AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

Forecasting with the Standardized Self-Perturbed Kalman Filter

Stefano Grassi, Nima Nonejad and Paolo Santucci de Magistris

CREATES Research Paper 2014-12

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Forecasting with the Standardized Self-Perturbed Kalman Filter *

Stefano Grassi†
University of Kent and CREATES

Nima Nonejad ‡
Aalborg University and CREATES

Paolo Santucci de Magistris§
Aarhus University and CREATES.

Abstract

We propose and study the finite-sample properties of a modified version of the self-perturbed Kalman filter of Park and Jun (1992) 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 estimate of the measurement error variance. This avoids the calibration of a design parameter as the perturbation term is scaled by the amount of uncertainty in the data. It is shown by Monte Carlo simulations that this perturbation method is associated with a good tracking of the dynamics of the parameters compared to other on-line algorithms and to classical and Bayesian methods. The standardized self-perturbed Kalman filter is adopted to forecast the equity premium on the S&P 500 index under several model specifications, and determine the extent to which realized variance can be used to predict excess returns.

Keywords: TVP models, Self-Perturbed Kalman Filter, Forecasting, Equity Premium, Realized Variance.

JEL Classification: C10, C11, C22, C80.

*We thank Dimitris Korobilis, Tommaso Proietti and Francesco Ravazzolo, the participants to the CFE'2012 conference (Oviedo) and to the SNDE 2013 conference (Milan) for helpful conversations and comments on previous versions of this paper. We also thank the Editor and three anonymous referees for very useful suggestions which have improved the earlier version of the paper. The authors acknowledge the research support of CREATES, funded by the Danish National Research Foundation (DNRF78).

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

‡Department of Mathematical Sciences, Aalborg University. Fredrik Bajers Vej 7G 9220 Aalborg, Denmark. E-mail: nimanonejad@gmail.com

§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 estimating and predicting structural breaks in the parameters governing the relationships between macroeconomic and financial variables. In particular, TVP models are attractive since they allow for empirical insights which are not available within the traditional framework with constant coefficients. Recently, TVP models have shown to be successful in macroeconomics, see for instance Primiceri (2005), Cogley and Sargent (2005) and Koop et al. (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, Stock and Watson (2007), Cogley et al. (2010) and Grassi and Proietti (2010) focus on the US inflation series. They all find strong evidence of a reduction in the volatility of the inflation rate over the last 25 years, a well known phenomenon called the *Great Moderation*. Moreover, the coefficients on the predictors of inflation are also found to vary over time and to be subject to structural breaks. This phenomenon is referred to as the *time-varying* Phillips curve. In finance, the interest for models with time-varying parameters dates back to the 1980's, when the successful class of ARCH-GARCH models was introduced by Engle (1982) and Bollerslev (1986). Together with stochastic volatility models, they can be thought of as two alternative ways to generate time-varying standard deviations of returns. Time-varying parameter models have also been successfully applied in studying how the stock return predictability has been changing over time, see among others Paye and Timmermann (2006), Timmermann (2008) and Henkel et al. (2011). Recently, Liu and Maheu (2008) have provided empirical evidence that allowing for structural breaks in the model parameters leads to sensible improvements in modeling and forecasting realized variance.

Although TVP models have proven to be successful in describing the changing behavior of macroeconomic variables, stock returns and volatility, most of the estimation methods employed so far are computationally intensive, since they generally require simulation based algorithms, such as MCMC or sequential Monte Carlo methods. Recently, Raftery et al. (2010) and Koop and Korobilis (2012, 2013) have proposed a simple method to estimate TVP models within a state-space framework, that does not involve the optimization of any objective function. Following Fagin (1964) and Jazwinsky (1970), they suggest estimating TVP models using a modified Kalman filter algorithm based on an approximation of the updating step of the covariance matrix of the latent states. In particular, the updating equation of the states covariance matrix is restricted to depend on the past by a decay rate that is function of a design parameter, the so called *forgetting factor*.

Similarly to Koop and Korobilis (2012, 2013), we propose an alternative method for the estimation of TVP models based on an extension of the self-perturbed Kalman filter of Park and Jun (1992). Specifically, the original 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. We introduce a modification of the perturbation function by standardizing the squared prediction errors by an estimate of the measurement error variance. Doing so not only avoids the calibration of a design parameter, but also makes the perturbation scheme dependent on the amount of uncertainty in the measurement errors at each point in time. 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 a simple exponential weighted moving average (EWMA). The *standardized self-perturbed Kalman filter* (SSP-KF) still relies on the calibration of two parameters, the sensitivity to the weighted squared prediction error, ς, and the decay parameter in the EWMA of the the error variance, κ. Given ς and κ, the SSP-KF method returns filtered trajectories of the latent processes assumed to evolve as random walks. Although the random-walk assumption of the regression coefficients may appear rather restrictive, the updating mechanism in the SSP-KF proves to be very flexible and able to accommodate many forms of parameter instability, such as structural breaks, in the form of rapid and large increments/decrements, or smooth transitions. Indeed, the parameters ς and κ are dynamically chosen over a grid of values by means of a model selection method based on the predictive likelihood, such that the response to large or small parameter variations can be determined endogenously. The main advantage of the proposed method lies in its *on-line* nature, i.e. the SSP-KF efficiently processes new information as soon as it becomes available and it produces real-time forecasts without the need of numerical optimization and the selection of an in-sample period. Compared to classical methods, like the Kalman filter or its Bayesian extensions, the SSP-KF turns out to be particularly useful under model uncertainty, i.e. when the best model among J alternative specifications must be selected over time.

We study the finite-sample performance of the SSP-KF by means of a large set of Monte Carlo simulations, and compare its ability in tracking the dynamics of the model parameters with other established methods which either involve maximizing the likelihood function or generating the model parameters and latent states from their respective conditional posteriors. The results indicate that the SSP-KF is characterized by small efficiency losses compared to the standard Kalman filter routine coupled with maximum likelihood estimation or its Bayesian extensions. Notably, when the error term contains outliers, SSP-KF improves the tracking of the parameters with respect to the Kalman filter, as the latter strongly relies on the Gaussianity assumption. In many cases, the SSP-KF improves over the on-line methods based on the forgetting factor, especially when the parameters are characterized by structural breaks in the form of sharp level changes, or when the error contains outliers. The average computational time of the SSP-KF is analogous to that of the method based on the forgetting factor, and it is several times shorter than the classical and Bayesian ones. This makes the SSP-KF particularly useful for dynamic model selection or averaging as illustrated in the empirical section.

Finally, we adopt the SSP-KF to study equity premium predictability over time, with a particular focus on how and when realized variance can be used to improve the quality of the forecasts. The papers by Pettenuzzo and Timmermann (2011), Dangl and Halling (2012) and Johannes et al. (2014) acknowledge the importance of accounting for time-varying parameters, especially time-varying volatility, when predicting excess returns. Similarly to Dangl and Halling (2012), we add to Johannes et al. (2014)'s framework the model uncertainty dimension, i.e. at each point in time the prediction of future excess returns is done by selecting among a number of possible explanatory variables. We find that dynamic model selection often includes realized variance among the relevant regressors, consistently with the finding of volatility feedback effect studied in Bollerslev et al. (2006) among others. Interestingly, we also find evidence that realized variance can be used as a driver of the prediction error variance in the SSP-KF method, thus not only having a non-linear effect on the future excess returns but also offering a more sophisticated control of parameter variability over time via the self-perturbation mechanism. The reason for this modification of the baseline SSP-KF routine lies in the efficiency of the realized variance as an estimator of the total return variance that exploits the information coming from returns at higher frequencies. We find some empirical support for this modification, not only in terms of statistical fitting but also in terms of utility gains for a risk averse investor who has to choose which portion of his wealth to invest into a risky asset on the basis of the predictions of a given model.

To conclude, the contributions of this paper are threefold. First, an extension of the self-perturbed Kalman filter of Park and Jun (1992) where the squared prediction errors are standardized by their variance in the perturbation term, thus avoiding the calibration of the design parameter controlling the size of the squared errors. Second, the proposed algorithm is compared to many other estimation methods for TVP models through Monte Carlo simulations. It emerges that the SSF-KF has very limited efficiency losses compared to the Kalman filter regardless the level of noise-to-signal ratio. Third, a linear TVP model with explanatory variables is proposed to forecast equity premium exploiting the information coming from the realized variance, both in conditional mean and in the conditional variance. The paper is organized as follows. Section 2 introduces the general TVP model and discusses the proposed estimation method. Section 3 presents a Monte Carlo study to assess the efficiency loss of the SSP-KF compared to other methods. The empirical application on the forecast of the monthly excess returns of S&P 500 is presented in Section 4. Finally Section 5 concludes.

2 The Standardized Sef-Perturbed Kalman Filter

The state-space representation of the 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 papers, see among others Primiceri (2005), Koop et al. (2009), Dangl and Halling (2012) and Koop and Korobilis (2012, 2013).

Starting from initial values of the states, θ_0, and of the covariance matrix of the state, P_0, the Kalman filter routine is based on a prediction and an updating step.

Prediction

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

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 Kalman gain. 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. The time-varying parameters are then automatically filtered as latent state variables, once that H_t and Q_t are estimated. The Bayesian estimation on the other hand requires generating from the conditional posterior distributions of H_t, Q_t and the latent states through MCMC methods, see Koop (2003). Although classical and Bayesian algorithms are reliable in the TVP context, they become computationally very intensive as the number of parameters increases. Indeed, estimating the parameters in the $m \times m$ matrix Q_t becomes unfeasible when the number of state variables grows, i.e. when the number of regressors in the measurement equation is very large. For the same reasons, standard methodologies cannot easily be adopted in a context characterized by model uncertainty, i.e. when carrying out dynamic averaging and/or selection over K candidate models at each point in time.

We therefore propose an alternative way to efficiently process the new information at each point in time, where the estimation of the TVP models is carried out by a modification of the updating equation of the covariance matrix $\mathrm{P}_{t|t}$ as suggested in Park and Jun (1992). 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

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

while the updating step (3) becomes

$$
\mathrm{P}_{t|t} = \mathrm{P}_{t|t-1} - \mathrm{P}_{t|t-1} \mathrm{Z}_t^{'} \mathrm{F}_{t|t-1}^{-1} \mathrm{Z}_t \mathrm{P}_{t|t-1} + \varsigma \cdot \mathrm{NINT}\left[\gamma \nu_t^2\right] \cdot \mathrm{I}_m, \tag{5}
$$

where ς is a design constant, γ is the sensitivity gain parameter and I_m is an $m \times m$ 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 mechanism in the sense that it revitalizes the adaptation gain by perturbing the matrix $P_{t|t}$. 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 $t = 1, \dots, T$, then the parameters remain constant. Conversely, when γ is large, such that $\text{NINT}\left[\gamma\nu_t^2\right] \neq 0$ for $t = 1, \dots, T$, then the parameters tend to change rapidly. Substituting equation (5) in equations (2)-(3), it follows that $Q_t = \varsigma\text{NINT}\left[\gamma\nu_t^2\right] \cdot I_m$. In other words, the matrix Q_t is diagonal and dependent on the squared prediction errors through two design parameters, ς and γ. Indeed, the setup of the self-perturbed Kalman filter of Park and Jun (1992) requires the selection of two hyper-parameters, γ and ς, that can be chosen over a grid of values minimizing some penalty criterion. This can be cumbersome, especially when many models are estimated and combined at each point in time.

Therefore, we propose the following modification of equation (5):

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}Z_t'F_{t|t-1}^{-1}Z_tP_{t|t-1} + \varsigma \cdot \text{MAX}\left[0, \text{FL}\left(\frac{\nu_t^2}{\hat{H}_t} - 1\right)\right] \cdot I_m, \tag{6}$$

where $\text{FL}(\cdot)$ is the floor operator rounding to the smallest integer and $\hat{H}_t$ is an online estimator of H_t. The quantity $\xi_t = \dfrac{\nu_t^2}{\hat{H}_t} - 1$ 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, γ, can be dropped as the ratio $\dfrac{\nu_t^2}{\hat{H}_t}$ automatically rescales the impact of the squared innovation by the estimate of the measurement error 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 with no parameter updating. Alternatively, when $\xi_t > 0$, the updating of the parameters is activated. Substituting equation (7) in the denominator of ξ_t and rearranging the terms, it follows that $\xi_t = \frac{\kappa(\nu_t^2 - \hat{H}_{t-1})}{\hat{H}_t}$. Hence, if $\kappa(\nu_t^2 - \hat{H}_{t-1})$ is such that ξ_t is greater than 0, the updating is switched on. 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. The updating mechanism automatically weights the variation in the parameters θ_t by the amount of variability in the data, thus avoiding that periods characterized by high volatility spuriously lead to variations in θ_t. Similarly, the updating mechanism is expected to provide protection against outliers. Indeed, if ν_t at time t is affected by an outlier, it follows that, with high probability, $\kappa(\nu_t^2 - \hat{H}_{t-1})$ will be large relative to H_t. Therefore, the perturbation mechanism will be activated at time t. However, in $t+1$ and in absence of large shocks, the term $\kappa(\nu_{t+1}^2 - \hat{H}_t)$ will be small or negative, such that, most likely, the perturbation mechanism will be switched off again. On the other hand, if the the parameters are subject to a structural break at time t, then the term $\text{FL}\left(\dfrac{\nu_t^2}{\hat{H}_t} - 1\right)$ remains greater than zero until the effect of the structural break is offset by the evolution of the estimated parameters. The speed of adjustment is determined by the parameter ς. Intuitively, the larger ς, the faster is the adaptation once a structural break hits the system.

As it is clear from the previous comments, the variance of the measurement error $\hat{\mathrm{H}}_t$ plays a crucial role in determining the activation of the perturbation scheme and it needs to be carefully estimated. Similarly to Koop and Korobilis (2012, 2013), H_t is estimated by the following exponentially weighted moving average (EWMA henceforth)

$$\hat{\mathrm{H}}_t = \kappa \hat{\mathrm{H}}_{t-1} + (1-\kappa)\,\nu_t^2, \tag{7}$$

which is a weighted sum of past squared prediction errors whose weight depends on κ, which determines the level of smoothness of the process. An alternative method to estimate H_t could be similar to the one outlined in Raftery et al. (2010) which subtracts the term related to the parameter uncertainty $(\mathrm{Z}_t \mathrm{P}_{t|t-1} \mathrm{Z}_t')$ from the squared prediction error. This difference can be negative when there is a large break in the parameters so that there is no updating of $\hat{\mathrm{H}}_t$ when $\nu_t^2 - \mathrm{Z}_t \mathrm{P}_{t|t1} \mathrm{Z}_t' < 0$. Alternatively, one could replace the term ν_t^2 in (7) with $\max[0, \nu_t^2 - \mathrm{Z}_t \mathrm{P}_{t|t1} \mathrm{Z}_t']$ and use $\mathrm{F}_{t|t-1}$ in the perturbation term in equation (6). For sake of comparison with the method of Koop and Korobilis (2012, 2013), we adopt the updating rule of equation (7) in the rest of the paper.

2.1 Selection of ς and κ

The SSP-KF method requires the calibration of two design parameters ς and κ. A simple solution is to assign a pre-specified value to ς and κ. For example, κ is generally set equal to 0.94 by practitioners working with daily financial data. Alternatively, a more sensible way to select these parameters is through a dynamic grid search procedure that chooses the optimal values of ς and κ at each point in time. Therefore, we dynamically select ς and κ based on the predictive likelihood associated to each possible combination of ς and κ within a given grid of values. Hence, the choice of ς and κ is fully data-driven. Given that a total of J possible combinations of ς and κ are considered, the goal is to calculate $\pi_{t|t-1,j}$, which is the probability that j-th combination of ς and κ is used to forecast y_t, given information through time $t-1$. Define $\mathrm{L}_t \in \{1, 2, ..., J\}$ the set of possible models at each point in time, and $\mathrm{Y}_t = \{y_1, \ldots, y_t\}$, the information set at time t, then using the same approximation as in Raftery et al. (2010) and Koop and Korobilis (2012, 2013),

$$\pi_{t|t-1,j} \quad = \quad \frac{\pi_{t-1|t-1,j}^{\alpha}}{\sum_{j=1}^{J} \pi_{t-1|t-1,j}^{\alpha}}, \qquad j = 1, \ldots, J, \tag{8}$$

where $0 < \alpha \leq 1$ acts as a *smoothing* factor that controls how much weight will be assigned to the model that has performed best in the recent past. The updating equation of (8) is then given by:

$$\pi_{t|t,j} \quad = \quad \frac{\pi_{t|t-1,j}\, p^{(j)}\left(y_t \mid \mathrm{Y}_{t-1}\right)}{\sum_{j=1}^{J} \pi_{t|t-1,l}\, p^{(j)}\left(y_t \mid \mathrm{Y}_{t-1}\right)}, \tag{9}$$

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

$$p^{(j)}(y_t \mid \mathrm{Y}_{t-1}) \sim \mathrm{N}(\mathrm{Z}_t^{(j)}\theta_{t|t-1}^{(j)}, \mathrm{H}_t^{(j)} + \mathrm{Z}_t^{(j)}\mathrm{P}_{t|t-1}^{(j)}\mathrm{Z}_t^{(j),'}). \tag{10}$$

Therefore, at each step, the optimal values for ς and κ are associated with the highest value of $\pi_{t|t-1,j}$. This method is called dynamic model selection, DMS henceforth.

3 Monte Carlo Simulations

The ability of the SSP-KF to correctly model the evolution of the parameters is analyzed by means of a set of Monte Carlo simulations. The purpose of this Monte Carlo analysis is to assess the efficiency loss of the SSP-KF compared to the estimates obtained with the Kalman filter and other commonly adopted routines under different data generating processes. We consider the following DGP for y_t:

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

where Z_t is a $1 \times m$ vector of iid standard Gaussian variates, and θ_t is the vector of time-varying parameters. At the same time, the parameters θ_t are assumed to vary according to different specifications. Table 1 summarizes all specifications adopted in the Monte Carlo for the DGP. Given

Table 1: Setup of the the Monte Carlo simulations. Table reports: the variation type in the parameters, the breaking dates and the parameter values. For the random walk case, table reports the initial values of the parameters $\theta_{1,0}$ and $\theta_{2,0}$ as well as the standard-deviations and the correlation of their innovations. For each case, we consider five different noise-to-signal ratios (σ), different error types and sample sizes.

Type	**Values**	**Break Dates**	σ	**Error Distribution**	**Sample**
No Breaks	$\theta_1 = 0.5$, $\theta_2 = -0.3$	——	0.1	Gaussian, constant variance	T=250
One Break	$\theta_1 = [0.2, 0.8]$	$\tau_1 = 55\%$	0.5	Student's t, dgf=3	T=500
	$\theta_2 = [0.4, -0.4]$	$\tau_2 = 35\%$	1.0	Gaussian, GARCH(1,1) variance	T=1000
Three Breaks	$\theta_1 = [0.1, 0.6, 1.2, 0.4]$	$\tau_1 = 35\%, 65\%, 85\%$	5.0		
	$\theta_2 = [0.5, -0.3, 0.3, 0.8]$	$\tau_2 = 25\%, 70\%, 80\%$	10.0		
Random Walk	$\theta_{1,0} = 0.5$	$\sigma_{\eta,1} = 0.0158$			
	$\theta_{2,0} = -0.3$	$\sigma_{\eta,2} = 0.0224$			
		$\rho_{1,2} = -0.2828$			

that the main assumption of the on-line estimation methods is that the variation in the parameters is driven by the squared prediction errors and its variance, a crucial quantity is represented by the noise-to-signal ratio, σ, i.e. the ratio between H_t and the variance of the signal, $\mathrm{Z}_t\theta_t$. Therefore, the Monte Carlo simulations are conducted for small values of σ, i.e. 0.1, for moderate values, 0.5 and 1, and for large values, i.e. 5 or 10. In particular, the variance H_t is set according to the following formula $\mathrm{H}_t = \sigma \cdot \mathrm{Var}\left(\mathrm{Z}_t\theta_t\right)$, where $\mathrm{Var}(\cdot)$ is the sample variance operator. In other words, the error variance, H_t, is assumed proportional to the variance of the signal. We also consider alternative setups for the measurement error term, ε_t in (11), in order to study the robustness to GARCH effects and to outliers, where the latter are generated by a Student's t distribution with 3 degrees of freedom.

The Monte Carlo results are contained in Table 2.[1] The table reports the Monte Carlo average of the absolute parameter distance, APD, of the estimators relative to the standard Kalman filter coupled with maximum likelihood estimation (KF-ML), for $T = 500$ observations based on $S = 1000$ Monte Carlo replications. The APD is given by this formula

$$APD = \frac{1}{mT} \sum_{i=1}^{m} \sum_{t=t_0+1}^{T} |\theta_{i,t} - \hat{\theta}_{i,t}|. \tag{12}$$

The set of alternative estimators includes the simple OLS as well as the on-line algorithms based on the forgetting factor with constant design parameters, λ and κ. For a fair comparison, we include the forgetting factor method of Koop and Korobilis (2013) with dynamic selection of λ and κ, for different choices for α in the DMS. Similarly, Table 2 reports the APD of the baseline self-perturbed Kalman filter of Park and Jun (1992), with dynamic selection of γ, κ and ς. We also consider the Bayesian MCMC-Kalman filter and its version robust to stochastic volatility with priors set at common value in the literature, see Koop and Korobilis (2010) for a discussion on the role of the prior hyperparameter values. Finally, also the change-point model of Pesaran et al. (2006) and Liu and Maheu (2008) is considered for different expected number of shifts, N_s. In particular, N_s is set proportional to the sample size and equal to either 0.2%, 1% and 10% of the sample size.

As expected, the OLS estimator is associated with the lowest APD for all values of σ when the true parameters are constant. Indeed, the APD of OLS relative to the KF-ML is always smaller than 1 and the lowest across all estimators. On the other hand, OLS is outperformed by other methods when the parameters are subject to structural breaks or vary as random walk processes. Interestingly, when the parameters evolve as random walks and the level of σ is extremely high, then OLS performs better than the KF-ML. Generally, all estimators perform rather similarly when σ is equal to 10. The on-line estimators based on the forgetting factor without optimal selection tend to under-perform when the true parameters contain structural breaks since the algorithm smooths the parameter dynamics when λ is close to unity. When the optimal values of the forgetting factor, λ, and κ are optimally selected as in Koop and Korobilis (2013), then the efficiency loss reduces sensibly, especially when the DGP contains structural breaks. Looking instead at the on-line methods based on the perturbation mechanism, the self-perturbed Kalman filter of Park and Jun (1992), with dynamic selection of γ, ς and κ, performs very well, especially when the true parameters contain one structural break. This evidence provides a first justification for the use of the perturbation scheme in the updating step of $\mathrm{P}_{t|t}$. Unfortunately, the method is also four to five times slower than the proposed SSP-KF due to the search on an additional grid of values for γ. When instead the contribution of the squared prediction error in the perturbation term is endogenously normalized by the SSP-KF algorithm, then the relative APD takes values very close to 1 for almost all DGPs and

[1] A training sample period, $\mathrm{T}_0 = [1, ..., t_0]$, for the parameters, based on the 10% initial observations, is used. We have evaluated the robustness and sensitivity to the initial conditions on H_0, θ_0 and P_0 and to the prior distribution by Monte Carlo simulations and the results are reported in a PDF document with the supplementary material. The document also reports Monte Carlo results for different sample sizes, $T = 250$ and $T = 1000$, and for larger number of regressors, $m = 10$.

Table 2: Monte Carlo. Table reports the 1-step ahead absolute parameter distance relative to that of the Kalman Filter of several estimators of TVP models. The considered estimators are the following: 1) OLS; 2) forgetting factor with constant parameters (CFF); 3) Forgetting factor with the dynamic selection of λ *and* κ *(KK), with* $\lambda \in [0.9, 0.91, ..., 0.99]$ *and* $\kappa \in [0.94, 0.96, 0.98]$ *as in Koop and Korobilis (2013); 4) the self-perturbed Kalman filter of Park (1992) (SP) with dynamic selection of* $\varsigma, \kappa, \gamma$ *with* $\varsigma \in [0.01, 0.02, 0.03, 0.04]$, $\kappa \in [0.94, 0.96, 0.98]$ *and* $\gamma \in [0.01, 0.21, 0.41, 0.61, 0.81, 1.01, 1.21, 1.41]$*; 5) the standardized self-perturbed Kalman filter, (SSP), with dynamic selection of* ς, κ *with* $\varsigma \in [0.01, 0.02, 0.03, 0.04]$ *and* $\kappa \in [0.94, 0.96, 0.98]$*; 6) MCMC with Kalman Filter for TVP model (KF-MCMC); 7) MCMC with Kalman Filter for TVP model under stochastic volatility (KF-MCMC-SV); 8) Change-Point model of Pesaran et al (2006) with different number of breaks percentages. The dynamic selection of the design parameters* λ, ς, κ *and* γ *has been performed with DMS for different values of* $\alpha \in [0.001, 0.95, 1]$*. Last column reports the CPU time relative to that of the Kalman Filter.*

	No Breaks					**One Break**					**Three Breaks**					**Random Walk**					**CPU**
	0.1	**0.5**	**1**	**5**	**10**	**0.1**	**0.5**	**1**	**5**	**10**	**0.1**	**0.5**	**1**	**5**	**10**	**0.1**	**0.5**	**1**	**5**	**10**	
iid Gaussian:																					
OLS	0.65	0.66	0.66	0.66	0.66	5.63	3.58	2.93	1.79	1.43	4.56	2.93	2.41	1.52	1.24	2.64	1.80	1.52	1.04	0.89	0.00
$\text{CFF}_{\lambda=0.96,\kappa=0.94}$	2.08	2.14	2.08	2.07	2.07	1.54	1.08	1.02	1.08	1.16	2.02	1.22	1.08	1.00	1.04	1.23	1.04	1.06	1.24	1.36	0.01
$\text{CFF}_{\lambda=0.98,\kappa=0.94}$	1.54	1.68	1.55	1.54	1.51	2.72	1.41	1.19	0.99	0.97	3.55	1.88	1.51	1.05	0.97	1.67	1.20	1.12	1.04	1.08	0.01
KK, $\alpha = 0.001$	1.29	1.43	1.31	1.32	1.27	2.62	1.45	1.26	1.04	0.99	3.46	1.88	1.57	1.15	1.04	1.67	1.27	1.18	1.04	1.04	0.10
KK, $\alpha = 0.95$	1.11	1.24	1.14	1.15	1.10	2.68	1.44	1.23	1.08	1.05	2.99	1.66	1.39	1.13	1.07	1.53	1.23	1.18	1.06	1.03	0.10
KK, $\alpha = 1$	1.10	1.27	1.17	1.17	1.11	3.19	1.57	1.32	1.21	1.18	3.06	1.68	1.39	1.18	1.13	1.47	1.24	1.22	1.12	1.09	0.10
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 0.001$	3.44	2.77	2.55	2.31	2.29	0.88	0.93	1.03	1.47	1.80	1.17	1.01	1.04	1.36	1.63	1.84	1.22	1.24	1.84	2.27	0.97
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 0.95$	1.63	1.52	1.48	1.44	1.43	0.80	0.90	0.96	1.04	1.05	1.07	0.99	1.02	1.05	1.04	1.85	1.19	1.11	1.09	1.11	0.96
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 1$	1.49	1.38	1.31	1.25	1.19	0.80	1.00	1.07	1.14	1.12	1.05	1.06	1.08	1.15	1.09	1.84	1.19	1.15	1.12	1.10	0.96
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 0.001$	1.13	1.82	2.28	2.93	3.27	0.96	1.02	1.06	1.21	1.29	0.98	1.05	1.06	1.10	1.16	1.15	1.17	1.22	1.42	1.53	0.23
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 0.95$	1.07	1.45	1.52	1.43	1.39	0.86	0.96	1.00	1.07	1.08	0.95	1.01	1.03	1.05	1.03	1.09	1.08	1.08	1.11	1.13	0.23
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 1$	1.07	1.35	1.31	1.23	1.18	1.26	1.14	1.12	1.19	1.19	1.18	1.12	1.10	1.12	1.14	1.12	1.16	1.17	1.11	1.09	0.23
KF-MCMC	3.17	2.46	2.23	1.84	1.74	0.82	0.84	0.86	0.91	0.93	0.78	0.78	0.78	0.80	0.82	0.80	0.87	0.92	1.04	1.10	5.65
KF-MCMC-SV	3.82	2.93	2.64	2.10	1.92	0.97	0.99	1.00	1.01	1.01	0.85	0.86	0.87	0.88	0.89	0.98	1.05	1.09	1.17	1.18	10.12
ChagePoint 0.2%	2.69	1.64	1.81	1.30	1.01	1.90	1.25	1.14	0.89	0.97	3.91	2.50	2.06	1.39	1.25	1.75	1.27	1.05	1.11	1.01	6.07
ChagePoint 2%	3.71	3.59	3.10	2.84	2.84	1.16	1.19	1.42	1.59	1.63	1.01	1.06	1.29	1.65	1.60	1.88	1.91	1.74	2.06	1.80	6.40
ChagePoint 10%	9.82	6.67	6.20	5.74	5.03	2.29	2.41	2.49	2.60	2.52	1.78	2.01	2.15	2.18	2.02	2.72	2.77	2.95	3.23	3.00	7.65
Student's t(3):																					
OLS	0.59	0.58	0.58	0.58	0.58	4.28	2.61	2.09	1.22	0.97	3.53	2.18	1.76	1.09	0.91	2.02	1.35	1.13	0.78	0.68	0.00
$\text{CFF}_{\lambda=0.96,\kappa=0.94}$	1.50	1.50	1.48	1.46	1.46	1.15	0.90	0.89	0.96	1.02	1.46	0.98	0.90	0.88	0.95	1.00	0.93	0.96	1.12	1.19	0.01
$\text{CFF}_{\lambda=0.98,\kappa=0.94}$	1.11	1.14	1.11	1.07	1.06	1.76	1.06	0.94	0.81	0.81	2.48	1.37	1.13	0.83	0.81	1.28	0.98	0.91	0.90	0.90	0.01
KK, $\alpha = 0.001$	0.93	0.98	0.97	0.92	0.91	1.78	1.13	1.02	0.85	0.80	2.41	1.41	1.22	0.92	0.85	1.33	1.07	0.98	0.88	0.83	0.11
KK, $\alpha = 0.95$	0.79	0.83	0.82	0.78	0.76	1.82	1.13	1.01	0.89	0.83	2.16	1.28	1.11	0.93	0.85	1.24	1.05	0.98	0.87	0.80	0.11
KK, $\alpha = 1$	0.80	0.86	0.85	0.79	0.77	2.07	1.20	1.10	0.99	0.91	2.24	1.28	1.13	0.98	0.89	1.22	1.08	1.02	0.90	0.82	0.11
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 0.001$	2.09	1.78	1.71	1.64	1.63	0.84	0.92	1.01	1.44	1.73	0.98	0.94	0.99	1.33	1.63	1.19	1.05	1.16	1.73	2.04	1.00
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 0.95$	1.11	1.04	1.02	0.97	0.97	0.83	0.87	0.88	0.88	0.87	0.95	0.94	0.93	0.90	0.87	1.18	0.97	0.94	0.90	0.90	1.00
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 1$	0.99	0.94	0.90	0.83	0.81	0.94	0.98	0.97	0.92	0.89	1.03	1.01	1.00	0.93	0.88	1.19	1.00	0.97	0.89	0.84	1.00
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 0.001$	1.25	1.61	1.76	2.23	2.43	0.90	0.96	0.99	1.10	1.16	0.96	0.98	0.98	1.03	1.09	1.04	1.08	1.12	1.26	1.34	0.25
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 0.95$	1.08	1.07	1.04	1.00	0.96	0.85	0.90	0.90	0.90	0.88	0.93	0.94	0.93	0.90	0.87	0.98	0.97	0.95	0.91	0.91	0.25
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 1$	1.04	0.95	0.92	0.88	0.85	1.12	1.02	1.00	0.97	0.86	1.11	1.00	0.99	1.00	0.91	1.05	1.03	0.99	0.88	0.82	0.25
KF-MCMC	2.30	1.85	1.71	1.48	1.42	0.83	0.86	0.87	0.93	0.97	0.79	0.79	0.79	0.84	0.89	0.81	0.91	0.96	1.07	1.11	5.91
KF-MCMC-SV	2.36	1.80	1.62	1.28	1.15	0.86	0.84	0.83	0.80	0.79	0.77	0.75	0.74	0.72	0.72	0.87	0.91	0.93	0.91	0.89	10.46
ChagePoint 0.2%	1.07	1.20	0.99	0.83	1.00	1.47	1.05	0.99	1.08	1.07	3.03	1.91	1.63	1.21	1.23	1.39	1.07	1.02	0.97	1.00	3.23
ChagePoint 2%	3.31	2.42	2.22	2.49	2.40	0.99	1.12	1.38	1.70	1.56	1.11	1.29	1.42	1.48	1.52	1.56	1.76	1.69	1.61	1.61	3.43
ChagePoint 10%	6.04	5.06	4.84	4.55	4.21	2.14	2.16	2.32	2.64	2.65	1.75	1.95	2.18	2.34	2.04	2.44	2.47	2.67	2.78	2.59	4.11
GARCH(1,1):																					
OLS	0.68	0.68	0.68	0.69	0.69	5.61	3.52	2.87	1.73	1.37	4.53	2.86	2.34	1.45	1.18	2.93	1.96	1.65	1.11	0.95	0.00
$\text{CFF}_{\lambda=0.96,\kappa=0.94}$	1.98	1.90	1.90	1.89	1.90	1.55	1.07	1.01	1.06	1.12	2.05	1.23	1.08	0.98	1.01	1.25	1.04	1.04	1.24	1.40	0.01
$\text{CFF}_{\lambda=0.98,\kappa=0.94}$	1.59	1.45	1.46	1.39	1.39	2.75	1.43	1.21	0.99	0.96	3.57	1.87	1.50	1.04	0.96	1.75	1.20	1.08	1.02	1.07	0.01
KK, $\alpha = 0.001$	1.38	1.28	1.30	1.22	1.20	2.62	1.47	1.29	1.06	1.00	3.47	1.88	1.56	1.14	1.03	1.71	1.25	1.14	0.99	0.97	0.13
KK, $\alpha = 0.95$	1.19	1.10	1.12	1.06	1.04	2.66	1.44	1.25	1.08	1.04	3.00	1.65	1.38	1.11	1.05	1.60	1.22	1.13	1.00	0.96	0.13
KK, $\alpha = 1$	1.17	1.10	1.14	1.07	1.05	3.19	1.58	1.35	1.20	1.17	3.11	1.70	1.39	1.15	1.11	1.55	1.21	1.15	1.03	0.98	0.13
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 0.001$	3.33	2.68	2.47	2.27	2.22	0.90	0.93	1.02	1.46	1.78	1.20	1.01	1.04	1.34	1.61	1.75	1.27	1.28	2.00	2.52	1.17
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 0.95$	1.61	1.45	1.43	1.39	1.36	0.83	0.92	0.97	1.06	1.05	1.08	1.01	1.03	1.05	1.04	1.78	1.25	1.12	1.14	1.16	1.18
$\text{SP}_{\varsigma,\kappa,\gamma}$, $\alpha = 1$	1.45	1.24	1.26	1.17	1.12	0.82	0.99	1.05	1.11	1.09	1.08	1.06	1.08	1.11	1.06	1.83	1.26	1.10	1.04	1.02	1.18
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 0.001$	1.13	1.69	2.06	2.78	3.09	0.98	1.03	1.07	1.21	1.29	1.00	1.04	1.06	1.11	1.16	1.19	1.19	1.24	1.45	1.59	0.29
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 0.95$	1.06	1.42	1.46	1.38	1.32	0.88	0.96	1.01	1.07	1.08	0.97	1.01	1.03	1.04	1.03	1.12	1.12	1.11	1.16	1.19	0.29
$\text{SSP}_{\varsigma,\kappa}$, $\alpha = 1$	1.06	1.25	1.28	1.17	1.10	1.25	1.12	1.10	1.15	1.15	1.18	1.12	1.09	1.08	1.09	1.18	1.19	1.17	1.08	1.05	0.29
KF-MCMC	3.50	2.68	2.42	1.94	1.76	0.97	0.98	0.99	0.99	0.98	0.86	0.86	0.86	0.85	0.85	0.97	1.04	1.09	1.21	1.26	7.43
KF-MCMC-SV	2.94	2.28	2.06	1.71	1.60	0.82	0.84	0.85	0.89	0.92	0.79	0.78	0.78	0.79	0.81	0.80	0.86	0.91	1.05	1.13	12.57
ChagePoint 0.2%	1.78	1.20	0.85	0.97	0.97	1.91	1.29	1.24	0.92	0.94	3.90	2.45	2.01	1.32	1.18	1.80	1.30	1.28	0.98	0.91	6.63
ChagePoint 2%	3.81	3.10	2.70	2.83	2.76	1.11	1.24	1.42	1.47	1.62	1.20	1.21	1.29	1.42	1.68	1.85	1.58	1.65	1.65	1.58	6.95
ChagePoint 10%	8.15	7.10	6.14	5.42	4.96	2.15	2.34	2.34	2.64	2.58	1.73	1.97	2.03	2.16	2.14	3.19	2.78	2.84	3.01	3.49	8.45

for most choices of σ. Looking at the choice of α, the best results are obtained when $\alpha = 0.95$, while the computational time is much lower compared to the KF-ML method. As expected, the SSP-KF has the best relative performances in the setup characterized by structural breaks, a feature that KF-ML cannot easily accommodate. In presence of structural breaks, also the Bayesian methods, i.e. those based on the MCMC algorithm, generally display the best performances as the *APD* relative to that of the standard Kalman filter is smaller than 1. On the contrary, we observe that the change-point models are almost always outperformed by the standard Kalman filter, also when the true DGP contains structural breaks. The reason is that the correct percentage of shifts should also be optimally selected when working with change-point models, see Liu and Maheu (2008) and the discussion in Pettenuzzo and Timmermann (2011). However, the computational time for carrying out the optimal selection of the number of breaks would be several times larger than that of the Kalman filter. Note that the CPU time is already six to seven time larger than that of KF-ML although the number of shifts is kept fixed.[2]

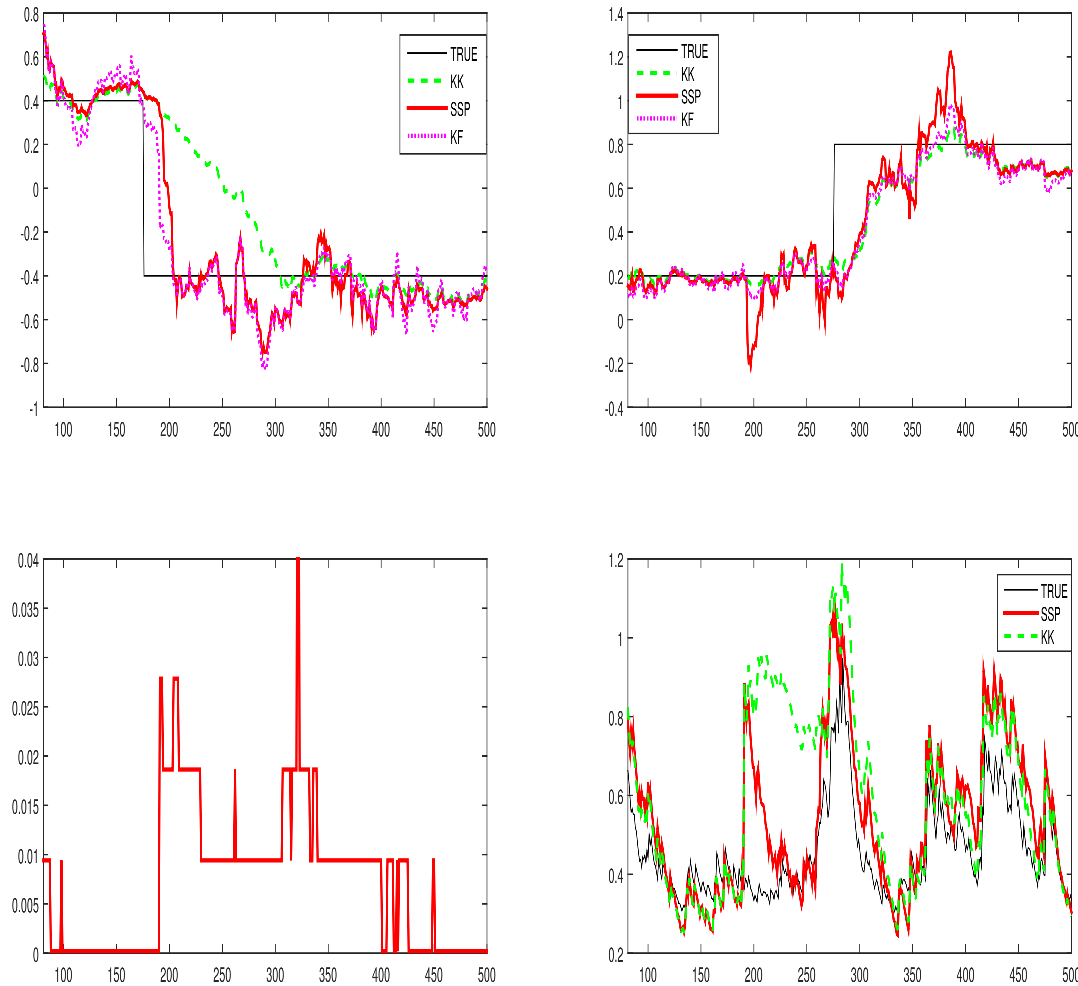
Notably, the proposed perturbation method also offers some degree of protection against outliers compared to the standard Kalman filter, as the average *APD* is smaller than 1 in many cases when the errors are generated from a Student's t distribution with 3 degrees of freedom. Similarly, under GARCH dynamics for the volatility of the error term, the results for the SSP-KF are analogous to those obtained under the constant volatility specification. The GARCH dynamics are generated as

$$\mathrm{H}_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta\mathrm{H}_{t-1},$$

where ω is set to guarantee that H_t has the same level of long-run (unconditional) mean as in the case with constant volatility. In other words, $\mathrm{E}(\mathrm{H}_t) = \frac{\omega}{1-\alpha-\beta} = \sigma \cdot \mathrm{Var}\,(\mathrm{Z}_t\theta_t)$. The dynamics of volatility are also rather persistent as the parameter β is set equal to 0.9. Perhaps, under more noisy dynamics of H_t, i.e. with a smaller choice of β, the results would be different to those obtained under constant variance. However, large values of β are empirically found to characterize financial time series such as returns, interest-rates, exchange rates or realized measures of variance. For illustrative purposes, Figure 1 reports the estimated parameters, together with the latent true parameters, when the latter are characterized by one break and $\sigma = 1$ under GARCH dynamics. It clearly emerges that the estimates of the parameters dynamics obtained with the standard Kalman filter algorithm and with the SSP-KF are analogous. This is manly due to the adjusting behavior of the parameter ς, bottom-left panel, which is higher after the break dates to increase the speed of adjustment. On the other hand, the tracking of the parameters associated with the method with the forgetting factor, although optimally selected as in Koop and Korobilis (2013), is too smooth, especially for the first parameter. This leads to generally larger *APD* than those obtained under SSP-KF and KF-ML. The estimate of the latent volatility process, H_t, is also very good, especially for the SSP-KF. Notably,

[2]Figure 1 in the document with the supplementary material displays the tracking of the parameters under the change-point method. If the number of breaks is correctly selected, then the change-point method is able to provide a good estimate of the break dates, although with some spurious effects on the other parameters. However, the levels of the parameter are not always correctly estimated and this may lead to large values of the *APD*.

Figure 1: Parameter estimates for the model with one break. The top panels of the figure report the true parameters (solid black lines) together with the estimates obtained with forgetting factor of Koop and Korobilis (2013) (dashed-green line), SSP-KF (solid-red line) and standard Kalman filter (purple-dotted line). The bottom-left panel reports the optimal choice of ς at each point in time for the SSP-KF method. The bottom right panel reports the true values of H_t (solid-black line) together with its estimates relative to the forgetting factor methods of Koop and Korobilis (2013) (dashed-green line) and with the SSP-KF (solid-red line).



after a shift the estimated matrix $\hat{\mathrm{H}}_t$ increases compared to the true one as ν_t^2 also depends on the variation of the parameters, but it reverts to the correct levels as soon as the break in the underlying parameter is *absorbed* by the adjustment mechanism. This provides a further insight on the validity of the proposed standardization of the self-perturbed Kalman filter. Similarly, the method based on the forgetting factor leads to an estimated H_t that also reverts to the correct levels after a break, although at a slower rate than SSP-KF. Based on the evidence that arises from the Monte Carlo simulation, we now show how the SSP-KF can be used to predict the equity premium in a framework characterized by model uncertainty.

4 Return Predictability: Does Realized Variance Matter?

The analysis of the extent of equity returns predictability is of primary interest in finance. Predicting the direction and the size of the fluctuations in the stock prices is indeed a central issue not only for portfolio allocation but also for risk management. Since the early 1980's, a number of articles have been dedicated to return predictability, finding evidence that excess stock returns could be predicted in-sample by regressing them on lagged financial variables. A number of econometric techniques have been adopted in the empirical studies of return predictability, see for an overview Malkiel (2003) and Campbell (2008). Traditionally, predictability in long-horizon (multi-year) returns has been shown using variance-ratio tests. Similarly, the short vs long-run dependence with financial variables, such as the dividend-price ratio or the earnings-price ratio, has been widely studied; see among many others Goyal and Welch (2003), Ang and Bekaert (2007), and Cochrane (2008). Since the paper of Welch and Goyal (2008), a number of studies have investigated if the amount of return predictability is likely to change, depending on the business cycle conditions. For example, Dangl and Halling (2012) find that return predictability can mostly be exploited during recessions and if this feature is properly captured by a model with time-varying parameters, it can lead to substantial utility gains. Similar evidence in favor of models with time-varying parameters is presented in Pettenuzzo and Timmermann (2011) and recently in Johannes et al. (2014).

In this section, we contribute to the large existing literature on return predictability trying to understand to which extend realized variance has predictive power for the conditional density of excess returns. As noted by Jensen and Maheu (2013), the early literature found conflicting results on the sign and significance of the conditional variance from GARCH models in the conditional mean of market excess returns, see also Lettau and Ludvingson (2010), an effect called *volatility spillover*. At the same time, the last 15 years have witnessed a substantial development and an increasing interest in the theory of realized variance, RV henceforth, as an efficient ex-post measure of the volatility of a financial returns, see Andersen and Bollerslev (1998), Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) among many others. Therefore, we study if the sign and the significance of the relation between excess returns and volatility, as measured by RV, is likely to change over time in a context characterized by model uncertainty. Hence, RV is used as an explanatory variable in a dynamic regression of returns under several model specifications. In particular, we propose the following model to predict the excess returns

$$\begin{aligned} r_t^* &= \alpha_t + \delta_t RV_{t-1} + \beta_t' X_{t-1} + \varepsilon_t, \quad t = 1, ..., T, \\ \alpha_t &= \alpha_{t-1} + \eta_{1,t}, \quad \delta_t = \delta_{t-1} + \eta_{2,t}, \quad \beta_t = \beta_{t-1} + \eta_{3,t}, \\ \varepsilon_t &\sim \mathrm{N}(0, \sigma_\varepsilon^2), \quad \eta_t \equiv [\eta_{1,t}, \eta_{2,t}, \eta_{3,t}] \sim \mathrm{N}(0, \mathrm{Q}_t), \end{aligned} \tag{13}$$

where $r_t^* = r_t - r_{f,t}$ is the log-return in excess of the risk free rate, denoted as $r_{f,t}$, and X_t contains a number of explanatory variables that are expected to have predictive power for excess returns. Following Welch and Goyal (2008) and Dangl and Halling (2012), the variables contained in the matrix X_t are: dividend yield (dy), earnings-to-price ratio (ep), dividend-payout ratio (dpayr), book-to-market ratio (bmr), net equity expansion (ntis), long-term government bond yields (lty), long-term government bond returns (ltr), T-bill rate (tbl), default return spread (dfr) and default yield spread (dfy), inflation (inf).[3] The dataset consists of monthly total excess returns of the S&P500 index from May 1937 to December 2013, and it is available on Amit Goyal's webpage. RV is computed using daily excess returns. Since most of the the explanatory variables have a strong non-stationary dynamic behavior and this can lead to compensatory and spurious dynamic effects in the time-varying parameters of the model, then the variables in X_t (with the exception of ltr) are considered in first differences, $\tilde{X}_t = \Delta X_t$. Moreover, since there is a strong evidence of long-memory in RV and in inflation, we fractionally difference both series as $\widetilde{RV}_t = \Delta^{d_{RV}}(RV_t - \mu_{RV})$ and $\widetilde{inf}_t = \Delta^{d_{inf}}(inf_t - \mu_{inf})$ and use them as regressors in (13). The parameters d_{RV} and d_{inf} are estimated with the semi-parametric method of Shimotsu (2008) that is robust to deterministic terms in the data. Therefore, the predictive regression of the excess returns is

$$r_t^* = \alpha_t + \delta_t \widetilde{RV}_{t-1} + \beta_t' \tilde{X}_{t-1} + \varepsilon_t, \quad t = 1, ..., T. \tag{14}$$

We also investigate if the information contained in RV can be exploited to improve the quality of the estimation of the prediction error variance. Since RV is known to be a very efficient estimator of the total return variation over a given period, see Barndorff-Nielsen and Shephard (2002), and given that the parameter variability in the SSP-KF is driven by a mechanism based on the ratio between ν_t^2 and $\hat{\mathrm{H}}_t$, we also consider the possibility of using RV_t instead of ν_t^2 in (7), i.e. as a forcing variable for the dynamics of $\hat{\mathrm{H}}_t$. Since RV is much more efficient than the squared daily returns innovations as a proxy for the total variance, we expect a more precise inference on the parameter variations. Therefore, the modified updating equation for the measurement variance is

$$\hat{\mathrm{H}}_t^* = \kappa \hat{\mathrm{H}}_{t-1}^* + (1 - \kappa)\, RV_t^*, \tag{15}$$

where $RV_t^* = RV_t \cdot \varphi$. The rescaling term $\varphi = \frac{\frac{1}{T}\sum_{t=1}^T \hat{\varepsilon}_t^2}{\frac{1}{T}\sum_{t=1}^T RV_t}$ accounts for the return variability explained by the regressors, since $\hat{\varepsilon}_t$ are the residuals of the OLS regression of r_t^* on X_t.[4] The on-line method based on the updating equation (15) is named SSP-KF-RV. In the next paragraphs, we will provide statistical and financial evaluation of the alternative specifications of model (13).

4.1 Empirical results

We consider several specifications of model (13) for the prediction of excess returns. They are described in Table 3. In particular, when the model involves the estimation of time-varying parameters with SSP-KF or SSP-KF-RV, i.e. specifications from III to XII, the optimal values of κ and ς must be selected trough a grid search as outlined in section 2.1. We assume that $\kappa \in [0.94, 0.95, ..., 0.99]$

[3] See Welch and Goyal (2008) and Dangl and Halling (2012) for a more detailed discussion of these variables.

[4] This scaling method does not account for the possible time-variation in the return predictability. More sophisticated time-dependent rescaling schemes can be adopted, and this is left to future research.

and $\varsigma \in [0.00001, 0.0022, 0.0043, 0.0065, 0.0087]$.[5]

Table 3: Summary of model specifications for the prediction of the excess returns.

Model	Regressors	Estimation Method
I	Intercept only	OLS
II	Intercept and $\widetilde{RV}_{t-1}$	OLS
III	Time-varying intercept	SSP-KF-DMS$_{\alpha=0.95}$ for κ and and ς
IV	Model III plus $\widetilde{RV}_{t-1}$ in the mean	SSP-KF-DMS$_{\alpha=0.95}$ for κ and and ς
V	Time-varying intercept	SSP-KF-RV-DMS$_{\alpha=0.95}$ for κ and ς
VI	Model V plus $\widetilde{RV}_{t-1}$ in the mean	SSP-KF-RV-DMS$_{\alpha=0.95}$ for κ and ς
VII	All explanatory variables	SSP-KF-DMS$_{\alpha=0.95}$ only for κ and ς
VIII	All explanatory variables	SSP-KF-RV-DMS$_{\alpha=0.95}$, only for κ and ς
IX	All explanatory variables	SSP-KF-DMA$_{\alpha=0.95}$, for all regressors, κ and ς
X	All explanatory variables	SSP-KF-DMS$_{\alpha=0.95}$, for all regressors, κ and ς
XI	All explanatory variables	SSP-KF-RV-DMA$_{\alpha=0.95}$, for all regressors, κ and ς
XII	All explanatory variables	SSP-KF-RV-DMS$_{\alpha=0.95}$, for all regressors, κ and ς
Rolling	All explanatory variables	Rolling OLS with window of 120 months
KF	All explanatory variables	Rolling Kalman filter with window of 120 months

Note that, when model uncertainty is accounted for, i.e. we evaluate the fit of the model for all possible combinations of the variables in X_t, then $K = dim(\kappa) \times dim(\varsigma) \times 2^m = 122,800$ models must be estimated at each point in time, where $dim(\kappa)$ and $dim(\varsigma)$ are the number of elements in the grids of κ and ς respectively, and $m = 12$ is the number of regressors including $\widetilde{RV}_t$. Figure 2 plots a summary of the estimates relative to model specification VI, i.e. when only the intercept and $\widetilde{RV}_{t-1}$ are used in the conditional mean of the excess returns and RV_t^* is adopted in the SSP-KF-RV to estimate H_t^*. The variations α_t and δ_t are quite evident compared to OLS estimates based on the full sample. In particular, δ_t is positive and significant during the early post-war period, and it becomes negative in the 1960's and 1980's as a consequence of two large breaks. On the other hand, it remains relatively stable and slightly positive from the late 1980's onward. Interestingly, the parameter δ_t displays a negative drop right after the recent financial crisis in 2008-2009, so that the impact of the RV innovations on the excess returns switches from positive to negative. The estimate of H_t is very smooth, due to the rather high value of the optimal κ which often lies above 0.98. Moreover, ς also changes over time to increase the speed of adaptation of the parameters. Specifically, it lies on the lower bound for long periods, for example between the years 2001-2008, thus implying a very limited variability in the parameters, and it suddenly increases to accelerate the variability in the parameters as in the early 1970's or at the end of the sample. Figure 3 reports the estimates of the prediction error variance and of δ_t obtained with SSP-KF and SSP-KF-RV relative to model specifications V and VI. First, it emerges that $\hat{\mathrm{H}}_t$ and $\hat{\mathrm{H}}_t^*$ are on a similar scale and they follow similar patterns, especially from the mid 1980's to the early 2000's. Interestingly, $\hat{\mathrm{H}}_t^*$ sharply increases in 2009 reaching

[5]The values for the grids are calibrated on the basis of the results of preliminary estimates. Increasing the size of the grid does not lead to significant changes in the parameter dynamics nor in the fitting. Appendix A provides additional details on DMS and dynamic model averaging (DMA) when jointly combining the grid of ς and κ with all possible combinations of the regressors.

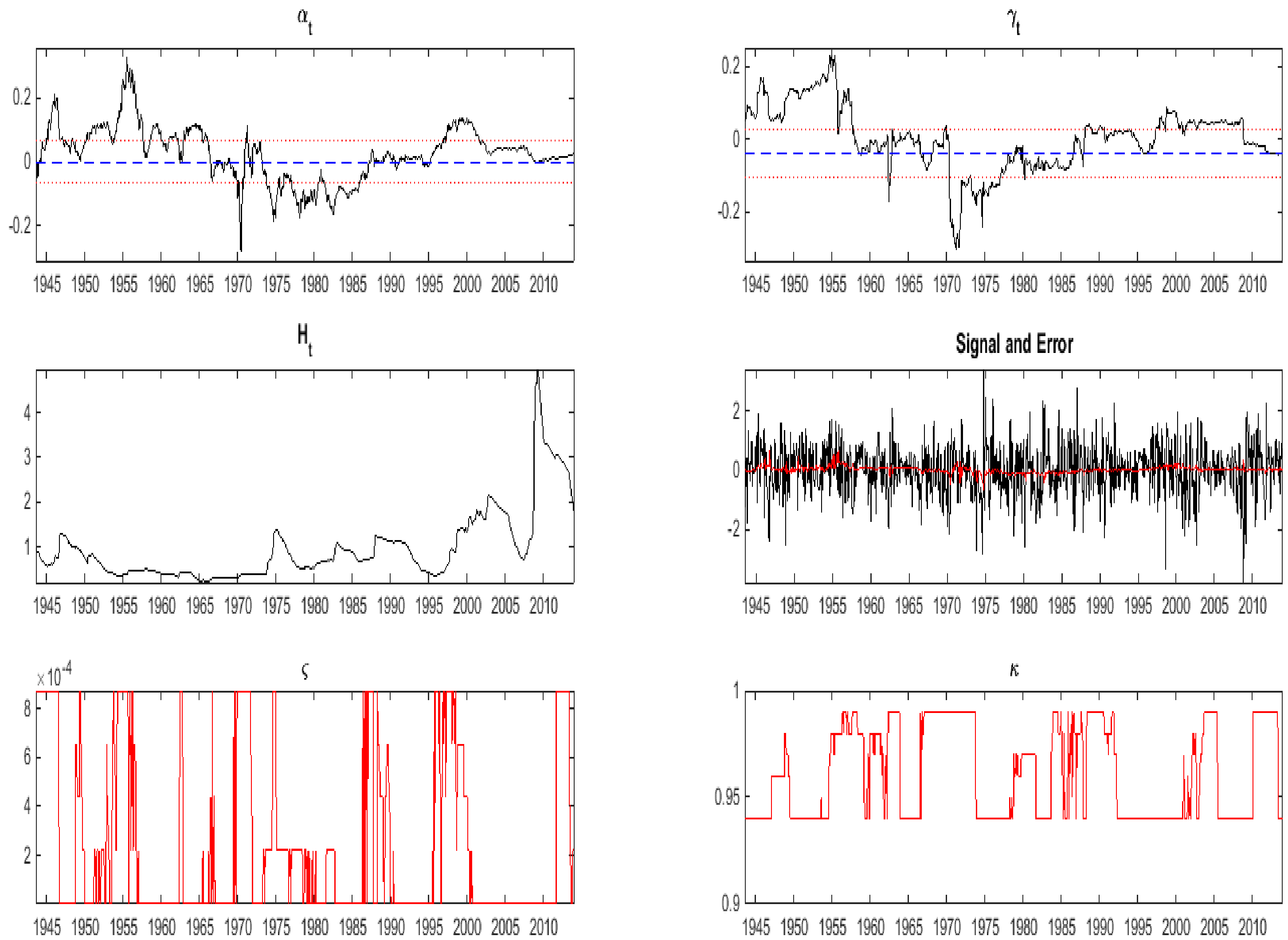
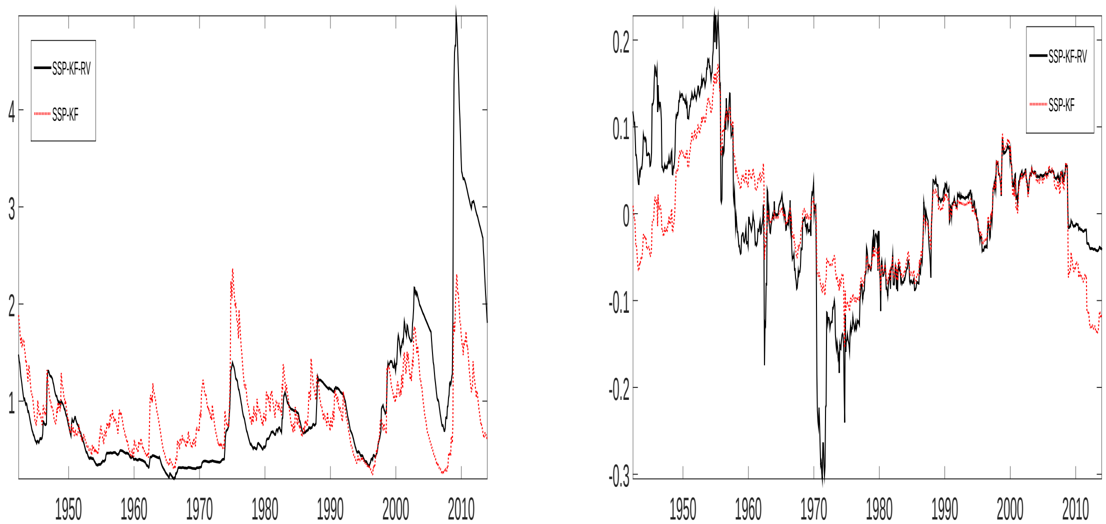


Figure 2: Parameter estimates for the model specification VI. The top panels of the figure report the estimate of the intercept (left) and of the parameter δ_t (right) together with the corresponding OLS estimates based on the full sample and their 95% confidence intervals. The central panels report the estimates of H_t (left) and the predicted returns together with the ex-post realized monthly excess returns (right). The bottom panels contain the selected values of ς (left) and κ (right).

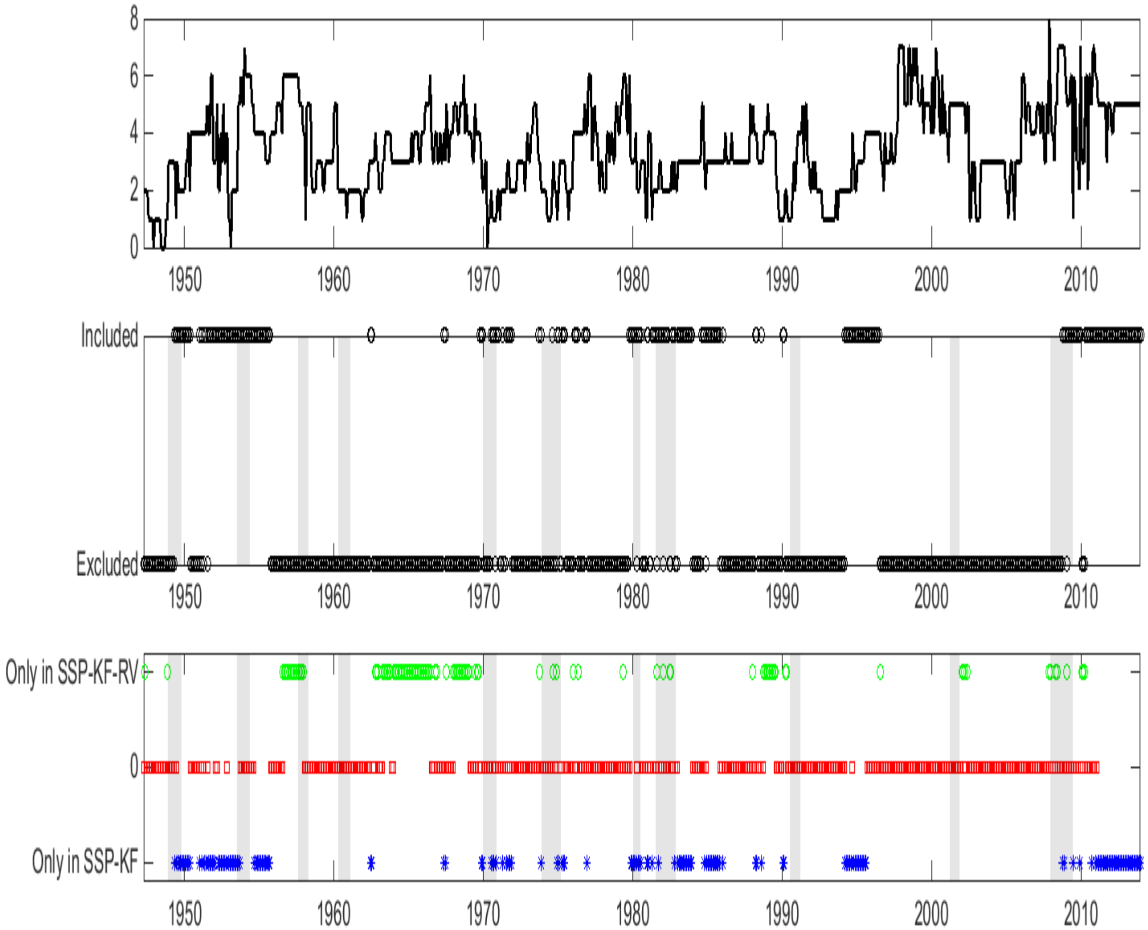
Figure 3: Figures report the different estimates of the prediction error variance, $\hat{H}_t$, and of the parameter δ_t, obtained under SSP-KF (in dotted-red line) and SSP-KF-RV (in solid-black line).



(a) Estimated H_t and H_t^* (b) Estimated δ_t by SSP-KF

abnormal levels, while the growth of $\hat{\mathrm{H}}_t$ after 2009 is much more limited. As a consequence, the size of the break of δ_t after 2009 is much more limited for the SSP-KF-RV model since large values of $\hat{\mathrm{H}}_t^*$ are associated with a lower parameter variability, through the parameter perturbation mechanism $\frac{\nu_t^2}{\hat{\mathrm{H}}_t^*} - 1$, that is most likely smaller than 0. On the other hand, when H_t is used in the SSP-KF the variation in δ_t is more pronounced after 2009. The top panel of Figure 4 reports the number of selected regressors by the DMS method at each point in time, relative to the model specification XII. In most cases, a number between 2 and 6 explanatory variables is selected by DMS, meaning that the size of the model is never too large. This should help avoiding the over-fitting problem thus potentially increasing the out-of-sample predictability, see Sections 4.1.1 and 4.1.2. For what concerns the inclusion probability of $\widetilde{RV}_{t-1}$, the latter belongs to the best model specification in 31% of the cases when the SSP-KF is adopted, and in 23% of the cases under SSP-KF-RV. The central panel of Figure 4 displays the periods in which $\widetilde{RV}_{t-1}$ is included/excluded from the best model specification for model X. In general, $\widetilde{RV}_{t-1}$ tends to be a relevant explanatory variable right after the financial crises or the recession periods, especially after the oil crisis in early 1970's and after the 2008-2009 crisis. $\widetilde{RV}_{t-1}$ is also included for a long period in the early 1980's, that is a recession phase. In other words, during financial crises, past RV has a non-linear effect on the future excess returns through the conditional variance of r_t^*, but not a linear impact in the conditional mean. This is analogous to the findings of Jensen and Maheu (2013). The bottom panel of Figure 4 signals if there is any difference in the inclusion of $\widetilde{RV}_{t-1}$ among the relevant regressors when the estimation is performed with SSP-KF or with SSP-KF-RV. The red squares imply coherence

Figure 4: Inclusion Probabilities. The top panel reports the number of selected variables in the best model specification at each point in time relative to the case XII. The central panel reports the inclusion/exclusion periods of $\widetilde{RV}_{t-1}$ in the best model implied by specification X. The bottom panel reports the difference in the inclusion of RV_{t-1} in the best model between specification X and specification XII. The red squares are the months in which $\widetilde{RV}_{t-1}$ is included/excluded in both cases. The green dots are the months in which $\widetilde{RV}_{t-1}$ is included in model XII but not in model X. The blue star are the months in which $\widetilde{RV}_{t-1}$ is included in model X but not in model XII. The gray areas are the recessions periods from NBER.



in the inclusion/exclusion of $\widetilde{RV}_{t-1}$ at time t under SSP-KF and SSP-KF-RV. Notably, there is accordance on the inclusion/exclusion of $\widetilde{RV}_{t-1}$ in 64% of the cases. In the remaining 36% of the cases, the indications on the inclusions of $\widetilde{RV}_{t-1}$ in the best model for SSP-KF and SSP-KF-RV are not coherent (green and blue dots). It is not simple to find a pattern in the discrepancies between the inclusions of $\widetilde{RV}_{t-1}$ obtained under SSP-KF and SSP-KF-RV. However, if we focus on the most recent financial crisis in the years 2008-2009, where we also observe the largest discrepancies in $\hat{\mathrm{H}}_t$ and $\hat{H}_t^*$, it emerges that $\widetilde{RV}_{t-1}$ is only included in the specification that adopts the SSP-KF, while

$\widetilde{RV}_{t-1}$ is included in the model under SSP-KF-RV just in the at the beginning of 2010.

Finally, in the same spirit of Dangl and Halling (2012), we perform a variance decomposition to disentangle all the sources of uncertainty in the excess returns implied by a given model specification. Compared to the decomposition in Dangl and Halling (2012), we integrate out the uncertainty on the hyper-parameters ς and κ as done in Koop and Korobilis (2013), so the model uncertainty depends only on the choice of the relevant regressors. Collecting the hyper-parameters selected by DMS at time t in the vector $\zeta_{t,DMS} = (\varsigma_{t,DMS}, \kappa_{t,DMS})$, then the variance decomposition is

$$\begin{aligned}\text{Var}(r^*_{t+1}) = &\sum_{i=1}^{I} p(\text{H}_t|\text{M}_i, \zeta_{t,DMS}, \mathcal{F}_{t-1}) \cdot p(\text{M}_i|\zeta_{t,DMS}, \mathcal{F}_{t-1}) \\ &+ \sum_{i=1}^{I} p(\tilde{X}_t' \text{P}_{t|t-1} \tilde{X}_t|\text{M}_i, \zeta_{t,DMS}, \mathcal{F}_{t-1}) \cdot p(\text{M}_i|\zeta_{t,DMS}, \mathcal{F}_{t-1}) \\ &+ \sum_{i=1}^{I} p(\hat{r}^{*,DMS}_{t+1,i} - \hat{r}^{*,DMS}_{t+1})^2 \cdot p(\text{M}_i|\zeta_{t,DMS}, \mathcal{F}_{t-1}),\end{aligned} \tag{16}$$

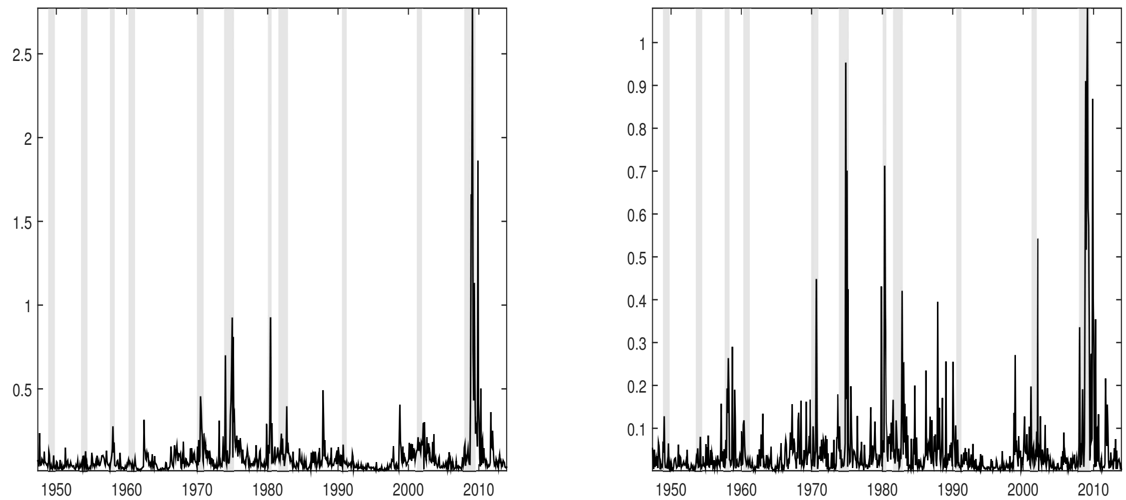
where $\mathcal{F}_{t-1}$ defined the information set at $t-1$, $I = 2^{12} = 4,096$ is the number of potential models considered and $\text{M}_i, i = 1, \ldots, I$ indicates the i-th model. The first term is the average expected variance, $\hat{\text{H}}_t$, with respect to the i-th model. The second term is the average expected variance from errors in the estimation of the coefficient vector, i.e. the estimation uncertainty. The last term is related to the model's uncertainty. Figure 5 displays the dynamics of the second and third components of the variance decomposition related to the model specification XII.[6] Interestingly, both components, i.e. the one related to the estimation uncertainty and the one related to the uncertainty about the model, increase during all recession periods starting already from the 1970's. This not only means that it is relatively more difficult to conduct a precise inference on the parameters when the volatility is high, i.e. during financial crisis or recessions, but also that it becomes more difficult to precisely select the relevant regressors. In the following sections, we evaluate the ability of each model specification in predicting excess returns from a statistical and a financial point of view.

4.1.1 Statistical Evaluation

We firstly focus on the point forecasts. Table 4 reports a comparison of the ability of each model specification to provide good point forecasts of the excess-returns. We focus on the accuracy of the point forecast, as measured by the mean-squared-prediction-error, MSPE, relative to the model with constant intercept (i.e. model I). It emerges from Table 4 that most specifications have performances in terms of point-forecast that are non-statistically superior to the simplest model with constant mean and variance. Some specifications, e.g. VII and $VIII$, even under-perform compared to model I. This is not fully surprising as the predictability of the equity premium is known to be

[6] A plot with the first variance component is also available. The dynamics of the first component are very close to those of $\hat{\text{H}}_t$ and $\hat{\text{H}}^*_t$, that are reported in Figure 3.

Figure 5: Figures report the second and third components of the return variance, obtained by the decomposition in 16 for model XII. Panel a) reports the dynamics of the second component, $\sum_{i=1}^{I} p(\tilde{X}_t' P_{t|t-1} \tilde{X}_t | M_i, \zeta_{t,DMS}, \mathcal{F}_{t-1}) p(M_i|\zeta_{t,DMS}, \mathcal{F}_{t-1})$, that is related to estimation uncertainty. Panel c) reports the dynamics of the third component, $\sum_{i=1}^{I} p(\hat{r}_{t+1,i}^{*,DMS} - \hat{r}_{t+1}^{*,DMS})^2 p(M_i|\zeta_{t,DMS}, \mathcal{F}_{t-1})$, that is related to the model's uncertainty. The gray areas are the recessions periods from NBER.



(a) Second Component

(b) Third Component

limited, see among many others Welch and Goyal (2008). On the other hand, when accounting for model uncertainty, the constant drift specification tends to be significantly outperformed. In particular, when DMS is used to select among all regressors at each point in time, the difference in the point prediction turns out to be positive and strongly statistically significant. This means that for a correct characterization of the predictability in the excess returns it is not only necessary to allow the parameters governing $\mathrm{E}(r_t^*|\mathcal{F}_{t-1})$ and $\mathrm{Var}(r_t^*|\mathcal{F}_{t-1})$ to vary over time, but it also required to select the relevant explanatory variables to avoid over-fitting.

The analysis of the quality of the forecasts can also be done by looking at the ability of each specification to provide a good description of the conditional density of the monthly excess returns. In this case, we are interested in the empirical fitting of the entire excess return distribution as well as parts of it. For example, the ability of a model to assign the right probability to tail events may be exploited for risk management purposes. In order to evaluate the quality of the predictive density of returns, we consider the method introduced by Berkowitz (2001), which allows to test for the adequacy of the proposed conditional density with the realization of the modeled variable. The test is flexible and can be applied to evaluate the fit of the entire density as well as over specific segments of the density support. In details, given the density of r_t^*, we compute the conditional CDF of r_t^* as

$$y_t = F\left(r_t^*|\mathcal{F}_{t-1}\right) = \int_0^{r_t^*} f\left(x|\mathcal{F}_{t-1}\right) \mathrm{d}x,$$

Table 4: Relative MSPE. The table contains the differences in MSPEs, Δ, (multiplied by 100) between the Model I benchmark and the other models. The table also provide the value of the one-sided test that the difference is greater than zero. In bold, significance at 5% level.

	1947+		**1965+**		**1976+**		**1988+**		**2000+**		**Expansions**		**Recessions**	
	Δ	test	Δ	test	Δ	test	Δ	test	Δ	test	Δ	test	Δ	test
Model II	0.00	0.12	0.00	0.13	0.00	0.15	0.00	0.19	0.00	0.26	0.00	0.18	0.00	-0.07
Model III	0.00	-1.13	0.00	-1.26	0.00	-1.25	0.00	-1.24	0.00	-2.16	0.00	-1.94	0.00	0.57
Model IV	0.00	0.04	0.00	0.23	0.01	0.44	-0.01	-2.56	-0.01	-1.49	0.00	0.16	-0.01	-1.46
Model V	0.00	0.00	0.00	0.15	0.00	0.40	-0.01	-3.68	-0.01	-4.21	0.00	0.12	-0.01	-1.74
Model VI	0.00	-0.24	0.00	-0.06	0.00	0.24	-0.01	-2.89	-0.01	-2.41	0.00	-0.02	-0.01	-1.47
Model VII	-0.01	-1.27	-0.01	-0.97	-0.01	-0.44	-0.02	-1.43	-0.03	-1.20	-0.01	-0.83	-0.04	-1.92
Model VIII	-0.01	-1.65	-0.02	-1.38	-0.01	-0.43	-0.02	-1.83	-0.03	-1.53	-0.01	-0.89	-0.06	-2.12
Model IX	0.01	0.89	0.01	0.91	0.01	0.99	0.00	0.04	0.00	0.15	0.01	0.88	0.00	0.05
Model X	0.03	**3.92**	0.03	**3.34**	0.04	**2.71**	0.02	**2.89**	0.03	**2.25**	0.03	**3.18**	0.05	**2.93**
Model XI	0.01	0.78	0.01	0.76	0.01	1.01	0.00	-0.24	0.00	-0.07	0.01	0.89	0.00	-0.21
Model XII	0.03	**3.77**	0.04	**3.13**	0.04	**2.54**	0.02	**1.62**	0.02	1.00	0.03	**3.21**	0.05	**2.61**
Rolling	-0.03	-3.15	-0.03	-2.53	-0.03	-1.78	-0.04	-2.55	-0.06	-2.30	-0.02	-2.12	-0.08	-2.40
KF	-0.03	-3.08	-0.03	-2.52	-0.03	-1.77	-0.04	-2.54	-0.06	-2.29	-0.02	-2.02	-0.08	-2.33

where $F(r_t^*|\mathcal{F}_{t-1})$ is Gaussian with $\mathrm{E}(r_t^*|\mathcal{F}_{t-1})$ and $\mathrm{Var}(r_t^*|\mathcal{F}_{t-1})$ dependent on the specific model specification at hands. Under correct model specification, the empirical CDF values should be distributed according to the standard uniform, i.e. $y_t \sim U(0,1)$, which are further transformed as

$$z_t = \Phi^{-1}(y_t)$$

where $\Phi(\cdot)$ is the standard normal CDF, so that z_t are distributed as a standardized normal. To test the correct coverage for each quantiles, q, we calculate a new truncated variable

$$z_t^* = \begin{cases} z_t & \text{if} \quad z_t \leq \mathrm{q} \\ \mathrm{q} & \text{if} \quad z_t > \mathrm{q}. \end{cases} \tag{17}$$

For example, if we are interested in the coverage of left tail, the quantile q corresponding to the $P_q = 1\%$ probability level is q=-2.326. A tail coverage test can be derived using the LR principle. Under the null, the mean and the variance of z_t^* are 0 and 1, respectively, while under the alternative they are unrestricted. Under the null of correct tail coverage the test statistic is distributed as $\chi^2(2)$. See Berkowitz (2001) for further details on this test.

Table 5 reports the p-values of the Berkowitz test of the alternative model specifications for different quantiles. The first evidence that emerges is that simple specifications with constant parameters, i.e. model I and II, are unable to provide a good fit of the distribution of the returns for any of the quantiles selected. This is somehow expected, as it is well known that the distribution of the returns is likely to vary over time. Indeed, when allowing the parameters in the mean and variance to vary over time (models III and IV), the fit improves significantly, especially for the left tail (P_q=1%,5%). However, when looking at the fit of almost the entire distribution, i.e. P_q=99%, the p-values are below 10%, meaning that the fitting is not perfect. Interestingly, when all the covariates in X_t are used to predict the excess returns, the fit is extremely poor (models V and VI). This is a direct consequence of the over-fitting problem and of the spurious variation induced in all parameters. On

Table 5: Berkowitz test. The Table reports the p-values *of the Berkowitz (2001) test for probability levels,* $P_q = Pr(z_t \leq q)$*, associated to different quantiles,* q*. In bold p-values greater than 10%.*

Model↓ P_q →	**1%**	**5%**	**15%**	**25%**	**35%**	**45%**	**55%**	**65%**	**75%**	**85%**	**95%**	**99%**
I	0.01	0.03	0.00	0.00	0.01	0.01	**0.13**	**0.15**	0.07	0.00	0.00	0.00
II	0.02	0.02	0.01	0.00	0.02	0.01	**0.12**	**0.13**	0.05	0.00	0.00	0.00
III	**0.98**	**0.81**	**1.00**	**0.18**	0.06	**0.20**	**0.25**	**0.53**	**0.52**	**0.38**	0.09	0.05
IV	**0.18**	**0.20**	**0.44**	**0.37**	**0.31**	**0.28**	**0.39**	**0.42**	**0.41**	**0.14**	0.04	0.03
V	**0.21**	**0.14**	0.08	0.08	0.05	0.05	0.05	0.03	0.01	0.01	0.00	0.00
VI	**0.50**	**0.55**	**0.34**	**0.23**	0.08	0.04	**0.32**	**0.77**	**0.79**	**0.43**	**0.53**	**0.59**
VII	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.09	0.09	**0.34**	**0.53**	**0.27**
VIII	0.06	0.04	0.01	0.03	0.01	0.02	0.09	**0.24**	**0.32**	**0.62**	**0.79**	**0.56**
IX	0.03	0.18	**0.15**	**0.10**	0.07	0.06	0.08	0.05	0.02	0.00	0.00	0.00
X	**0.88**	**0.89**	**0.77**	**0.55**	**0.36**	**0.17**	**0.11**	**0.36**	**0.53**	**0.83**	**0.96**	**0.88**
XI	**0.16**	**0.95**	**0.58**	**0.15**	**0.27**	0.09	**0.27**	**0.52**	**0.45**	**0.16**	**0.11**	**0.16**
XII	**0.28**	**0.29**	**0.25**	0.05	0.03	0.02	0.02	**0.25**	**0.73**	**0.77**	**0.88**	**0.77**
Rolling	0.02	0.01	0.00	0.00	0.05	0.03	0.04	0.02	0.043	0.00	0.00	0.00
KF	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.06	0.00	0.00	0.00	0.00

the other hand, when selecting the optimal model via DMA or DMS, either with the SSP-KF or the SSP-KF-RV, the fitting is good for most quantiles. In particular, when SSP-KF and DMS are jointly used, the p-values of the Berkowitz test are above 10% for all quantiles.

4.1.2 Financial Evaluation

In the last section, we concentrated on the ability of TVP models to provide significant improvements over models with constant parameters in predicting the equity premium and its distribution. The most important result that arises from the statistical analysis is that allowing for time-varying parameters and selecting the best model specification at each point in time are both essential for a good statistical characterization of excess returns. This is in line with the results of Pettenuzzo and Timmermann (2011) and Dangl and Halling (2012).

We now study how an investor with mean-variance utility function can gain from the use of RV in predicting returns. Specifically, we think of an investor that learns about the models, the parameters, and the state variables sequentially in real time and updates his expectations about the future expected equity premium through the updating algorithm embedded in the SSP-KF algorithm. In particular, given a model specification, the investor is able to compute $\mathrm{E}(r^*_{t+1}|\mathcal{F}_t)$ and $\mathrm{Var}(r^*_{t+1}|\mathcal{F}_t)$ at time t. Given the conditional moments, the investor can choose how much of his wealth to allocate to the risk-free asset and how much to allocate to the risky asset by maximizing the expected utility, $\mathrm{E}(U_{t+1}|\mathcal{F}_t) = \mathrm{E}(R_{t+1}|\mathcal{F}_t) - \psi/2 \cdot \mathrm{Var}(R_{t+1}|\mathcal{F}_t)$ with $\psi = 4$. The term $R_{t+1} = \omega_{t+1|t} \cdot r^f_{t|t+1} + (1 - \omega_{t+1|t}) \cdot r_{t+1}$ with $\omega_{t+1|t} \in [0, 1]$ is the return on a portfolio with a risky asset (the S&P500 index) and a risk-free bond, whose return for period $[t, t+1]$ is known and equal to $r^f_{t|t+1}$. The assumption that $\omega_{t+1|t} \in [0, 1]$ rules out short selling. At the end of each period, the investor realizes gains and losses, updates the parameter and model estimates and computes new portfolio weights

Table 6: Dynamic Asset Allocation. Table reports the average certainty equivalent returns (CER) that is the annualized risk-free return that gives the investor the same utility as the portfolio with the risky asset, based on the ex-post realization of the returns and variance of the portfolio. Table also reports the average Sharpe ratios, SR. In bold the highest value for each column.

	1947+		**1965+**		**1976+**		**1988+**		**2000+**		**Expansions**		**Recessions**	
	CER	SR	CER	SR	CER	SR	CER	SR	CER	SR	CER	SR	CER	SR
Model I	0.42	0.02	0.05	0.01	0.32	0.02	0.55	0.04	-0.19	0.04	1.13	0.03	-3.35	0.00
Model II	0.53	0.09	0.28	0.06	0.51	0.09	0.85	0.13	0.33	0.09	0.99	0.14	-1.99	-0.13
Model III	0.60	0.10	-0.26	0.05	0.21	0.09	0.34	0.11	-0.99	0.03	1.30	0.14	-3.53	-0.16
Model IV	0.61	0.10	-0.19	0.05	0.30	0.09	0.42	0.11	-0.86	0.04	1.41	0.14	-4.27	-0.18
Model V	0.39	0.08	-0.54	0.02	0.37	0.08	0.76	0.11	-0.72	-0.04	0.96	0.11	-2.81	-0.10
Model VI	0.33	0.08	-0.67	0.01	0.11	0.06	0.34	0.09	-0.99	-0.04	0.85	0.10	-2.77	-0.07
Model VII	1.89	0.15	0.83	0.12	1.25	0.15	0.75	0.14	0.38	0.14	2.54	0.19	-3.74	-0.06
Model VIII	1.25	0.12	-0.20	0.06	0.68	0.11	0.32	0.10	-0.44	0.07	2.27	0.16	-6.11	-0.09
Model IX	2.46	0.18	1.65	0.15	2.33	0.19	2.10	0.19	1.28	0.16	3.20	0.22	-2.85	-0.04
Model X	6.74	**0.35**	5.48	**0.31**	6.07	**0.34**	5.62	0.34	4.70	**0.36**	7.17	0.36	2.89	0.20
Model XI	2.34	0.16	1.31	0.12	2.18	0.17	1.61	0.15	0.70	0.11	3.18	0.20	-3.47	-0.05
Model XII	**7.24**	**0.35**	**6.06**	**0.31**	**6.68**	**0.34**	**6.04**	**0.34**	**4.72**	0.34	**7.83**	**0.37**	**3.37**	**0.21**
Rolling	-0.66	0.08	-1.56	0.06	-0.69	0.11	-0.54	0.13	-2.66	0.07	1.35	0.15	-10.93	-0.19
KF	-0.57	0.08	-1.54	0.06	-0.66	0.11	-0.50	0.13	-2.59	0.08	1.47	0.15	-10.40	-0.18

$\omega_{t+2|t+1}$. This procedure is repeated for each time period, generating a time series of out-of-sample realized returns and variances of the portfolio. We follow Dangl and Halling (2012) and we use the monthly RV, based on daily S&P500 returns as an ex-post estimate of the total variance over monthly horizons. Given the time series of realized returns and variances, then standard summary statistics such as certainty equivalent returns (CER) and Sharpe ratios are computed to summarize the portfolio performance. Table 6 reports the results of the optimal portfolio allocation analysis. The reported evidence strongly support the specifications that involve model selection among all regressors. Interestingly, the average CER remains positive in all sub-periods for models X and XII, and the highest average CER is always associated with model XII. This results support the idea that exploiting the information on past RV in both the conditional mean and variance of the excess returns leads to utility gains for a risk averse investor. Notably, the CER associated with the other model specifications are quite low and sometimes negative, especially after 2000 and during the recession periods. An analogous evidence arises by looking at the Sharpe ratios (SR), whose highest values are generally associated with the specifications X and XII. Differently from Dangl and Halling (2012), we find that the utility gain for models X and XII does not increase during the recessions, although it remains positive as opposed to the other model specifications. This difference is probably due to the fact that the period with the financial crisis 2008-2009 is included in our sample but not in the sample of Dangl and Halling (2012). Since we are also ruling out the possibility of short-selling, it turns out to be hard to generate very large returns during recessions. The fact that portfolios based on models X and XII can still generate positive CER and Sharpe ratios also during recessions is a very strong evidence in favor of combining DMS with the SPP-KF approach to predict excess returns and to provide the correct buy and sell signals.

5 Conclusion

This paper introduces a novel methodology to estimate TVP models in economics and finance, namely the standardized self-perturbed Kalman filter, that extends the method proposed by Park and Jun (1992). In the standardzied self-perturbed Kalman filter, the measurement error variance enters directly in the updating step, so that the activation of the updating process of the parameters becomes endogenously determined by the amount of uncertainty in the data. This method has the advantage, over the traditional Kalman filter of being computationally very fast, thus making it an useful tool in frameworks characterized by model uncertainty where the correct specification must be chosen among a large number of alternatives. A Monte Carlo study shows that the efficiency loss of the SSP-KF in tracking the true parameters variation is generally small compared to the traditional methods when the design parameters, ς and κ, are optimally selected by DMS. Concluding, the standardized self-perturbed Kalman filter proves to be a valid alternative to online methods based on forgetting factors. We believe that the relative advantage of this method compared to traditional methods increases when the model at hand is extended to the multivariate case and hundreds of variables are jointly modeled, see also Koop and Korobilis (2013). An extension of the standardized self-perturbed Kalman filter to the multivariate case, possibly adapting the perturbation term to account for spillover effects between equations and different perturbation speeds in each equation, is a topic of future research. The SSP-KF is used to forecast the monthly equity premium series of the $S\&P$ 500 index from 1937 to 2013, with the purpose of studying how the realized variance can be exploited both in the conditional mean and in the conditional variance. The SSP-KF allows to precisely extract the variation in the parameters and, hence, to provide the right signals for the optimal selection of the relevant explanatory variables. We show that accounting for model uncertainty and time-variation in the model parameters leads to utility gains for an investor, especially when the realized variance is used as a driver of the time-varying measurement error variance.

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

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20:651–707.

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

Berkowitz, J. (2001). The accuracy of density forecasts in risk management. *Journal of Business and Economic Statistics*, 19:465–474.

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

Bollerslev, T., Litvinova, J., and Tauchen, G. (2006). Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 4:353–384.

Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium. *Canadian Journal of Economics/Revue canadienne d'conomique*, 41:1–21.

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, 21:1533–1575.

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.

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 analysis of optimal systems. *IEEE International Convention Record Part*, pages 216–240.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49:639–654.

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

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

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics*, 99:560–580.

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

Jensen, M. J. and Maheu, J. M. (2013). Risk, Return and Volatility Feedback: A Bayesian Nonparametric Analysis. MPRA paper, University Library of Munich, Germany.

Johannes, M., Kortweg, A., and Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69:611–644.

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

Koop G., and Korobilis, D. (2010). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *in Foundations and Trends in Econometrics*, 3:267–358.

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

Koop, G. and Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177:185–198.

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

Lettau, M. and Ludvingson, S. (2010). Measuring and modeling variation in the risk-return tradeoff. In Ait-Shalia, Y. and Hansen, L.-P., editors, *Handbook of Financial Econometrics*, pages 617–690. North Holland.

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

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17:59–82.

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

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

Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. (2006). Forecasting Time Series Subject to Multiple Structural Breaks. *Review of Economic Studies*, 73:1057–1084.

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

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.

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

Timmermann, A. (2008). Elusive return predictability. *International Journal of Forecasting*, 24:1–18.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21:1455–1508.

A Model Averaging and Model Selection

One of the advantages of the on-line Kalman filter is the possibility to carry out the dynamic model averaging (DMA) and dynamic model selection (DMS) in a computationally feasible way. Define $\mathrm{L}_t \in \{1, 2, \dots, K\}$ the set of possible models at each point in time t, given by $K = dim(\varsigma) \times dim(\kappa) \times 2^m$. Where ς and κ are the design parameter discussed in the paper and m is the number of explanatory variables considered. Since the model can change over time, then the set of possible models is $G = K^T$ where T is the number of observations. Define $\mathrm{Y}_T = \{y_1, \dots, y_t\}$ the information set, then the state space form can be written as follows:

$$\begin{aligned} y_t &= \mathrm{Z}_t^{(k)} \theta_t^{(k)} + \varepsilon_t^{(k)}, & \varepsilon_t^{(k)} &\sim \mathrm{N}\left(0, \mathrm{H}_t^{(k)}\right), \\ \theta_{t+1}^{(k)} &= \theta_t^{(k)} + \eta_t^{(k)}, & \eta_t^{(k)} &\sim \mathrm{N}\left(0, \mathrm{Q}_t^{(k)}\right), \end{aligned} \tag{18}$$

where $k = 1, \dots, K$ indicates each possible model specification at time t, such that a different set of predictors and design parameters is associated with each k. The SSP-KF for the k-th model becomes:

$$\theta_{t|t}^{(k)} = \theta_{t|t-1}^{(k)} + \mathrm{P}_{t|t-1}^{(k)} \mathrm{Z}_t^{(k)\prime} \left(\hat{\mathrm{H}}_t^{(k)} + \mathrm{Z}_t^{(k)} \mathrm{P}_{t|t-1}^{(k)} \mathrm{Z}_t^{(k)\prime}\right)^{-1} \nu_t^{(k)} \tag{19}$$

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

$$\hat{\mathrm{H}}_t^{(k)} = \kappa^{(k)} \hat{\mathrm{H}}_{t-1}^{(k)} + \left(1 - \kappa^{(k)}\right) \nu_t^{2,(k)}. \tag{20}$$

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

$$p\left(\Theta_{t-1|t-1} \mid \mathrm{Y}_{t-1}\right) = \sum_{k=1}^{K} p\left(\theta_{t-1|t-1}^{(k)} \mid \mathrm{L}_{t-1} = k, \mathrm{Y}_{t-1}\right) p\left(\mathrm{L}_{t-1} = k \mid \mathrm{Y}_{t-1}\right), \tag{21}$$

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

$$\Theta_{t-1|t-1} \mid \mathrm{L}_{t-1} = k, \mathrm{Y}_{t-1} \sim \mathrm{N}(\theta_{t-1|t-1}^{(k)}, \mathrm{P}_{t-1|t-1}^{(k)}), \tag{22}$$

and $p(\mathrm{L}_{t-1} = k \mid Y_{t-1})$ is the probability to be at model k at time $t-1$. The predictive likelihood for model k given by

$$p^{(k)}(y_t \mid \mathrm{Y}_{t-1}) \sim \mathrm{N}(\mathrm{Z}_t^{(k)} \theta_{t|t-1}^{(k)}, \hat{\mathrm{H}}_t^{(k)} + \mathrm{Z}_t^{(k)} \mathrm{P}_{t|t-1}^{(k)} \mathrm{Z}_t^{(k),\prime}). \tag{23}$$

Using the same approximation as in Raftery et al. (2010) and Koop and Korobilis (2012), we assume that the probability $\pi_{t|t-1,k}$ that the k-th combination of ς, κ and the explanatory variables is used to forecast y_t, given information through time $t-1$, is

$$\pi_{t|t-1,k} \quad = \quad \frac{\pi^{\alpha}_{t-1|t-1,k}}{\sum_{k=1}^{K} \pi^{\alpha}_{t-1|t-1,k}}, \tag{24}$$

where $0 < \alpha \leq 1$ is set to a fixed value slightly less than one and is interpreted as a smoothing factor. The updating equation of (24) is then given by:

$$\pi_{t|t,k} \quad = \quad \frac{\pi_{t|t-1,k} p^{(k)} \left(y_t \mid \mathrm{Y}_{t-1}\right)}{\sum_{k=1}^{K} \pi_{t|t-1,k} p^{(k)} \left(y_t \mid \mathrm{Y}_{t-1}\right)}. \tag{25}$$

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

$$p\left(y_t \mid \mathrm{Y}_{t-1}\right) \quad = \quad \sum_{k=1}^{K} p^{(k)} \left(y_t \mid \mathrm{Y}_{t-1}\right) \pi_{t|t-1,k}. \tag{26}$$

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

$$\mathrm{E}\left[y_t \mid \mathrm{Y}_{t-1}\right] \quad = \quad \sum_{k=1}^{K} \mathrm{Z}_t^{(k)} \theta_{t|t-1}^{(k)} \pi_{t|t-1,k}. \tag{27}$$

On the contrary, DMS requires the selection of the single model with the highest probability value at each point in time. Koop and Korobilis (2012) find that both DMA and DMS forecast inflation very well.

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

1. In $t = 0$, initialize the inclusion probabilities to $\pi_{0|0,k} = 1/2^m \ \forall k$ and the design parameters $\varsigma = 0.00001$ and $\kappa = 0.94$. We set $\theta_{0|0} = 0$ and $P_{0|0} = 100 \times I_m$.

2. At time $t \geq 1$, run the predicting steps of the SSP-KF for each model.

3. At the end of the period t, y_t is observed. Hence run the updating steps of the SSP-KF and use equation (23) to compute the predictive likelihood for each model k.

4. Use equation (25) to compute the updated inclusion probabilities for each combination of ς, κ and the included regressors. In the case of DMA, produce DMA forecasts using (26) and (27). In the case of DMS, use the forecasts based on the best performing model, i.e. the one with the highest model probaility.

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

Research Papers 2013

CREATES
Center for Research in Econometric Analysis of Time Series

2013-46: Peter Christoffersen, Kris Jacobs, Xisong Jin and Hugues Langlois: Dynamic Diversification in Corporate Credit

2013-47: Peter Christoffersen, Mathieu Fournier and Kris Jacobs: The Factor Structure in Equity Options

2013-48: Peter Christoffersen, Ruslan Goyenko, Kris Jacobs and Mehdi Karoui: Illiquidity Premia in the Equity Options Market

2013-49: Peter Christoffersen, Vihang R. Errunza, Kris Jacobs and Xisong Jin: Correlation Dynamics and International Diversification Benefits

2013-50: Georgios Effraimidis and Christian M. Dahl: Nonparametric Estimation of Cumulative Incidence Functions for Competing Risks Data with Missing Cause of Failure

2013-51: Mehmet Caner and Anders Bredahl Kock: Oracle Inequalities for Convex Loss Functions with Non-Linear Targets

2013-52: Torben G. Andersen, Oleg Bondarenko, Viktor Todorov and George Tauchen: The Fine Structure of Equity-Index Option Dynamics

2014-01 Manuel Lukas and Eric Hillebrand: Bagging Weak Predictors

2014-02: Barbara Annicchiarico, Anna Rita Bennato and Emilio Zanetti Chini: 150 Years of Italian CO2 Emissions and Economic Growth

2014-03: Paul Catani, Timo Teräsvirta and Meiqun Yin: A Lagrange Multiplier Test for Testing the Adequacy of the Constant Conditional Correlation GARCH Model

2014-04: Timo Teräsvirta and Yukai Yang: Linearity and Misspecification Tests for Vector Smooth Transition Regression Models

2014-05: Kris Boudt, Sébastien Laurent, Asger Lunde and Rogier Quaedvlieg: Positive Semidefinite Integrated Covariance Estimation, Factorizations and Asynchronicity

2014-06: Debopam Bhattacharya, Shin Kanaya and Margaret Stevens: Are University Admissions Academically Fair?

2014-07: Markku Lanne and Jani Luoto: Noncausal Bayesian Vector Autoregression

2014-08: Timo Teräsvirta and Yukai Yang: Specification, Estimation and Evaluation of Vector Smooth Transition Autoregressive Models with Applications

2014-09: A.S. Hurn, Annastiina Silvennoinen and Timo Teräsvirta: A Smooth Transition Logit Model of the Effects of Deregulation in the Electricity Market

2014-10: Marcelo Fernandes and Cristina M. Scherrer: Price discovery in dual-class shares across multiple markets

2014-11: Yukai Yang: Testing Constancy of the Error Covariance Matrix in Vector Models against Parametric Alternatives using a Spectral Decomposition

2014-12: Stefano Grassi, Nima Nonejad and Paolo Santucci de Magistris: Forecasting with the Standardized Self-Perturbed Kalman Filter