mathrm{Q}_t), \end{aligned} \tag{1}$$

where y_t is the observed time series, Z_t is an $1 \times m$ vector containing explanatory variables and θ_t is an $m \times 1$ vector of time varying parameters (states), which are assumed to follow random-walk dynamics. Finally the errors, ε_t and η_t are assumed to be mutually independent at all leads and lags. The model (1) is used in a number of recent paper, see Primiceri (2005), Koop and Strachan (2009), Koop and Korobilis (2012) and Koop and Korobils (2013). Traditionally, the model in equation (1) is estimated with both classical and Bayesian approaches. In the first case, the likelihood is efficiently calculated with the Kalman filter routine, see Durbin and Koopman (2001) and Harvey and Proietti (2005) for an introduction, and the time-varying parameters are automatically filtered as latent state variables, once that H_t and Q_t are estimated. The Bayesian estimation method, which requires simulation based methods, such as the MCMC, involves the specification of H_t and Q_t together with the initial condition, $\theta_{0|0}$, of the model parameters, see for an introduction Koop (2003). Although classical and Bayesian algorithms are reliable in this context, they become computational intensive as the number of parameters increases.

For this reason, Raftery et al. (2010) and Koop and Korobilis (2012) propose a fast algorithm, adopting the on-line methodology of Fagin (1964), to extract the dynamics of the parameters in a linear state-space framework. Next section presents alternative on-line procedures to capture the evolution of the parameters of model (1).

2.1 Online Estimation with the Forgetting Factor

A standard recursive estimation of model (1) proceeds as follows. Starting from a initial values $\theta_{0|0}$ and $\mathrm{P}_{0|0}$ (the updated covariance matrix of the state), the Kalman filter routine is constituted by a prediction and an updating step.

Prediction

$$\begin{aligned} \theta_{t|t-1} &= \theta_{t-1|t-1} \\ \mathrm{P}_{t|t-1} &= \mathrm{P}_{t-1|t-1} + \mathrm{Q}_t \\ \nu_t &= y_t - \mathrm{Z}_t\theta_{t|t-1} \\ \mathrm{F}_{t|t-1} &= \mathrm{Z}_t\theta_{t|t-1}\mathrm{Z}_t^{'} + \mathrm{H}_t. \end{aligned} \tag{2}$$

Updating

$$\begin{aligned} \theta_{t|t} &= \theta_{t|t-1} + \mathrm{P}_{t|t-1}\mathrm{Z}_t^{'}\mathrm{F}_{t|t-1}^{-1}\nu_t \\ \mathrm{P}_{t|t} &= \mathrm{P}_{t|t-1} - \mathrm{P}_{t|t-1}\mathrm{Z}_t^{'}\mathrm{F}_{t|t-1}^{-1}\mathrm{Z}_t\mathrm{P}_{t|t-1}. \end{aligned} \tag{3}$$

where the term $\mathrm{P}_{t|t-1}\mathrm{Z}_t^{'}\mathrm{F}_{t|t-1}^{-1}$ is the well known Kalman gain. As mentioned in the previous section, the $m \times m$ matrix Q_t requires computational intensive algorithms in order to be estimated, especially when the number of state variables is large.

Raftery et al. (2010) suggest to substitute $\mathrm{P}_{t|t}$ in the updating equation (2) with an approximation:

$$\mathrm{P}_{t|t-1} = \frac{1}{\lambda}\mathrm{P}_{t-1|t-1}, \tag{4}$$

where $\lambda \in [0; 1]$ is the *forgetting factor*. This implies that there is no longer need to estimate or simulate Q_t, while the latter is simply obtained from the following relation $\mathrm{Q}_t = (\lambda^{-1} - 1)\,\mathrm{P}_{t-1|t-1}$. This method is called Kalman filter with constant forgetting factor (KF-CFF henceforth).

The tuning parameter λ plays a crucial role in adjusting the effective memory of the algorithm, leading to a weighted estimation where data at i time points in the past has weight λ^i. For example, in the case of daily data, setting $\lambda = 0.99$ implies that observations ten days ago will receive 90% as much weight as last periods observation. Whereas for $\lambda = 0.92$, observations ten days ago will receive 43% as much weight as last periods observation. The first case, $\lambda = 0.99$, is consistent with models where changes in θ_t are gradual, the second, $\lambda = 0.92$, is consistent with models where changes in θ_t are quite rapid and abrupt.

Finally, the measurement error variance H_t needs also to be estimated, as it is well known that both macroeconomic and financial time series are characterized by heteroskedastic effects. Therefore, following Koop and Korobilis (2012) H_t is assumed to follow an exponentially weighted moving average (EWMA henceforth):

$$\mathrm{H}_t = \kappa \mathrm{H}_{t-1} + (1-\kappa)\,\nu_t^2. \tag{5}$$

The EWMA estimator requires to select a value for κ. As suggested in Koop and Korobilis (2012), the value of κ is set to 0.94 for daily data.

2.2 Kalman Filter with Time-Varying Forgetting Factor

The estimation procedures outlined above assumes the *forgetting factor* λ to be time invariant, this is a restrictive assumption that can be relaxed. For example, Koop and Korobils (2013) propose to estimate λ and κ at each point in time, by selecting their optimal values, over a grid of possible values, by means of the predictive likelihood. Alternatively, following Park et al. (1991), the parameter λ can be assumed to vary according to the following law of motion:

$$\begin{aligned} \lambda_{t|t-1} &= \lambda_{min} + (1-\lambda_{min})\,2^{\mathrm{L}_t}, \\ \mathrm{L}_t &= -\mathrm{NINT}\left[\rho\nu_{t-1}^2\right], \end{aligned} \tag{6}$$

where $\mathrm{NINT}[\cdot]$ rounds the argument to the nearest integer, ρ is a design parameter which controls the width of a unity zone. In this case the prediction (2) and the updating (3) equations do not change. The only term that evolves is the $P_{t|t}$ matrix that becomes:

$$\mathrm{P}_{t|t-1} = \frac{1}{\lambda_{t|t-1}} \mathrm{P}_{t-1|t-1}. \tag{7}$$

Equation (6) relates the magnitude of forgetting factor, at each point in time, to the squared prediction errors ν_{t-1}^2. The minimum value of the forgetting factor is obtained as ν_{t-1}^2 tends to infinity, while as ν_{t-1}^2 decreases to zero, the forgetting factor converges to unity at an exponential rate controlled by the parameter ρ. For small values of ρ, the parameter vector θ_t evolves smoothly, since λ_t remains close to 1. On the other hand, if ρ is high, such that $\lambda_t = \lambda_{min}$ then the parameters tend to be updated at a fast rate. The main disadvantage of this method is that ρ and λ_{min} are unknown quantities and they need to be calibrated. This method is named Kalman filter with time varying forgetting factor, henceforth KF-TFF.

2.3 Standardized Self-Perturbing Kalman Filter

Following Park and Jun (1992), the estimation of the TVP models can be carried out also by a modification of the updating equation of the covariance matrix $\mathrm{P}_{t|t}$. In particular, the updating equation of $\mathrm{P}_{t|t}$ in (3) is perturbed by a function of the squared prediction errors. Formally, the prediction equation (2) for $\mathrm{P}_{t|t-1}$ is replaced by

$$\mathrm{P}_{t|t-1} = \mathrm{P}_{t-1|t-1}, \tag{8}$$

while the updating step (3) becomes

$$\mathrm{P}_{t|t} = \mathrm{P}_{t|t-1} - \mathrm{P}_{t|t-1}\mathrm{Z}_t^{\prime}\mathrm{F}_{t|t-1}^{-1}\mathrm{Z}_t\mathrm{P}_{t|t-1} + \beta \cdot \mathrm{NINT}\left[\gamma\nu_t^2\right] \cdot \mathrm{I} \tag{9}$$

where β is a design constant, γ is the sensitivity gain parameter and I is the identity matrix. The term added to the updating equation of $\mathrm{P}_{t|t}$ acts as a feedback driving force and it is interpreted as a self-perturbation in the sense that it revitalizes the adaptation gain perturbing the $\mathrm{P}_{t|t}$. Indeed, the squared prediction error, ν_t^2, plays a crucial role in the algorithm. If $\gamma\nu_t^2 < 0.5$, the self-perturbing term is set to zero by the round-off operator. Hence, γ controls the maximum error bound set up for starting the self-perturbing action. If γ is low, such that

$\text{NINT}\left[\gamma\nu_t^2\right] = 0$ for $t = 1, \ldots, T$, then the parameters remain constant. Conversely, when γ is large, such that $\text{NINT}\left[\gamma\nu_t^2\right] \neq 0$ for $t = 1, \ldots, T$, then the parameters tend to change rapidly. Substituting equation (9) in equations (2)-(3), under the assumption that $\text{P}_{t|t-1} = \text{P}_{t-1|t-1}$, it follows that $\text{Q}_t = \beta\text{NINT}\left[\gamma\nu_t^2\right] \cdot \text{I}$. In other words, the matrix Q_t is assumed to be diagonal and dependent on the squared prediction errors through two design parameters, β and γ. Indeed, the setup of the self-perturbing Kalman filter requires that two parameters, γ and β, need to be calibrated, for example by means of a grid search. This can be cumbersome, especially when many models are estimated and averaged by dynamic model averaging, henceforth DMA, see Section 3.

Therefore, the equation (9) is modified as follows:

$$\text{P}_{t|t} = \text{P}_{t|t-1} - \text{P}_{t|t-1}\text{Z}_t'\text{F}_{t|t-1}^{-1}\text{Z}_t\text{P}_{t|t-1} + \beta \cdot \text{MAX}\left[0, \text{FL}\left(\frac{\nu_t^2}{\text{H}_t} - 1\right)\right] \cdot \text{I} \qquad (10)$$

where $\nu_t = y_t - z_t\theta_{t|t-1}$, $\beta \geq 0$ and $\text{FL}(\cdot)$ is the floor operator rounding to the smallest integer. This setup is called standardized self-perturbing Kalman filter, SSP-KF henceforth. The quantity $\xi_t = \dfrac{\nu_t^2}{\text{H}_t} - 1$ plays a crucial role in the proposed estimator. Indeed, the squared innovation is weighted by the innovation variance, avoiding the need to calibrate the sensitivity parameter γ. More specifically, the sensitivity parameter is replaced by the ratio $\dfrac{\nu_t^2}{\text{H}_t}$ that measures the relative impact of the squared innovation to the innovation variance. If the squared innovation is small relative to the variance, i.e. $\xi_t \leq 0$, then the self-perturbing term is null by the round off operator, and there is no updating of the parameters. Alternatively, when $\xi_t > 1$, the updating of the parameters is activated. Substituting equation (5) in the denominator of ξ_t and rearranging the terms, it follows that $\xi_t = \frac{\kappa(\nu_t^2 - \text{H}_{t-1})}{\text{H}_t}$. Hence, if $\kappa(\nu_t^2 - \text{H}_{t-1})$ is such that ξ_t is larger than 1, then the updating is activated. In other words, if the size of the shock at time t, as measured by ν_t^2, is larger than the past innovation variance H_{t-1}, then ξ_t is positive. Interestingly, the updating mechanism provides protection against outliers. Indeed, if ν_t at time t is affected by an outlier, it follows that, with high probability, $\kappa(\nu_t^2 - \text{H}_{t-1})$ will be large relative to H_t. Therefore, the perturbation mechanism is activated at time t. However, in $t+1$ and in absence of large shocks, the term $\kappa(\nu_{t+1}^2 - \text{H}_t)$ will be small or negative, such that the perturbation mechanism is not activated. On the other hand, if the the parameters are subject to a structural break at time t, then the term $\text{FL}\left(\dfrac{\nu_t^2}{\text{H}_t} - 1\right)$ is expected to be larger than zero until the effect of the structural break is offset by the evolution of the parameters. The speed of adjustment is determined by the parameter β, the larger the β, the faster is the adaptation once a structural break hits the system. The SSP-KF method requires the calibration of the parameter β. In the rest of the paper, the parameter β is chosen with a grid search procedure.

2.4 Monte Carlo Simulations

The ability of the KF-TFF and SSP-KF to correctly capture the evolution of the parameters is analysed by means of a Monte Carlo study. The data generating process is

$$y_t = X_t\theta_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}\left(0, \mathrm{H}_t\right), \tag{11}$$

where X_t is a 1×2 vector of iid standard Gaussian variates, and θ_t is the vector of time-varying parameters. Several specifications are considered for the variations in θ_t. In particular, we consider the constant specification, i.e. 0 breaks, as well as specifications with 1, 4 and 8 structural breaks. Finally, also the random walk dynamics for θ_t are considered.

Given that the main assumption of the on-line estimation methods is that the variation in the parameters is driven by the measurement error, then a crucial quantity in this context is the noise-to-signal ratio, τ, i.e. the ratio between the variance of ν_t, H_t, and the signal, $X_t\theta_t$. Therefore, the Monte Carlo simulations are conducted for moderate values of τ, i.e. 0.1 and 1, and for extreme values of τ, i.e. 5 or 10. In particular, the variance H_t is set according to the following formula $\mathrm{H}_t = \tau \cdot Var\left(X_t\theta_t\right)$, where $Var(\cdot)$ denotes the sample variance. In other words, the error variance, H_t, is assumed proportional to the variance of the signal.

Table 1 reports the average RMSFE of several on-line estimators relative to the OLS for different sample sizes, $T = 500, 1000, 2000$, based on $M = 1000$ Monte Carlo replications. As expected, the OLS estimator provides the lowest RMSFE for all values of τ, when the parameters, θ_t, are constant. Indeed, in this case, the RMSFE of the other estimator relative to that of OLS is always larger than 1. On the other hand, the on-line methods over-perform the OLS, in terms of RMSFE, when the parameters are subject to structural breaks or vary as a random walk process. More precisely, it emerges that the on-line methods with a constant forgetting factor often under-perform with respect to the OLS, especially when $\lambda = 0.75$. Indeed, when $\lambda = 0.75$, the parameters are expected to vary significantly in each period t, so that their trajectories are extremely noisy. This evidence becomes particularly clear as τ increases, since most of the variability in the error term is transferred to the parameters. On the other hand, both KF-TFF and SSP-KF provide excellent performances for moderate values of τ, while for large values of τ only SSP-KF performs slightly better than the OLS. When τ is 10, the data do not provide sufficient information to extract the dynamics in the parameters, so that the on-line methods provide similar results to those obtained with the OLS.

Figure 1 plots the true parameters, θ_1 and θ_2, when they are subject to 1 structural break with τ is equal to 1. The figures report the on-line estimates obtained respectively with recursive OLS, Panel a), KF-TFF, Panel b), and with SSP-KF, Panel c).[1] It clearly emerges that the recursive OLS are poorly designed for this kind of problems, as they adapt too smoothly after the structural break. On the other hand, both KF-TFF and SSP-KF provide a good tracking of the variation in the parameters. The point estimates obtained with the SSP-KF are smoother than those obtained with the KF-TFF and, the confidence interval of the SSP-KF is slightly larger than that obtained with the KF-TFF. However, in both cases, the variation in the parameters is well captured by the on-line methods.

Finally, in order to assess the robustness of the on-line methods to deviations from the Gaussian state space in (1), we add an outliers process to model (11). Hence y_t is generated as

$$y_t = X_t\theta_t + \Psi_t + \varepsilon_t \quad \varepsilon_t \sim \mathrm{N}\left(0, \mathrm{H}_t\right), \tag{12}$$

where $\Psi_t = sign(x) \cdot Ber(p) \cdot \psi$. The operator $sign(\cdot)$ returns the sign of the argument. In this case, x is extracted from a standard Gaussian distribution such that there is the same probability of obtaining negative or positive signs. $Ber(p)$ is a Bernoulli random variable, where $p = 0.015$ in the simulations. Finally, ψ determines the size of the outlier and it is set equal to 3. Table A.1 in the Appendix reports the results of the Monte Carlo simulations when y_t is contaminated by additive outliers. The evidence of Table 1 is confirmed also when outliers are present. The RMSFE of the SSP-KF is generally the lowest compared to the other methods. Figure 2 reports the estimates of the parameter θ_1 when the latter is constant and the series is contaminated by outliers. It emerges that the parameter estimate obtained with the KF-TFF method over-react after the occurrence of an outlier, while the SSP-KF seems more robust to

[1]The hyper-parameters ρ and β are calibrated according to a grid search at each point in time.

the presence of outliers.

3 Model Averaging and Model Selection

One of the advantages of the on-line Kalman filter is the possibility to carry out the DMA and the dynamic model selection, henceforth DMS, in a computationally feasible way. Define $\mathrm{L}_t \in \{1, 2, \ldots, K\}$ the set of possible models at each point in time t, where $K = 2^m$, m is the number of variables included in the model. Since the model can change over time, then $\mathrm{L} = 1, 2, \ldots, G$ is the set of possible models over time where $G = 2^{mT}$ and T is the number of observations. Define $\mathrm{Y}_T = \{y_1, \ldots, y_t\}$ the information set, then the state space form can be written as follows:

$$\begin{aligned} y_t &= \mathrm{Z}_t^{(k)} \theta_t^{(k)} + \varepsilon_t^{(k)}, \qquad \varepsilon_t^{(k)} \sim \mathrm{N}\left(0, \mathrm{H}_t^{(k)}\right), \\ \theta_{t+1}^{(k)} &= \theta_t^{(k)} + \eta_t^{(k)}, \qquad \eta_t^{(k)} \sim \mathrm{N}\left(0, \mathrm{Q}_t^{(k)}\right), \end{aligned} \tag{13}$$

where $k = 1, \ldots, K$ indicates what is the selected model at time t. At each different k corresponds a different set of predictors and parameters. For example, the SSP-KF for the k-th model becomes:

$$\theta_{t|t}^{(k)} = \theta_{t|t-1}^{(k)} + \mathrm{P}_{t|t-1}^{(k)} \mathrm{Z}_t^{(k)\prime} \left(\mathrm{H}_t^{(k)} + \mathrm{Z}_t^{(k)} \mathrm{P}_{t|t-1}^{(k)} \mathrm{Z}_t^{(k)\prime}\right)^{-1} \nu_t^{(k)} \tag{14}$$

$$\mathrm{P}_{t|t}^{(k)} = \mathrm{P}_{t|t-1}^{(k)} - \mathrm{P}_{t|t-1}^{(k)} \mathrm{Z}_t^{(k)\prime} \left(\mathrm{H}_t^{(k)} + \mathrm{Z}_t^{(k)} \mathrm{P}_{t|t-1}^{(k)} \mathrm{Z}_t^{(k)\prime}\right)^{-1} \mathrm{Z}_t^{(k)} \mathrm{P}_{t|t-1}^{(k)} + \beta \cdot \mathrm{MAX}\left[0, \mathrm{FL}\left(\frac{\nu_t^{2,(k)}}{\mathrm{H}_t^{(k)}} - 1\right)\right] \cdot \mathrm{I}. \tag{15}$$

Following Koop and Korobilis (2012) the DMA and DMS proceed as follows. Define $\Theta_t = \{\theta_1^{(1)}, \ldots, \theta_t^{(k)}\}$ the set of parameters at time t then it holds that

$$p\left(\Theta_{t-1|t-1} \mid \mathrm{Y}_{t-1}\right) = \sum_{k=1}^{K} p\left(\theta_{t-1|t-1}^{(k)} \mid \mathrm{L}_{t-1} = k, \mathrm{Y}_{t-1}\right) p\left(\mathrm{L}_{t-1} = k \mid \mathrm{Y}_{t-1}\right) \tag{16}$$

where $p\left(\theta^{(k)}_{t-1|t-1} \mid \mathrm{L}_{t-1} = k, \mathrm{Y}_{t-1}\right)$ is given by:

$$\Theta_{t-1|t-1} \mid \mathrm{L}_{t-1} = k, \mathrm{Y}_{t-1} \sim \mathrm{N}(\hat{\theta}^{(k)}_{t-1|t-1}, \mathrm{P}^{(k)}_{t-1|t-1}) \tag{17}$$

and $p(\mathrm{L}_{t-1} = k \mid Y_{t-1})$ is the probability to be at model k at time $t-1$. Define $\pi_{t|s,k} = p\left(\mathrm{L}_t = k \mid \mathrm{Y}_s\right)$ such that right-hand-side of equation (16) is $\pi_{t-1|t-1,k}$ then we get

$$\pi_{t|t-1,k} = \sum_{l=1}^{K} \pi_{t-1|t-1,l} p_{kl}, \tag{18}$$

where p_{kl} is the element of the transition probability in P with elements p_{kl}.

Using the same approximation as in Raftery et al. (2010) and Koop and Korobilis (2012), it follows that

$$\pi_{t|t-1,k} \quad = \quad \frac{\pi^{\alpha}_{t-1|t-1,k}}{\sum_{l=1}^{K} \pi^{\alpha}_{t-1|t-1,l}} \tag{19}$$

where $0 < \alpha \leq 1$ is set to a fixed value slightly less than one and is interpreted in a similarly to λ in Section 2.1. The updating equation of (19) is then given by:

$$\pi_{t|t,k} \quad = \quad \frac{\pi_{t|t-1,k} p^{(k)}\left(y_t \mid \mathrm{Y}_{t-1}\right)}{\sum_{l=1}^{K} \pi_{t|t-1,l} p^{(l)}\left(y_t \mid \mathrm{Y}_{t-1}\right)} \tag{20}$$

where $p^{(l)}\left(y_t \mid \mathrm{Y}_{t-1}\right)$ is the predictive likelihood for model l, given by

$$p(y_t \mid \mathrm{Y}_{t-1}) \sim \mathrm{N}(\mathrm{Z}^{(l)}_t \hat{\theta}^{(l)}_{t-1|t-1}, \mathrm{H}^{l}_t + \mathrm{Z}^{l}_t \mathrm{P}^{(l)}_{t|t-1} \mathrm{Z}^{(l),'}_t). \tag{21}$$

The predictive likelihood of DMA is a weighted average of each of the individual model predictive likelihoods

$$p\left(y_t \mid \mathrm{Y}_{t-1}\right) \quad = \quad \sum_{l=1}^{K} p^{(l)}\left(y_t \mid \mathrm{Y}_{t-1}\right) \pi_{t|t-1,l}, \tag{22}$$

similarly, the predictive mean of y_t is a weighted average of model specific predictions, where the weights are equal to the posterior model probabilities

$$\mathrm{E}\left[y_t \mid \mathrm{Y}_{t-1}\right] \;=\; \sum_{l=1}^{K} \mathrm{Z}_t^{(l)} \theta_{t|t-1}^{(l)} \pi_{t|t-1,l}. \tag{23}$$

On the other hand DMS involves selecting at each point in time the single model with the highest value for $p(\mathrm{L}_t = k \mid \mathrm{Y}_s)$ and using this to forecast. Koop and Korobilis (2012) find that both DMA and DMS forecast inflation very well.

The KF-TFF and SSP-KF methods require the parameters ρ and β to be calibrated.[2] One possibility is to select a grid of values where all possible model combinations are estimated. Then the model with the smallest prediction error, $\nu_t^{(k)} = y_t - \mathrm{Z}_t^{(k)} \theta_{t|t-1}^{(k)}$ is selected. Although the grid search for ρ and β is one dimensional, the calibration becomes computationally cumbersome as the number of models grows, because the grid search must be carried out at each time point t for each model k.

The following strategy in therefore used in the forecasting exercise presented in Section 4:

1. In $t = 1$, initialize the inclusion probabilities to $\pi_{1|1,k} = 1/2^m \ \forall k$ and the design parameters $\beta = 0.001$ and $\rho = 1$.

2. At time $t > 1$, equation (20) is used to compute the updated inclusion probabilities and to find the best performing model.

3. Conditional on Y_t and the exogenous variable corresponding to the best selected model, $\left\{\mathrm{Z}_1^{(*)}, \ldots, \mathrm{Z}_t^{(*)}\right\}$, find by means of a grid search an estimate of β and ρ. Denote them by $\hat{\beta}$ and $\hat{\rho}$.

4. Use the values $\hat{\beta}$ and $\hat{\rho}$ for the other models.

5. Iterate points 2-4 for $t = 1, ..., T$.

This allows the parameters $\hat{\rho}$ and $\hat{\beta}$ to adjust with time as the best model changes.

4 Online Forecast of Realized Volatility

Predicting the extent of the fluctuations of the stock prices is a primary issue in finance. Strong empirical evidence, dating back to the seminal papers of Engle (1982) and Bollerslev

[2]The parameter λ_{min} for the KF-TFF is set equal to 0.94 for all $t = 1, ..., T$.

(1986), supports the idea that the volatility of financial returns is time varying, stationary and long-range dependent. This evidence is confirmed by the statistical analysis of the ex-post volatility measures, such as realized volatility, RV henceforth, which are precise estimates of latent integrated variance and are obtained from intradaily returns, see Andersen and Bollerslev (1998), Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) among many others. For instance, Andersen et al. (2003), Giot and Laurent (2004), Lieberman and Phillips (2008) and Martens et al. (2009) report evidence of long memory and model RV as a fractionally integrated process. As noted by Ghysels et al. (2006) and Forsberg and Ghysels (2007) mixed data sampling approaches are also empirically successful in accounting for the observed strong serial dependence. In particular, Corsi (2009) approximates long range dependence by means of a long lagged autoregressive process, called heterogeneous-autoregressive model (HAR). The main feature of the HAR model is its interpretation as a volatility cascade, where each volatility component is generated by the actions of different types of market participants with different investment horizons. HAR type parameterizations are also suggested by Corsi et al. (2008), Andersen et al. (2007) and Andersen et al. (2011). In its simplest version, the HAR model of Corsi (2009) is defined as

$$y_t = \alpha + \phi^d y_{t-1} + \phi^w y_{t-1}^w + \phi^m y_{t-1}^m + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2), \tag{24}$$

where $y_t = \log(RV_t)$, $y_t^w = \frac{1}{5}\sum_{j=0}^{4} y_{t-j}$, $y_t^m = \frac{1}{22}\sum_{j=0}^{21} y_{t-j}$, and $\theta = \left[\phi^d, \phi^w, \phi^m\right]$.

In light of the recent global financial crisis, and the different behaviour of RV series during periods of high and low trading activity, a time-varying coefficients model may lead to a better understanding of the volatility dynamics. The HAR parameters ϕ^d, ϕ^w and ϕ^m in equation (24) are assumed to follow random walk dynamics, so that they measure the proportion of the total variance that is captured by each volatility component at time t. Hence, the TV-HAR parameters are interpreted as time varying weights for each volatility component and the model is given by

$$\begin{aligned} y_t &= c_t + \phi_t^d y_{t-1} + \phi_t^w y_{t-1}^w + \phi_t^m y_{t-1}^m + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, \mathrm{H}_t) \\ c_t &= c_{t-1} + \eta_t^{\alpha}, \qquad \phi_t^d = \phi_{t-1}^d + \eta_t^{\phi^d}, \\ \phi_t^w &= \phi_{t-1}^w + \eta_t^{\phi^w}, \qquad \phi_t^m = \phi_{t-1}^m + \eta_t^{\phi^m}, \end{aligned} \tag{25}$$

where $\eta_t \equiv [\eta_t^{\alpha}, \eta_t^{\phi^d}, \eta_t^{\phi^2w}, \eta_t^{\phi^m}] \sim \mathrm{N}(0, \mathrm{Q}_t)$, Q_t is the 4×4 covariance matrix of the state innovations while H_t is the time-varying variance of ε_t, see equation (5), which reflects the dynamics in the volatility of volatility. In contrast to Liu and Maheu (2008) and McAleer and Medeiros (2008), model (25) allows for a potentially large number of changing points of the HAR parameters.

Apart from the pure autoregressive structure of the HAR, we are interested in evaluating if some key financial and macroeconomic variables have additional explanatory power for $S\&P$ 500 realized volatility. Therefore, a number of predictors is added as explanatory variables to equation (25). Following Fernandes et al. (2007), the lags of the following explanatory variables are supposed to carry incremental information for the future values of RV: the foreign value of the US dollar, S_t , the term spread, TS_t, the difference between the effective and target Federal Fund rates, FF_t, the credit default swap of the US bank sector, CDS_t, the CBOE VIX index, VIX_t, the trading volume on the $S\&P$ 500 index, V_t, the negative returns of the $S\&P$ 500 index, r_t^-.[3]

With three autoregressive terms and seven explanatory variables, the number of potential models is $K = 2^{10} = 1024$ at each point in time. This calls for some simplifying assumptions. In particular, we assume that the three autoregressive terms are always included, thus reducing the number of potential models to $K = 2^7 = 128$. Hence, the predictions for each model are combined with DMA or DMS as described in Section 3, with different values of the parameter α. Table 2 reports a description of all the 45 models/methods that have been considered for forecasting realized volatility. The performances are compared at different forecasting horizons, short $h = 1$, medium $h = 5, 10$ and long $h = 22$. When forecasting $h > 1$ periods ahead, the direct forecasting method is used. Therefore, the TV-HAR-X model for a given h and for a given set of predictors, X_t, has the following form

$$y_{t+h} = c_t + \phi_t^d y_t + \phi_t^w y_t^w + \phi_t^m y_t^m + \zeta_t X_t + \varepsilon_{t+h}, \quad t = 1, \ldots, T-h \tag{26}$$

where ζ_t is an $1 \times N$ vector where N is the number of columns of X_t. The following criteria are adopted to evaluate the quality of the out-of-sample forecasts:

1. The root-mean squared error, RMSE.

[3] All the explanatory variables are included with logarithmic transformation.

2. Log of predictive likelihood log(PL), see equation (22);

3. Continuous ranked probability score (CRPS), see the recent article of Groen et al. (2013) for a discussion.[4]

The empirical analysis is carried out on the RV series of the $S\&P$ 500 index computed from returns sampled at 5-minutes intervals. The sample period starts on January 2, 2004 and it ends on December 31, 2012 for a total of 2256 daily observations. Tables 3 and 4 report the values of the performance criteria for all models at all forecasting horizons for the out-of-sample period that starts on January 03, 2007. This out-of-sample period includes the subprime financial crisis, with an high probability of occurrence of one or more structural breaks in the HAR model. From Table 3 it emerges that, at short forecasting horizons, the exogenous variables have small predictive power, especially when they are not optimally combined, by DMA. For example, for $h = 1$, the value of $\log(PL)$ of the TVP-HAR model without explanatory variables (first row) is close to that of the TVP-HAR with the explanatory variables (rows 2-8). This suggests that the relative contribution of the explanatory variables to the forecast is negligible. The same conclusion holds when looking at other criteria and at longer forecasting horizons. When focusing on the performance of the forecast combination obtained with DMA for different values of α, it emerges a slightly better forecasting accuracy, especially at longer horizons, than that obtained without averaging. It should be noted that, the best performance in terms of forecasting accuracy for DMA is achieved when the SSP-KF method is employed for $h > 1$. This result is independent of the choice of α. It is also interesting to note that the DMA at short horizons does not produce better predictions than those obtained without forecast combination. This is probably due to the fact that many models, with a small predictive power, are averaged at each point in time thus increasing the prediction uncertainty.

The forecasting performances improve substantially, both at short and long horizons, when the predictions of the best model, out of the 128 considered at each point in time, are obtained by DMS with different values of α, see Table 4. Indeed, the DMS with $\alpha = 0.95$ is always contained in the model confidence set, see Hansen et al. (2011). For short forecasting horizons, the best performance is obtained with DMS and KF-CFF when λ is set at high values, i.e. $\lambda = 0.99$. This means that the trajectories of the parameters need to be smooth in order to

[4]The prediction with the smallest CRPS is preferred.

compensate the noise emerging from the innovations in the measurement equations.

At longer forecasting horizons there is a strong evidence that the best performance is obtained when the SSP-KF is employed to capture the evolution of the parameters, and the best model is chosen by DMS with $\alpha = 0.95$. Relatively to the case without explanatory variables and forecast combination, i.e. case 1, the point values of the forecast evaluation criteria are always below by a factor larger than 20%. This gives an idea of the additional power of the financial covariates when they are correctly included in the model. It should also be noted that the DMS provides good forecasts also for other values of α when the SSP-KF method is adopted. This evidence is robust to the choice of the forecasting period. Indeed, the results do not change when the out-of-sample window starts on January 2, 2009, thus excluding most part of the sub-prime crisis, see Tables A.2 and A.3 in the Appendix. The best forecasting performance is achieved when DMS with $\alpha = 0.95$ is adopted together with the SSP-KF method.

4.1 Which Variables Are Good Predictors of Realized Volatility?

A great benefit of DMA and DMS over standard forecasting methods is that they allow the forecasting model to change over time. Moreover, the inclusion probabilities give important insights on which are the relevant predictors of the dependent variable. Although the TV-HAR-X model presented in Section 4 has 7 potential predictors, DMA and DMS generally select more parsimonious models with only a subset of all the included predictors. Following Koop and Korobilis (2012), a measure of the dimension of the selected model is given by the expected size, $\mathrm{E}(S_t)$. The expected size, i.e. the expected optimal number of predictors,[5] is obtained at each point in time by the following formula

$$\mathrm{E}(S_t) = \sum_{k=1}^{K} \pi_{t|t-1,k} S_{t,k} \tag{27}$$

where $S_{t,k}$ is the number of predictors in the k-th model and $\pi_{t|t-1,k}$ is the inclusion probability defined in equation (19). In other words, the expected size is the weighted average of the size of each individual model, and it carries important information for the degree of shrinkage achieved by DMA. On the other hand, DMS only selects the model with the highest inclusion

[5]Excluding the intercept and the HAR parameters.

probability. Hence, the measure of the size that is relevant for DMS is

$$S_t^B = S_{t,k^*} \quad \text{with} \quad k^* \quad \text{such that} \quad \pi_{t|t-1,k^*} = max(\pi_{t|t-1,k}) \quad \forall k = 1, \ldots, K \tag{28}$$

where S_t^B is the size of the *best* model. Figure 3 plots the evolution over time of $\mathrm{E}(S_t)$ and S_t^B for different forecasting horizons, $h = 1, 5, 10, 22$. [6] When $h = 1$ there is a strong evidence that 3 predictors are selected by both DMA and DMS. As the forecasting horizon increases, more predictors are selected. $\mathrm{E}(S_t)$ generally ranges between 3 and 5, and more predictors are selected during the period of the sub-prime crisis until the beginning of 2009. A similar evidence emerges also when looking at the dimension of the best model. It is interesting to note that S_t^B, although more noisy than $\mathrm{E}(S_t)$ by construction, remains constant for long periods. This evidence may suggest that an high probability is generally associated to the best model, such that the same number of predictors are selected for long periods. This is confirmed by a visual inspection of Figure 4. Indeed, the inclusion probability of the best model when $h = 1$ is very high and it is higher than 0.85 after 2009. When the forecasting horizon increases, the choice of the best model becomes less sharp, $\pi_{t|t-1,k*} \approx 0.15$. However, during the sub-prime financial crisis, the probability of the best model increases up to 0.45 for $h = 5$. This is a very high value obtained for a single model, considering that 128 competing models are evaluated and contrasted. This may explain the good forecasting performance of the DMS compared to DMA, as reported in Tables 3-4, since the uncertainty on the selection of the optimal model is reduced when there is an high probability associated to a single model.

Figure 5 reports the inclusion probabilities for the seven explanatory variables obtained with DMS for the four forecasting horizons considered. The plot for $h = 1$ confirms the evidence arising from Figures 3 and 4 depicting a situation where, after the 2007-2008 financial crisis, only 3 variables have predictive power for $\log(RV)$. These variables are the dollar index, the VIX and the past negative returns, thus confirming the importance of the market expectations and the leverage effect in forecasting volatility. On the other hand, variables like FF and CDS decrease their predicting power throughout the sample and they are almost excluded from the model at the end of the period. At longer horizons is difficult to find clear dynamic patterns

[6] The reported values of $\mathrm{E}(S_t)$ and S_t^B are relative to the TV-HAR-X model (26) estimated by SSP-KF with $\alpha = 1$. The large value of α guarantees an high degree of smoothness in the inclusion probabilities and hence in the model size.

for the inclusion probabilities of the explanatory variables. It appears that the dollar index, the trading volume and the CDS have high inclusion probability when $h > 1$, especially during the period 2008-2009.

Finally, Figure 6 reports the dynamic evolution of the HAR parameters obtained with the SSP-KF. There is a marked difference between the estimates obtained with the baseline TV-HAR model without explanatory variables, red line, and those obtained with DMS, blue line. Interestingly, the estimates of the HAR parameters obtained with DMS are more stable than those obtained with a pure autoregressive specification. For example, the parameter loading the weekly volatility factor, ϕ_t^w, is rather constant under DMS, while it has noisy trajectories and fluctuates significantly when the explanatory variables are not included in the model. The variation in the autoregressive parameters, when the relevant explanatory variables are not included, compensates the predictive power of the missing variables and the gap between the two line is relevant. This confirms once more the role of the macro-financial variables in predicting realized volatility.

5 Conclusion

This paper introduces a novel method to estimate TVP models in economics and finance. In particular, a new estimation procedure is proposed, the standardized self-perturbed Kalman filter, extending the on-line method proposed by Park and Jun (1992). This method has the advantage, over the traditional Kalman filter routine that it is computationally very fast, as the innovation in the parameters is driven by a non-linear function of the innovations in the measurement equation. With respect to other on-line methods, such as the Kalman filter with time-varying forgetting factor, there is no need to specify a low of motion for the forgetting factor, λ_t. In the self-perturbed Kalman filter, the perturbation enters directly in the updating step, so that the updating process of the parameters is endogenous. The updating mechanism induces the parameters to change gradually, as it is also assessed via Monte Carlo simulations.

The proposed estimator is used to forecast the realized variance series of the $S\&P$ 500 index. The self-perturbed Kalman filter allows to precisely extract the variation in the parameters and, hence, to provide better signals for the optimal selection of the relevant explanatory variables. It is found that the dollar index, the negative returns and VIX have predictive power for realized volatility at short horizons.

References

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89:701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71:579–625.

Andersen, T. G., Bollerslev, T., and Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics*, 160:176–189.

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17:457–477.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Cogley, T., Primiceri, G. E., and Sargent, T. J. (2010). Inflation-gap persistence in the us. *American Economic Journal: Macroeconomics*, 2:43–69.

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.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7:174–196.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27:46–78.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106:157–181.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, UK.

Eklund, J. and Karlsson, S. (2007). Forecast combination and model averaging using predictive measures. *Econometric Reviews*, 26:329–363.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50:987–1008.

Fagin, S. (1964). Recursive linear regression theory, optimalter theory, and error analyses of optimal systems. *IEEE International Convention Record Part*, pages 216 – 240.

Fernandes, M., Medeiros, M. C., and Scharth, M. (2007). Modeling and predicting the cboe market volatility index. Textos para discussão, Department of Economics PUC-Rio (Brazil).

Forsberg, L. and Ghysels, E. (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*, 5:31–67.

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

Giot, P. and Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and arch type models. *Journal of Empirical Finance*, 11:379–398.

Grassi, S. and Proietti, T. (2010). Has the volatility of u.s. inflation changed and how? *Journal of Time Series Econometrics*, 2:1–26.

Groen, J. J. J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, 31:29–44.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79:453–497.

Harvey, A. C. and Proietti, T. (2005). *Readings in Unobserved Components Models*. Advanced Texts in Econometrics. Oxford University Press, Oxford, UK.

Jazwinsky, A. (1970). *Stochastic Processes and Filtering Theory*. New York: Academic Press.

Koop, G. (2003). *Bayesian Econometrics*. John Wiley and Sons Ltd, England.

Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, 53:867–886.

Koop, G. and Korobils, D. (2013). Large time-varying parameter vars. Forthcoming journal of econometrics.

Koop, G., L.-G. R. and Strachan, R. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, 33:997 – 1017.

Lieberman, O. and Phillips, P. C. B. (2008). Refined inference on long-memory in realized volatility. *Econometric Reviews*, 27:254–267.

Liu, C. and Maheu, J. M. (2008). Are there structural breaks in realized volatility? *Journal of Financial Econometrics*, 1:1–35.

Martens, M., Van Dijk, D., and de Pooter, M. (2009). Forecasting s&p 500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, 25:282–303.

McAleer, M. and Medeiros, M. C. (2008). A multiple regime somooth transition heterogeneous autoregressive model for long memory and asymmetries. *Journal of Econometrics*, 147:104–119.

Park, D. J. and Jun, B. E. (1992). Selfperturbing recursive least squares algorithm with fast tracking capability. *Electronics Letters*, 28:558–559.

Park, D. J., Jun, B. E., and H., K. J. (1991). Fast tracking rls algorithm using novel variable forgetting factor with unity zone. *Electronics Letters*, 27:2150–2151.

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

Raftery, A., Karny, M., and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52:52–66.

Watson, W. M. and Stock, J. H. (2007). Why has u.s. inflation become harder to forecast? *Journal of Money, Banking and Credit*, 39:3–33.

Table 1: Monte Carlo results for $T = 500, 1000, 2000$. The table reports the average RMSFE, relative to OLS, for several online estimation methods. For the SSP-KF and KF-TFF, the hyper-parameters γ and ρ are selected according to a grid search at each point in time. The KF-FF is calculated for two values of λ. Criteria: RMSFE (ratio to OLS).

	$T = 500$					$T = 1000$					$T = 2000$				
	No break	1 break	4 breaks	8 breaks	RW	No break	1 break	4 breaks	8 breaks	RW	No break	1 break	4 breaks	8 breaks	RW
$\tau = 0.1$															
KF-TFF	**1.0140**	0.5374	0.4306	**0.4306**	**0.7607**	**1.0052**	0.4068	0.4039	0.4035	**0.7188**	**1.0024**	0.3334	0.3716	0.3745	**0.3477**
SSP-KF	1.0886	**0.4611**	**0.4285**	0.4412	0.7647	1.0968	**0.3991**	**0.3940**	**0.3968**	0.7322	1.0263	**0.3319**	**0.3625**	**0.3661**	0.3490
KF-CFF $\lambda = 0.75$	1.4627	0.4865	0.4767	0.4858	0.9537	1.4882	0.4549	0.4669	0.4611	0.9331	1.4376	0.4121	0.4467	0.4399	0.4551
KF-CFF $\lambda = 0.99$	1.0933	1.2005	0.8218	0.9158	0.8857	1.1012	0.8951	0.7176	0.7688	0.8791	1.0397	0.6339	0.5849	0.6417	0.4099
$\tau = 1$															
KF-TFF	**1.0114**	0.8033	**0.8076**	**0.8165**	**0.9838**	**1.0064**	**0.7518**	0.7841	0.7850	**0.9734**	**1.0027**	**0.7145**	**0.8076**	0.7580	0.7510
SSP-KF	1.0268	**0.8029**	0.8097	0.8227	0.9901	1.0246	0.7607	**0.7827**	**0.7837**	0.9775	1.0134	0.7217	0.8097	**0.7560**	**0.7500**
KF-CFF $\lambda = 0.75$	1.4162	1.0352	1.0397	1.0455	1.3377	1.4308	1.0122	1.0394	1.0315	1.3321	1.4294	0.9807	1.0397	1.0130	1.0380
KF-CFF $\lambda = 0.99$	1.0338	0.9215	0.9115	0.9501	0.9975	1.0335	0.8317	0.8408	0.8693	0.9925	1.0193	0.7480	0.9115	0.8040	0.7570
$\tau = 5$															
KF-TFF	**1.0093**	**0.9609**	0.9840	0.9993	1.0124	**1.0061**	**0.9406**	0.9773	0.9902	1.0154	**1.0032**	**0.9216**	0.9840	0.9755	1.0035
SSP-KF	1.0174	0.9632	**0.9666**	**0.9762**	**1.0116**	1.0138	0.9433	**0.9533**	**0.9554**	**1.0067**	1.0083	0.9252	**0.9666**	**0.9413**	**0.9359**
KF-FF $\lambda = 0.75$	1.4117	1.3017	1.3035	1.3066	1.3971	1.4252	1.3002	1.3125	1.3086	1.3987	1.4275	1.2893	1.3035	1.3037	1.3152
KF-FF $\lambda = 0.99$	1.0211	0.9768	0.9829	0.9944	1.0127	1.0180	0.9492	0.9591	0.9678	1.0096	1.0146	0.9255	0.9829	0.9455	0.9362
$\tau = 10$															
KF-TFF	**1.0093**	1.0009	1.0109	1.0093	1.0119	**1.0076**	0.9899	1.0092	1.0079	1.0113	**1.0032**	0.9774	1.0109	1.0051	1.0190
SSP-KF	1.0143	**0.9893**	**0.9909**	**0.9977**	**1.0111**	1.0113	**0.9762**	**0.9820**	**0.9841**	**1.0074**	1.0069	**0.9646**	**0.9909**	**0.9746**	**0.9698**
KF-CFF $\lambda = 0.75$	1.4105	1.3524	1.3527	1.3548	1.4040	1.4241	1.3570	1.3642	1.3618	1.4074	1.4270	1.3529	1.3527	1.3605	1.3683
KF-CFF $\lambda = 0.99$	1.0181	0.9937	0.9959	1.0017	1.0128	1.0167	0.9776	0.9829	0.9866	1.0099	1.0128	0.9646	0.9959	0.9753	0.9706

Table 2: Model used in the analysis. TVPHAR is the time varying HAR model. TVPHARX is the time varying HAR with explanatory variables. DMA is the dynamic model average and DMS is the dynamic model selection. DMA and DMS are reported with time varying forgetting factor (TFF), standardized self-perturbed kalman filter (SSP) and with a constant forgetting factor (CFF). For the CFF case we report the results for different values of λ and α.

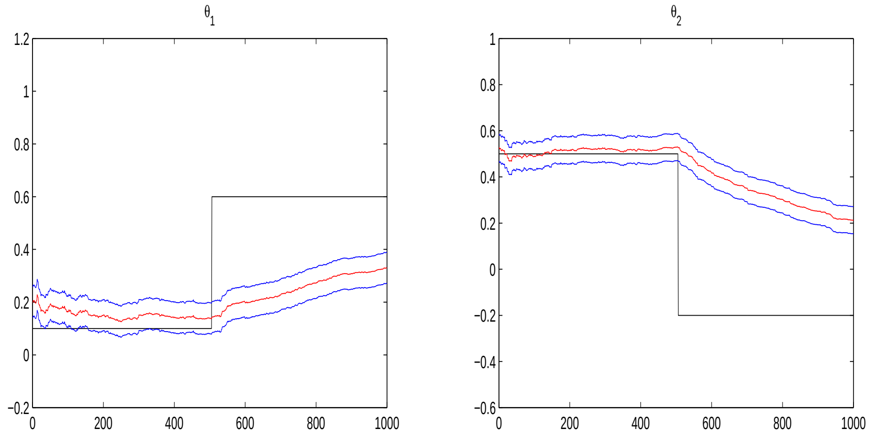
Set of Forecasts Considered			
1	TVPHAR: Time-varying parameter HAR model with CFF and $\lambda = 0.99$.	24	DMA with $\lambda = 0.99$ and $\alpha = 1$.
2	TVPHARX1: TV-HAR model with lags of $\triangle S_t$, with CFF and $\lambda = 0.99$	25	DMA-CFF with $\lambda = 1$ and $\alpha = 1$.
3	TVPHARX2: TV-HAR model with lags of FF_t, with CFF and $\lambda = 0.99$.	26	DMA-KF-TFF with $\alpha = 1$.
4	TVPHARX3: TV-HAR model with lags of TS_t, with CFF and $\lambda = 0.99$.	27	DMA-SSP-KF with $\alpha = 1$.
5	TVPHARX4: TV-HAR model with lags of CDS_t, with CFF and $\lambda = 0.99$.	28	DMS-CFF with $\lambda = 0.95$ and $\alpha = 0.95$.
6	TVPHARX5: TV-HAR model with lags of (VIX_t) , with CFF and $\lambda = 0.99$.	29	DMS-CFF with $\lambda = 0.98$ and $\alpha = 0.95$.
7	TVPHARX6: TV-HAR model with lags of $\log(V_t)$, with CFF and $\lambda = 0.99$.	30	DMS-CFF with $\lambda = 0.99$ and $\alpha = 0.95$.
8	TVPHARX7: TV-HAR model with lags of r_t^-, with CFF and $\lambda = 0.99$.	31	DMS-CFF with $\lambda = 1$ and $\alpha = 0.95$.
9	TVPHARX1-7: TV-HAR model with lags of all explanatory variables and $\lambda = 0.99$.	32	DMS-KF-TFF with $\alpha = 0.95$.
10	DMA-CFF with $\lambda = 0.95$ and $\alpha = 0.95$.	33	DMS-SSP-KF with $\alpha = 0.95$.
11	DMA-CFF with $\lambda = 0.98$ and $\alpha = 0.95$.	34	DMS-CFF with $\lambda = 0.95$ and $\alpha = 0.99$.
12	DMA-CFF with $\lambda = 0.99$ and $\alpha = 0.95$.	35	DMS-CFF with $\lambda = 0.98$ and $\alpha = 0.99$.
13	DMA-CFF with $\lambda = 1$ and $\alpha = 0.95$.	36	DMS-CFF with $\lambda = 0.99$ and $\alpha = 0.99$.
14	DMA-KF-TFF with $\alpha = 0.95$.	37	DMS-CFF with $\lambda = 1$ and $\alpha = 0.99$.
15	DMA-SSP-KF with $\alpha = 0.95$.	38	DMS-KF-TFF with $\alpha = 0.99$.
16	DMA-CFF with $\lambda = 0.95$ and $\alpha = 0.99$.	39	DMS-SSP-KF with $\alpha = 0.99$.
17	DMA-CFF with $\lambda = 0.98$ and $\alpha = 0.99$.	40	DMS-CFF with $\lambda = 0.95$ and $\alpha = 1$.
18	DMA-CFF with $\lambda = 0.99$ and $\alpha = 0.99$.	41	DMS-CFF with $\lambda = 0.98$ and $\alpha = 1$.
19	DMA-CFF with $\lambda = 1$ and $\alpha = 0.99$.	42	DMS-CFF with $\lambda = 0.99$ and $\alpha = 1$.
20	DMA-KF-TFF with $\alpha = 0.99$.	43	DMS-CFF with $\lambda = 1$ and $\alpha = 1$.
21	DMA-SSP-KF with $\alpha = 0.99$.	44	DMS-KF-TFF with $\alpha = 1$.
22	DMA-CFF with $\lambda = 0.95$ and $\alpha = 1$.	45	DMS-SSP-KF with $\alpha = 1$.
23	DMA-CFF with $\lambda = 0.98$ and $\alpha = 1$.		

Table 3: Out-of-sample forecast performances with DMA. The table reports the measures of forecasting performance relative to several online predictions, including forecast combinations by means of DMA with alternative values of α. The out-of-sample period begins on January 3, 2007 and includes 1400 daily observations. An asterisk (∗) implies that the model belongs to the 5% model confidence set of Hansen et al. (2011).

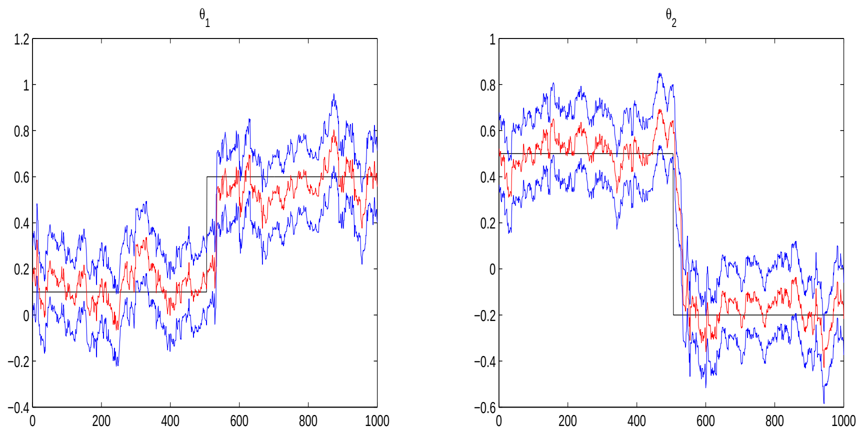
	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE
		$h = 1$			$h = 5$			$h = 10$			$h = 22$	
TVPHAR	-0.8041	0.3136	0.5735	-0.9903	0.3846	0.7101	-1.0706	0.4221	0.7862	-1.1529	0.4667	0.8885
TVPHARX1	-0.8039	0.3129	0.5724	-0.9816	0.3793	0.6988	-1.0527	0.4120	0.7643	-1.1073	0.4408	0.8344
TVPHARX2	-0.8081	0.3150	0.5766	-0.9929	0.3849	0.7098	-1.0700	0.4202	0.7818	-1.1494	0.4626	0.8774
TVPHARX3	-0.8080	0.3142	0.5749	-0.9922	0.3847	0.7106	-1.0672	0.4189	0.7806	-1.1323	0.4545	0.8618
TVPHARX4	-0.8060	0.3135	0.5736	-0.9858	0.3815	0.7030	-1.0658	0.4185	0.7789	-1.1403	0.4569	0.8622
TVPHARX5	-0.7932	0.3099	0.5671	-0.9846	0.3815	0.7063	-1.0597	0.4168	0.7792	-1.1388	0.4601	0.8808
TVPHARX6	-0.8031	0.3133	0.5724	-0.9905	0.3843	0.7102	-1.0674	0.4201	0.7829	-1.1465	0.4642	0.8814
TVPHARX7	-0.7866	0.3068	0.5617	-0.9928	0.3843	0.7111	-1.0737	0.4221	0.7862	-1.1566	0.4681	0.8898
TVPHARX1-7	-0.7762	0.3024	0.5523	-0.9770	0.3759	0.6916	-1.0445	0.4047	0.7474	-1.0997	0.4270	0.7853
$\alpha = 0.95$												
DMA-$\lambda = 0.95$	-0.8704	0.3317	0.5820	-0.9739	0.3726	0.6540	-0.9954	0.3815	0.6717	-0.9825	0.3756	0.6597
DMA-$\lambda = 0.98$	-0.7967	0.3120	0.5601	-0.9587	0.3744	0.6715	-1.0043	0.3929	0.7065	-1.0063	0.3934	0.7045
DMA-$\lambda = 0.99$	-0.7783	0.3070	0.5539	-0.9649	0.3768	0.6833	-1.0249	0.4022	0.7330	-1.0446	0.4112	0.7452
DMA-$\lambda = 1$	-0.7902	0.3113	0.5668	-0.9779	0.3835	0.7002	-1.0568	0.4195	0.7746	-1.1395	0.4641	0.8781
DMA-KF-TFF	-0.7772	0.3064	0.5546	-0.9653	0.3764	0.6802	-0.9975	0.3877	0.6906	-0.9830	0.3806	0.6751
DMA-SSP-KF	-0.8100	0.3126	0.5662	-0.8938	0.3414	0.6167	-0.8928	0.3399	0.6128	-0.8980	0.3412	0.6151
$\alpha = 0.99$												
DMA-$\lambda = 0.95$	-0.8540	0.3278	0.5826	-0.9642	0.3717	0.6596	-0.9856	0.3806	0.6793	-0.9728	0.3742	0.6669
DMA-$\lambda = 0.98$	-0.7908	0.3094	0.5588	-0.9632	0.3774	0.6787	-1.0080	0.3955	0.7163	-1.0136	0.3980	0.7191
DMA-$\lambda = 0.99$	-0.7723	0.3039	0.5516	-0.9695	0.3795	0.6907	-1.0298	0.4056	0.7439	-1.0533	0.4178	0.7607
DMA-$\lambda = 1$	-0.7885	0.3110	0.5666	-0.9811	0.3857	0.7051	-1.0633	0.4240	0.7840	-1.1515	0.4713	0.8960
DMA-KF-TFF	-0.7736	0.3042	0.5522	-0.9687	0.3785	0.6866	-0.9976	0.3884	0.6974	-0.9852	0.3830	0.6864
DMA-SSP-KF	-0.8032	0.3109	0.5644	-0.8923	0.3412	0.6166	-0.8891	0.3386	0.6117	-0.8948	0.3400	0.6147
$\alpha = 1$												
DMA-$\lambda = 0.95$	-0.8442	0.3198	0.5805	-0.9675	0.3710	0.6670	-0.9898	0.3794	0.6857	-0.9762	0.3761	0.6765
DMA-$\lambda = 0.98$	-0.7816	0.3045	0.5564	-0.9683	0.3780	0.6818	-1.0109	0.3949	0.7192	-1.0153	0.3999	0.7239
DMA-$\lambda = 0.99$	-0.7684	0.3016	0.5503	-0.9762	0.3819	0.6962	-1.0371	0.4091	0.7511	-1.0571	0.4216	0.7682
DMA-$\lambda = 1$	-0.7862	0.3094	0.5654	-0.9855	0.3862	0.7102	-1.0692	0.4266	0.7910	-1.1614	0.4765	0.9084
DMA-KF-TFF	-0.7716	0.3022	0.5509	-0.9734	0.3805	0.6890	-1.0020	0.3886	0.7025	-0.9881	0.3845	0.6930
DMA-SSP-KF	-0.7958	0.3085	0.5611	-0.8948	0.3419	0.6205	-0.8915	0.3388	0.6154	-0.8976	0.3409	0.6199

Table 4: Out-of-sample forecast performances with DMS. The table reports the measures of forecasting performance relative to several online predictions with forecast combinations by means of DMS with alternative values of α. The out-of-sample period begins on January 3, 2007 and includes 1400 daily observations. An asterisk $(*)$ implies that the model belongs to the 5% model confidence set of Hansen et al. (2011).

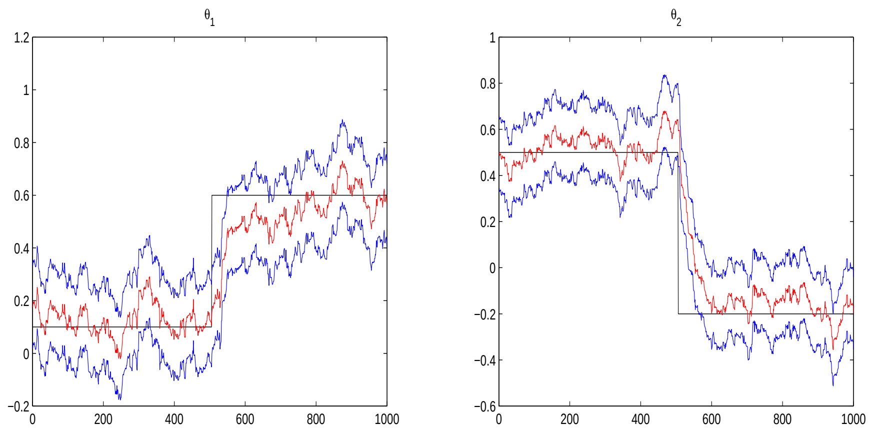
	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE
		$h=1$			$h=5$			$h=10$			$h=22$	
$\alpha = 0.95$												
DMS-$\lambda = 0.95$	-0.7776	0.2954	0.5338	-0.8787	(*) 0.3307	(*) 0.6009	-0.8977	0.3374	0.6094	-0.8918	0.3353	(*) 0.6107
DMS-$\lambda = 0.98$	(*) −0.7205	(*) **0.2826**	(*) **0.5172**	-0.8800	0.3376	0.6183	-0.9254	0.3547	0.6515	-0.9309	0.3558	0.6524
DMS-$\lambda = 0.99$	(*) **−0.7147**	(*) 0.2835	(*) 0.5193	-0.9015	0.3480	0.6390	-0.9589	0.3707	0.6847	-0.9779	0.3769	0.6980
DMS-$\lambda = 1$	-0.7490	0.2969	0.5445	-0.9303	0.3618	0.6679	-1.0101	0.3966	0.7393	-1.0889	0.4381	0.8341
DMS-KFTFF	-0.7212	(*) 0.2868	0.5255	-0.8964	0.3451	0.6333	-0.9151	0.3481	0.6369	-0.8993	0.3414	0.6246
DMS-SSP-KF	-0.7685	0.2995	0.5454	(*) **−0.8486**	(*) **0.3244**	(*) **0.5914**	(*) **−0.8495**	(*) **0.3242**	(*) **0.5896**	(*) **−0.8505**	(*) **0.3233**	(*) **0.5894**
$\alpha = 0.99$												
DMS-$\lambda = 0.95$	-0.8228	0.3132	0.5695	-0.9351	0.3532	0.6429	-0.9526	0.3611	0.6584	-0.9448	0.3579	0.6549
DMS-$\lambda = 0.98$	-0.7650	0.2976	0.5446	-0.9296	0.3580	0.6554	-0.9758	0.3765	0.6924	-0.9851	0.3816	0.7035
DMS-$\lambda = 0.99$	-0.7510	0.2950	0.5403	-0.9438	0.3659	0.6734	-1.0079	0.3927	0.7292	-1.0286	0.4014	0.7452
DMS-$\lambda = 1$	-0.7714	0.3040	0.5567	-0.9656	0.3761	0.6943	-1.0485	0.4143	0.7739	-1.1360	0.4627	0.8847
DMS-KF-TFF	-0.7527	0.2960	0.5411	-0.9403	0.3627	0.6670	-0.9642	0.3687	0.6766	-0.9556	0.3654	0.6695
DMS-SSP-KF	-0.7865	0.3052	0.5564	-0.8772	0.3338	0.6077	-0.8738	0.3324	0.6042	-0.8792	0.3332	0.6074
$\alpha = 1$												
DMS-$\lambda = 0.95$	-0.8442	0.3189	0.5807	-0.9615	0.3649	0.6671	-0.9845	0.3738	0.6882	-0.9716	0.3705	0.6812
DMS-$\lambda = 0.98$	-0.7801	0.3034	0.5555	-0.9611	0.3689	0.6763	-1.0046	0.3877	0.7162	-1.0081	0.3904	0.7220
DMS-$\lambda = 0.99$	-0.7680	0.3010	0.5506	-0.9691	0.3747	0.6896	-1.0334	0.4034	0.7485	-1.0494	0.4113	0.7670
DMS-$\lambda = 1$	-0.7858	0.3084	0.5649	-0.9834	0.3837	0.7105	-1.0666	0.4219	0.7887	-1.1539	0.4721	0.9031
DMS-KF-TFF	-0.7699	0.3013	0.5499	-0.9677	0.3723	0.6845	-0.9987	0.3824	0.7018	-0.9797	0.3770	0.6943
DMS-SSP-KF	-0.7950	0.3081	0.5605	-0.8919	0.3393	0.6184	-0.8907	0.3381	0.6145	-0.8943	0.3393	0.6172



(a) Recursive OLS



(b) Estimates based on KF-TFF



(c) Estimates based on SSP-KF

Figure 1: Online estimation of time-varying parameter model. The black solid line is the true parameter. The red line is the Monte Carlo average for each t, while the blue solid lines are the 90% Monte Carlo confidence bands.

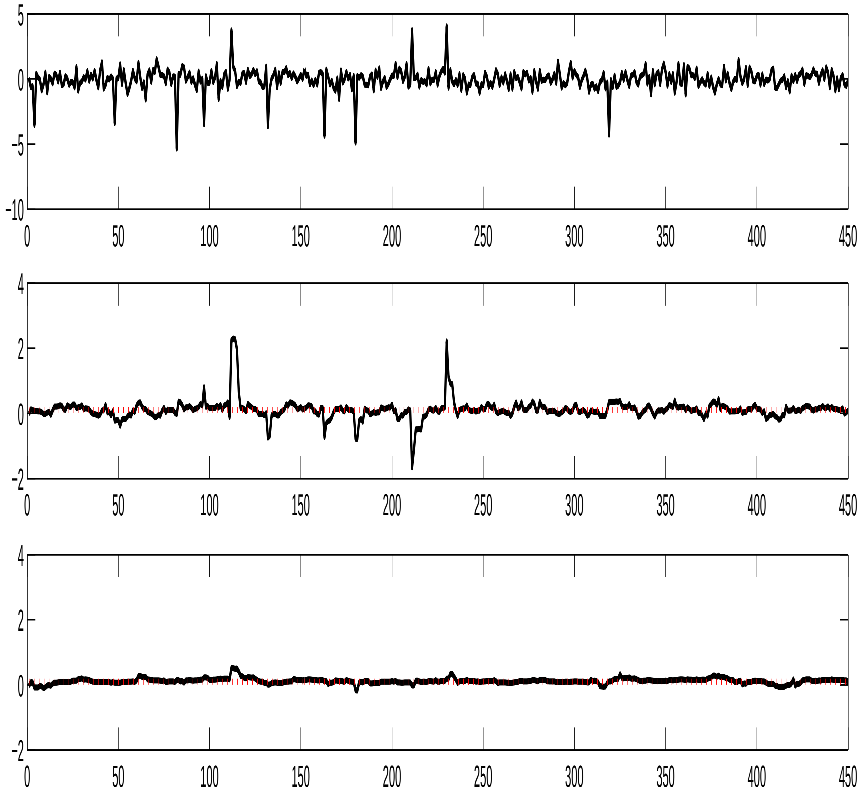
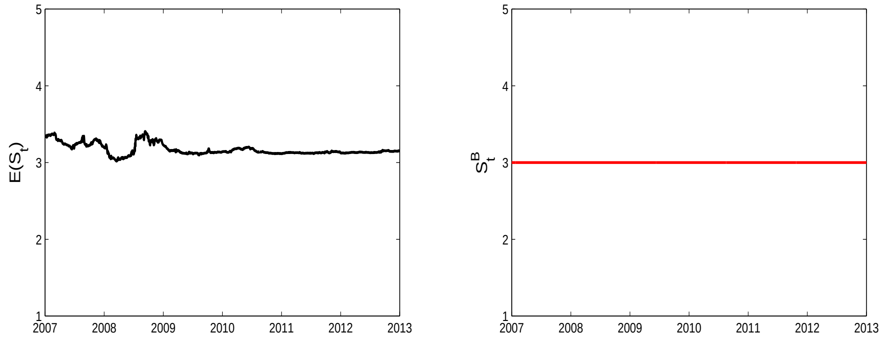
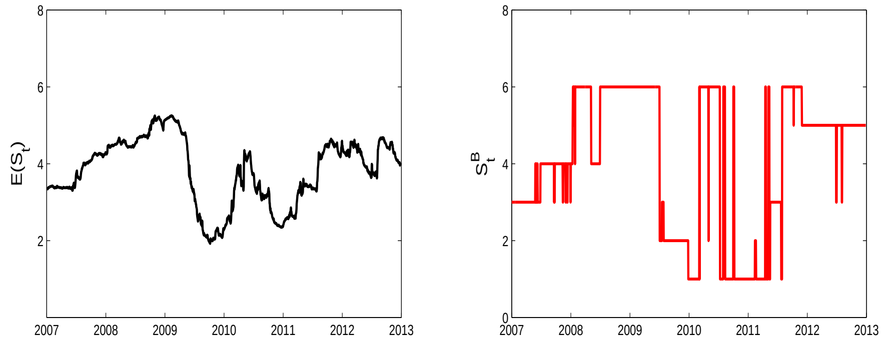


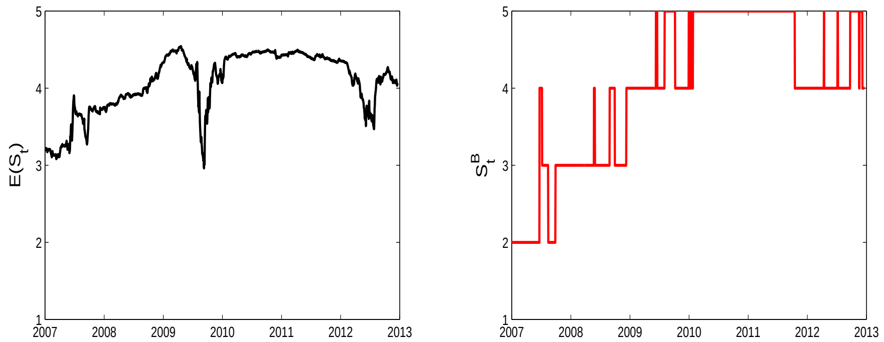
Figure 2: Online estimation of time-varying parameter model when the series is subject to additive outliers. The top panel reports a simulated trajectory of y_t from equation (13) with $\tau = 0.1$. The parameter estimates are reported in the middle and in the bottom panels. The red dotted line is the true parameter, $\theta_1 = 0.1$. The black solid line is the estimated parameter obtained with KF-TFF (middle panel) and with SSP-KF (bottom panel).



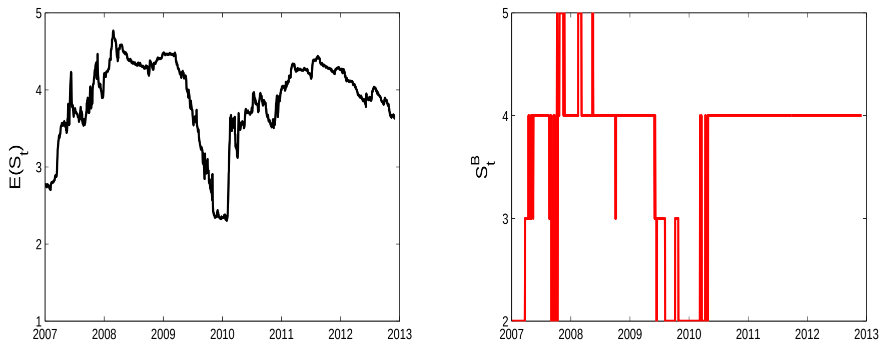
(a) $h = 1$



(b) $h = 5$

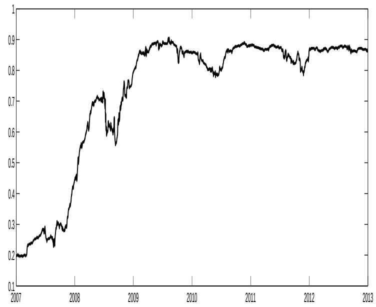


(c) $h = 10$

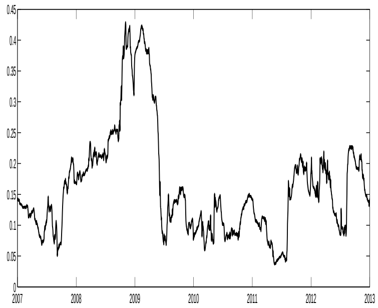


(d) $h = 22$

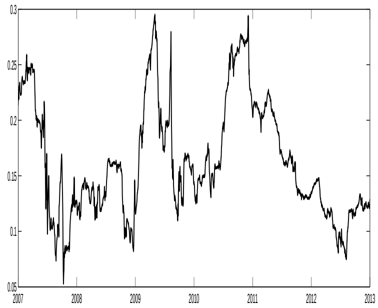
Figure 3: Expected size, $E(S_t)$, and size of the best model, S_t^B, for different forecasting horizons, $h = 1, 5, 10, 22$.



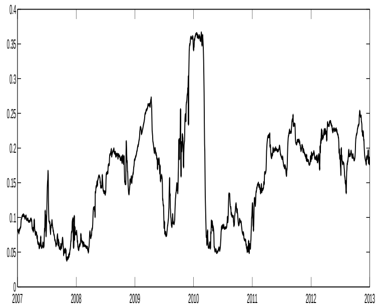
(a) $h = 1$



(b) $h = 5$

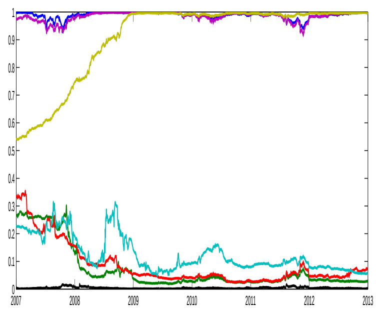


(c) $h = 10$

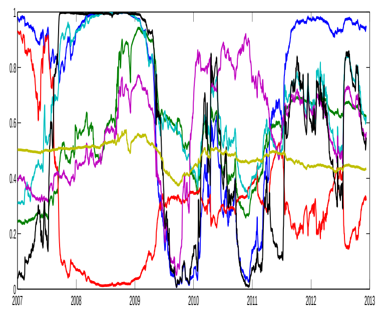


(d) $h = 22$

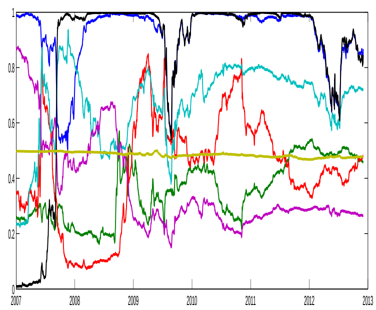
Figure 4: Probability of the best model. The model reports the probability of the best model for



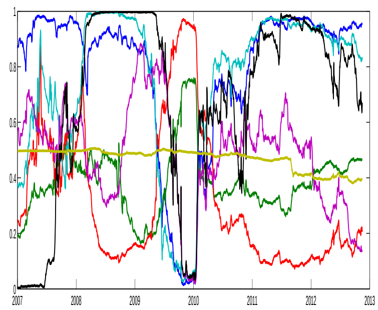
(a) $h = 1$



(b) $h = 5$

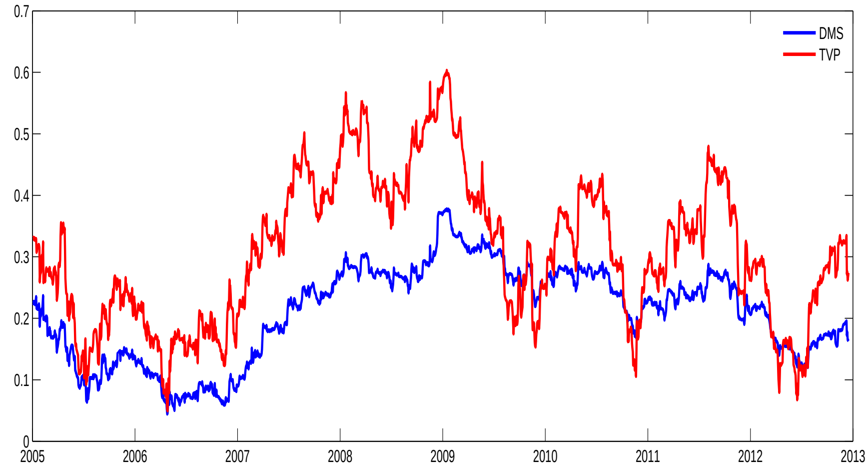


(c) $h = 10$

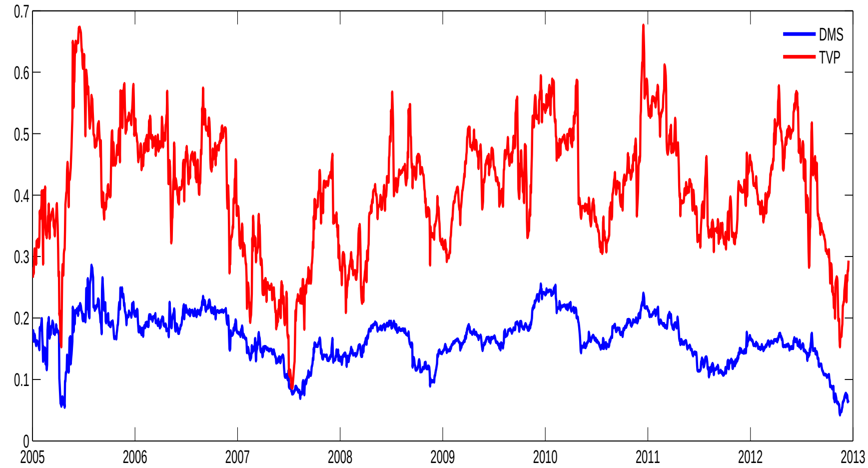


(d) $h = 22$

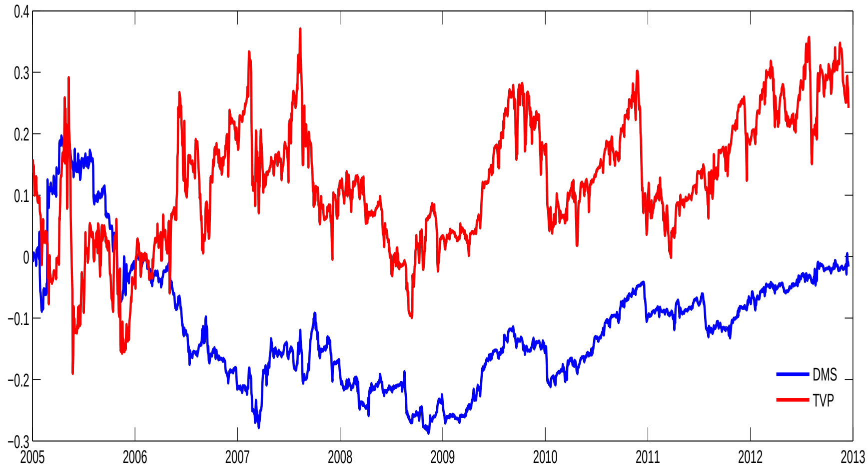
Figure 5: Inclusion probabilities for DMS with $h = 1, 5, 10, 22$. The plot reports the inclusion probabilities for the seven explanatory variables. The inclusion probabilities are relative to the DMS model with $\alpha = 1$. Each coloured line corresponds to the inclusion probability of the each explanatory variable. The colors are: Blue-S_t; Green-FF_t; Red-TS_t; Azure-CDS_t; Purple-VIX_t; Black-V_t; Yellow-r_t^-.



(a) Daily Volatility Factor, ϕ_t^d



(b) Weekly Volatility Factor, ϕ_t^w



(c) Monthly Volatility Factor, ϕ_t^m

Figure 6: Time Varying HAR parameters for $h = 1$. The plots report the evolution of the HAR parameters. In each figure, the blue solid line is the estimate obtained with DMS with $\alpha = 1$. The red solid line is the estimate obtained with the TV-HAR model without explanatory variables.

1 Appendix

This appendix contains additional results relative to the paper ”Forecasting with the Standardized Self-Perturbed Kalman Filter” by Stefano Grassi, Nima Nonejad and Paolo Santucci de Magistris.

Table 1: Monte Carlo results for $T = 500, 1000, 2000$ when the process y_t contains outliers and it is generated as in equation (12). The table reports the average RMSFE, relative to OLS, for several online estimation methods. For the SSP-KF and KF-TFF, the hyper-parameters γ and ρ are selected according to a grid search at each point in time. The KF-FF is calculated for two values of λ. Criteria: RMSFE (ratio to OLS).

	T=500					T=1000					T=2000				
	No break	1 break	4 break	8 break	RW	No break	1 break	4 break	8 break	RW	No break	1 break	4 break	8 break	RW
$\tau = 0.1$															
KF-TFF	**1.0048**	0.6627	**0.5252**	**0.5214**	**0.8385**	**0.9985**	**0.6049**	0.5138	0.4924	**0.8641**	**0.9998**	0.5931	0.5070	0.4783	0.5230
SSP-KF	1.0317	**0.6438**	0.5277	0.5284	0.8470	1.0206	0.6108	**0.5117**	**0.4881**	0.8709	1.0024	**0.5887**	**0.5015**	**0.4717**	**0.5225**
KF-CFF-$\lambda = 0.75$	1.2282	0.7238	0.6045	0.6005	1.0227	1.2096	0.6984	0.6049	0.5704	1.0963	1.2017	0.6909	0.6101	0.5701	0.6680
KF-CFF-$\lambda = 0.99$	1.0332	1.0635	0.8408	0.9070	0.9184	1.0211	0.8757	0.7452	0.7813	0.9154	1.0038	0.7166	0.6427	0.6739	0.5945
$\tau = 1$															
KF-TFF	**1.0094**	0.7972	0.8185	0.8008	0.8910	**1.0043**	**0.7743**	0.7934	0.7942	0.9857	**1.0021**	0.7776	0.7895	0.7885	0.8635
SSP-KF	1.0236	**0.8016**	**0.8112**	**0.7993**	**0.9025**	1.0252	0.7890	**0.7802**	**0.7784**	**0.9438**	1.0050	**0.7776**	**0.7669**	**0.7594**	**0.8564**
KF-CFF-$\lambda = 0.75$	1.3841	1.0289	1.0511	1.0160	1.1079	1.3851	1.0271	1.0260	1.0163	1.2884	1.3801	1.0369	1.0330	1.0168	1.1498
KF-CFF-$\lambda = 0.99$	1.0262	0.9186	0.9005	0.9439	1.0139	1.0309	0.8534	0.8423	0.8664	0.9492	1.0098	0.8034	0.7952	0.8044	0.8820
$\tau = 5$															
KF-TFF	**1.0108**	1.0191	1.0137	1.0258	0.9919	**1.0097**	1.0243	1.0153	1.0167	1.0067	**1.0034**	1.0250	1.0112	1.0095	1.0171
SSP-KF	1.0190	**0.9492**	**0.9617**	**0.9614**	**0.9825**	1.0162	**0.9387**	**0.9457**	**0.9479**	**0.9958**	1.0056	**0.9319**	**0.9370**	**0.9374**	**0.9696**
KF-CFF-$\lambda = 0.75$	1.4216	1.2865	1.3079	1.2907	1.3091	1.4144	1.2806	1.2915	1.2883	1.3926	1.4201	1.2936	1.3003	1.2960	1.3455
KF-CFF-$\lambda = 0.99$	1.0215	0.9671	0.9755	0.9886	1.0122	1.0214	0.9479	0.9552	0.9634	0.9962	1.0117	0.9333	0.9392	0.9422	0.9707
$\tau = 10$															
KF-TFF	**1.0115**	1.0209	1.0109	1.0179	1.0310	1.0144	1.0170	1.0124	1.0128	1.0059	1.0025	1.0125	1.0070	1.0060	1.0096
SSP-KF	1.0167	**0.9822**	**0.9884**	**0.9909**	**1.0023**	**1.0128**	**0.9725**	**0.9772**	**0.9799**	**1.0018**	**1.0061**	**0.9674**	**0.9712**	**0.9725**	**0.9892**
KF-CFF-$\lambda = 0.75$	1.4218	1.3470	1.3598	1.3488	1.3614	1.4169	1.3417	1.3483	1.3466	1.4082	1.4231	1.3526	1.3567	1.3547	1.3834
KF-CFF-$\lambda = 0.99$	1.0193	0.9878	0.9920	0.9993	1.0168	1.0199	0.9756	0.9802	0.9851	1.0030	1.0121	0.9681	0.9716	0.9733	0.9894

Table 2: Out-of-sample forecast performances with DMA. The table reports the measures of forecasting performance relative to several online predictions, including forecast combinations with alternative values of α. The out-of-sample period begins on January 2, 2009 and includes 901 daily observations. An asterisk ($*$) implies that the model belongs to the 5% model confidence set of Hansen et al. (2011).

	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE	log(PL)	CRPS	RMSE
		$h = 1$			$h = 5$			$h = 10$			$h = 22$	
TVPHAR	-0.8243	0.3190	0.5825	-0.9658	0.3735	0.6883	-1.0319	0.4037	0.7492	-1.0893	0.4309	0.8050
TVPHARX1	-0.8243	0.3185	0.5820	-0.9604	0.3704	0.6813	-1.0187	0.3968	0.7326	-1.0463	0.4086	0.7577
TVPHARX2	-0.8263	0.3198	0.5837	-0.9668	0.3735	0.6887	-1.0297	0.4019	0.7445	-1.0815	0.4252	0.7878
TVPHARX3	-0.8272	0.3197	0.5835	-0.9645	0.3724	0.6861	-1.0272	0.4010	0.7450	-1.0740	0.4242	0.7959
TVPHARX4	-0.8260	0.3190	0.5830	-0.9643	0.3720	0.6861	-1.0268	0.4003	0.7431	-1.0822	0.4258	0.7957
TVPHARX5	-0.8151	0.3160	0.5765	-0.9655	0.3727	0.6872	-1.0283	0.4012	0.7479	-1.0866	0.4287	0.8029
TVPHARX6	-0.8237	0.3194	0.5820	-0.9635	0.3724	0.6874	-1.0274	0.4013	0.7447	-1.0868	0.4296	0.8011
TVPHARX7	-0.8086	0.3139	0.5728	-0.9686	0.3738	0.6902	-1.0349	0.4044	0.7499	-1.0925	0.4316	0.8069
TVPHARX1-7	-0.7977	0.3097	0.5633	-0.9574	0.3686	0.6785	-1.0082	0.3896	0.7207	-1.0411	0.4036	0.7477
$\alpha = 0.95$												
DMA-$\lambda = 0.95$	-0.8855	0.3368	0.5920	-0.9618	0.3672	0.6457	-0.9727	0.3714	0.6521	-0.9801	0.3739	0.6561
DMA-$\lambda = 0.98$	-0.8181	0.3184	0.5723	-0.9406	0.3664	0.6601	-0.9724	0.3784	0.6834	-0.9810	0.3825	0.6880
DMA-$\lambda = 0.99$	-0.8023	0.3140	0.5670	-0.9514	0.3696	0.6751	-1.0010	0.3911	0.7177	-1.0167	0.3991	0.7299
DMA-$\lambda = 1$	-0.8097	0.3159	0.5742	-0.9596	0.3740	0.6831	-1.0243	0.4034	0.7424	-1.0754	0.4286	0.7945
DMA-KF-TFF	-0.7992	0.3128	0.5662	-0.9502	0.3690	0.6709	-0.9655	0.3728	0.6643	-0.9725	0.3754	0.6658
DMA-SSP-KF	-0.8333	0.3188	0.5770	-0.8822	0.3353	0.6045	-0.8880	0.3366	0.6035	-0.8950	0.3374	0.6050
$\alpha = 0.99$												
DMA-$\lambda = 0.95$	-0.8706	0.3336	0.5924	-0.9564	0.3677	0.6533	-0.9640	0.3706	0.6596	-0.9744	0.3742	0.6642
DMA-$\lambda = 0.98$	-0.8143	0.3171	0.5709	-0.9436	0.3683	0.6649	-0.9742	0.3796	0.6901	-0.9874	0.3861	0.7000
DMA-$\lambda = 0.99$	-0.7972	0.3119	0.5640	-0.9522	0.3707	0.6772	-1.0015	0.3919	0.7233	-1.0206	0.4021	0.7394
DMA-$\lambda = 1$	-0.8076	0.3153	0.5731	-0.9594	0.3734	0.6838	-1.0283	0.4053	0.7474	-1.0870	0.4349	0.8061
DMA-KF-TFF	-0.7952	0.3112	0.5631	-0.9504	0.3699	0.6726	-0.9633	0.3723	0.6677	-0.9748	0.3775	0.6748
DMA-SSP-KF	-0.8272	0.3173	0.5751	-0.8809	0.3352	0.6038	-0.8853	0.3357	0.6034	-0.8916	0.3365	0.6040
$\alpha = 1$												
DMA-$\lambda = 0.95$	-0.8632	0.3269	0.5913	-0.9626	0.3685	0.6638	-0.9702	0.3711	0.6667	-0.9771	0.3744	0.6727
DMA-$\lambda = 0.98$	-0.8058	0.3119	0.5684	-0.9501	0.3698	0.6680	-0.9806	0.3816	0.6965	-0.9884	0.3885	0.7031
DMA-$\lambda = 0.99$	-0.7930	0.3093	0.5620	-0.9558	0.3727	0.6802	-1.0056	0.3942	0.7281	-1.0220	0.4042	0.7447
DMA-$\lambda = 1$	-0.8043	0.3140	0.5705	-0.9607	0.3722	0.6852	-1.0324	0.4062	0.7529	-1.0977	0.4399	0.8166
DMA-KF-TFF	-0.7944	0.3097	0.5617	-0.9523	0.3718	0.6701	-0.9723	0.3759	0.6766	-0.9783	0.3793	0.6803
DMA-SSP-KF	-0.8215	0.3151	0.5721	-0.8824	0.3367	0.6089	-0.8838	0.3347	0.6043	-0.8912	0.3368	0.6086

Table 3: Out-of-sample forecast performances with DMS. The table reports the measures of forecasting performance relative to several online predictions with forecast combinations by means of DMS with alternative values of α. The out-of-sample period begins on January 2, 2009 and includes 901 daily observations. An asterisk $(*)$ implies that the model belongs to the 5% model confidence set of Hansen et al. (2011).

$\alpha = 0.95$												
DMS-$\lambda = 0.95$	-0.7903	0.3005	0.5423	-0.8666	0.3261 (*)	0.5924 (*)	-0.8775	0.3296	0.5929	-0.8912	0.3350	0.6085 (*)
DMS-$\lambda = 0.98$	−0.7485 (*)	**0.2916** (*)	**0.5330** (*)	-0.8671	0.3327	0.6106	-0.8988	0.3456	0.6355	-0.9105	0.3495	0.6414
DMS-$\lambda = 0.99$	**−0.7482** (*)	0.2944 (*)	0.5376 (*)	-0.8972	0.3468	0.6378	-0.9444	0.3660	0.6782	-0.9564	0.3710	0.6873
DMS-$\lambda = 1$	-0.7749	0.3039	0.5559	-0.9180	0.3569	0.6575	-0.9810	0.3838	0.7123	-1.0266	0.4054	0.7582
DMS-KF-TFF	−0.7520 (*)	0.2964 (*)	0.5412	-0.8870	0.3413	0.6279	-0.8843	0.3361	0.6145	-0.8898	0.3372	0.6156
DMS-SSP-KF	-0.7962	0.3074	0.5584	**−0.8448** (*)	**0.3218** (*)	**0.5855** (*)	**−0.8461** (*)	**0.3220** (*)	**0.5824** (*)	**−0.8561** (*)	**0.3231** (*)	**0.5859** (*)
$\alpha = 0.99$												
DMS-$\lambda = 0.95$	-0.8367	0.3180	0.5760	-0.9254	0.3492	0.6352	-0.9327	0.3523	0.6406	-0.9442	0.3572	0.6504
DMS-$\lambda = 0.98$	-0.7903	0.3059	0.5574	-0.9089	0.3496	0.6394	-0.9442	0.3633	0.6688	-0.9618	0.3716	0.6844
DMS-$\lambda = 0.99$	-0.7781	0.3032	0.5530	-0.9287	0.3589	0.6592	-0.9816	0.3819	0.7113	-1.0002	0.3897	0.7254
DMS-$\lambda = 1$	-0.7932	0.3097	0.5651	-0.9474	0.3672	0.6767	-1.0155	0.3981	0.7408	-1.0719	0.4254	0.7964
DMS-KF-TFF	-0.7778	0.3041	0.5529	-0.9233	0.3552	0.6520	-0.9292	0.3536	0.6476	-0.9456	0.3606	0.6579
DMS-SSP-KF	-0.8138	0.3125	0.5683	-0.8656	0.3286	0.5964	-0.8691	0.3300	0.5963	-0.8753	0.3300	0.5967
$\alpha = 1$												
DMS-$\lambda = 0.95$	-0.8632	0.3257	0.5916	-0.9593	0.3649	0.6678	-0.9629	0.3632	0.6625	-0.9739	0.3688	0.6735
DMS-$\lambda = 0.98$	-0.8058	0.3117	0.5684	-0.9452	0.3618	0.6635	-0.9770	0.3769	0.6956	-0.9813	0.3797	0.7024
DMS-$\lambda = 0.99$	-0.7921	0.3086	0.5618	-0.9482	0.3662	0.6724	-1.0044	0.3916	0.7282	-1.0183	0.3983	0.7453
DMS-$\lambda = 1$	-0.8044	0.3130	0.5703	-0.9597	0.3716	0.6847	-1.0311	0.4042	0.7532	-1.0939	0.4356	0.8171
DMS-KF-TFF	-0.7920	0.3085	0.5599	-0.9439	0.3613	0.6600	-0.9713	0.3703	0.6757	-0.9671	0.3701	0.6761
DMS-SSP-KF	-0.8216	0.3153	0.5720	-0.8797	0.3340	0.6074	-0.8840	0.3341	0.6037	-0.8882	0.3353	0.6058