BANK OF ENGLAND

Staff Working Paper No. 577

Adaptive models and heavy tails

Davide Delle Monache and Ivan Petrella

January 2016

BANK OF ENGLAND

Staff Working Paper No. 577

Adaptive models and heavy tails

Davide Delle Monache[(1)] and Ivan Petrella[(2)]

Abstract

This paper introduces an adaptive algorithm for time-varying autoregressive models in presence of heavy tails. The evolution of the parameters is driven by the score of the conditional distribution. The resulting model is observation-driven and is estimated by classical methods. Meaningful restrictions are imposed on the model parameters, so as to attain local stationarity and bounded mean values. In particular, we consider time variation in both coefficients and volatility, emphasizing how the two interact. The model is applied to the analysis of inflation dynamics. Allowing for heavy tails leads to significant improvements in terms of fit and forecast. The adoption of the Student-t distribution proves to be crucial in order to obtain well-calibrated density forecasts. These results are obtained using US CPI inflation rate and are confirmed for other indicators of inflation as well as the CPI inflation of the other G7 countries. Finally, we show how the proposed approach generalizes various adaptive algorithms used in the literature.

Key words: Adaptive algorithms, student-t, inflation, score driven models, time-varying parameters.

JEL classification: C22, C51, C53, E31.

(1) Banca d'Italia. Email: dellemonachedavide@gmail.com
(2) Bank of England, Birkbeck University London, and CEPR. Email: ivan.petrella@bankofengland.co.uk

The views expressed are those of the authors and do not necessarily reflect those of Banca d'Italia, the Bank of England or the Monetary Policy Committee. While assuming the scientific responsibility for any error in the paper, the authors would like to thank Michele Caivano, Anthony Garratt, Emmanuel Guerre, Dennis Kristensen, Haroon Mumtaz, Zacharias Psaradakis, Barbara Rossi, Emiliano Santoro, Tatevik Sekhposyan, Ron Smith, Brad Speigner and Fabrizio Venditti for their useful suggestions; to the participants of the workshop 'Economic Modelling and Forecasting, Warwick Business School — 2013', the EABCN Conference 'Inflation Developments after the Great Recession, Eltville — 2013', the '7th International Conference on Computational and Financial Econometrics, London — 2013', the workshop on 'Dynamic Models Driven by the Score of Predictive Likelihoods, Tenerife — 2014', the 'IAAE Annual Conference, London — 2014', the '25th EC2 Conference — Advances in Forecasting, Barcelona —2014', the 'European Winter Meeting of the Econometric Society, Madrid —2014', and the seminar participants at Queen Mary University of London, University of Glasgow, the Bank of England and Banca d'Italia for their comments.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

In the last two decades there has been an increasing interest in models with time-varying parameters (TVP). Attempts to take into account the well-known instabilities in macroeconomic time series can be traced back to the 1970s (see e.g. Cooley and Prescott, 1973, 1976, Rosenberg, 1972, and Sarris 1973). Stock and Watson (1996) renewed interest in this area by documenting widespread forecasting gains for models with TVP in macroeconomic variables.[1] Recently, Cogley and Sargent (2005), Primiceri (2005), Stock and Watson (2007) have highlighted the importance of allowing for both time variation in the volatility as well as in the coefficients in the analysis of macroeconomic data. Yet, most of the studies so far have considered TVP models under the assumption that the errors are Normally distributed. Although this assumption is convenient, it limits the ability of the model to capture the tail behavior that characterizes a number of macro variables. As the recent recession has shown, departure from Gaussianity is important so as to properly account for the risks associated with black swans (see e.g. Curdia et al. 2013).

This paper considers an adaptive autoregressive model with Student-t distribution of the errors. Specifically, the parameters' variation is driven by the score of the conditional distribution (Creal et al., 2013, and Harvey, 2013). In this framewok, the distribution of the innovations not only modifies the likelihood function (as e.g. in the t-GARCH of Bollerslev, 1987), but also implies a different updating mechanism for the TVP. In fact, Harvey and Chakravarty (2009) highlight that the score driven model for time-varying scale with Student-t innovations leads to a filter that is robust to outliers. Harvey and Luati (2014) show that the same intuition holds true in models for time-varying location.

As stressed by Stock (2002) in his discussion of Cogley and Sargent (2002), estimating TVP models without controlling for the possible heteroscedasticity is likely to overstate the time-variation in the coefficients (see also Benati, 2007). In this paper, we consider time variation in both coefficients and volatility, emphasizing how the two interact in a score driven model. Moreover, we show how to impose restrictions to the model so as to achieve local stationarity and bounded long-run mean. Both of these restrictions, commonly used in applied

[1] D'Agostino et al. (2013) have restated the relative gains in terms of forecast accuracy of TVP models compared to the traditional constant parameter models in a multivariate setting. Koop and Korobilis (2013) highlighted the usefulness of the traditional adaptive algorithms to deal with TVP in large VAR models.

macroeconomics, have not yet been considered in the context of score driven models.[2]

The adaptive model in this paper is related to an extensive literature that has investigated ways of improving the forecasting performance in presence of instability. Pesaran and Timmerman (2007), Pesaran and Pick (2011) and Pesaran et al. (2013) focus on optimal weighting scheme in the presence of structural breaks. Giraitis et al. (2011) propose a non-parametric estimation approach of time-varying coefficient models. The implied weights of these models are typically monotonically decreasing with time, a feature which they share with traditional exponential weighted moving average forecasts (see e.g. Cogley, 2002). Our model features time variation in the location and scale parameters with Student-t errors. This implies a non-linear filtering process with a weighting pattern that cannot be replicated by the procedures proposed in the literature. The benefit of this approach is that observations that are perceived as outliers, based on the estimated time-varying location and scale of the process, have effectively no weights in updating the TVP. The resulting pattern of weights is both non-monotonic and time varying since this is a function of the estimated TVP. Our adaptive model implies a faster update of the coefficients in periods of high volatility. Furthermore, in periods of low volatility, even deviations from the mean that are not extremely large in absolute terms are more likely to be 'classified' as outliers. As such, they are disregarded by the filter, which by construction is robust to extreme events. These characteristics of the model are important in the study of macroeconomic time series that display instability and changes in volatility. This is demostrated empirically with an application to inflation dynamics.

Understanding inflation dynamics is key for policy makers. In particular, modern macroeconomic models highlight the importance of forecasting inflation for the conduct of monetary policy (see e.g. Svensson, 2005). There are at least three reasons why our model is particularly suitable for inflation forecasting. Firstly, simple univariate autoregressive models have been shown to work well in the context of forecasting inflation (Faust and Wright, 2013). Secondly, Pettenuzzo and Timmermann (2015) show that TVP models outperform constant-parameter models. Furthermore, they show that models with small/frequent changes, like the model proposed in this paper, produce more accurate forecasts than models whose parameters exhibit large/rare changes. Thirdly, while important changes in the dynamic properties of inflation

[2] Koop and Potter (2012) and Chan et al. (2013) deal with local stationarity and bounded trend in the context of TVP models, and they discuss the computational costs associated with those restrictions in a Bayesian setting.

are well documented (see e.g. Stock and Watson, 2007), most of the empirical studies of the time variation of inflation dynamics are typically framed in a Bayesian setup and presents a number of shortcomings: (i) it is computationally demanding, (ii) when restrictions are imposed to achieve stationarity, a large number of draws need to be discarded, therefore leading to potentially large inefficiency, and (iii) Normally distributed errors are usually assumed. The latter point is particularly relevant as it is well known, at least since the seminal work of Engle (1982), that the distribution of inflation displays non-Gaussian features. The adaptive model presented in this paper tackles all these shortcomings.

When used to analyze inflation, our model produces reasonable patterns for the long-run trend and the underlying volatility. By introducing the Student-t distribution, we make the model more robust to short lived spikes in inflation (for instance in the last part of the sample). At the same time, the specifications with Student-t innovation display substantially more variation in the volatility. In practice, with Student-t innovations the variance is less affected by the outliers and it can better adjust to accommodate changes in the dispersion of the central part of the distribution. The introduction of heavy tails improves the fit and the out-of-sample forecasting performance of the model. The density forecasts produced under Student-t distribution are improved substantially with respect to those produced by both its Gaussian counterpart and the benchmark model of Stock and Watson (2007). In fact, well calibrated density forecasts are obtained only when we allow for heavy tails. While the baseline analysis is centered on CPI inflation, which is noticeably noisier and harder to forecast than other measures of inflation, we show that an improvement in the performance of density forecasting is also obtained for other inflation measures, such as those derived from the PCE and GDP deflators. Given the different inflation dynamics across countries (Cecchetti et al., 2007), we also examine the performance of the model in the analysis of CPI inflation for different countries. We confirm that allowing for heavy tails provides substantial improvements in terms of density forecasting performance for all the G7 countries.

The model proposed here is also closely related to the adaptive algorithms that have been extensively used in the engineering literature (Fagin, 1964, Jazwinski, 1970, Ljung and Soderstrom, 1985), as well as in econometrics (Stock and Watson, 1996, Koop and Korobilis, 2012). Since the work of Marcet and Sargent (1989), these adaptive algorithms have been widely used in macroeconomics to describe the learning mechanism of expectation formation (see, e.g., Sargent, 1999, and Evans and Honkapohja, 2001). The algorithm proposed here generalizes the previous ones. Therefore, our adaptive model could potentially be used to analyse learning dynamics in the presence of time-variation in the volatility of the structural innovations (see, e.g., Justiniano and Primiceri, 2008), and/or when a non-Gaussian distribution is introduced into a structural model (see Curdia et al., 2013).

The paper is organized as follows. Section 2 describes the score-driven autoregressive model with Student-t distribution. Section 3 shows how to impose restrictions on the parameters to guarantee stationarity and a bounded long-run mean. Section 4 applies the model to the study of inflation. Section 5 shows how the model proposed here nests the traditional adaptive algorithm used in the literature, and Section 6 concludes.

2 Autoregressive model with heavy tails

Consider the following regression model with TVP and Student-t distributed residuals,

$$y_t = x_t' \phi_t + \varepsilon_t, \qquad \varepsilon_t \sim t_v \left(0, \sigma_t^2\right), \qquad t = 1, ..., n. \tag{1}$$

We consider an autoregressive (AR, hereafter) model of order p with intercept. Thus, the vector of regressors is $x_t = (1, y_{t-1}, ..., y_{t-p})'$, and $\phi_t = (\phi_{0,t}, \phi_{1,t}, ..., \phi_{p,t})'$ is the vector of time-varying coefficients.[3] The disturbance ε_t follows a Student-t distribution with v degrees of freedom and has conditional mean $E(\varepsilon_t|Y_{t-1}) = 0$ and variance $Var(\varepsilon_t|Y_{t-1}) = \sigma_t^2$. The information set at time t is denoted by $Y_t = \{y_t, y_{t-1}, ..., y_1\}$. Following Creal et al. (2013) and Harvey (2013), we postulate score-driven dynamics for the paramters vector $f_t = (\phi_t', \sigma_t^2)'$. Specifically, we opt for a random walk law of motion

$$f_{t+1} = f_t + Bs_t, \tag{2}$$

where the matrix B contains the static parameters which regulate the updating speed. We denote by $\ell_t = \log p(y_t|\mathcal{F}_t, \theta)$ the predictive log-likelihood for the $t-$th observation, conditional on $\mathcal{F}_t = \{f_t, Y_{t-1}\}$ and θ, where the latter is a vector of static parameters[4]. The driving force

[3] The results derived here are valid for additional regressors in x_t.

[4] The vector θ contains the static in B as well as the degrees of freedom v characterizing the t-distribution.

in (2) is represented by the scaled score vector, $s_t = \mathcal{S}_t^{-1}\nabla_t$, where

$$\nabla_t = \frac{\partial \ell_t}{\partial f_t} \text{ and } \mathcal{S}_t = -E\left[\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'}\right]. \tag{3}$$

The scaling matrix is chosen to be equal to the inverse of the Fisher Information matrix, $S_t = \mathcal{I}_t$. Other scaling matrices can be also used (see for details Creal et al., 2013). The scaled score vector, s_t, is the sole driving force characterizing the dynamics of f_t and is determined only by present and past observations. As a result, the model (1)-(3) is observation driven.[5]

Let's focus for a moment on the updating rule (2). The TVP are updated so as to maximise the local fit of the model at each point in time. Specifically, the size of the update depends on the slope and curvature of the likelihood function. As such, the updating law of motion (2) can be rationalized as a stochastic analog of the Gauss–Newton search direction for estimating the TVP (Ljung and Soderstrom, 1985). Blasques et al. (2014) show that updating the parameters using the score is optimal, as it locally reduces the Kullback-Leibler divergence between the true conditional density and the one implied by the model.

The law of motion (2) could have been defined more generally (see Creal et al., 2013, and Harvey, 2013), but this would have implied estimating a larger number of static parameters.[6] At the same time, the use of a random walk law of motion is supported by a large consensus in macroeconomics. As shown in Lucas (1973), most policy changes will permanently alter the agents' behaviour. As such, the model's parameters will systematically drift away from the initial value without returning to the mean value (see also Cooley and Prescott, 1976). Furthermore, in a context of learning expectations (Marcet and Sargent, 1989) the parameters would be updated as postulated in (2).

Our specification (1) elaborates on previous work. In particular, Harvey and Chakravarty (2009) consider time-varying volatility with Student-t errors, highlighting how the score driven model leads to a filtering method which is robust to a few large errors. Harvey and Luati (2014) uncover a similar mechanism in models for time varying location. More recently, Blasques at al. (2014) consider an AR(1) specification without an intercept and with constant variance,

[5] For details on the observation driven model as opposed to the parameter driven modes see Cox (1981) and Creal et al (2013).

[6] In the empirical section we find useful to set restrictions to avoid the proliferation of the static parameters. In particular, in order to mantain the model as parsimonious as possible, we restrict the matrix B to be diagonal were the first $p+1$ elements are equal to κ_ϕ and the last one is equal to κ_σ.

focusing on the stochastic properties of the implied non-linear model. Our specification features time variation for both coefficients and volatility, emphasizing the interaction between the two and their relevance for modelling macroeconomic data.

2.1 The score vector

Following the parametrization in Fiorentini et al. (2003), the conditional log-likelihood of model (1) is equal to

$$\ell_t\left(y_t|\mathcal{F}_t,\theta\right) = c\left(\eta\right) - \frac{1}{2}\log\sigma_t^2 - \left(\frac{\eta+1}{2\eta}\right)\log\left[1+\frac{\eta}{1-2\eta}\frac{\varepsilon_t^2}{\sigma_t^2}\right], \tag{4}$$

where

$$c\left(\eta\right) = \log\left[\Gamma\left(\frac{\eta+1}{2\eta}\right)\right] - \log\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \frac{1}{2}\log\left(\frac{1-2\eta}{\eta}\right) - \frac{1}{2}\log\pi,$$

$\eta = 1/\upsilon$ is the reciprocal of the degrees of freedom ($\upsilon > 2$), and $\Gamma(\cdot)$ is the Gamma function. A score-driven model with non-Gaussian innovations not only modifies the likelihood function, as in the t-GARCH of Bollerslev (1987), but it also implies a different filtering process for the TVP. Specifically, the scaled score, which drives the dynamics of the TVP, can be specialized in two sub-vectors, i.e. $s_t = (s_{\phi t}', s_{\sigma t})'$,

$$\begin{aligned} s_{\phi t} &= \frac{\left(1-2\eta\right)\left(1+3\eta\right)}{\left(1+\eta\right)}\frac{1}{\sigma_t^2}\mathcal{S}_t^{-1}x_t w_t\varepsilon_t, & (5)\\ s_{\sigma t} &= \left(1+3\eta\right)\left(w_t\varepsilon_t^2-\sigma_t^2\right), & (6)\end{aligned}$$

where $s_{\phi t}$ drives the coefficients while $s_{\sigma t}$ drives the volatility and $S_t = \frac{1}{\sigma_t^2}\left(x_t x_t'\right)$, see Appendix A for details.[7] A crucial role in the score vector (5)-(6) is played by the weights

$$w_t = \frac{\left(1+\eta\right)}{\left(1-2\eta+\eta\zeta_t^2\right)}, \tag{7}$$

which are function of the (squared) standardized prediction error $\zeta_t = \varepsilon_t/\sigma_t$.

Figure 1 provides intuition for the role of these weights in the updating scheme that governs the model parameters. The left panel plots the magnitude of w_t as a function of the standardized prediction error ζ_t. Whereas the right panel shows the so called influence function (see

[7] $\mathcal{S}_t^{-1} = \sigma_t^2\left(x_t x_t'\right)^+$ denotes the Moore-Penrose generalized inverse.

e.g. Maronna et al.), which is given by the product of the weights and of the standardized error itself. The magnitude of w_t depends on how close the observation y_t is to the center of the distribution. A small value of w_t is more likely with low degrees of freedom and low dispersion of the distribution. The weights *robustify* the updating mechanism because they downplay the effect of large (standardized) forecast errors given that, in the presence of heavy tails: such forecast errors are not informative of changes in the location of the distribution. The right panel in Figure 1 shows that the score is a bounded function of the prediction errors. Another feature of the weights w_t is that the volatility, σ_t, plays a role in *re-weighting* the observations, and as such the time varying variance has a direct impact on the coefficients' updating rule.[8] Therefore, the score vector (5)-(6) implies a *double weighting scheme*: the observation are weighted both in time and across realizations.

[insert Figure 1]

A simplified version of model (1) helps clarify the impact of this double weighting mechanism. Assume that $x_t = 1$ and w_t is exogenously given. This specification leads to an IMA(1,1) model with moving average coefficient equal to $(1 - \kappa_\theta w_t)$, and time-varying variance. Therefore, the estimated mean can be expressed as

$$\mu_{t+1} = \kappa_\theta \sum_{j=0}^{t} \gamma_j \tilde{y}_{t-j}, \quad \text{with} \quad \tilde{y}_{t-j} = w_{t-j} y_{t-j},$$

and $\kappa_\theta = \kappa_\phi \frac{(1-2\eta)(1+3\eta)}{(1+\eta)}$. Therefore observations are weighted to be robust to the impact of extreme events, through the weights w_t, as well as being discounted (in time) by $\gamma_j = \prod_{k=t-j+1}^{t} (1 - \kappa_\theta w_k)$, $\gamma_0 = 1$. Similarly, the estimated variance is

$$\sigma_{t+1}^2 = \kappa_\zeta \sum_{j=0}^{t} (1 - \kappa_\zeta)^j \tilde{\varepsilon}_{t-j}^2,$$

where $\tilde{\varepsilon}_{t-j}^2 = w_{t-j} \varepsilon_{t-j}^2$ are the weighted prediction errors, and $\kappa_\zeta = \kappa_\sigma (1 + 3\eta)$ regulates how past observations are discounted.[9] In the presence of an outlier (i.e. $w_t = 0$) the moving

[8] This is not the case under Gaussian distibution.

[9] For large t we can interpret the estimated trend as the low-pass filter $\kappa_\theta / [1 - (1 - \kappa_\theta w_t) L]$ applied to the weighted observations $\tilde{y}_t$. Similarly, the estimated variance is obtained by the following filter $\kappa_\zeta / [1 - (1 - \kappa_\zeta) L]$

average coefficients collapses to one, and the model becomes a pure random walk.

In practice the weights w_t depend non-linearly on the current observations and the past estimated parameters through $\zeta_t = \varepsilon_t/\sigma_t$. Therefore, under the Student-t distribution the score driven model leads to a non-linear filter which cannot be analytically expressed from the last two formulae. It is worth noticing that, since coefficients and volatility are simultaneously updated, prediction errors of the same size are weighted differently according to the conditional mean and volatility. Specifically, in periods of low volatility, a given prediction error is more likely to be categorized as part of the tails and therefore it is downweighted. This mechanism reinforces the smoothness of the filter in periods of low volatility. Conversely, the updating is quicker in periods of high volatility with prediction errors reflecting to a greater extent into paramters changes. As such, the weighting pattern is non monotonic and time-varying, and it cannot be easily replicated by the weighting schemes which are meant to improve the forecasts under structural breaks such as the ones proposed by Giraitis et al. (2011) or Pesaran et al. (2013).

Blasques et al. (2014) show that a similar feature applies when the model is autoregressive of order one. In general, for autoregressive models the updating process depends on the prediction error as well as on the conditional mean relatively to the long-run mean, i.e. deviations from the unconditional mean are relevant to qualify the signal provided by the prediction error.

2.2 Gaussian distribution

The Gaussian case is recovered setting $\eta = 0$ (i.e. $\upsilon \to \infty$, and $w_t = 1, \forall t$), which results in the following TVP model

$$
\begin{aligned}
y_t &= x_t'\phi_t + \varepsilon_t, \qquad \varepsilon_t \sim N\left(0, \sigma_t^2\right), \qquad t = 1, ..., n, && (8)\\
\phi_{t+1} &= \phi_t + \kappa_\phi \frac{1}{\sigma_t^2} \mathcal{S}_t^{-1} x_t \varepsilon_t, && (9)\\
\sigma_{t+1}^2 &= \sigma_t^2 + \kappa_\sigma (\varepsilon_t^2 - \sigma_t^2), && (10)
\end{aligned}
$$

where $\mathcal{S}_t = \frac{1}{\sigma_t^2}\left(x_t x_t'\right)$.

Equation (9) resembles the Kalman filter (KF), since the updated parameters react to the

applied to the weighted (squared) prediction errors $\tilde{\varepsilon}_t^2$.

prediction error ε_t scaled by a gain depending on $\frac{1}{\sigma_t^2}x_t$. Equation (10) is the integrated GARCH model. In contrast to the Student-t case, the volatility cancels out from the coefficients' dynamics. Therefore, the time variation in the volatility does not directly affect the time variation in the coefficients.[10]

As opposed to the more common parameter-driven model, both the signal (8) and the parameters (9)-(10) are driven by the prediction error. The model is therefore similar to the single-source error model of Casalas et al. (2002) and Hyndman et al. (2008).[11] Furthermore, it is worth stressing that the Gaussian model (8)-(10) implies exponential discounting of past observations. As such, it nests several adaptive algorithms used in the literature. This is discussed in section 5.

2.3 Estimation

The static parameters of (1)-(4) are estimated by maximum likelihood (ML), i.e. $\widehat{\theta} = \arg\max \mathcal{L}$. Once the updating equation (2) is implemented together with the predictive likelihood (4), the log-likelihood function $\mathcal{L} = \sum_{t=1}^{n} \ell_t \left(y_t | \mathcal{F}_t, \theta\right)$ is maximized numerically with respect to the static parameter θ. Following Creal et al. (2013, sec. 2.3), we conjecture that $\sqrt{n}(\widehat{\theta} - \theta) \rightarrow N(0, \Omega)$, where Ω is evaluated by numerical derivative at the optimum.[12]

3 Model restrictions

Applications of TVP models often require imposing restrictions on the parameters space. For instance, an AR model such as (1) is usually restricted so that the implied roots lie within the unit circle. In the Bayesian framework, such constraints are usually imposed by rejection sampling. This, however, leads to heavy inefficiencies (see e.g. Koop and Potter, 2012, and Chan el al., 2013). When restrictions are implemented within a score-driven setup, the resulting model can still be estimated by MLE without the need of computational demanding simulation

[10] Note that this feature is not shared by the equivalent parameter driven models (see e.g. Stock, 2002).

[11] In the single source of error model, the state space has perfectly correlated disturbances, and as such it leads to an observation driven model.

[12] A formal proof of these results is beyond the scope of this paper. Harvey (2013, sec 4.6) derives the asymptotic theory of a model with time-varying volatility only. Harvey and Luati (2014) prove the asymptotic properties for a model with time-varying level only. Blasques et al (2014) studied the asymptotic properties of AR(1) model with constant volatility.

methods. This requires reparametrizing the vector of TVP as follows

$$\tilde{f}_t = g(f_t), \tag{11}$$

where f_t is the unrestricted vector of parameters we model, and $\tilde{f}_t$ is the vector of interest restricted through the function $g(\cdot)$. The latter is a time invariant, continuous and twice differentiable function, often called *link function* (Creal et al., 2013, and Harvey, 2013). The vector f_t continues to follow the updating rule (2), but the score needs to be amended as follows

$$s_t = (\Psi_t' \mathcal{I}_t \Psi_t)^{-1} \Psi_t' \nabla_t, \tag{12}$$

where $\Psi_t = \frac{\partial \tilde{f}_t}{\partial f_t}$ is the Jacobian of $g(.)$, and ∇_t and $\mathcal{I}_t$ are the score and the scaling matrix previously computed with respect to $\tilde{f}_t$. In practice, we model $f_t = h(\tilde{f}_t)$, where $h(\cdot)$ is the inverse function of $g(\cdot)$. Given a continuous and differentiable function $g(\cdot)$, Ψ_t is a deterministic function given past information, whose role is to re-weight the original score such that the restrictions are satisfied at each point in time.

Given model (1), the vector of interest is partitioned as follows, $\tilde{f}_t = (\phi_{0,t}, \underline{\phi}_t', \sigma_t^2)'$, and its unrestricted counterpart, $f_t = (\alpha_{0,t}, \underline{\alpha}_t', \gamma_t)'$, is the vector we model. The following subsections describe in detail how to impose restrictions on the autoregressive coefficients, $\underline{\phi}_t = (\phi_{1,t}, ..., \phi_{p,t})'$, and on the intercept, $\phi_{0,t}$, in order to achieve stationarity and bounded (long-run) mean of the process. Moreover, the variance is always constrained to be positive using the exponential link function, $\sigma_t^2 = \exp(2\gamma_t)$, implying that $\gamma_t = \log(\sigma_t)$.

It is useful to specialize the Jacobian matrix as follows

$$\Psi_t = \begin{bmatrix} \frac{\partial \phi_{0,t}}{\partial \alpha_{0,t}} & \frac{\partial \phi_{0,t}}{\partial \underline{\alpha}_t'} & \\ \frac{\partial \underline{\phi}_t}{\partial \alpha_{0,t}} & \frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'} & 0_{(p+1)\times 1} \\ \multicolumn{2}{c}{0_{1\times(p+1)}} & 2\sigma_t^2 \end{bmatrix}, \tag{13}$$

where $\frac{\partial \phi_{0,t}}{\partial \alpha_{0,t}}$ and $2\sigma_t^2$ are scalars. $\frac{\partial \phi_{0,t}}{\partial \underline{\alpha}_t'}$ and $\frac{\partial \underline{\phi}_t}{\partial \alpha_{0,t}}$ are $1 \times p$ and $p \times 1$ vectors respectively, and $\frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'}$ is a $p \times p$ matrix.

3.1 Imposing stationarity

This sub-section describes the way we achieve a locally stationary model. To this end we use the function mapping the AR coefficients to the partial autocorrelations (PACs). Stationarity is then imposed by restricting the latter to the interval $(-1, 1)$.[13]

Proposition 1 *Let $\underline{\phi}_t = (\phi_{1,t}, ..., \phi_{p,t})'$ denote the vector of AR coefficients, $\rho_t = (\rho_{1,t}, ..., \rho_{p,t})'$ is the corresponding vector of PACs and $\underline{\alpha}_t = (\alpha_{1,t}, ..., \alpha_{p,t})'$ is the vector of unrestricted coefficients. In a locally stationary model $\underline{\phi}_t \in S^p$, where S^p is the hyperplane with all the roots, z_t, inside the unit circle, i.e. $\underline{\phi}_t(z_t) = 0$, $z_t \in C^p$ and $|z_{j,t}| < 1$ for $j = 1, ..., p$. It is possible to show that $\underline{\phi}_t \in S^p$, if and only if, $\rho_t \in R^p$ and $|\rho_{j,t}|<1$. The link function mapping the AR coefficients to the PACs is $\underline{\phi}_t = \Phi(\rho_t)$, and it is obtained by the last recursion of the Durbin-Levinson algorithm*

$$\phi_t^{j,k} = \phi_t^{j,k-1} - \rho_{k,t}\phi_t^{k-j,k-1} \quad \text{for } j = 1, ..., k-1 \text{ and } k = 2, ..., p, \tag{14}$$

with $\phi_t^{1,1} = \rho_{1,t}$ and $\phi_t^{k,k} = \rho_{k,t}$. Moreover, $\rho_t = \Upsilon(\underline{\alpha}_t)$ is the function restricting the PACs to lie in $(-1, 1)$, and this can be obtained by any monotonic and differentiable function

$$\rho_{j,t} = \Upsilon(\alpha_{j,t}), \text{ such that } \rho_{j,t} \in (-1,1), \text{ for } j = 1, ..., p. \tag{15}$$

Finally, the composite function, $g(\cdot) = \Phi[\Upsilon(\cdot)]$, maps the unrestricted parameters into the stationary coefficients, i.e. $\underline{\phi}_t = g(\underline{\alpha}_t)$, where $\underline{\alpha}_t \in R^p$ and $\phi_t \in S^p$.

Proof. See Bandorff-Nielsen and Schou (1973) and Monahan (1984). ■

The functions $\Phi(\cdot)$ and $\Upsilon(\cdot)$ are continuous and differentiable, as such the sub-matrix $\frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'}$ in (13) is equal to

$$\frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'} = \frac{\partial \Phi(\rho_t)}{\partial \rho_t'} \frac{\partial \Upsilon(\underline{\alpha}_t)}{\partial \underline{\alpha}_t'}, \tag{16}$$

where $\frac{\partial \Upsilon(\underline{\alpha}_t)}{\partial \underline{\alpha}_t'}$ is a diagonal matrix with elements $\frac{\partial \Upsilon(\alpha_{jt})}{\partial \alpha_{j,t}}$ for $j = 1, ..., p$, and $\frac{\partial \Phi(\rho_t)}{\partial \rho_t'}$ is provided by the theorem below.

[13] It is worth noting that one of the specifications considered in Blasques et al. (2014) is a special case of our setting. They use the logistic transformation to restrict the coefficient of the AR(1) model.

Theorem 1 *The Jacobian matrix* $\Gamma_t = \frac{\partial\Phi(\rho_t)}{\partial\rho_t'}$ *is obtained from the last iteration of the recursion*

$$\begin{aligned}\Gamma_{k,t} &= \begin{bmatrix} \tilde{\Gamma}_{k-1,t} & b_{k-1,t} \\ 0_{k-1}' & 1 \end{bmatrix}, \\ \tilde{\Gamma}_{k-1,t} &= J_{k-1,t}\Gamma_{k-1,t}, \qquad k = 2, ..., p,\end{aligned} \tag{17}$$

with

$$b_{k-1,t} = -\begin{bmatrix} \phi_t^{k-1,k-1} \\ \phi_t^{k-2,k-1} \\ \vdots \\ \phi_t^{2,k-1} \\ \phi_t^{1,k-1} \end{bmatrix}, \quad J_{k-1,t} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -\rho_{k,t} \\ 0 & 1 & 0 & -\rho_{k,t} & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & -\rho_{k,t} & 0 & 1 & 0 \\ -\rho_{k,t} & 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{18}$$

If k is even, the central element of $J_{k-1,t}$ is equal to $(1-\rho_{k,t})$. The recursion is initialized with $J_{1,t} = (1-\rho_{2,t})$ and $\Gamma_{1,t} = 1$.

Proof. See Appendix A. ■

When the time-varying intercept is included without imposing any restrictions, the remaining elements of the Jacobian matrix (13) are: $\frac{\partial\phi_{0,t}}{\partial\alpha_{0,t}} = 1$, $\frac{\partial\underline{\phi}_t}{\partial\alpha_{0,t}} = 0_{p\times 1}$, and $\frac{\partial\phi_{0,t}}{\partial\underline{\alpha}_t'} = 0_{1\times p}$.

3.2 Bounded trend

It is also often the case that in practice one wants to discipline the model so as to have a bounded conditional mean. Following Beveridge and Nelson (1981), a stochastic trend can be expressed in terms of long-horizon forecasts. For a driftless random variable, the Beveridge-Nelson trend is defined as the value to which the series is expected to converge once the transitory component dies out (see e.g. Benati, 2007 and Cogley et al., 2010).

Specifically, the local-to-date t approximation implies that the unconditional time-varying mean is equal to

$$\mu_t = \frac{\phi_{0,t}}{1-\sum_{j=1}^{p}\phi_{j,t}}, \tag{19}$$

following Chan et al. (2013), we restrict $\mu_t \in [\underline{b}, \bar{b}]$.

In line with Cogley et al. (2010), our specification implies that the detrended component, $\tilde{y}_t = (y_t - \mu_t)$, follows a locally stationary AR(p) model, i.e. $\Pr\left\{\lim_{h\to\infty}\mathbb{E}_t\left(\tilde{y}_{t+h}\right) = 0\right\} = 1$.

Proposition 2 *Let $h(\cdot)$ be any continuous and differential function so that $h(\cdot) \in [\underline{b}, \overline{b}]$. The link function allowing $\mu_t \in [\underline{b}, \overline{b}]$ is*

$$\phi_{0,t} = h(\alpha_{0,t}) \left(1 - \sum_{j=1}^{p} \phi_{j,t}\right). \tag{20}$$

Therefore, the elements of the Jacobian matrix (13) are:

$$\begin{aligned} \frac{\partial \phi_{0,t}}{\partial \alpha_{0,t}} &= \frac{\partial h(\alpha_{0,t})}{\partial \alpha_{0,t}} \left(1 - \textstyle\sum_{j=1}^{p} \phi_{j,t}\right), \\ \frac{\partial \phi_{0,t}}{\partial \underline{\alpha}_t'} &= -h(\alpha_{0,t})\, \iota' \frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'}, \\ \frac{\partial \underline{\phi}_t}{\partial \alpha_{0,t}} &= 0_{p\times 1}, \end{aligned} \tag{21}$$

where ι' is a $1 \times p$ vector of ones and $\frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'}$ is equal to (16).

4 Application to inflation dynamics

TVP models have been widely used for the analysis of inflation. The following features have been documented in the literature: (i) substantial time variation in trend inflation (e.g. Cogley, 2002, and Stock and Watson, 2006), (ii) changes in persistence (Cogley and Sargent, 2002, and Pivetta and Reis, 2007) and (iii) time varying volatility (e.g. Stock and Watson, 2006, and Clark and Doh, 2011). Here we aim to capture these features by the p-th order autoregressive model with time-varying parameters

$$\pi_t = \phi_{0,t} + \sum_{j=1}^{p} \phi_{j,t} \pi_{t-j} + \varepsilon_t, \qquad \varepsilon_t \sim t_v\left(0, \sigma_t^2\right). \tag{22}$$

In the case with $p = 0$, we have a specification were the time varying constant captures trend inflation. In particular, with Gaussian innovations the trend is estimated by exponential smoothing as in Cogley (2002).[14] Autoregressive models have been shown to work well in the context of forecasting inflation (Stock and Watson, 2007, and Faust and Wright, 2013). Many empirical works have been framed in a Bayesian setup,[15] Here instead we use of the observation-

[14] Notice that Cogley (2002) does not include time-variation in the variance. As shown in Section 2, under Gaussian distribution the time-varying variance does not affect directly the estimation of the trend, but it does affect the estimate of the smoothing parameter. Differently from Cogley (2002), the smoothing parameter here is estimated at the value that minimizes the (standardized) one-step ahead prediction error.

[15] A noticeable exception is the work of Pivetta and Reis (2007).

driven model introduced in section 2, and we emphasize the importance of allowing for Student-t distribution.

Various specifications of model (22) are considered in terms of lags ($p = 0, 1, 2, 4$) and restrictions. The model is reparameterized so that the variance is positive and, for $p > 0$, the model is locally stationary as shown in sub-section 3.1. For every specification we also consider a counterpart with bounds (between 0 and 5) on the long-run trend as shown in sub-section 3.2. This follows the work by Chan et al. (2013) arguing that a level of the trend inflation that is too low (or too high) is inconsistent with the central bank's inflation target.[16] Finally, for all specifications we consider both Gaussian and Student-t distribution of the innovations.

[Insert Table 1]

Table 1 reports the estimates of the various specifications for the annualized quarterly US CPI inflation over the period 1955Q1–2012Q4. Besides the estimates of the parameters and their associated standard error, we also report the value of the log likelihood function, the Akaike (AIC) and Bayesian Information Criterion (BIC).

The trend-only specification ($p = 0$) features a high estimated value of the smoothing parameter κ_ϕ implying that past observations are discounted more heavily. This is also true for the specification with Student-t distribution. By adding the autoregressive component we obtain substantially smaller estimates of κ_ϕ, and this is due to the fact that part of the persistence of inflation is captured by the autoregressive terms. In contrast, the smoothing parameter associated to the variance, that is κ_σ, is stable and typically higher than κ_ϕ. This result supports to the idea that changes in the volatility is an important feature of inflation (see e.g. Pivetta and Reis, 2007). Noticeably, the specifications with Student-t distribution considerably outperform the ones with Gaussian innovations both in terms of the likelihood values and information criteria. The estimates of the degrees of freedom υ, between 4 and 6, depict a remarkable difference between the Gaussian and the Student-t specification and underline the presence of pronounced variations of inflation at the quarterly frequency. Those variations either arise from measurement errors or are due to the presence of rare events that structural macroeconomics should explicitly account for (as recently advocated by Curdia et

[16] The bounds correspond to the upper and lower bounds of the posterior in Chen et al. (2013).

al., 2013). Notice that $\upsilon = 5$ is also consistent with the calibrated density forecast in Corradi and Swanson (2006). Overall, the AR(1) model without bounds on the long-run mean and Student-t distribution slightly outperforms all the other specifications in terms of fitting.

4.1 Trend Inflation and Volatility

In this sub-section we show that our model is able to capture the salient features of inflation dynamics in terms of trend inflation and volatility. Furthermore, we highlight the main differences between the specifications with Gaussian and Student-t distribution.

Figure 2 compares the estimates of the long run trend for the different specifications.[17] The trend-only specification follows inflation very closely through the ups and downs, whereas including lags of inflation leads to a smoother long-run trend estimate. Therefore, when we allow for intrinsic persistence a substantial part of inflation fluctuations during the high inflation period (in the early part of the sample and in the 70s) is attributed to deviations from the trend. These results indicate that various autoregressive specifications are likely to deliver very similar long-run forecasts and the choice of the lag length impacts only on the shape of the dynamics toward the long-run level, i.e. the short to medium horizon forecasts. For all specifications we find that since the mid 90s, the long-run trend is stable between 2-3%, going slightly over 3% on the run up to the recent recession. Figure 3 presents the estimates of the long-run trend for the trend-only model and the AR(1) specification, focussing on the differences between Gaussian and Student-t innovations. The left panel highlights that the trend-only model with Student-t is generally less affected by the sharp transitory movements, as it is evident in the last part of the sample. This is a direct consequence of the way the algorithm modifies the updating mechanism under Student-t. In particular, it downplays the relevance of the forecast error when this is perceived as an outlier. Once lagged inflation is included, the differences between the two specifications are attenuated. Both of them deliver a very smooth outline of trend inflation but some differences are apparent in the last part of the sample. In this latter case the outliers still have an impact on the parameters' estimates for the

[17] Imposing the upper bound on the long-run mean implies a qualitatively similar picture for the trend-inflation across all specifications (see Appendix D). In this case the trend estimates are consistent with the idea of a central bank anchoring expectations of trend-inflation to a fairly stable level over the sample. Trend-inflation rises above 3% in the early '70s and then decreases back to a slightly lower level only in the mid '90s. Worth noting that the pattern in the long-run trend is quite similar to the one found by Chan et al. (2012) despite the fact that they use a different model specification and different estimation techniques.

Gaussian model, whereas they have a smaller effect under Student-t. However, the variation in the time varying intercept is offset by the variation in the autoregressive coefficients and the model ends up delivering rather smooth long run forecasts.[18]

[Insert Figure 2 and Figure 3]

Figure 4 reports measures of the changes in volatility. Only few differences can be appreciated when comparing the trend-only model to the ARs specifications. For all the specifications it is true that the variance was substantially higher in the 50s, in the 70s and then again in the last decade. This pattern of the volatility is consistent with Chan et al. (2013) and Cogley and Sargent (2014). However, a comparison between the left and right panels reveals some interesting differences between the models under Gaussian and Student-t distribution. First, the model based on the Student-t distribution is more robust to single outliers. In fact, under Gaussian distribution the volatility seems to be disproportionately affected by very few observations in the last part of the sample.[19] Second, although the volatility shows very similar low-frequency variation across different specifications, under Student-t the model displays substantially more high frequency movements in the volatility. Note that under Student-t the observations are weighted such that large deviations are heavily down-weighted and small deviations are instead magnified. In other words, under Student-t the variance is less affected by the outliers and it can better adjust to accommodate changes in the dispersion of the central part of the distribution. The latter result is particularly important in light of the superior in sample fit of the Student-t specification reported in the previous sub-section. It is worth noting that most of the literature, which has mainly focused on the Gaussian distribution, has only

[18] In order to clarify this point it is instructive to look at what happens to the autoregressive coefficients in the Gaussian model in response to the inflation shift in 2008 (see Appendix D). The 2008:Q3 observation (approximately -9%) is clearly a tail event given the usual inflation variability. This single observation leads to a shift of the autoregressive coefficient from approx. 0.8 to -0.5. At the same time the shift in the long-run trend is slightly less than 1%, as a result of a simultaneous jump in the intercept. Conversely, the long-run trend under Student-t barely varies as a result of the same episode.

[19] Again it is worth to report what happens as a result of a single tail event in 2008:Q3. The log-volatility shifts from approx. 1 to 2-2.5 for the Gaussian model, whereas it moves only up to 1.5-1.7 with the Student-t distribution. Therefore, with a Gaussian model one would have a misleading picture of the inflation uncertainty, which surpasses by far the level reached in the 70s. This result maps into a severely biased estimate of the densities that is going to be significantly fatter as a result of a single observation.

reported and emphasized the importance of the low frequency variation in the volatility.

[Insert Figure 4]

4.2 Forecasting Evaluation

In this section we assess the forecasting performance of the model. Specifically, we evaluate the forecasts over the period 1973Q1–2012Q4, where the model is estimated recursively over an expanding window. Consistent with a long-standing tradition in the learning literature (referred to as anticipated-utility by Kreps, 1998), we update the coefficients period by period and we treat the updated values as if they remained constant going forward in the forecast. We first assess the point forecast using both the root mean squared error (RMSE) and the absolute mean error (MAE). Later on, we will evaluate the performance of the models in terms of their density forecasts.

Forecasts are evaluated versus the Stock and Watson (2007, SW thereafter) model that is usually considered to be a good benchmark for inflation forecasting.[20] In SW both the conditional mean and the measurement error are driven by two independent shocks with stochastic volatility. This implies that the reduced form model follows an integrated moving-average of order one (IMA(1,1) hereafter) where the parameters are driven by a convolution of the two independent stochastic volatilities. This model bears some similarity to our trend-only model that also implies an IMA(1,1) with TVP. Specifically, under Gaussian distribution the variance is time-varying whereas the MA coefficient is constant. In contrast, under the Student-t distribution the score-driven model produces an IMA(1,1) in which both the parameters are time-varying as discussed in section 2. Whereas in SW the time-varying MA coefficient drifts smoothly as a result of the random walk specification for the stochastic volatilities, in our score-driven model (with Student-t) the time-varying MA coefficient is more volatile. In fact, the time variation of this coefficient derives from the weights, w_t, and therefore is due to the fact that the model discounts the signal from the observations that are perceived as outliers.

Table 2 reports the results for the point forecast. Despite the well-known performance of the

[20]The SW model is estimated by Bayesian MCMC method and the Gibbs Sampling algorithm is broken into the following steps: (i) sample the variance of the noise component using the independent Metropolis Hastings as in Jacquier et al (2004); (ii) sample the variance of trend component as in (i); (iii) sample the trend component using the Carter and Kohn (1994) algorithm.

benchmark model, many of the alternative specifications we consider tend to have lower RMSE and MAE. However, the differences tend to disappear at longer horizons. The difference in forecasting performance is also statistically significant for many specifications.[21] For instance, the AR(4) model reduces the loss by roughly 15% both for the one quarter and one year ahead forecast.[22] Imposing bounds on the long-run mean does not seem to improve the performance of the various specifications.[23] Most importantly, a comparison between the Gaussian and Student-t models reveals little differences in terms of point forecast.

[Insert Table 2]

An important element of any forecast lies in the ability to quantify and convey the outcome's uncertainty. This requires a forecast of the whole density of inflation. For instance, Cogley and Sargent (2014), highlight the relevance of deflation risk and a prediction of the latter requires an estimation of the overall density. Table 3 reports the results from the density forecast exercise in which we focus on the one-step-ahead forecast. The first two columns report the results of two tests for the calibration of the densities. One is Berkowitz's (2001) LR test on the inverse transformation of the probability integral transforms (PITs) and the other is the nonparametric test of Rossi and Sekhposyan (2014, RS hereafter). The latter is still valid also in the presence of parameter estimation error. The results suggest that the density forecasts of all the specifications with Gaussian innovations, as well as the SW model, are not well calibrated. In order to understand why this is the case Figure 5 plots the empirical distribution function (p.d.f.) of the PITs. In addition to the PITs, we also provide the 95% confidence interval (broken lines) using a Normal approximation to a binomial distribution as in Diebold et al. (1998).[24] From both figures it is evident that the models with Gaussian innovations tend to

[21] We report the test of Giacomini and White (2006). Despite the expanding window, this test is approximately valid as our model implicitly discount the observations, so that earlier observations are in practice discarded for the estimates in the late part of the sample that is used to forecast.

[22] Looking at the subsample reveals that most of the gains are achieved at the beginning and at the end of sample, while the SW model seems to be slightly better in the low volatile period (from mid-80s to early 2000). None of this differences are significant using the fluctuation tests of Giacomini and Rossi (2010) highlighting a relatively high volatility of the forecast errors.

[23] The trend-only model with restricted long-run mean is outperformed by the alternative ones, in particular for the short horizon. However, the relative performance of this specification is severely biased by the inclusions of the great inflation period (mid 70s-80s).

[24] In appendix D we report the cumulative distribution function (c.d.f.) of the PITs for each realization (see Rossi and Sekhposyan, 2014). This figure reveals same results as the one given by the plot of the PITs, i.e. only the densities from the adaptive models with Student-t innovations are well calibrated.

produce densities in which too many realizations fall in the middle of the forecast densities relative to what we would expect if the data were really Normally distributed.[25] Density forecasts are instead well calibrated for the models under Student-t distribution (see Table 3 and Figure 5). Two features of the Student-t model explain this result. First, the volatility is not affected by the observations in the tail of distribution, as a result of this it varies in a way that better capture the changes in the dispersion of the central part of the density. Second, the distribution by nature has a slower decay in the tail and as such it allows for higher probability of extreme events. The last two columns of Table 3 report the average logarithm score of the various models and the p-value of the test proposed by Amisano and Giacomini (2007) comparing the performance of the various specifications with respect to SW. The models with Student-t distribution significantly improve the accuracy of the density forecast and outperform considerably both the SW benchmark as well as all the specifications with Gaussian errors.

[Insert Table 3]

[Insert Figure 5]

In order to understand whether the improvements under Student-t are stable over time, Figure 6 reports the fluctuation test of Giacomini and Rossi (2010). We consider the trend-only model with Student-t versus the Gaussian case and SW model. The value of the statistic is always positive suggesting that the densities produced by the heavy tails model delivers a consistently higher log-score on average throughout the sample. The differences between Gaussian versus Student-t innovations are however not statistically significant in the 90s. This is not surprising since in this period inflation has been quite stable and as such we would not expect considerably different densities produced by the two models.[26]

[Insert Figure 6]

The adaptive model developed in this paper delivers a model-consistent way to deal with

[25] The histogram of the PITs for the SW model is quite similar to the one obtained for the score-driven models with Gaussian distribution, i.e. this model produces densities that are overall too wide relative to the realizations.

[26] The results are qualitatively similar when any other autoregressive specifications is considered. Those results are not reported, but are available upon request.

time-variation in presence of heavy tailed distribution. Appendix B explores the importance of using an updating mechanism for the parameters consistent with the score-driven approach as opposed to some ad-hoc specifications. We show that the score-driven specification outperforms the alternative ones, and the degree of freedom as well as the score-driven updating mechanism are both important ingredients to achieve well calibrated density forecasts.

4.3 Additional empirical evidence

In the previous section we have shown how the model with Student-t errors produces time variation in the parameters which is robust to the presence of heavy tails. Furthermore, the volatility is less affected by the behavior in the tail of the distribution so that it can better reflect the changes in the spread of the central part of the density. These aspects of the model are key in order to retrieve a well calibrated density forecast for the US CPI inflation over the sample analyzed. However, CPI inflation is notoriously more volatile than other inflation indicators such as the GDP deflator and the PCE deflator.[27] In order to assess whether the improvements of the heavy tail model carry through more generally for other measures of inflation we repeat the forecasting exercise using these two measures of inflation. The results we obtain are in line with the evidence reported in the previous section. To preserve space we focus on the density forecast while the results for the point forecast are reported in Appedix D.[28] Table 4 shows how for these two additional indicators of inflation the densities are well calibrated only under the Student-t specification. Furthermore, the models under Student-t tend to outperform the SW benchmark, as well as all the Gaussian specifications, by a considerable margin. Interestingly, since those two indicators are smoother than CPI inflation adding lags of inflation can deliver significant improvements in the density forecast.

[Insert Table 4]

Cecchetti et al. (2007) highlight the presence of similarities in inflation dynamics across countries. Therefore, as in the last exercise, we investigate the performance of our model

[27] In fact, SW report that their model is better suited for these smoother series.

[28] The point forecast assessment confirms that various specifications outperform on average the SW model. However, the differences are statistically significant for few specifications only and mainly for short forecast horizons.

for CPI inflation of the remaining G7 countries. Table 5 reports a summary of the density evaluation focussing on the trend-only model.[29] The results are in line with those reported for the US. Specifically, the model with heavy tails results in large gains in terms of log-score and the densities are well calibrated only when the Student-t distribution is allowed.

[Insert Table 5]

5 Relation with the existing adaptive algorithms

This section highlights the relation between the score-driven model proposed in this paper and other algorithms used in the literature. The engineering literature has a long tradition of modelling parameter instability by means of *adaptive algorithms* (Fagin, 1964, Jazwinski, 1970, Ljung and Soderstrom, 1985). Doan, Litterman, and Sims (1984) have been the first to explore the estimation of time-varying coefficients models using the Kalman filter. Stock and Watson (1996) use a similar approach arguing that the TVP model is superior to the fixed coefficients models for economic forecasting. Recently, Koop and Korobilis (2012) revisit the adaptive algorithm in the context of a large VAR model. We illustrate that those approaches are nested in the score-driven algorithm developed in this paper. Furthermore, since Marcet and Sargent (1989), adaptive algorithms have been extensively used in the macroeconomic literature to describe the learning mechanism of forming expectations (see, e.g., Sargent, 1999 and Evans and Honkapohja, 2001). Those learning algorithms can also be obtained as a special case of the one developed in this paper. Therefore, we pave the way to analysing learning models in the presence of time varying volatility in the structural innovations (see, e.g., Justiniano and Primiceri, 2008) and/or in the context of rare events (see e.g. Curdia et al., 2013).

Consider the case in which we smooth the scaling matrix in (3), namely

$$\mathcal{S}_t = (1-\lambda_t)\mathcal{S}_{t-1} + \lambda_t \left(\frac{1}{\sigma_t^2} x_t x_t' \right), \tag{23}$$

where $\lambda_t = 1/t$ results in the recursive estimator of the second moment matrix, $\lambda_t = \lambda$ produces

[29] For the remaining G7 countries the trend-only model with Gaussian distribution is the benchmark specification; this because in the previous section we have documented how this model performs very similar to the SW model. In Appendix C, we report the results for the other specifications excluding the bounded-trend models as it is not clear a priori what should be the upper and lower bounds for those countries.

a discounted estimator, and $\lambda_t = 1$ leads to the information matrix at time t, that is $\mathcal{S}_t = \frac{1}{\sigma_t^2} x_t x_t'$. To facilitate the comparison with the existing algorithms, it is convenient to start with a model with constant variance. Setting $\kappa_\phi = \lambda_t = \lambda$, the Gaussian score-driven filter (9), with scaling matrix (23), collapses to the Constant Gain Learning (CGL) widely used in the learning literature[30]

$$\begin{aligned} \mathcal{S}_t &= \mathcal{S}_{t-1} + \lambda \left(\frac{1}{\sigma^2} x_t x_t' - \mathcal{S}_{t-1} \right), \\ \phi_{t+1} &= \phi_t + \lambda \mathcal{S}_t^{-1} x_t \frac{1}{\sigma^2} \varepsilon_t. \end{aligned} \tag{24}$$

Remark 1 *The CGL weights the observations y_{t-j} exponentiallly at the rate $(1-\lambda)^j$, where $0 < \lambda < 1$ gives a trade-off between the tracking capability and the volatility. The CGL is a forgetting factor algorithm that can be derived from the discounted least squares principle (see details in Appendix A).*

To further establish the bridge between our model and the adaptive algorithms used in the literature, it is useful to recall some known results. Starting with the following parameters-driven model

$$\begin{aligned} y_t &= x_t' \phi_t + \varepsilon_t, \ \varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right), \\ \phi_{t+1} &= \phi_t + \eta_t, \ \eta_t \sim N\left(0, Q_t\right). \end{aligned} \tag{25}$$

Koop and Korobilis (2012) propose to estimate the TVP model using the so-called *'forgetting factor'* algorithm. In practice, this is obtained from the KF applied to (25) with restrictions on σ_ε^2 and Q_t as described in the following Lemma.

Lemma 1 *In the model (25), $P_{t|t}$ is the MSE of the real-time filter $\phi_{t|t}$ obtained by the KF. If we set $\sigma_\varepsilon^2 = \frac{\sigma^2}{1-\lambda}$ and $Q_t = P_{t|t} \frac{\lambda}{1-\lambda}$, where λ is the gain parameter, the predictive filter $\phi_{t+1|t}$ (obtained by the KF) is exactly the CGL (24).*

Remark 2 *Under a Gaussian distribution, the estimated volatility (10) is an exponential smoothing of the squared prediction errors. This filter is used in Ljung and Soderstrom (1985,*

[30] See, among others, Evan and Honkaphoja (2001), Sargent and William (2005), Branch and Evans (2006) and Carceles-Poveda and Giannitsarou (2007).

sec. 3.4.3) and Koop and Korobilis (2012) to capture the time-varying volatility.

Given Lemma 1 and Remark 2, Koop and Korobilis (2012) is a special case of our score-driven model.

Lemma 2 *In model (25), if we set $Q_t = \kappa^2\sigma^2[E(x_t x_t')]^{-1}$, the KF converges to the score-driven filter (9) with scaling matrix $\mathcal{S}_t = \frac{1}{\sigma^2}\mathbb{E}(x_t x_t')$. Similarly, setting $Q_t = \kappa^2\frac{1}{\sigma^2}\mathbb{E}(x_t x_t')$, the KF converges to the same filter with scaling matrix $\mathcal{S}_t = I$. See details in the Appendix A.*

Interestingly, the specification in the Lemma 2 has been used among others by Stock and Watson (1996), Sargent and William (2005), Branch and Evans (2006) and Li (2008). Evans et al. (2010) named it as Stochastic Gradient algorithm, whereas Slobodyan and Wouters (2012) refer to it as "KF learning".

Remark 3 *Ljung and Soderstrom (1985) show that the CGL (24) can be obtained from a recursive solution of a quadratic loss function. In particular, given a sequence of random variables $\epsilon = \{\varepsilon_1, ..., \varepsilon_T\}$, the optimal choice of the full coefficients' path, that is $\phi = \{\phi_1, ..., \phi_T\}$, can be obtained optimizing with respect to a quadratic criterion function and this leads to the stochastic analog of a Gauss-Newton search direction method*

$$\phi_{t+1} = \phi_t + \kappa_t[H(\phi_t, \varepsilon_t)]^{-1}G(\phi_t, \varepsilon_t),$$

where $G(\phi_t, \varepsilon_t)$ and $H(\phi_t, \varepsilon_t)$ are the Gradient vector and the Hessian matrix respectively, and κ_t is a sequence of appropriately chosen gain parameters. The recursive Gauss-Newton solution for a quadratic criterion function of Ljung and Soderstrom (1985) is equivalent to the Gaussian score-driven filter.

Ljung and Soderstrom (1985, sec. 3.5) consider the possibility of departing from quadratic loss functions. Yet, it is not clear how one should choose the non-linear function in practice. Remark 3 makes the point that the MSE loss function is in fact equivalent to a Gaussian likelihood function. Therefore, the score-driven model (1)-(2) extends the traditional adaptive algorithms by allowing for non-Gaussian distribution, and changes in volatility. This results in a recursive algorithm for a non-quadratic loss function.

In section 3 we have shown how to impose restrictions to the model by amending the score as in (12). Note that constrained algorithms have been usually implemented in the literature by means of the *'projection facility'* (see Ljung and Soderstrom, 1985, Timmermann, 1996, and Evans and Honkapohja, 1998).[31] The adaptive model (1), with filtering (2) and (12), progressively shrinks the incremental step until the restriction is satisfied. In fact, the matrix Ψ_t re-weights the Gauss-Newton search direction so that the restrictions are always satisfied. With respect to the traditional projection facility, the re-weighting here varies at different points of the recursion and, most importantly, shrinks the search in a model consistent way as opposed to the usual ad-hoc shrinkage. The theorem below formalizes this point.

Theorem 2 *Consider the Gaussian model (8). The non-linear transformation $\phi_t = g(\alpha_t)$, introduced in section 3, leads to the score-driven filter (2) with (12) which is equal to the Extended KF of Anderson and Moore (1979, sec. 8.2).*

(Proof in the Appendix A.)

6 Conclusion

In this paper we derive an adaptive algorithm for time-varying autoregressive models in presence of heavy tails. Following Creal et al. (2012) and Harvey (2013), the score of the conditional distribution is the driving process for the evolution of the parameters. In this context we emphasize the importance of allowing for time variation in both parameters and volatilities. Furthermore, the algorithm is extended to incorporate restrictions which are popular in the empirical literature. Specifically, the model is allowed to have a bounded long-run mean and the coefficients are restricted so that the model is locally stationary. The model introduced in this paper does not require the use of simulation techniques and thus has a clear computational advantage especially when restrictions on the parameters are imposed. Moreover, we show that the algorithm obtained under Student-t distribution extends traditional adaptive algorithms well known in the literature.

We apply the algorithm to the study of inflation dynamics. Several alternative specifications are shown to track the data very well, so that they give a parsimonious characterization of

[31] The projection facility is a procedure that constrains the TVP in the neighborhood of a particular solution. In the context of adaptive algorithms, the parameters are restricted so that the model produces stable predictions; see Ljung and Soderstrom (1985, Sections 3.4.4, and 6.6). In practice this is often implemented by skipping the updating each time the restrictions are violated.

the inflation dynamics and produce good forecasts. In particular, allowing for heavy-tails is found to be a key ingredient to obtain well calibrated density forecasts over the analyzed sample. The dynamics of the parameters under Student-t innovations are more robust to short lived variations in inflation, especially in the last decade. Furthermore, the use of heavy-tails highlights the presence of high-frequency variations in volatility on top of the well documented low-frequency variations.

The results of this paper can be extended along various directions. Whereas the empirical analysis is centered around the study of inflation dynamics we suspects that similar gains in forecasting performance extend to other macroeconomic time series. Furthermore, the model can be extended (along the lines of Koop and Korobilis, 2012) to the multivariate case where the dimensions of the model might be so large that the use of MCMC methods is infeasible and imposing stationarity is problematic.

References

Abraham, B. and J. Ledolter (1983). Statistical Model for forecasting. Wiley.

Amisano, G. and R. Giacomini (2007). Comparing Density Forecasts via Weighted Likelihood Ratio Tests. Journal of Business and Economic Statistics, 25, 177-190

Anderson, B.D.O. and J.B. Moore (1979). Optimal Filtering. Prentice Hall Englewood.

Bandorff-Nielsen, O. and G. Shou (1973). On the Parametrization of Autoregressive Models by Partial Autocorrelations. Journal of Multivariate Analysis, 3, 408-419.

Benati, L. (2007). Drift and breaks in labor productivity, Journal of Economic Dynamics and Control, 31(8), 2847-2877.

Benati L. and H. Mumtaz (2007) The U.S. evolving macroeconomic dynamics: a structural investigation. ECB- Working Paper Series 0746.

Benveniste, A., Metivier, M. and P. Priouret (1990). Adaptive algorithms and stochastic approximations. Springer-Verlag.

Berkowitz, J. (2001). Testing Density Forecasts with Applications to Risk Management. Journal of Business and Economic Statistics, 19, 465-474.

Beveridge, S. and C.R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. Journal of Monetary Economics, 7, 151-174.

Blasques F., Koopman S.J. and A. Lucas (2014). Optimal Formulations for Nonlinear Autoregressive Processes. Tinbergen Institute Discussion Papers TI 2014-103/III.

Brainard, W. and G. Perry (2000). Making policy in a changing world. In: Perry, G., Tobin, J. (Eds.), Economic Events, Ideas, and Policies: and Policies: The 1960s and After. Brookings Institution Press, Washington.

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. Review of Economics and Statistics, 69, 542-547.

Branch, W.A. and G.W. Evans (2006). Simple Recursive Forecasting model. Economic Letter, 91, 158-166.

Brown, R.G. (1963). Smoothing, forecasting and prediction. Prentice Hall.

Canova, F. and L. Gambetti (2009). Structural changes in the US economy: Is there a role for monetary policy? Journal of Economic Dynamics and Control, Elsevier, 33(2), 477-490.

Carceles-Poveda, E. and C. Giannitsarou (2007). Adaptive learning in practice. Journal of Economic Dynamics and Control, 3, 2659-2669.

Casals, J., Jerez, M. and S. Sotoca (2002). An Exact Multivariate Model-Based Structural Decomposition. Journal of the American Statistical Association, 97, 553-564.

Chan, J.C.C., Koop, G. and S.M. Potter (2013). A New Model of Trend Inflation. Journal of Business and Economic Statistics, 31(1), 94-106.

Chiu, Ching-Wai, H. Mumtaz and G. Pinter (2014). Fat-tails in VAR Models. Queen Mary, University of London, School of Economics and Finance, WP-714.

Chow, G.C. (1984). Random and Changing Coefficient Models. In Handbook of Econometrics, Vol.2, eds. Z. Griliches and M. D. Intrilligator, Amsterdam North-Holland, 1213-1245

Clark, T.E. and T. Doh (2011). A Bayesian Evaluation of Alternative Models of Trend Inflation. Kansas City FED WP2011-16.

Clark, T.E. and F. Ravazzolo (2012). The macroeconomic forecasting performance of autoregressive models with alternative specifications of time-varying volatility. Cleveland FED, WP-1218.

Cogley, T. (2002). A Simple Adaptive Measure of Core Inflation. Journal of Money, Credit and Banking, 34(1), 94-113.

Cogley, T. and T.J. Sargent (2002). Evolving Post World War II U.S. In Inflation Dynamics. NBER Macroeconomics Annual,16.

Cogley, T. and T.J. Sargent (2005). Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US. Review of Economic Dynamics, 8(2), 262-302.

Cogley, T. and T.J. Sargent (2014). Measuring Price-Level Uncertainty and Instability in the U.S., 1850-2012. Review of Economics and Statistics. Forthcoming.

Cogley, T., Primiceri, G. and T.J. Sargent (2010). Inflation-Gap persistence in the US. American Economic Journal: Macroeconomics, 2(1), 43-69.

Cooley, T. F. and E.C. Prescott (1973). An Adaptive Regression Model. International Economic Review, 14(2), 364-71.

Cooley, T. F. and E.C. Prescott (1976). Estimation in the Presence of Stochastic Parameter Variation. Econometrica, 44(1), 167-84.

Corradi, V. and N.R. Swanson (2006). Predictive Density and Conditional Confidence Interval Accuracy Tests. Journal of Econometrics, 135, (1-2),187-228.

Creal, D., Kopman, S.J. and A. Lucas (2013). Generalised Autoregressive Score Models with Applications. Journal of Applied Econometrics, 28, 777-795.

Cox, D.R. (1981). Statistical analysis of time series: some recent developments. Scandinavian Journal of Statistics, 8(2), 93-115.

Cúrdia, V., Del Negro, M. and D.L. Greenwald (2013). Rare shocks, Great Recessions. Journal of Applied Econometrics. Forthcoming.

D'Agostino, A., Gambetti, L. and D. Giannone (2013). Macroeconomic forecasting and structural change. Journal of Applied Econometrics, 28, 82-101.

Dahlhaus, R. (2012). Locally stationary processes. Handbook of Statistics, Time Series Analysis: Methods and Applications. Ed. T.S. Rao, S.S. Rao and C.R. Rao, (30), 351-413.

Diebold, F.X , Gunther, T.A. and S.A. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. International Economic Review, 39(4), 863-83.

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

Engle, , R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987-1007.

Evans, G.W. and S. Honkapohja (1998). Convergence of learning algorithms without a projection facility. Journal of Mathematical Economics, 30, 59-86.

Evans, G.W. and S. Honkapohja (2001). Learning and Expectations in Macroeconomics. Princeton University Press.

Evans, G.W., Honkapohja, S. and N. Williams (2010). Generalized Stochastic Gradient Learning. International Economic Review, 51, 237-262.

Fagin, S.L. (1964). Recursive linear regression theory, optimal filter theory, and error analysis of optimal systems. IEEE International Conv. Record, 12, 216-240.

Faust, J. and J.H Wright (2013). Forecasting Inflation. Handbook of Economic Forecasting, forthcoming, Vol 2A, Elliott G. and A. Timmermann Eds. North Holland.

Fiorentini, G., Calzolari, G. and L. Panattoni (1996). Analytical Derivatives and the Computation of GARCH Models. Journal of Applied Econometrics, 11, 399-417.

Fiorentini, G., Sentana, E. and G. Calzolari (2003). Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models with Student t Innovations. Journal of Business and Economic Statistics, 21(4), 532-546.

Giraitis, L., Kapetanios, G. and T. Yates (2011). Inference on stochastic time-varying coefficient models. Queen Mary Univeristy of London, WP540.

Giacomini, R. and B. Rossi (2010). Forecast comparisons in unstable environments. Journal of Applied Econometrics, vol. 25(4), pages 595-620.

Giacomini, R. and H. White (2006). Tests of Conditional Predictive Ability. Econometrica, 74(6), 1545-1578.

Hamilton, J.D. (1986). A standard error for the estimated state vector of a state-space model. Journal of Econometrics, 33(3), 387-397.

Harvey, A.C. (1989). Forecasting, structural time series models and Kalman filter. Cambridge University Press.

Harvey, A.C. (2013). Dynamic Models for Volatility and Heavy Tails. With Applications to Financial and Economic Time Series. Cambridge University Press.

Harvey, A.C. and T. Chakravarty, (2008). Beta-t-(E)GARCH. Cambridge University, Faculty of Economics, WP-8340.

Harvey, A.C. and A. Luati (2014). Filtering with Heavy-Tails. Journal of American Statistical Assciation. Forthcoming.

Hyndman, R.J., Koehler, A.B., Ord, J.K. and R.D. Snyder (2008). Forecasting with Exponential Smoothing: The State Space Approach. Springer.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, San Diego.

Justiniano, A. and G. Primiceri (2008). The Time-Varying Volatility of Macroeconomic Fluctuations. American Economic Review, 98(3), 604-641.

Kim, C-J. and C. Nelson (1999). State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. MIT Press.

Koop, G. and D. Korobilis (2009). Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. Foundations and Trends in Econometrics, 3(4), 267–358.

Koop, G. and D. Korobilis (2013). Large Time Varying Parameter VARs. Journal of Econometrics, forthcoming.

Koop, G. and S.M. Potter (2011). Time varying VARs with inequality restrictions. Journal of Economic Dynamics and Control, 35(7), 1126-1138.

Kozicki, S. and P.A. Tinsley (2001). Term structure views of monetary policy under alternative models of agent expectations. Journal of Economic Dynamics and Control, 25(1-2), 149-184.

Kreps, D. (1998). Anticipated Utility and Dynamic Choice. 1997 Schwartz Lecture, in Frontiers of Research in Economic Theory, Edited by D.P. Jacobs, E. Kalai, and M. Kamien. Cambridge University Press.

Li, H. (2008). Estimation and testing of Euler equation models with time-varying reduced-form coefficients. Journal of Econometrics, 142, 425-448.

Ljung, L. (1992). Stochastic approximation and optimization of random system. Georg Pflug, Harro Walk - Basel [etc.]. Birkhauser.

Ljung, L. (1999) System Identification. Theory for the User. Prentice Hall. 2nd Eds.

Ljung, L. and T. Soderstrom (1985). Theory and Practice of Recursive Identification. MIT Press.

Lucas, R. Jr (1976). Econometric policy evaluation: A critique. Carnegie-Rochester Conference Series on Public Policy, 1(1),19-46.

Marcet, A and T. Sargent (1989). Convergence of least squares learning mechanisms in self-referential linear stochastic models. Journal of Economic Theory, 48(2), 337-368.

Maronna, R.A., Martin, R. D. and V.J. Yohai (2006). Robust Statistics: Theory and Methods. Wiley Series in Probability and Statistics.

Monahan, J.F. (1984). A note on enforcing stationarity in autoregressive-moving average models. Biometrika, 71(2), 403-404.

Montgomery, D.C. and L.A. Johnson (1976). Forecasting and time series analysis. Mc Graw-Hill.

Nichols, D.F, and A.R. Pagan (1985). Varying Coefficient Regression. In Handbook of Statistics, Vol. 5, eds. E. J. Hannan P, . R. Krishnaiah, and M. M. Rao, Amsterdam N orth-Holland, 413-449

Pesaran, M.H. and A. Timmermann (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics, Elsevier, 137(1), 134-161.

Pesaran, M.H. and A. Pick (2011). Forecast Combination across Estimation Windows. Journal of Business Economics and Statistics, 29, 307-318.

Pesaran, M.H., Pick, A. and M. Pranovich (2013). Optimal Forecasts in the Presence of Structural Breaks. Journal of Econometrics. Forthcoming.

Pettenuzzo, F. and A. Timmermann (2015). Forecasting Macroeconomic Variables Under Model Instability. Journal of Economics and Business Statistics. Forthcoming

Pivetta, F. and R. Reis (2007). The persistence of inflation in the United States. Journal of Economic Dynamics and Control, 31(4), 1326-1358.

Primiceri, G. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. The Review of Economic Studies, 72, 821-852.

Rossi, B. and T. Sekhposyan (2014). Alternative Tests for Correct Specification of Conditional Predictive Densities. MIMEO. http://www.econ.upf.edu/docs/papers/downloads/1416.pdf

Rosenberg B. (1972) The Estimation of Stationary Stochastic Regression Parameters Reexamined. Journal of the American Statistical Association, 67, 650-654.

Sargent, T.J. (1999). The Conquest of American Inflation. Princeton University Press.

Sargent, T.J., and N. William (2005). Impacts of priors on convergence and escapes from Nash inflation. Review of Economic Dynamics, 8, 360-391.

Sargent, T.J., William, N. and T. Zha (2006). Shocks and Government Beliefs: The Rise and Fall of American Inflation, American Economic Review, 96(4), 1193-1224.

Sarris, A.H. (1973). A Bayesian Approach to Estimate Time Varying Regression Coefficients. Annals of Economic and Social Measurement, 2, 501-526

Slobodyan S. and R. Wouters (2012). Learning in an estimated medium-scale DSGE model. Journal of Economic Dynamics and Control, 36(1), 26-46.

Stock, J. (2002). Comments on Evolving Post World War II U.S. In Inflation Dynamics. NBER Macroeconomics Annual, 16.

Stock, J. and M. Watson (1996). Evidence on Structural Instability in Macroeconomic Time Series Relations. Journal of Business and Economic Statistics, 14(1), 14-30.

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

Svensson, L. E. O. (2005). Monetary Policy with Judgment: Forecast Targeting, International Journal of Central Banking, 1(1), May.

Taylor, J. (2000). Low inflation, pass-through, and the pricing power of firms. European Economic Review 44 (7), 1389–1408.

Timmermann, A. (1996). Excess volatility and predictability of stock prices in autoregressive dividend models with learning. The Review of Economic Studies, 63, 523-557.

Watson, M. W and Engle, R. F. (1985). Testing for Regression Coefficient Stability with a Stationary AR(1) Alternative. The Review of Economics and Statistics, 67(2), 341-46.

Figures and Tables

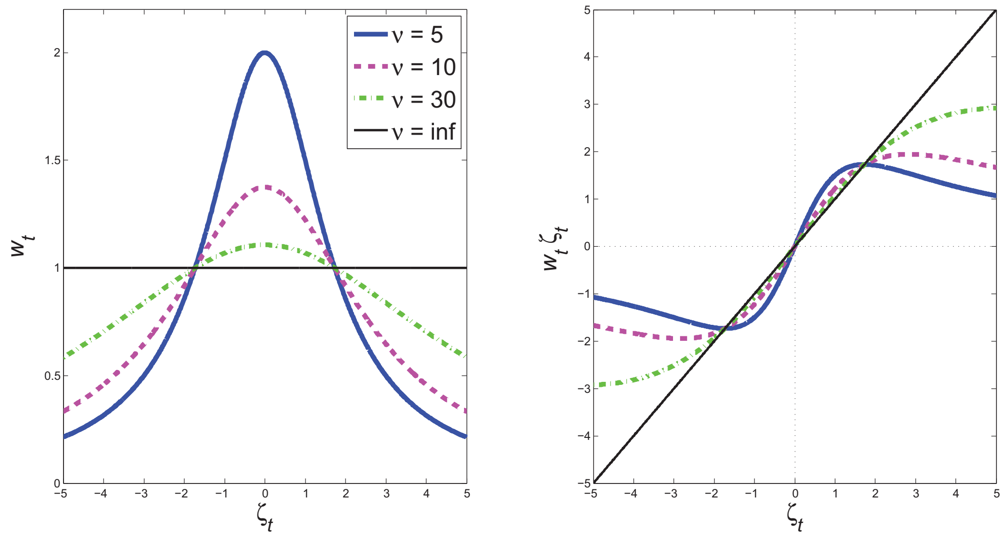

Figure 1: The left panel plots of the weights w_t against the standardized errors $\zeta_t = \varepsilon_t/\sigma_t$ for different values of the degrees of freedom υ. The right panel, instead, plots of the weighted standardized erros, $w_t\zeta_t$, knonw as influnce function, against the standardized errors $\zeta_t = \varepsilon_t/\sigma_t$.

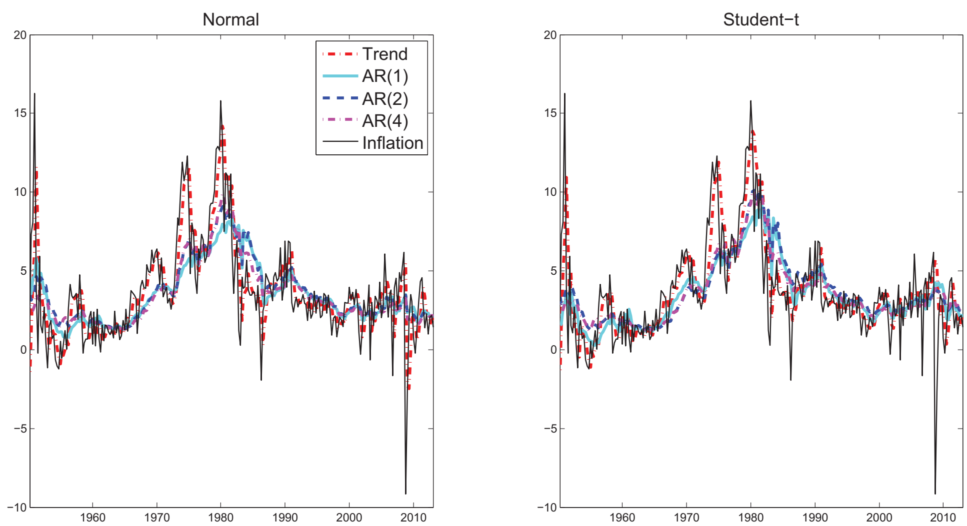

Figure 2: Implied "long-run" inflation, $\mu_t = \phi_{0,t}/(1 - \sum_{j=1}^{p} \phi_{j,t})$, together with the realized inflation: left panel Gaussian models, right panel Student-t models.

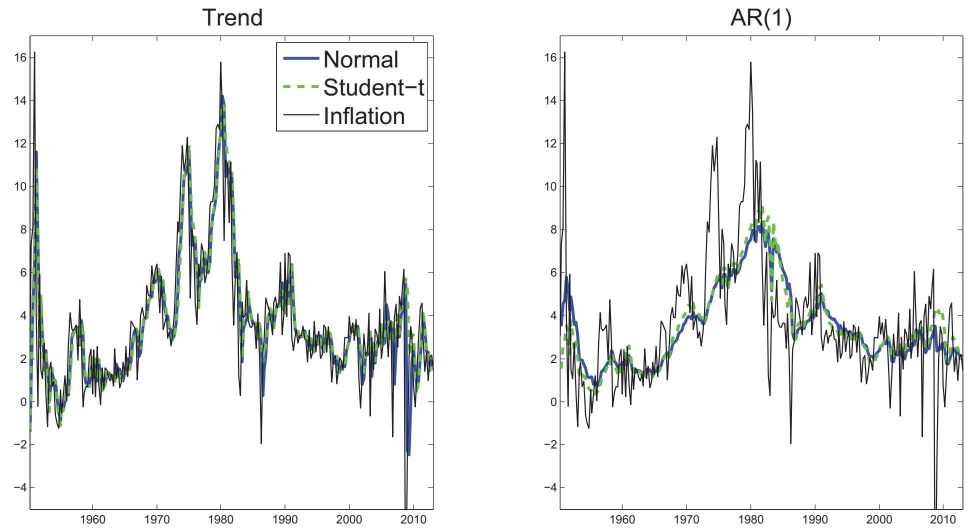

Figure 3: The implied "long-run" inflation $\mu_t = \phi_{0,t}/(1 - \sum_{j=1}^{p} \phi_{j,t})$, together with the realized inflation for various specifications: left panel trend-only models, right panel AR(1) models.

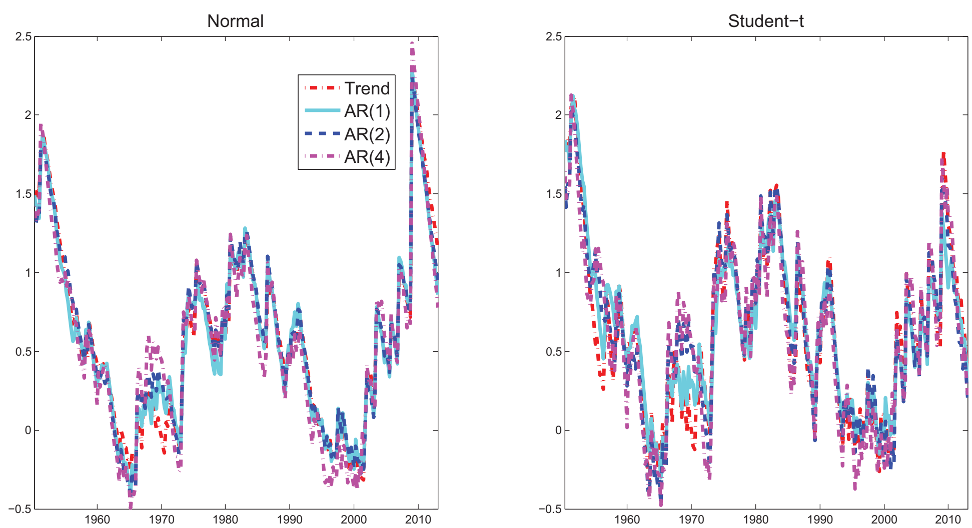

Figure 4: Implied volatility, $\log \sigma_t$, for different specifications: left panel Gaussian models, right panel Student-t models.

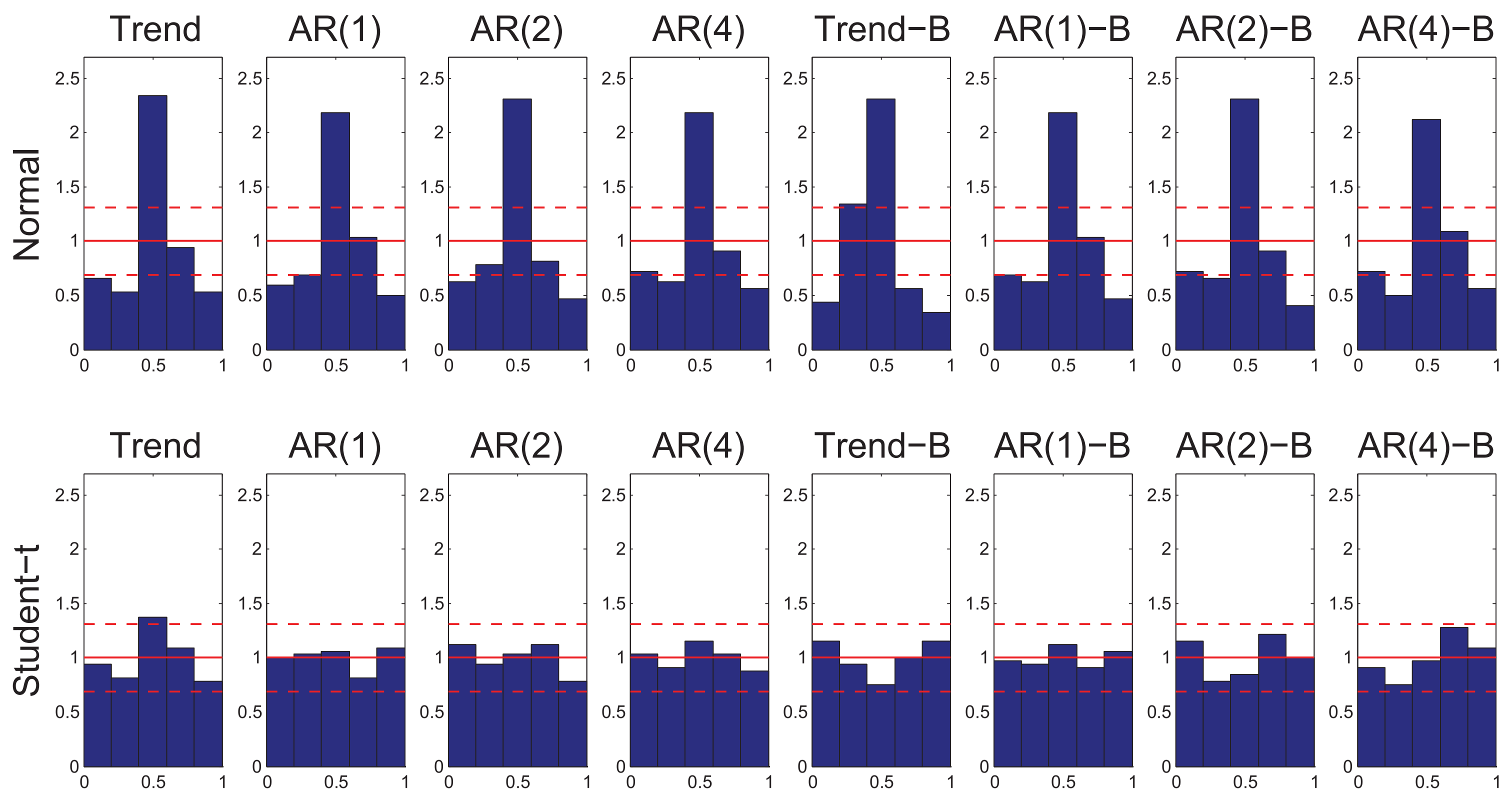

Figure 5: The p.d.f. of the PITs (normalized) and the 95% critical values (dashed lines) constructed using a Normal approximation as in Diebold et al. (1998).

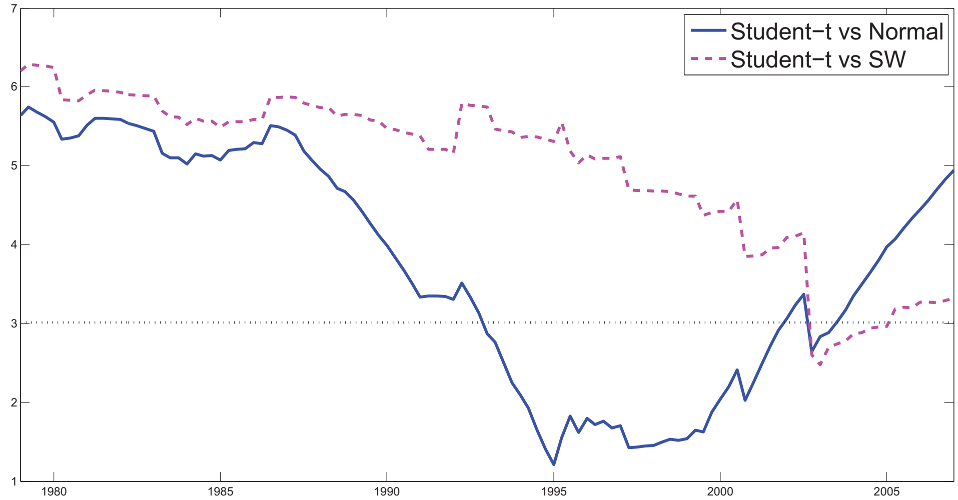

Figure 6: Fluctuation test statistics (Giacomini and Rossi, 2010) and the 5% critical value of the two sided test. The test is computed over a window of 4 years, the dates on the x-axis correspond to the mid-point of the window. Positive values of the fluctuation statistic imply that the Student-t model outperform the alternative (Normal, continuous line, and SW, broken line).

	Normal					Student-t					
	κ_c	κ_σ	LL	AIC	BIC	κ_c	κ_σ	v	LL	AIC	BIC
Trend	0.5309	0.1467	-549.1139	1102.2277	1109.3491	0.4707	0.2116	5.3309	-523.1822	1052.3645	1063.0465
	(0.0214)	(0.0225)				(0.0700)	(0.0847)	(1.2107)			
AR(1)	0.1483	0.1830	-541.1469	1086.2937	1093.4151	0.1704	0.1697	5.1371	-519.5975	1045.1951	1055.8771
	(0.0227)	(0.0233)				(0.0003)	(0.0003)	(0.0019)			
AR(2)	0.1168	0.1661	-551.3741	1106.7481	1113.8695	0.0978	0.2171	5.7377	-526.5179	1059.0359	1069.7179
	(0.0215)	(0.0229)				(0.0001)	(0.0001)	(0.0005)			
AR(4)	0.1360	0.2144	-544.2799	1092.5598	1099.6811	0.1381	0.2598	6.2070	-520.5114	1047.0227	1057.7048
	(0.0224)	(0.0233)				(0.0001)	(0.0001)	(0.0006)			
Trend-B	0.2055	0.4870	-604.3270	1212.6541	1219.7754	0.8197	0.2461	5.8753	-561.7474	1129.4949	1140.1769
	(0.0221)	(0.0203)				(0.1649)	(0.1137)	(1.4062)			
AR(1)-B	0.0861	0.2831	-535.9191	1075.8381	1082.9595	0.0735	0.2624	4.2080	-520.5671	1047.1342	1057.8163
	(0.0201)	(0.0232)				(0.0000)	(0.0000)	(0.0001)			
AR(2)-B	0.0976	0.2669	-535.4122	1074.8245	1081.9458	0.0683	0.2527	4.7426	-520.7939	1047.5879	1058.2699
	(0.0209)	(0.0233)				(0.0000)	(0.0000)	(0.0001)			
AR(4)-B	0.0960	0.2729	-545.2302	1094.4603	1101.5817	0.0867	0.2952	5.6393	-521.3150	1048.6299	1059.3120
	(0.0206)	(0.0230)				(0.0000)	(0.0000)	(0.0003)			

Table 1: Estimation of the score-driven model $\pi_t = \phi_{0,t} + \sum_{j=1}^{p} \phi_{j,t}\pi_{t-j} + \varepsilon_t$, where $\pi_t = \Delta \log p_t$ is annualized quarterly US CPI inflation 1955Q1-2012Q4. "Trend" denotes the specification with $p = 0$, "B" denotes the specifications with restricted long-run mean, and "AR(p)" denotes the specification with lags order equal to p. "Normal" denotes $\varepsilon_t \sim N(0, \sigma_t^2)$ and "Student-t" that $\varepsilon_t \sim t_v(0, \sigma_t^2)$. The estimated static parameters are κ_c, κ_σ and v (std. error in brackets). "LL" is the Log-likelihood, AIC and BIC are the information criteria.

	RMSE			MAE		
	h=1	h=4	h=8	h=1	h=4	h=8
SW	2.4028	2.9892	3.1921	1.6170	2.0763	2.2990
<u>Normal</u>						
Trend	**0.9085**	0.9384	1.0383	**0.8985**	0.9783	1.0422
	(0.0288)	(0.2461)	(0.5011)	**(0.0373)**	(0.6768)	(0.3542)
AR(1)	**0.8768**	**0.8856**	0.8989	**0.8961**	0.9291	0.9034
	(0.0038)	**(0.0615)**	(0.2619)	**(0.0361)**	(0.2867)	(0.2837)
AR(2)	**0.8891**	**0.8700**	0.9509	**0.8947**	0.9246	0.9501
	(0.0067)	**(0.0323)**	(0.5338)	**(0.0239)**	(0.2397)	(0.5370)
AR(4)	**0.8623**	**0.8502**	0.9232	**0.8709**	0.9086	0.9192
	(0.0036)	**(0.0162)**	(0.2545)	**(0.0158)**	(0.1319)	(0.2023)
Trend-B	1.3063	1.0742	0.9434	1.4363	1.1597	1.0192
	(0.0002)	(0.4447)	(0.6514)	(0.0000)	(0.1394)	(0.8839)
AR(1)-B	0.9468	0.9023	0.9169	0.9813	0.9280	0.9231
	(0.3051)	(0.1773)	(0.4783)	(0.7570)	(0.3464)	(0.4878)
AR(2)-B	0.9316	**0.8795**	0.9100	0.9520	0.9137	0.8950
	(0.1089)	**(0.0602)**	(0.3752)	(0.3594)	(0.1953)	(0.2582)
AR(4)-B	1.0428	0.9024	0.9258	0.9729	0.9061	0.8963
	(0.7414)	(0.1395)	(0.5375)	(0.7004)	(0.1411)	(0.3153)
<u>Student-t</u>						
Trend	0.9258	0.9270	1.0277	**0.8855**	0.9580	1.0146
	(0.1430)	(0.2127)	(0.7077)	**(0.0275)**	(0.4431)	(0.8111)
AR(1)	**0.8783**	**0.9005**	0.9176	**0.8885**	0.9492	0.9109
	(0.0035)	**(0.0638)**	(0.3237)	**(0.0257)**	(0.3925)	(0.2912)
AR(2)	**0.9180**	**0.8973**	0.9902	**0.9153**	0.9495	0.9759
	(0.0418)	**(0.0684)**	(0.9007)	**(0.0598)**	(0.4131)	(0.7684)
AR(4)	**0.8686**	**0.8523**	0.9188	**0.8728**	0.9040	0.9063
	(0.0049)	(0.0116)	(0.2621)	**(0.0170)**	(0.1145)	(0.2151)
Trend-B	1.2731	1.0424	0.9137	1.3233	1.0696	0.9626
	(0.0010)	(0.6623)	(0.4913)	(0.0001)	(0.4806)	(0.7584)
AR(1)-B	**0.8694**	0.8818	0.8975	**0.8910**	0.8822	0.8550
	(0.0041)	(0.1002)	(0.4041)	**(0.0397)**	(0.1136)	(0.1976)
AR(2)-B	**0.9219**	0.8976	0.9476	0.9599	0.9143	0.9078
	(0.0558)	(0.1145)	(0.6079)	(0.3904)	(0.1879)	(0.3403)
AR(4)-B	0.9292	0.9018	0.9645	0.9453	0.9228	0.9362
	(0.2149)	(0.1215)	(0.7448)	(0.3176)	(0.2035)	(0.4991)

Table 2: Point forecast for US CPI inflation (forecast sample: 1973Q1–2012Q4). The Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are expressed in relative term with respect to SW's model. The forecast horizon is "h". In brackets are the p-values of Giacomini and White's (2006) test (in bold when it is significant at 10% level).

	LR	RS	ALogS	AG
SW	0.0000	9.8010	-2.6837	–
Normal				
Trend	0.1445	4.4223	-2.5591	0.1281
AR(1)	0.0041	3.7823	-2.4857	0.0033
AR(2)	0.0058	4.6923	-2.5357	0.0340
AR(4)	0.1455	3.4810	-2.4663	0.0037
Trend-B	0.0581	7.8323	-3.1073	0.0000
AR(1)-B	0.0046	3.1923	-2.4543	0.0001
AR(2)-B	0.4572	3.7210	-2.6275	0.6807
AR(4)-B	0.9784	4.0960	-2.6022	0.4025
Student-t				
Trend	0.7000	0.7023	-1.5897	0.0000
AR(1)	0.9976	0.1323	-1.6325	0.0000
AR(2)	0.5005	0.5522	-1.6249	0.0000
AR(4)	0.7914	0.2560	-1.5976	0.0000
Trend-B	0.0048	0.2723	-1.5688	0.0000
AR(1)-B	0.9946	0.0423	-1.6766	0.0000
AR(2)-B	0.7590	0.3610	-1.6671	0.0000
AR(4)-B	0.7549	1.3323	-1.6272	0.0000

Table 3: Density Forecast for US CPI inflation (forecast sample: 1973Q1-2012Q4). The first column reports the p-values of the LR test of Berkowitz (2001). The second column reports Rossi and Sekhposyan's (RS) test (whose critical values are 2.25 at 1%, 1.51 at 5% and 1.1 at 10%). The third column reports the Average Log Score (ALogS), whereas the forth column reports the p-values of the test by Amisano and Giacomini (2007) (AG) of the difference in the ALogS of each of the models against SW's model.

	PCE Deflator				GDP Deflator			
	LR	RS	ALogS	AG	LR	RS	ALogS	AG
SW	0.0000	11.7052	-2.4525	–	0.0000	13.36824	-2.31589	–
Normal								
Trend	0.0228	2.8438	-2.2493	0.1049	0.0000	0.4257	-2.2520	0.8587
AR(1)	0.2778	2.5959	-2.1650	0.0014	0.0000	3.7489	-1.8543	0.0019
AR(2)	0.2341	2.7755	-2.1993	0.0179	0.0000	2.8148	-1.8785	0.0137
AR(4)	0.1344	3.6585	-2.2365	0.0763	0.0000	1.0162	-3.1160	0.3351
Trend-B	0.0675	3.1976	-2.6070	0.1922	0.0000	0.7450	-1.8103	0.0050
AR(1)-B	0.2527	2.7755	-2.1260	0.0002	0.0000	0.6661	-3.2007	0.2495
AR(2)-B	0.6046	3.3220	-2.1131	0.0019	0.0000	5.4737	-1.7326	0.0000
AR(4)-B	0.0010	2.1156	-2.2863	0.2371	0.0000	5.1134	-1.7403	0.0000
Student-t								
Trend	0.4297	0.3752	-1.9762	0.2622	0.7261	0.1678	-1.5666	0.0000
AR(1)	0.7831	0.0597	-1.6270	0.0000	0.1902	2.8942	-1.4467	0.0000
AR(2)	0.9807	0.2194	-1.6399	0.0000	0.1929	1.9591	-1.4768	0.0000
AR(4)	0.9768	0.1640	-1.6966	0.0000	0.4625	0.7262	-1.6837	0.0000
Trend-B	0.0325	0.4773	-1.5388	0.0000	0.1201	0.2450	-1.5648	0.0000
AR(1)-B	0.2698	0.3656	-1.6297	0.0000	0.6121	0.6998	-1.4700	0.0000
AR(2)-B	0.8188	0.6919	-1.8845	0.0000	0.2336	1.2059	-1.4848	0.0000
AR(4)-B	0.4439	0.1802	-1.6182	0.0000	0.0082	0.6471	-1.5115	0.0000

Table 4: Density Forecast for US PCE and GDP Deflators (forecast sample: 1973Q1-2012Q4). For each of them The first column reports the p-values of the LR test of Berkowitz (2001). The second column reports Rossi and Sekhposyan's (RS) test (whose critical values are 2.25 at 1%, 1.51 at 5% and 1.1 at 10%). The third column reports the Average Log Score (ALogS), whereas the forth column reports the p-values of the test by Amisano and Giacomini (2007) (AG) of the difference in the ALogS of each of the models against SW's model.

	Normal			Student-t			
	LR	RS	ALogS	LR	RS	ALogS	AG
CA	0.0000	5.7263	-2.4568	0.5268	0.3567	-1.4456	0.0000
FR	0.0000	1.6459	-2.3982	0.4827	0.8434	-1.4296	0.0109
DE	0.0606	1.7418	-1.9100	0.0624	0.4414	-1.4569	0.0000
IT	0.0000	2.3119	-2.5871	0.1605	1.2523	-1.5183	0.0003
JP	0.0048	5.1273	-2.5986	0.4313	1.0006	-1.5212	0.0000
UK	0.0336	7.5950	-2.7366	0.8862	0.3836	-1.6319	0.0000

Table 5: Density Forecast of the "Trend" specification for CPI inflation of the other G7 countries (forecast sample: 1973Q1-2012Q4). 'LR' denotes Berkowitz's (2001) test, 'RS' denotes Rossi and Sekhposyan's (RS) test (whose critical values are 2.25 at 1%, 1.51 at 5% and 1.1 at 10%), 'ALogS' denotes the Average Log Score. The last column reports the p-values of the test by Amisano and Giacomini (2007) (AG) of the difference in the ALogS between the Student-t and the Gaussian model.

Appendix A: Proofs

Proofs for Section 2

The score vector: Following Fiorentini et al. (2003), we re-write the predictive log-likelihood (4) as follows

$$\ell_t\left(\mathcal{F}_t, \theta\right) = c + d_t + g_t$$

with

$$c = \log\left[\Gamma\left(\frac{\eta+1}{2\eta}\right)\right] - \log\left[\Gamma\left(\frac{1}{2\eta}\right)\right] - \frac{1}{2}\log\left(\frac{1-2\eta}{\eta}\right) - \frac{1}{2}\log\pi,$$

and

$$d_t = -\frac{1}{2}\log\sigma_t^2,\ g_t = -\left(\frac{\eta+1}{2\eta}\right)\log\left[1 + \frac{\eta}{1-2\eta}\zeta_t^2\right],$$

where $\zeta_t = \varepsilon_t/\sigma_t$ and $\Gamma(.)$ is the Euler's gamma function. Let $\nabla_t = \partial\ell_t(\mathcal{F}_t, \theta)/\partial f_t$ denote the gradient function and partition it in two blocks, ∇_ϕ and ∇_σ, the first one depend upon g_t and ζ_t, while the second upon d_t, g_t and ζ_t. We have to compute $\frac{\partial g_t}{\partial\phi_t'} = \frac{\partial g_t}{\partial\zeta_t^2}\frac{\partial\zeta_t^2}{\partial\phi_t'}$, where

$$\begin{aligned}
\frac{\partial g_t}{\partial\zeta_t^2} &= -\left(\frac{\eta+1}{2\eta}\right)\frac{\partial\ln\left[1+\frac{\eta}{1-2\eta}\zeta_t^2\right]}{\partial\zeta_t^2} = -\left(\frac{\eta+1}{2\eta}\right)\left[1+\frac{\eta}{1-2\eta}\zeta_t^2\right]^{-1}\left(\frac{\eta}{1-2\eta}\right) \\
&= -\frac{\eta+1}{2\left(1-2\eta\right)}\left[\frac{1-2\eta+\eta\zeta_t^2}{1-2\eta}\right]^{-1} = -\frac{\eta+1}{2(1-2\eta+\eta\zeta_t^2)}
\end{aligned}$$

and $\frac{\partial\zeta_t^2}{\partial\phi_t'} = -\frac{2x_t'\varepsilon_t}{\sigma_t^2}$. The score for the coefficients of the model is then equal to

$$\nabla_\phi = \frac{\partial g_t}{\partial\zeta_t^2}\frac{\partial\zeta_t^2}{\partial\phi_t} = x_t\frac{\left(\eta+1\right)\varepsilon_t/\sigma_t^2}{(1-2\eta+\eta\varepsilon_t^2/\sigma_t^2)}.$$

The gradient for the variance component is

$$\nabla_\sigma = \frac{\partial d_t}{\partial\sigma_t^2} + \frac{\partial g_t}{\partial\sigma_t^2} = \frac{\partial d_t}{\partial\sigma_t^2} + \frac{\partial g_t}{\partial\zeta_t^2}\frac{\partial\zeta_t^2}{\partial\sigma_t^2},$$

where $\frac{\partial\zeta_t^2}{\partial\sigma_t^2} = -\frac{\varepsilon_t^2}{\sigma_t^4}$ and thus we obtain

$$\nabla_\sigma = -\frac{1}{2\sigma_t^2} + \frac{(\eta+1)\varepsilon_t^2/\sigma_t^4}{2(1-2\eta+\eta\zeta_t^2)} = \frac{1}{2\sigma_t^4}\left[\frac{(\eta+1)}{(1-2\eta+\eta\zeta_t^2)}\varepsilon_t^2 - \sigma_t^2\right].$$

We compute the information matrix as $I_t = -E(H_t)$, where H_t the Hessian matrix and it can be partitioned in four blocks

$$H_t = \begin{bmatrix} H_{\phi\phi,t} & H_{\phi\sigma,t} \\ H_{\sigma\phi,t} & H_{\sigma\sigma,t} \end{bmatrix}.$$

The first block $H_{\phi\phi,t}$ can be calculated as

$$H_{\phi\phi,t} = \frac{\partial \nabla_{\phi,t}}{\partial \phi_t'} = \frac{(1+\eta)\left[\eta\zeta_t^2 + 2\eta - 1\right]}{(1-2\eta+\eta\zeta_t^2)^2}\frac{x_t x_t'}{\sigma_t^2}.$$

Recalling that $\varepsilon_t/\sigma_t = \zeta_t \sim t_\upsilon(0,1)$ implies that $\zeta_t = \sqrt{\frac{(\upsilon-2)\varsigma_t}{\xi_t}}u_t$, where u_t is uniformly distributed on the unit set, ς_t is a chi-squared random variable with 1 degree of freedom, ξ_t is a gamma variate with mean $\upsilon > 2$ variance 2υ, and u_t, ς_t and ξ_t are mutually independent. Therefore, it is possible to show that

$$\mathcal{I}_{\phi\phi,t} = -\mathbb{E}(H_{\phi\phi,t}) = \frac{(1+\eta)}{(1-2\eta)(1+3\eta)}\frac{x_t x_t'}{\sigma_t^2}.$$

The Hessian with respect to the volatility is

$$H_{\sigma\sigma,t} = \frac{\partial \nabla_\sigma}{\partial \sigma_t^2} = \frac{1}{2\sigma_t^4} - \frac{\left[2(1-2\eta) + \eta\varepsilon_t^2/\sigma_t^2\right](\eta+1)\,\varepsilon_t^2/\sigma_t^6}{2[1-2\eta+\eta\varepsilon_t^2/\sigma_t^2]^2},$$

and

$$\mathcal{I}_{\sigma\sigma,t} = -\mathbb{E}(H_{\sigma\sigma,t}) = \frac{(1+\eta)}{2(3+\eta)\,\sigma_t^4} - \frac{\eta}{2(3+\eta)\,\sigma_t^4} = \frac{1}{2(1+3\eta)\,\sigma_t^4}.$$

The cross-derivative in the Hessian is $H_{\phi\sigma,t} = -x_t\frac{\varepsilon_t}{\sigma_t^4}$ and therefore $I_{\phi\sigma,t} = -E(H_{\phi\sigma,t}) = 0$. Finally, the information matrix is equal to

$$\mathcal{I}_t = \begin{bmatrix} \frac{(1+\eta)}{(1-2\eta)(1+3\eta)}\frac{1}{\sigma_t^2}x_t x_t' & 0 \\ 0' & \frac{1}{2(1+3\eta)\sigma_t^4} \end{bmatrix},$$

and the final expression for the scaled score vector is

$$s_t = \mathcal{I}_t^{-1}\nabla_t = \begin{bmatrix} s_{\phi t} \\ s_{\sigma t} \end{bmatrix} = \begin{bmatrix} \frac{(1-2\eta)(1+3\eta)}{(1-2\eta+\eta\zeta_t^2)} \mathcal{S}_t^{-1} \frac{1}{\sigma_t^2} x_t \varepsilon_t \\ (1+3\eta)\left[\frac{(1+\eta)}{(1-2\eta+\eta\zeta_t^2)}\varepsilon_t^2 - \sigma_t^2\right] \end{bmatrix}.$$

where $S_t = \frac{1}{\sigma_t^2} x_t x_t'$

Estimated trend: Considering the model (1) with time varying mean only

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim t_v(0, \sigma_t^2).$$

Let assume that w_t are exogenously give, the estimated level is

$$\begin{aligned} \mu_{t+1} &= \mu_t + \kappa_\theta w_t (y_t - \mu_t) = (1 - \kappa_\theta w_t)\mu_t + \kappa_\theta w_t y_t \\ &= \frac{\kappa_\theta}{1 - \kappa_\theta w_t L} w_t y_t = \kappa_\theta \sum_{j=0}^{\infty} \gamma_j w_{t-j} y_{t-j}, \end{aligned}$$

with $\kappa_\theta = \kappa_\phi \frac{(1-2\eta)(1+3\eta)}{(1+\eta)}$. After a bit of algebra, we can obtain explicit expression the weights across time that is

$$\gamma_0 = 1 \text{ and } \gamma_j = \prod_{k=t-j+1}^{t} (1 - \kappa_\theta w_k).$$

The same weighting pattern is obtained when regressors are included. Since the weights across time are affected by the cross sectional weights w_t, we can not obtained the robust filter for μ_{t+1} as solution of a re-weighted quadratic criterion function as Ljung and Sostrestrom (1985, sec. 2.6.2). In general, when we depart from Gaussianity the stochastic Newton-Gradient algorithm cannot be obtained as a recursive solution of a quadratic criterion function. For the variance is straightforward to obtain the expression for the variance its implied weighting pattern.

Proofs for Section 3

Theorem 1 For simplicity we drop the temporal subscript t such that the $p \times p$ Jacobian matrix is

$$\Gamma = \frac{\partial \Phi(\rho)}{\partial \rho'}.$$

The first $(p-1)$ coefficients are obtained from last recursion in (14), and the last coefficients is equal to the last partial autocorrelation ρ_p. We denote the final vector of coefficients as $\phi_p = (\phi^{1,p}, ..., \phi^{p-1,p}, \phi^{p,p})' = (a'_p, \rho_p)$, where $a_p = (\phi^{1,p}, ..., \phi^{p-1,p})$ and $\phi^{p,p} = \rho_p$. Therefore, we can express the last iteration of (14) in matrix form $a_p = J_{p-1}\phi_{p-1}$, where $\phi_{p-1} = (\phi^{1,p-1}, ..., \phi^{p-2,p-1}, \phi^{p-1,p-1})' = (a'_{p-1}, \rho_{p-1})'$ and

$$J_{p-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -\rho_p \\ 0 & \ddots & & & 0 \\ \vdots & & . & & \vdots \\ 0 & & & \ddots & 0 \\ -\rho_p & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Note that if p is even the central element of J_{p-1} is $1 - \rho_p$. Moreover, the vector $\widetilde{\phi}_p = (\phi'_{p-1}, \rho_p)'$ contains all the partial autocorrelations, i.e. $\widetilde{\phi}_p = (a'_{p-1}, \rho_{p-1}, \rho_p)$ and keep substituting we obtain $\widetilde{\phi}_p = \rho_p = (\rho_1, ..., \rho_{p-1}, \rho_p)$. The Jacobian matrix can be expressed as follows

$$\Gamma = \Gamma_p = \begin{bmatrix} \frac{\partial a_p}{\partial \phi'_{p-1}} & \frac{\partial a_p}{\partial \rho_p} \\ \frac{\partial \rho_p}{\partial \phi'_{p-1}} & \frac{\partial \rho_p}{\partial \rho_p} \end{bmatrix}.$$

The upper-left block is a $(p-1) \times (p-1)$ matrix and it can be computed using the definition $a_p = J_{p-1}\phi_{p-1}$; since J_{p-1} contains the last partial correlation ρ_p we have the recursive formulation

$$\frac{\partial a_p}{\partial \phi'_{p-1}} = J_{p-1}\Gamma_{p-1}$$

where $\Gamma_{p-1} = \partial\phi_{p-1}/\partial\rho_{p-1}$ is the Jacobian of the first $p-1$ coefficients with respect to the first $p-1$ partial autocorrelations. Finally, we have that the other three blocks are

$$\frac{\partial\rho_p}{\partial a'_{p-1}} = 0', \ \frac{\partial\rho_p}{\partial\rho_p} = 1 \text{ and } \frac{\partial a_p}{\partial\rho_p} = \frac{\partial J_{p-1}}{\partial\rho_p}\phi_{p-1} = \begin{bmatrix} -\phi^{p-1,p-1} \\ -\phi^{p-2,p-1} \\ \vdots \\ -\phi^{1,p-1} \end{bmatrix}.$$

Note that ϕ_{p-1} is a given and $\frac{\partial J_{p-1}}{\partial\rho_p} = antidiag(-1, ..., -1)$ inverts the order of elements in $\phi_{p-1} = (\phi^{1,p-1}, ..., \phi^{p-2,p-1}, \phi^{p-1,p-1})'$ with opposite sign.

Proofs for Section 5

Remark 1 The discounted regression model has been extensively used in the adaptive control literature (see Brown, 1963, Montgomery and Johnson, 1976, and Abraham and Ledolter, 1983). Similarly, in the engineering literature the same algorithm is known as forgetting factor algorithm. Fagin (1964) notes that a given linear state space model might be adequate for a time period but may not be for long time intervals and therefore proposes to robustify the KF using an exponentially decay forgetting factor labelled as fading memory (or limited memory) filter (see Jazwinski, 1970, p. 255). Following Ljung and Soderstrom (1985, section 2.6.2), the recursive estimation of the CGL can be obtained from an off-line identification approach that minimizes the weighted sum of squared errors

$$S_t(\phi_t) = \sum_{j=1}^{t} \gamma_j \left(y_{t-j} - x'_{t-j}\phi_t\right)^2,$$

where $\gamma_j = \prod_{k=j+1}^{t} \delta_k$ is a sequence of weights assign to the observation y_{t-j}. Setting $\delta = (1-\kappa)$, where $\delta \leq 1$ is known as the forgetting factor, the observations are weighted exponentially, i.e. $\gamma_j = (1-\kappa)^j$, and the gain parameter is equal to $\left[\sum_{j=1}^{t} \gamma_j\right]^{-1} \to \kappa$. Thus, the CGL can be seen as a recursive estimation of the discounted least squares and it generalizes the exponential smoothing of Hyndman et al. (2008) when explanatory variables are included. Under time-varying parameters model the constant gain κ regulates the tracking ability (large κ) and the noise insensitivity (small κ). On the other hand, for $\kappa = 1/t$ we obtain the recursive least squares and the parameters variation vanishes asymptotically.

Lemma 1 Ljung (1992, p. 99) and Sargent (1999, p. 115) show how to obtain the CGL from the KF applied to the restricted state space model. It is worth to show that the restrictions imply that $\eta_t = c(\phi_{t|t} - \phi_t)$, where $c = [\kappa/(1-\kappa)]^{1/2}$. Consequently, the transition equation in (25) is equal to $\phi_{t+1} = (1-c)\phi_t + c\phi_{t|t}$ and the true state vector can be expressed as exponential weighted average of past filter estimates

$$\phi_{t+1} = c\sum_{j=0}^{t-1}(1-c)^j \phi_{t-j|t-j}.$$

Moreover, the filter estimate can be expressed as

$$\phi_{t|t} = L_t\phi_{t-1|t-1} + K_t y_t = \sum_{j=0}^{t-1}\left(\prod_{i=0}^{j-1} L_{t-i}\right) K_{t-j} y_{t-j}$$

where

$$L_t = (I - K_t x_t'), \quad K_t = P_{t|t-1}x_t\left(x_t' P_{t|t-1} x_t + \frac{\sigma^2}{1-\kappa}\right)^{-1}.$$

Thus, differently from the parameter-driven model, the Kalman gain does not depend on any unobserved shock and it rather obtained from past observations only. Therefore, those restrictions leads to have time-varying coefficients that are driven by past observations only.

Lemma 2 Setting $Q_t := \kappa^2\Sigma$, with $\Sigma = \sigma^2 E[(x_t x_t')]^{-1}$, we have that the shock driving the time-varying coefficients is

$$\eta_t = \kappa(x_t x_t')^{-1} x_t \varepsilon_t = \kappa(x_t x_t')^{-1} x_t \varepsilon_t.$$

Therefore, the parameter-driven model collapses to an observation-driven model. Moreover, up to a scalar factor, the shock η_t is equal to the driving process of our score driven model. However, under the parameter driven framework the vector of coefficients is considered as unobserved state vector which is optimally estimated by the mean of KF which leads to

$$\begin{aligned}
\phi_{t+1|t} &= \phi_{t|t-1} + P_{t|t-1}x_t(x_t'P_{t|t-1}x_t + \sigma^2)^{-1}(y_t - x_t'\phi_{t|t-1}) \\
P_{t+1|t} &= P_{t|t-1} - P_{t|t-1}x_t(x_t'P_{t|t-1}x_t + \sigma^2)^{-1}x_t'P_{t|t-1} + \kappa^2\Sigma.
\end{aligned}$$

Following Benveniste et al. (1990, p. 139), for $\kappa^2 \ll \sigma^2$ meaning that the variance drifting parameters is much smaller than the variance model disturbances, for $t > \mathring{t}$, where $\mathring{t}$ is a given large value of t, one has the approximation $(x_t'P_{t|t-1}x_t + \sigma^2) \approx \sigma^2$, this implies that the conditional variance of the forecast error converges to the variance of model disturbances. For t large enough, the variation of $P_{t|t-1}$ is small with respect to x_t, and $x_t'P_{t|t-1}x_t$ can be neglected with respect to σ^2. Using these approximations, we obtain

$$\begin{aligned}
\phi_{t+1|t} &= \phi_{t|t-1} + P_{t|t-1}x_t\frac{1}{\sigma^2}(y_t - x_t'\phi_{t|t-1}) \\
P_{t+1|t} &= P_{t|t-1} - P_{t|t-1}x_t\frac{1}{\sigma^2}x_t'P_{t|t-1} + \kappa^2\Sigma.
\end{aligned}$$

Replacing $x_t x_t'/\sigma^2$ with its expected value Σ^{-1} we obtain $P_{t+1|t} = P_{t|t-1} - P_{t|t-1}\Sigma^{-1}P_{t|t-1} + \kappa^2\Sigma$. When $P_{t|t-1}$ is set to its steady-state value P as in Harvey (1989, p. 118), one has $P\Sigma^{-1}P = \Lambda\Sigma\Lambda \Rightarrow \kappa^{-2}P\Sigma^{-1}P = \Sigma \Rightarrow \frac{1}{\kappa}P = \Sigma$. Thus we obtain

$$\phi_{t+1|t} = \phi_{t|t-1} + \kappa\Sigma x_t\frac{1}{\sigma^2}(y_t - x_t'\phi_{t|t-1}),$$

which has the same asymptotic behavior of the CGL; see Sargent and William (2005) and Evans et al. (2010). Similarly, setting $Q_t := \kappa^2\Sigma^{-1}$, we have that $\eta_t = \kappa x_t \varepsilon_t$ and the parameter-driven model collapses to an observation-driven model. In the steady-state $\frac{1}{\kappa}P = I$ and we obtain

$$\phi_{t+1|t} = \phi_{t|t-1} + \kappa x_t\frac{1}{\sigma^2}(y_t - x_t'\phi_{t|t-1}).$$

which is a score based algorithm without the use of scaling matrix.

Theorem 2 Given the non-linear state space model

$$
\begin{aligned}
y_t &= x_t'\phi_t + \varepsilon_t,\ \varepsilon_t \sim N\left(0, \sigma^2\right), \\
\alpha_{t+1} &= \alpha_t + \eta_t,\ \eta_t \sim N\left(0, Q_t\right),
\end{aligned}
$$

with $\phi_t = g(\alpha_t)$. We can solve it by the mean of the Extended Kalman filter

$$
\begin{aligned}
v_t &= y_t - \widetilde{x}_t'\phi_{t|t-1}, \\
K_t &= P_{t|t-1}\widetilde{x}_t F_t^{-1}, \\
F_t &= \widetilde{x}_t' P_{t|t-1}\widetilde{x}_t + \sigma^2 \\
\alpha_{t+1|t} &= \alpha_{t|t-1} + K_t v_t, \\
P_{t+1|t} &= P_{t|t-1} - P_{t|t-1}\widetilde{x}_t F_t^{-1}\widetilde{x}_t' P_{t|t-1} + Q_t,
\end{aligned}
$$

where $\widetilde{x}_t' = x_t' \frac{\partial g(\alpha)}{\partial \alpha'}|_{\alpha=\alpha_{t|t-1}} = x_t'\Psi_t$. Setting $\sigma^2 = \frac{\sigma^2}{1-\kappa}$and $Q_t = P_{t|t}\frac{\kappa}{1-\kappa}$ and following same approach in Ljung (1992, p. 99) and Sargent (1999, p. 115), we obtain the modified version of the CGL algorithm

$$
\begin{aligned}
\alpha_{t+1|t} &= \alpha_{t|t-1} + \kappa R_t^{-1}\Psi_t' x_t \frac{1}{\sigma^2}(y_t - x_t'\phi_{t|t-1}), \\
R_t &= (1-\kappa)R_{t-1} + \kappa\left(\frac{1}{\sigma^2}\Psi_t' x_t x_t' \Psi_t\right).
\end{aligned}
$$

This is exactly the score-driven filter (2) with (12), where the information matrix $\frac{1}{\sigma^2}\Psi_t' x_t x_t' \Psi_t$ is replaced by its smoothed version R_t.

Appendix B: Robustness analysis

Section 2 shows that, in presence of heavy tails, the adaptive algorithm developed in this paper delivers a model-consistent penalization of the outliers. In fact, the estimated time variation in the parameters is such that the observations are downweighted when they are too large. In this appendix we assess the importance of using the law of motion of the parameters consistent with the score-driven model in presence of heavy-tails. In order to achieve this goal, we compare the density forecast of the specifications under Student-t innovations to two 'misspecified' cases. Firstly, we consider the case where the dynamic of the parameters is driven by the law of motion under Normal distribution (8)-(10) but we assume that the appropriate density is the Student-t; this is similar in spirit to the t-GARCH model of Bollerslev (1987) and it is labelled "Miss1". Secondly, we use the estimated time varying parameters obtained under Gaussian distribution and produce the density using a Student-t with calibrated degrees of freedom. Following Corradi and Swanson (2006) we choose $\upsilon = 5$. This second case is labelled "Miss2".

Table 6 reports the average log-scores for the above two specifications together with the benchmark Student-t specifications. Figures 7 and 8 report the empirical distribution of the PITs as in Diebold et al. (1998), and its cumulative distribution as in Rossi and Sekhposyan (2014). In both cases, we report the 95% confidence interval. Miss1 model delivers average log-scores which are comparable with the baseline Student-t specifications. However, an inspection of the PITs suggests that the densities from this model tend to be not well calibrated, slightly overstating the probability mass at the center of the density. Conversely, Miss2 model produces much better calibrated densities, but they perform rather poorly compared to the benchmark models as documented in the lower panel of Table 6. Those results suggest that both the low degree of freedom and the score-driven law of motion of the time-varying parameters, are important to achieve well calibrated density forecasts.

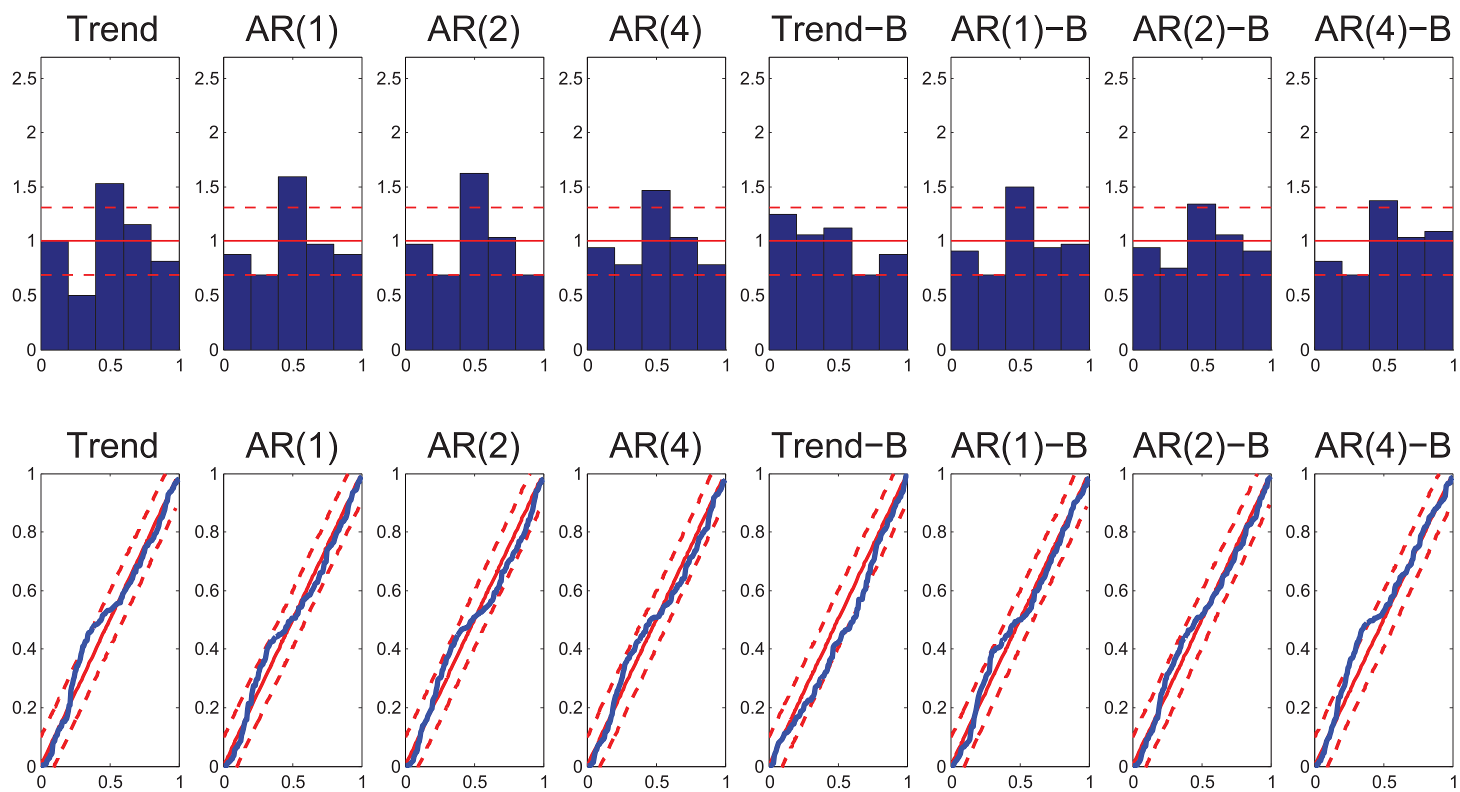

Figure 7: Density forecast for Miss-1. In the upper panel, the p.d.f. of the PITs (normalized) and the 95% critical values (dashed lines), obtaiined by binomial distribution constructed using a normal approximation as in Diebold et al. (1998). In the lower panel, the c.d.f. of the PITs with critical values based on Rossi and Sekhposyan (2013).

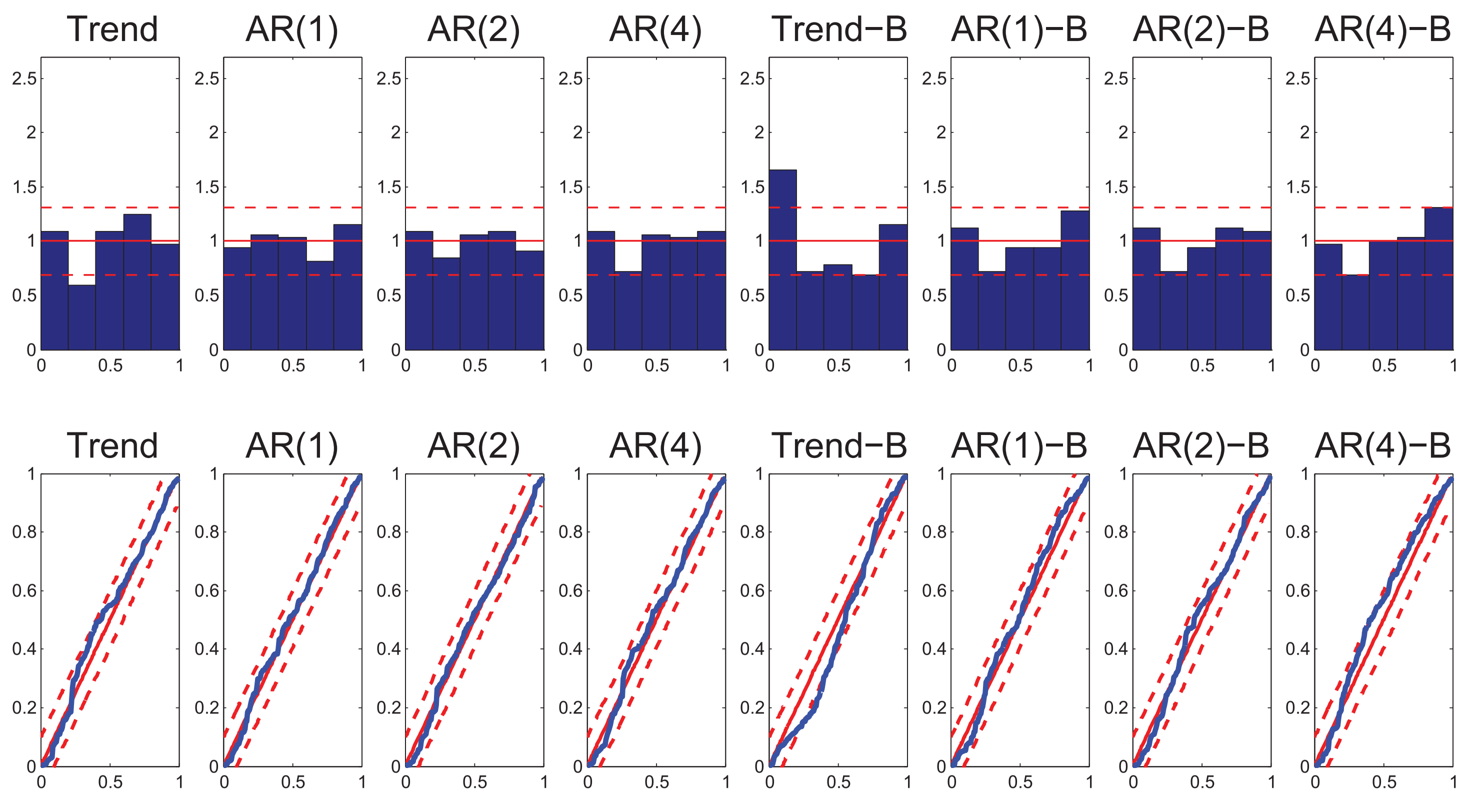

Figure 8: Density forecast for Miss-2. In the upper panel, the p.d.f. of the PITs (normalized) and the 95% critical values (dashed lines), obtaiined by binomial distribution constructed using the Normal approximation as in Diebold et al. (1998). In the lower panel, the c.d.f. of the PITs with critical values based on Rossi and Sekhposyan (2013)

		Trend	AR(1)	AR(2)	AR(4)	Trend-B	AR(1)-B	AR(2)-B	AR(4)-B
	ALogS	-1.5897	-1.6325	-1.6249	-1.5976	-1.5688	-1.6766	-1.6671	-1.6272
Miss-1									
Trend	-1.6546	0.0530	0.5386	0.3557	0.1611	0.1148	0.6230	0.6763	0.5172
AR(1)	-1.5713	0.6151	0.0177	0.1338	0.5013	0.9601	0.0043	0.0036	0.1975
AR(2)	-1.5248	0.0606	0.0024	0.0000	0.0210	0.4343	0.0016	0.0001	0.0209
AR(4)	-1.5266	0.1377	0.0149	0.0032	0.0007	0.4760	0.0045	0.0026	0.0144
Trend-B	-1.476	0.0493	0.0019	0.0116	0.0270	0.0244	0.0001	0.0001	0.0071
AR(1)-B	-1.5354	0.2296	0.0007	0.0322	0.1258	0.4723	0.0001	0.0005	0.0357
AR(2)-B	-1.5902	0.9907	0.3241	0.3895	0.8839	0.6932	0.0835	0.0384	0.4795
AR(4)-B	-1.5453	0.2534	0.0464	0.0749	0.1534	0.6111	0.0103	0.0060	0.0047
Miss-2									
Trend	-1.7339	0.0000	0.0072	0.0006	0.0012	0.0054	0.2327	0.0505	0.0162
AR(1)	-1.7260	0.0041	0.0000	0.0097	0.0029	0.0033	0.1052	0.1071	0.0430
AR(2)	-1.7171	0.0017	0.0115	0.0003	0.0012	0.0119	0.3333	0.1270	0.0556
AR(4)	-1.7354	0.0020	0.0099	0.0025	0.0000	0.0055	0.2026	0.0979	0.0048
Trend-B	-1.8480	0.0001	0.0000	0.0005	0.0001	0.0000	0.0005	0.0007	0.0002
AR(1)-B	-1.7896	0.0001	0.0000	0.0007	0.0001	0.0000	0.0011	0.0011	0.0009
AR(2)-B	-1.7747	0.0001	0.0003	0.0002	0.0003	0.0005	0.0256	0.0012	0.0048
AR(4)-B	-1.7703	0.0000	0.0022	0.0016	0.0000	0.0002	0.0605	0.0161	0.0000

Table 6: The first column reports the Average Log Scores (ALogS) for the two misspecified models. The first row refers to the ALogS for the score-driven models with Student-t. All the other entries correspond to the p-values of the Amisano and Giacomini (AG) test for the difference in the ALogS.

Appendix C: Data

The price data for the US are obtained from Federal Reserve economic database (FRED). US CPI: Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL). US GDP Deflator: Gross Domestic Product: Implicit Price Deflator (GDPDEF). US PCE Deflator: Personal Consumption Expenditures: Chain-type Price Index (PCECTPI). All data are seasonally adjusted at the origin.

The G7 CPI data instead are from the OECD Consumer Prices (MEI) dataset. The data have been seasonally adjusted using X11 prior to the analysis

Appendix D: Additional Results

Moving to the analysis of the persistence in inflation, for $p > 0$ we follow Pivetta and Reis (2007) and compute both the sum of the AR coefficients and the largest root as proxy of the overall persistence; those are shown in Figures 11 and 10. Similar to Cogley and Sargent (2001), most of our specifications tend to suggest that the persistence of inflation in the US rose in the early part of the sample to reach the pick during the great inflation of the 1970s, before starting a gradual decline from mid to late 1980s. Yet it is also interesting to note that allowing for a large number of lags tends to decrease the time variation in the estimated persistence profiles. This finding reconciles the results reported by Pivetta and Reis (2007), who use a time-varying AR model with three lags and report little variation in inflation persistence, to the one of most of the literature which instead allow for a smaller numer of lags. The fact that model with different lags map into different inflation profiles whereas at the same time the results in similar long-run forecast highlight that the specification of the number of lags is going to be relevant for the short horizon forecasts.

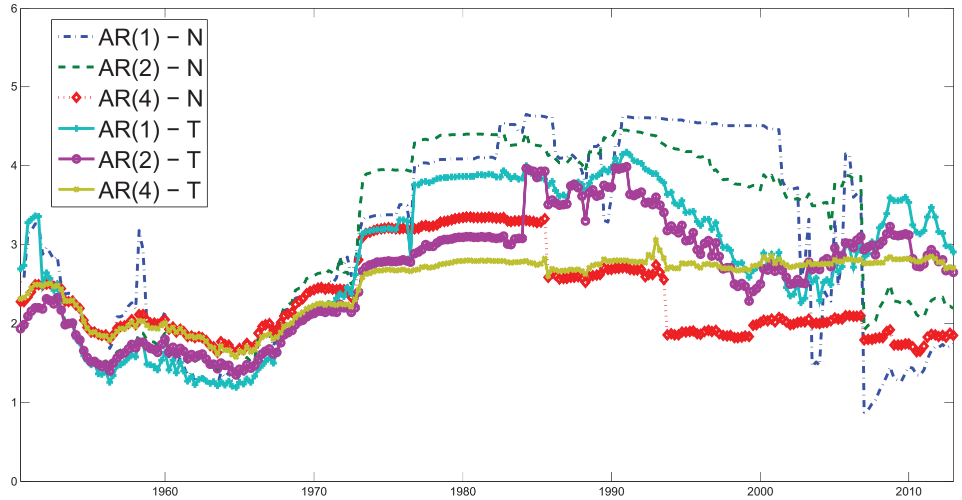

Figure 9: The implied "restricted" long-run trend for various specifications, "N" denotes Gaussian distribution and "T" for Student-t distribution.

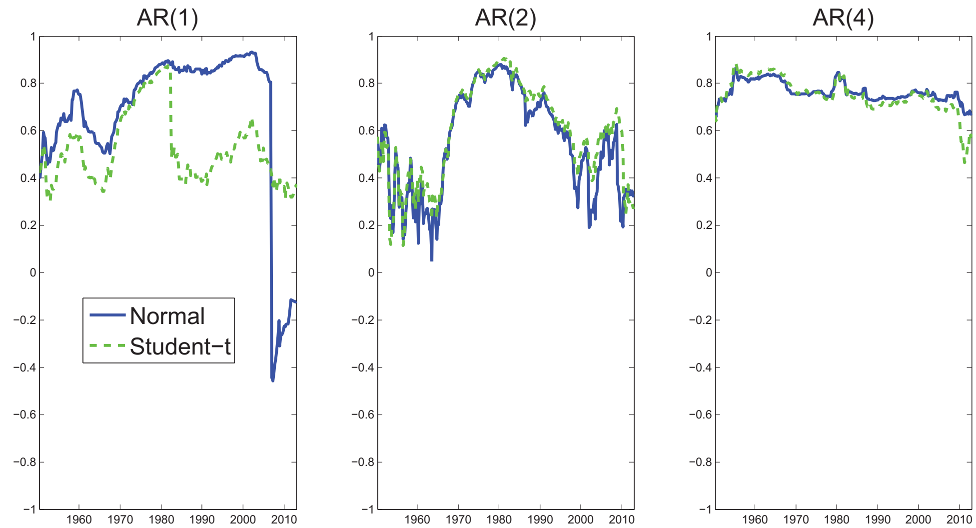

Figure 10: Largest eigenvalue for various specifications.

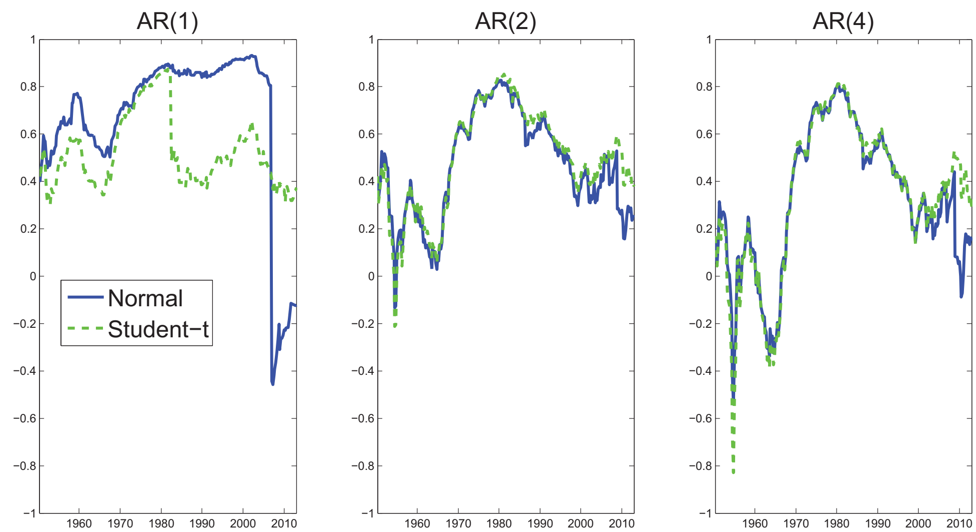

Figure 11: Sum of the ARs coefficients for various specifications.

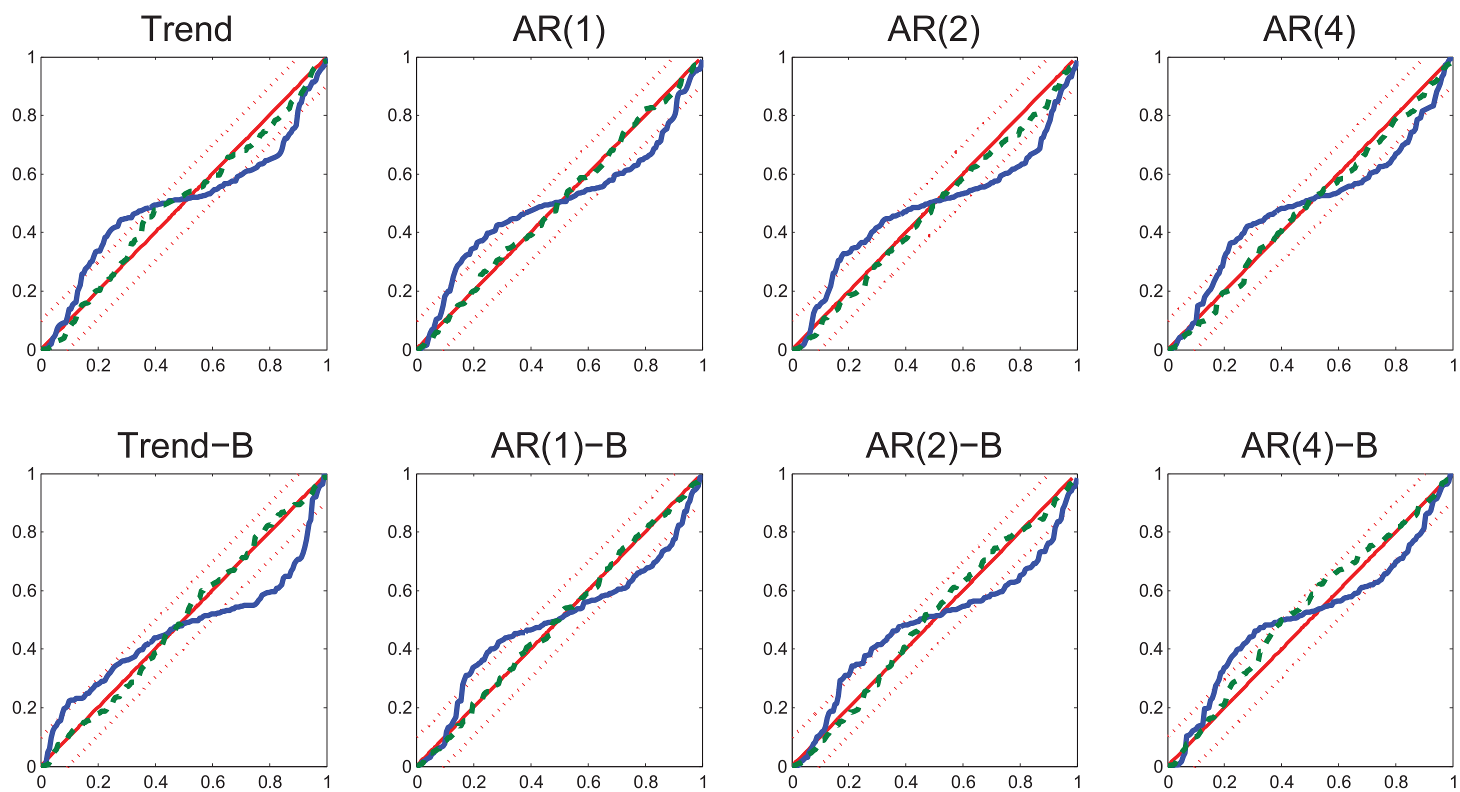

Figure 12: The c.d.f. of the PITs for various specifications. Critical values based on Rossi and Sekhposyan (2013). Solid (blue) line for Gaussian distribution and dashed (green) line for Student-t distribution.

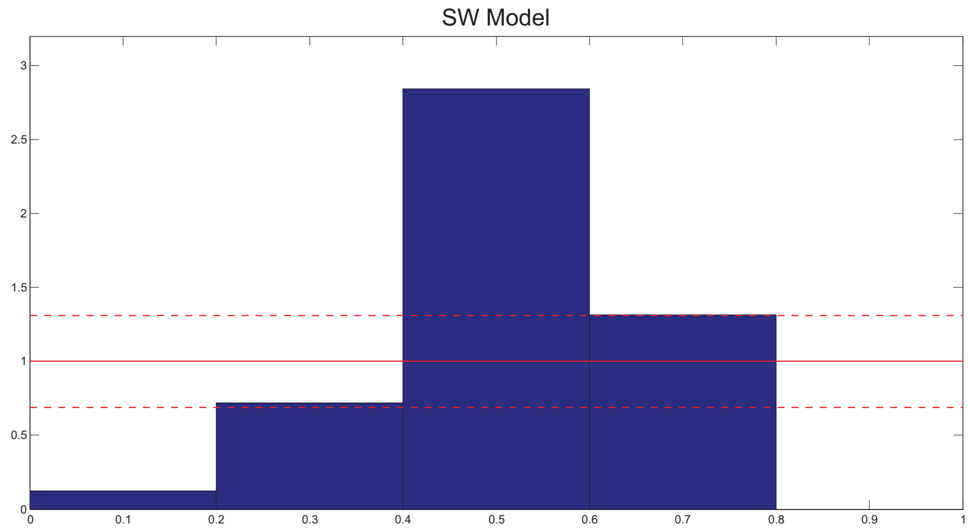

Figure 13: The p.d.f. of the PITs (normalized) and the 95% critical values (dashed lines), obtained by binomial distribution constructed using the Normal approximation as in Diebold et al. (1998).

	PCE Deflator						GDP Deflator					
	RMSE			MAE			RMSE			MAE		
	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8	h=1	h=4	h=8
SW	1.7399	2.1691	2.2040	1.1877	1.5723	1.6833	1.2666	1.6281	1.8037	0.8990	1.2012	1.3588
Normal												
Trend	**0.8793**	0.9663	1.0561	0.8724	0.9378	1.0352	**0.8576**	0.9689	1.0208	**0.8799**	0.9285	1.0025
	(0.0065)	(0.4892)	(0.4255)	**(0.0082)**	(0.2617)	(0.5813)	**(0.0142)**	(0.5783)	(0.7386)	**(0.0053)**	(0.1800)	(0.9670)
AR(1)	**0.8924**	1.0404	1.0738	**0.8978**	1.0113	1.0149	0.9312	1.2437	1.1822	0.9604	1.2390	1.2158
	(0.0251)	(0.6561)	(0.6757)	**(0.0799)**	(0.9143)	(0.9297)	(0.3251)	(0.1156)	(0.4227)	(0.5583)	(0.1132)	(0.3173)
AR(2)	**0.8863**	1.0106	1.0481	**0.8848**	0.9697	0.9766	0.9127	1.1752	1.1484	0.9406	1.1448	1.1603
	(0.0114)	(0.9022)	(0.7843)	**(0.034)**	(0.7626)	(0.8877)	(0.1908)	(0.231)	(0.5123)	(0.3445)	(0.3138)	(0.45)
AR(4)	0.9339	1.1231	1.2140	0.9579	1.1505	1.1899	**0.8339**	0.9412	1.0163	0.8677	0.9201	1.0044
	(0.2586)	(0.1933)	(0.2759)	(0.5252)	(0.1654)	(0.3053)	**(0.0189)**	(0.3450)	(0.7626)	(0.0115)	(0.1676)	(0.9300)
Trend-B	1.2473	1.0549	0.9510	1.2520	1.0247	0.9645	1.4628	1.1834	1.0224	1.4647	1.1937	1.0673
	(0.0010)	(0.5113)	(0.6798)	(0.0014)	(0.7816)	(0.7596)	(0.0000)	(0.1412)	(0.8901)	(0.0000)	(0.1090)	(0.6492)
AR(1)-B	**0.8806**	0.9852	1.0291	0.8849	0.9853	1.0250	**0.8746**	1.1576	1.2036	0.9385	1.2502	1.3660
	(0.0150)	(0.8389)	(0.8491)	(0.0521)	(0.8749)	(0.8736)	**(0.0802)**	(0.1666)	(0.2559)	(0.3412)	(0.0541)	(0.0657)
AR(2)-B	**0.8779**	0.9401	0.9805	0.8710	0.9390	0.9845	**0.8396**	0.9237	0.9476	**0.8749**	0.9280	1.0020
	(0.0050)	(0.3193)	(0.8806)	(0.0116)	(0.4165)	(0.9072)	**(0.015)**	(0.2869)	(0.659)	**(0.014)**	(0.2733)	(0.9856)
AR(4)-B	**0.8779**	0.9268	1.0147	**0.8794**	0.8972	0.9859	0.9252	1.1243	1.2457	0.9892	1.2303	1.4338
	(0.0355)	(0.2727)	(0.8752)	**(0.0473)**	(0.1274)	(0.8809)	(0.3018)	(0.2488)	(0.1724)	(0.871)	(0.0692)	(0.0343)
Student-t												
Trend	**0.9048**	0.9488	1.0289	**0.8831**	0.9080	1.0025	**0.8228**	0.9541	1.0293	**0.8578**	0.9350	1.0083
	(0.0969)	(0.3351)	(0.6963)	**(0.0312)**	(0.122)	(0.9699)	**(0.0081)**	(0.4808)	(0.6798)	**(0.0031)**	(0.2635)	(0.9060)
AR(1)	**0.8950**	1.0479	1.0787	**0.9006**	1.0131	1.0149	0.9345	1.2496	1.1870	0.9618	1.2387	1.2143
	(0.0283)	(0.6046)	(0.6593)	**(0.0882)**	(0.9025)	(0.9303)	(0.3479)	(0.1120)	(0.4167)	(0.5718)	(0.1157)	(0.3236)
AR(2)	**0.8892**	1.0083	1.0470	**0.8886**	0.9690	0.9788	0.9155	1.1793	1.1496	0.9398	1.1412	1.1527
	(0.0111)	(0.9214)	(0.7883)	**(0.0357)**	(0.7545)	(0.8982)	(0.2055)	(0.2275)	(0.5141)	(0.3391)	(0.3292)	(0.4738)
AR(4)	**0.8824**	**0.8914**	0.9737	**0.8999**	0.8816	0.9677	0.9820	1.3174	1.2367	0.9569	1.1876	1.1227
	(0.0117)	**(0.0489)**	(0.6694)	**(0.0445)**	(0.0445)	(0.5954)	(0.8063)	(0.0899)	(0.3755)	(0.5331)	(0.2273)	(0.5901)
Trend-B	1.2709	1.0751	0.9819	1.3129	1.0638	1.0117	**1.4343**	1.1484	0.9838	1.4196	1.1437	1.0164
	(0.0003)	(0.3722)	(0.8801)	(0.0001)	(0.4982)	(0.9242)	**(0.0001)**	(0.2347)	(0.9197)	(0.0000)	(0.2231)	(0.9081)
AR(1)-B	**0.8732**	0.9551	0.9928	**0.8769**	0.9517	0.9875	**0.8653**	0.9737	1.0034	**0.8865**	0.9311	1.0214
	(0.0111)	(0.5042)	(0.9613)	**(0.0402)**	(0.5842)	(0.9339)	**(0.0346)**	(0.7748)	(0.9752)	**(0.0394)**	(0.3525)	(0.8369)
AR(2)-B	**0.8995**	0.9294	0.9795	**0.9011**	0.9464	0.9997	**0.8169**	0.9169	0.9353	**0.8637**	0.9223	0.9970
	(0.0240)	(0.1787)	(0.8306)	**(0.0633)**	(0.3910)	(0.9977)	**(0.0065)**	(0.2986)	(0.6480)	**(0.0071)**	(0.2984)	(0.9812)
AR(4)-B	**0.8804**	**0.8884**	0.9746	**0.8888**	**0.8845**	0.9663	0.9265	0.9811	1.0515	0.9743	0.9821	1.0588
	(0.0371)	**(0.0703)**	(0.8308)	**(0.0625)**	**(0.0980)**	(0.7719)	(0.3153)	(0.7721)	(0.5263)	(0.666)	(0.7632)	(0.4704)

Table 7: Point forecast of the US PCE and GDP Deflator 1973Q1–2012Q4. The Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are expressed in relative term with respect to the SW model. The forecast horizon is "h", in brackets the p-values of the Giacomini and White (2006) test (in bold when it is significant at 10% level).

		Normal				Student-t			
		LR	RS	ALogS	AG	LR	RS	ALogS	AG
CA	Trend	0.0000	5.7263	-2.4568	–	0.5268	0.3567	-1.4456	0.0000
	AR(1)	0.0000	6.1251	-2.3703	0.1021	0.3624	0.8969	-1.4031	0.0000
	AR(2)	0.0000	5.3410	-2.3591	0.0566	0.3646	1.0757	-1.4470	0.0000
	AR(4)	0.0000	7.6040	-2.5042	0.3674	0.3699	0.9282	-1.4891	0.0000
FR	Trend	0.0000	1.6459	-2.3982	–	0.4827	0.8434	-1.4296	0.0109
	AR(1)	0.0000	1.1535	-2.2035	0.5409	0.0100	2.8370	-1.6271	0.0610
	AR(2)	0.0000	18.5406	-2.5520	0.7243	0.3268	1.8628	-1.5607	0.0383
	AR(4)	0.0000	18.5406	-2.7298	0.4486	0.0000	2.0931	-1.3412	0.0164
DE	Trend	0.0606	1.7418	-1.9100	–	0.0624	0.4414	-1.4569	0.0000
	AR(1)	0.0019	3.2809	-2.0028	0.0015	0.0641	0.7203	-1.5353	0.0000
	AR(2)	0.0000	3.4461	-2.0190	0.0035	0.0316	0.9842	-1.5365	0.0000
	AR(4)	0.0000	3.5844	-1.9632	0.2158	0.1461	1.1535	-1.4329	0.0000
IT	Trend	0.0000	2.3119	-2.5871	–	0.1605	1.2523	-1.5183	0.0003
	AR(1)	0.0000	5.9041	-2.5844	0.9802	0.0000	5.1906	-1.5887	0.0016
	AR(2)	0.0000	6.0846	-2.5218	0.5373	0.0000	4.1311	-1.5537	0.0010
	AR(4)	0.0000	12.2025	-2.6211	0.9175	0.0342	2.2255	-1.6434	0.0038
JP	Trend	0.0048	5.1273	-2.5986	–	0.4313	1.0006	-1.5212	0.0000
	AR(1)	0.0000	11.9010	-2.5882	0.9519	0.0000	5.9041	-1.3441	0.0000
	AR(2)	0.0000	19.2528	-2.8653	0.1583	0.0000	5.6872	-1.4016	0.0000
	AR(4)	0.0000	16.4845	-2.8101	0.2531	0.0000	10.4503	-1.4363	0.0000
UK	Trend	0.0336	7.5950	-2.7366	–	0.8862	0.3836	-1.6319	0.0000
	AR(1)	0.0184	6.5290	-2.8552	0.0820	0.5945	0.4199	-1.6731	0.0000
	AR(2)	0.0000	13.6571	-2.8285	0.1506	0.8235	0.8814	-1.5508	0.0000
	AR(4)	0.0000	15.1809	-3.1021	0.0000	0.0149	5.3032	-1.3970	0.0000

Table 8: Density Forecast of the CPI inflation for the other G7 countries 1973Q1-2012Q4. We report the p-values of the LR test (Berkowitz, 2001), the Rossi and Sekhposyan (RS) test with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%), the Average Log Score (ALogS), and p-values of the test by Amisano Giacomini (AG) for the difference in the ALogS with respect to the Gaussian model.