Adaptive models and heavy tails with an application to inflation forecasting

Delle Monache, Davide and Petrella, Ivan

Banca d'Italia, Warwick Business School and CEPR

1 September 2016

Adaptive models and heavy tails with an application to inflation forecasting*

Davide Delle Monache[†] Ivan Petrella[‡]
Bank of Italy WBS and CEPR

November, 2016

Abstract

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

JEL classification: C22, C51, C53, E31.

Keywords: adaptive algorithms, inflation, score-driven models, student-t, time-varying parameters.

*The views expressed in this paper are those of the authors and do not necessarily reflect those of Banca d'Italia. The authors would like to thank Michele Caivano, Ana Galvao, Anthony Garratt, Emmanuel Guerre, Andrew Harvey, Dennis Kristensen, Haroon Mumtaz, Zacharias Psaradakis, Barbara Rossi, Emiliano Santoro, Tatevik Sekhposyan, Ron Smith, Brad Speigner and Fabrizio Venditti for their useful suggestions. We also thank the participants at 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, University of Glasgow, Bank of England and Banca d'Italia.

[†]Banca d'Italia, Via Nazionale 91, 00184, Rome. Italy. *Email*: davide.dellemonache@bancaditalia.it
[‡]Warwick Business School, University of Warwick and CEPR, UK. *Email*: Ivan.Petrella@wbs.ac.uk

1 Introduction

In the last two decades there has been an increasing interest in models with time-varying parameters (TVP) for the analysis of macroeconomic variables. Stock and Watson (1996) have renewed the interest in this area by documenting the widespread forecasting gains of TVP models.[1] Recently, Cogley and Sargent (2005), Primiceri (2005), Stock and Watson (2007) have highlighted the importance of allowing for time variation in the volatility as well as in the coefficients.[2] 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 macroeconomic variables. As the recent recession has shown, departures from Gaussianity are important so as to properly account for the risks associated with black swans (see, e.g., Cúrdia et al. 2014).

This paper considers an adaptive autoregressive model where the errors are Student-t distributed. Following Creal et al. (2013) and Harvey (2013), the parameters' variation is driven by the score of the conditional distribution. In this framework, 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 this regard, 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, while Harvey and Luati (2014) show that the same intuition holds true in models with time-varying location. The resulting model is observation-driven. As opposed to parameter-driven models, the parameters' values are obtained as functions of the observations only, the likelihood function is available in closed form, and thus the model is estimated by classical methods.[3]

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 with each other in a score-driven model. Moreover, we show how to impose restrictions on the model's parameters so as to achieve local stationarity and bounded long-run mean. Both restrictions, commonly used in applied macroeconomics, have not yet been considered in the context of score-driven models.

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. (2014) propose a non-parametric estimation approach of time-varying coefficient models. The weighting function implied by these models are typically monotonically decreasing with time, a feature which they share with

[1] Attempts to take into account the well-known instabilities in macroeconomic time series can be traced back to Cooley and Prescott (1973, 1976), Rosenberg (1972), and Sarris (1973).

[2] D'Agostino et al. (2013) highlight the relative gains in terms of forecast accuracy of TVP models compared to the traditional constant parameter models in a multivariate setting.

[3] In the parameter-driven models the dynamics of the parameters is driven by an additional idiosyncratic innovations. Therefore, analytical expressions for the likelihood function are hardly available in closed form, and the use of computational-intensive simulation methods is usually required (see, e.g., Koopman et al., 2016).

traditional exponential weighted moving average forecasts (see e.g. Cogley, 2002). Our model features time variation in location and scale, and 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 conditional location and scale of the process, have effectively no weight in updating the TVP. The resulting pattern of the weights is both non-monotonic and time-varying since it is a function of the estimated TVP. Therefore, the 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 is robust to extreme events. Those model's features are important in the analyis of macroeconomic time series that display instability and changes in volatility. This is demonstrated 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. First, simple univariate autoregressive models have been shown to work well in the context of forecasting inflation (see Faust and Wright, 2013). Second, Pettenuzzo and Timmermann (2015) show that TVP models outperform constant-parameters models and 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. Third, while important changes in the dynamic properties of inflation are well documented (see e.g. Stock and Watson, 2007), most of the empirical studies are typically framed in a Bayesian setup and present 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[4]; (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 conveniently 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 a Student-t distribution improve substantially with respect to those produced by both its Gaus-

[4]Koop and Potter (2011) 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.

sian counterpart and the model of Stock and Watson (2007) and a TVP-VAR with stochastic volatility (see e.g. D'Agostino et al., 2013). In fact, well calibrated density forecasts are obtained only when we allow for heavy tails. While the baseline analysis is centered on US CPI inflation, which is noticeably noisier and harder to forecast than other measures of inflation, we show that the 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 of other countries, and we confirm that allowing for heavy tails provides substantial improvements in terms of density forecasting performance for all the G7 countries.

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 concludes.

2 Autoregressive model with heavy tails

Consider the following TVP model:

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

For $x_t = (1, y_{t-1}, ..., y_{t-p})'$ we have an autoregressive (AR) model of order p with intercept, where $\phi_t = (\phi_{0,t}, \phi_{1,t}, ..., \phi_{p,t})'$ is the vector of time-varying coefficients.[5] The disturbance ε_t follows a Student-t distribution with $\upsilon > 2$ degrees of freedom, zero conditional mean, and conditional variance σ_t^2.

Following Creal et al. (2013) and Harvey (2013), we postulate the score-driven dynamics for the TVP. Specifically, given $f_t = (\phi_t', \sigma_t^2)'$, we opt for a random walk law of motion:

$$f_{t+1} = f_t + Bs_t, \quad s_t = \mathcal{S}_t^{-1}\nabla_t, \tag{2}$$

where B contains the static parameters regulating the updating speed. The driving mechanism is represented by the scaled score vector:

$$\nabla_t = \frac{\partial \ell_t}{\partial f_t}, \quad \mathcal{S}_t = -E\left[\frac{\partial^2 \ell_t}{\partial f_t \partial f_t'}\right]^i, \quad i = 0, 1/2, 1, \tag{3}$$

$\ell_t = \log p(y_t|f_t, Y_{t-1}, \theta)$ is the predictive log-likelihood for the $t-$th observation conditional on the estimated vector of parameters f_t, the information set $Y_{t-1} = \{y_{t-1}, ..., y_1\}$, and the vector of static parameters θ. In the empirical application the scaling matrix is chosen to be equal to the inverse of the Fisher Information matrix, i.e. $i = 1$ and $\mathcal{S}_t = \mathcal{I}_t$. Other scaling matrix can be used as discussed in Creal et al. (2013).

[5]The results derived here are valid for additional regressors in the vector x_t.

The scaled score vector, s_t, is the sole driving mechanism characterizing the dynamics of f_t, this implies that the resulting model is observation-driven; i.e. the parameters' values are obtained as functions of the observations only, the likelihood function is available in closed form and the model can be estimated by maximum likelihood (ML). The vector f_t is 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 law of motion (2) can be rationalized as a stochastic analogue 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.

In principle, the law of motion (2) could have been defined more generally by letting the parameters follow $f_{t+1} = \omega + Af_t + Bs_t$, but this would have implied estimating a larger number of static parameters. The use of a random walk law of motion (2) 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 (Evans and Honkapohja, 2001), the parameters's updating rule is consistent with (2). In the empirical application we find it useful to set restrictions to avoid the proliferation of the static parameters. In particular, we restrict the matrix B to be diagonal, with the first $p+1$ elements are equal to κ_ϕ and the last one equal to κ_σ. For a given scaled score, these two scalar parameters regulate the speed of updating for the coefficients and volatility respectively.

Model (1) elaborates on previous works. Harvey and Chakravarty (2009) consider time-varying volatility with Student-t errors, highlighting how the score-driven model leads to a filtering 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 intercept and with constant variance, focusing on the stochastic properties of the implied non-linear model. Our specification features time variation for both coefficients and volatility, we emphasize the interaction between the two and their relevance for modelling macroeconomic data.

2.1 The score vector

The conditional log-likelihood of model (1) is equal to

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

with

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

where $\eta = 1/\upsilon$ is the reciprocal of the degrees of freedom, and $\Gamma(\cdot)$ is the Gamma function. The score-driven model with non-Gaussian innovations not only modifies the likelihood function, as in the t-GARCH of Bollerslev (1987) and Fiorentini et al. (2003), but it will also imply a different filtering process for the TVP. Given the specification (3), the score vector ($s_t = [s'_{\phi,t}, s_{\sigma,t}]'$) can be specialized in $s_{\phi,t}$ driving the coefficients, and $s_{\sigma,t}$ driving the volatility:

$$s_{\phi,t} = \frac{(1-2\eta)(1+3\eta)}{(1+\eta)} \frac{1}{\sigma_t^2} \mathcal{S}_t^{-1} x_t w_t \varepsilon_t, \tag{5}$$

$$s_{\sigma,t} = (1+3\eta)(w_t \varepsilon_t^2 - \sigma_t^2), \tag{6}$$

where $\mathcal{S}_t = \frac{1}{\sigma_t^2}(x_t x_t')$,[6] and

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

(see Appendix A for details).

A crucial role in the score vector is played by the weights, w_t, which are function of the (squared) standardized prediction error, that is $\zeta_t = \varepsilon_t/\sigma_t$. Figure 1 provides intuition for the role played by those weights in the updating mechanism governing the model's parameters. The left panel plots the magnitude of w_t as a function of the standardized prediction error, ζ_t. Whereas the right panel plots the influence function (see e.g. Maronna et al. 2006), obtained as the product of the weights and 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 of Figure 1 shows how the score is a bounded function of the prediction errors. Furthermore, the volatility, σ_t, plays a role in *re-weighting* the observations, and as such the past estimated variance has a direct impact on the coefficients' updating rule.[7] Therefore, the score vector (5)-(7) implies a *double weighting scheme*; i.e. the observation are weighted both over time and across realizations.

[Insert Figure 1]

2.2 An example: the time-varying level model

A simplified version of model (1) helps clarify the impact of such double weighting mechanism. Assume that $x_t = 1$ and w_t is exogenously given. This specification leads to an integrated moving-average (IMA(1,1), hereafter) model with TVP. In particular, the moving average (MA) coefficient is equal to $(1 - \kappa_\theta w_t)$, and the conditional mean can be expressed as

$$\mu_{t+1} = \kappa_\theta \sum_{j=0}^{t} \gamma_j \tilde{y}_{t-j}, \quad \tilde{y}_{t-j} = w_{t-j} y_{t-j}, \tag{8}$$

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

[7] This is not the case under Gaussian distribution (see section 2.3).

where

$$\gamma_j = \prod_{k=t-j+1}^{t}(1-\kappa_\theta w_k), \quad \gamma_0 = 1, \quad \kappa_\theta = \kappa_\phi \frac{(1-2\eta)(1+3\eta)}{(1+\eta)} \tag{9}$$

(see details in Appendix A). The observations are weighted to be robust to the impact of extreme events through the weights w_t, and simultaneously are discounted over time by γ_j. Similarly, the time-varying variance can be expressed as

$$\sigma^2_{t+1} = \kappa_\zeta \sum_{j=0}^{t}(1-\kappa_\zeta)^j \tilde{\varepsilon}^2_{t-j}, \quad \tilde{\varepsilon}^2_{t-j} = w_{t-j}\varepsilon^2_{t-j}, \tag{10}$$

where $\kappa_\zeta = \kappa_\sigma(1+3\eta)$ regulates how past observations are discounted.[8] In the presence of an outlier (i.e. $w_t = 0$), the MA coefficient collapses to one and the model becomes a pure random walk.

In practice, the weights w_t are not exogenously given but they depend non-linearly on the current observation 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 that cannot be analytically expressed as in 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 parameters 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 Pesaran et al. (2013) or Giraitis et al. (2014).

At this stage it is interesting to compare the filter implied by the our model to the one of Stock and Watson (2007, SW hereafter). In particular, SW consider the local level model with stochastic volatilities in both disturbances, this leads to a reduced form model equal to an IMA(1,1) with TVP driven by a convolution of the two stochastic volatilities. In the SW model the MA coefficient drifts smoothly as a result of the random walk specification for the stochastic volatilities, while in our score-driven model (with Student-t) the MA coefficient is more volatile, because the time variation depends on the weights, w_t, and the model discounts the signal from the observations that are perceived as outliers.

[8] 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 can be expressed by the filter $\kappa_\zeta/[1-(1-\kappa_\zeta)L]$ applied to the weighted (squared) prediction errors, $\tilde{\varepsilon}^2_t$.

2.3 The Gaussian case

The Gaussian case is recovered by setting $\eta = 0$ (i.e. $\upsilon \to \infty$). In this case $w_t = 1, \forall t$, and the TVP are estimated by the following filter:

$$\phi_{t+1} = \phi_t + \kappa_\phi \frac{1}{\sigma_t^2} \mathcal{S}_t^{-1} x_t \varepsilon_t, \tag{11}$$

$$\sigma_{t+1}^2 = \sigma_t^2 + \kappa_\sigma (\varepsilon_t^2 - \sigma_t^2). \tag{12}$$

Equation (11) resembles the Kalman filter, since the updated parameters react to the prediction error ε_t scaled by a gain, depending on σ_t^2 and x_t.[9] Equation (12) is equal to the integrated GARCH model. In contrast to the Student-t case, the volatility term vanishes in the law of motion of the coefficients, and the estimated conditional variance does not directly affect the estimated time-varying coefficients.[10]

2.4 Estimation

Given the score vector (5)-(7) at time t, all the elements of the model (1)-(2) are known. Hence, the conditional likelihood (4) can be analytically evaluated, and the log-likelihood function is constructed as $\mathcal{L}(\theta) = \sum_{t=1}^{n} \ell_t$, so that the vector θ is estimated as $\widehat{\theta} = \arg\max \mathcal{L}(\theta)$. Following Creal et al. (2013, sec. 2.3), we conjecture that $\sqrt{n}(\widehat{\theta} - \theta) \to \mathcal{N}(0, \Omega)$, where Ω is numerically evaluated at the optimum.[11]

3 Model restrictions

Applications of TVP models often require imposing restrictions on the parameter' space. For instance, an AR model is usually restricted so that the implied roots lie within the unit circle at each point in time. In the Bayesian framework, such constraints are usually imposed by rejection sampling, which however leads to heavy inefficiencies (see e.g. Koop and Potter, 2011, and Chan el al., 2013). When restrictions are implemented within a score-driven setup, the model can still be estimated by ML without the need of computational demanding simulation methods. Specifically, the TVP vector is reparametrized as follows

$$\tilde{f}_t = \psi(f_t), \tag{13}$$

[9] As opposed to standard parameter-driven models, both the signal and the parameters 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).

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

[11] Harvey (2013, sec 4.6) derives the consistency and asymptotic normality of a model with time-varying volatility only. Harvey and Luati (2014) prove the same for a model with time-varying level only. Blasques et al. (2014) studied the asymptotic properties of AR(1) model with time-varying coefficient and constant volatility. The asymptotic properties for a model with time variation in both the conditional mean and the conditional variance has not been established yet. This goes beyond the scope of this paper.

where f_t is the unrestricted vector of parameters we model, while $\tilde{f}_t$, is the restricted vector of interest with respect to which the likelihood function is expressed. The function $\psi(\cdot)$, also known as link function, is assumed to be time-invariant, continuous, invertible and twice differentiable. The vector f_t continues to follow the updating rule (2), and the corresponding score results in:

$$s_t = (\Psi_t' \mathcal{S}_t \Psi_t)^{-1} \Psi_t' \nabla_t, \tag{14}$$

where $\Psi_t = \frac{\partial \tilde{f}_t}{\partial f_t'}$ is the Jacobian of $\psi(\cdot)$, while ∇_t and $\mathcal{S}_t$ are the gradient and the scaling matrix expressed with respect to $\tilde{f}_t$. In practice, we model $f_t = h(\tilde{f}_t)$, where $h(\cdot)$ is the inverse function of $\psi(\cdot)$, and Ψ_t is needed to apply the chain rule when we compute the score. Given past information, Ψ_t is a deterministic function whose role is to re-weight the score such that the restrictions are satisfied at each point in time.

Specifically, for model (1) we have that $\tilde{f}_t = (\phi_{0,t}, \underline{\phi}_t', \sigma_t^2)'$, where $\underline{\phi}_t = (\phi_{1,t}, ..., \phi_{p,t})'$, and $f_t = (\alpha_{0,t}, \underline{\alpha}_t', \gamma_t)'$, where $\underline{\alpha}_t = (\alpha_{1,t}, ..., \alpha_{p,t})'$. The next sub-sections describe in detail how to impose restrictions on the autoregressive coefficients $\underline{\phi}_t$ and the intercept $\phi_{0,t}$, in order to achieve local stationarity and bounded long-run mean of the process. The variance is always constrained to be positive using the exponential function, $\sigma_t^2 = \exp(2\gamma_t)$, this implies that $\gamma_t = \log \sigma_t$. Given the the partition of f_t and $\tilde{f}_t$, it is useful to specialize the Jacobian matrix as follows:

$$\Psi_t = \begin{bmatrix} \frac{\partial \phi_{0,t}}{\partial \alpha_{0,t}} & \frac{\partial \phi_{0,t}}{\partial \underline{\alpha}_t'} & \\ \frac{\partial \underline{\phi}_t}{\partial \alpha_{0,t}} & \frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'} & 0 \\ & 0' & 2\sigma_t^2 \end{bmatrix}, \tag{15}$$

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, $\frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'}$ is a $p \times p$ matrix, and $0'$ is row vector of zeros.

3.1 Imposing local stationarity

The Durbin-Lenvinson (DL) algorithm maps the autoregressive coefficients (ARs) into the partial autocorrelations (PACs). Local stationarity, which implies that at each point in time the AR model has stable roots, is imposed by restricting the PACs within the unit circle.[12]

Notation 1 *Given the vector of ARs* $\underline{\phi}_t = (\phi_{1,t}, ..., \phi_{p,t})' \in R^p$, $z_t = (z_{1,t}, ..., z_{p,t})' \in C^p$ *denotes the corresponding vector of roots, and* $\rho_t = (\rho_{1,t}, ..., \rho_{p,t})' \in R^p$ *is the corresponding vector of PACs.* R^p *and* C^p *are the real and the complex domain respectively. Recall that* $\underline{\alpha}_t = (\alpha_{1,t}, ..., \alpha_{p,t})' \in R^p$ *is the unrestricted counterpart of* $\underline{\phi}_t$.

Definition 1 *Model (1) is locally stationary if* $\underline{\phi}_t \in S^p$, $\forall t$, *where* S^p *is the stationary hyperplane where all the roots are inside the unit circle, i.e.* $|z_{j,t}| < 1$, $\forall t$. *Furthermore,* $\underline{\phi}_t \in S^p$ *if and only if* $|\rho_{j,t}| < 1$, $\forall t$.

[12] Note that the logistic transformation considered by Blasques et al. (2014) is a special case of the general transformation considered here.

Definition 1 extends the results in Bandorff-Nielsen and Schou (1973) and Monahan (1984) to the case of time-varying coefficients. For more details on locally stationary processes see Dahlhaus (2012).

Assumption 1 $\Upsilon(\cdot)$ *is a continuous (invertible) and differentiable function mapping* $\underline{\alpha}_t \in R^p$ *into* $\rho_t \in S^p$*; i.e.* $\rho_{j,t} = \Upsilon(\alpha_{j,t})$*, such that* $\rho_{j,t} < |1|$, $\forall t$*. Consequently,* $\frac{\partial \Upsilon(\underline{\alpha}_t)}{\partial \underline{\alpha}_t'} = diag\left[\frac{\partial \Upsilon(\alpha_{j,t})}{\partial \alpha_{j,t}}\right]$.

In the application we use the Fisher transformation, i.e. $\Upsilon(\alpha_{j,t}) = \tanh \alpha_{j,t}$, this implies that $\alpha_{j,t} = \operatorname{arctanh} \rho_{j,t}$, and $\frac{\partial \Upsilon(\alpha_{j,t})}{\partial \alpha_{j,t}} = (1-\rho_{j,t}^2)$.

Definition 2 $\Phi(\cdot)$ *is the continuous (invertible) and differentiable function mapping the PACs into the ARs, i.e.* $\underline{\phi}_t = \Phi(\rho_t)$*. Such function is obtained by the DL algorithm:*

$$\phi_t^{i,k} = \phi_t^{i,k-1} - \rho_{k,t}\phi_t^{k-i,k-1}, \quad i = 1, ..., k-1, \quad k = 2, ..., p, \tag{16}$$

with $\phi_t^{1,1} = \rho_{1,t}$ *and* $\phi_t^{k,k} = \rho_{k,t}$*. For k=p we have that* $\phi_{j,t} = \phi_t^{i,p}$.

Corollary 1 *Given Definitions 1-2, and Assumption 1,* $\psi_s(\cdot) = \Phi[\Upsilon(\cdot)]$ *is the function ensuring the model is the locally stationarity. Specifically,* $\psi_s(\cdot)$ *maps* $\underline{\alpha}_t \in R^p$ *into* $\underline{\phi}_t \in S^p$*; i.e.* $\underline{\phi}_t = \psi_s(\underline{\alpha}_t)$*. The Jacobian of* $\psi_s(\cdot)$ *is equal to:*

$$\frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'} = \frac{\partial \underline{\phi}_t}{\partial \rho_t'}\frac{\partial \rho_t}{\partial \underline{\alpha}_t'} \quad \Rightarrow \quad \frac{\partial \psi_s(\underline{\alpha}_t)}{\partial \underline{\alpha}_t'} = \frac{\partial \Phi(\rho_t)}{\partial \rho_t'}\frac{\partial \Upsilon(\underline{\alpha}_t)}{\partial \underline{\alpha}_t'}. \tag{17}$$

Theorem 1 *Given Definition 2, the Jacobian* $\frac{\partial \underline{\phi}_t}{\partial \rho_t'} = \frac{\partial \Phi(\rho_t)}{\partial \rho_t'}$ *is equal to the last element of the following recursion:*

$$\Gamma_{k,t} = \begin{bmatrix} \tilde{\Gamma}_{k-1,t} & b_{k-1,t} \\ 0_{k-1}' & 1 \end{bmatrix}, \quad \tilde{\Gamma}_{k-1,t} = J_{k-1,t}\Gamma_{k-1,t}, \quad k = 2, ..., p, \tag{18}$$

where

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

$\phi_t^{j,k-1}$ *are the elements of (16), and for k 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. ■

Corollary 1 defines the link function through which we impose the stationarity restriction, and Theorem 1 defines the Jacobian of function $\Phi(\cdot)$ in Definition 2. When the time-varying intercept $\phi_{0,t}$ is included without restrictions, the remaining elements of the Jacobian matrix (15) are: $\frac{\partial \phi_{0,t}}{\partial \alpha_{0,t}} = 1$ and $\frac{\partial \phi_{0,t}}{\partial \underline{\alpha}_t'} = \left(\frac{\partial \underline{\phi}_t}{\partial \alpha_{0,t}}\right)' = 0'$.

3.2 Bounded trend

Sometimes it may be the case that one wants to discipline the model so as to have a bounded long-run mean (see e.g. Chan et al., 2013). 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).

Definition 3 *Given Definition 1, the local-to-date t approximation of the unconditional (time-varying) mean of model (1) is:*

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

As a result, the detrended component, $\tilde{y}_t = (y_t - \mu_t)$, follows a locally stationary process, i.e. $\Pr\{\lim_{h\to\infty} \mathbb{E}_t(\tilde{y}_{t+h}) = 0\} = 1$.

Assumption 2 $g(\cdot)$ *is a continuous (invertible) and differential function such that* $g(\cdot) \in [\underline{b}, \overline{b}]$.

In the application we use $g(\alpha_{0,t}, \underline{b}, \overline{b}) = \frac{\underline{b} + \overline{b} \exp \alpha_{0,t}}{1 + \exp \alpha_{0,t}}$. This implies that $\alpha_{0,t} = \log\left(\frac{\mu_t - \underline{b}}{\overline{b} - \mu_t}\right)$ and $\frac{\partial g(\alpha_{0,t}, \underline{b}, \overline{b})}{\partial \alpha_{0,t}} = \frac{(\overline{b} - \underline{b}) \exp \alpha_{0,t}}{(1 + \exp \alpha_{0,t})^2}$.

Corollary 2 *Given Definition 3 and Assumption 2, $\psi_b(\cdot)$ is the function mapping the unrestricted intercept $\alpha_{0,t}$ (and the restricted $\underline{\phi}_t$) into its restricted counterpart $\phi_{0,t}$, in order to have $\mu_t \in [\underline{b}, \overline{b}]$; i.e. $\phi_{0,t} = \psi_b\left(\alpha_{0,t}, \underline{b}, \overline{b}, \underline{\phi}_t\right)$. Specifically,*

$$\psi_b(\cdot) = g\left(\alpha_{0,t}, \underline{b}, \overline{b}\right)\left(1 - \sum_{j=1}^{p} \phi_{j,t}\right). \tag{20}$$

Given $\psi_b(\cdot)$, the elements of the Jacobian matrix (15) are equal to:

$$\frac{\partial \phi_{0,t}}{\partial \alpha_{0,t}} = \frac{\partial g\left(\alpha_{0,t}, \underline{b}, \overline{b}\right)}{\partial \alpha_{0,t}}\left(1 - \sum_{j=1}^{p} \phi_{j,t}\right), \quad \frac{\partial \phi_{0,t}}{\partial \underline{\alpha}_t'} = -g\left(\alpha_{0,t}, \underline{b}, \overline{b}\right)\iota' \frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'}, \quad \frac{\partial \underline{\phi}_t}{\partial \alpha_{0,t}} = 0, \tag{21}$$

where ι' is a p-dimensional row vector of ones, 0 is a p-dimensional column vector of zeros, and $\frac{\partial \underline{\phi}_t}{\partial \underline{\alpha}_t'}$ has been outlined in Corollary 1.

Corollary 2 defines the link function, and the corresponding Jacobian, through which the bounds on the long-run mean are imposed. Putting together (17) and (21), we have all the elements of the Jacobian matrix (15).

4 Application: inflation forecasting

Autoregressive models have been shown to work well in the context of forecasting inflation (Faust and Wright, 2013). Moreover, models with TVP have been widely used to highlight the following features of inflation dynamics: (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); (iii) time-varying volatility (e.g. Stock and Watson, 2007, and Clark and Doh, 2014).

In this section we use the score-driven model introduced in Section 2 to capture the key characteristics of inflation dynamics. In particular, we emphasize the importance of allowing for t-distributed innovations to the TVP autoregressive specification:

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

where π_t in the annualized quarterly inflation.[13] 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 each specification, we also consider a counterpart with bounds on the long-run trend, as shown in sub-section 3.2. The choice of the bounds (between 0 and 5) follows the work by Chan et al. (2013), who argue that a level of the trend inflation that is too low (or too high) is inconsistent with the central bank's inflation target.[14] Finally, for all specifications we consider both Gaussian and Student-t distribution of the innovations.

For the case with $p = 0$, we have a time-varying level that tracks trend inflation. In particular, with Gaussian innovations trend inflation is estimated by exponential smoothing, as in Cogley (2002).[15]

[Insert Table 1]

Table 1 reports the estimates of the various specifications for CPI inflation in the US, 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 the 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. By contrast, the smoothing parameter associated with the variance, κ_σ, is stable and typically higher than κ_ϕ. This

[13] The bulk of the analysis focuses on US CPI inflation. Section 4.3 shows that the superiority of the model with Student-t distributed errors (in terms of forecasting) carries over different indicators of US inflation (PCE and GDP deflators), as well as for the CPI inflation of the remaining G7 countries. The data sources are discussed in Appendix B.

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

[15] Notice that Cogley (2002) does not include time variation in the variance. However, as it has been shown in Section 2, under a 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.

result supports the idea that changes in the volatility are 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 Cúrdia et al., 2014. 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 trend-only and the AR(1) specifications.[16] The trend-only specification tracks 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 (i.e., in the early part of the sample and in the 70s) is attributed to deviations from the trend. For all specifications we find that, since the mid 90s, the long-run trend is stable between 2-3%, going slightly over 3% in the run up to the recent recession.

Focussing on differences between the models with Gaussian and Student-t innovations, one notes that the trend-only model with Student-t innovations is generally less affected by the sharp transitory movements in the underlying inflation rates. In fact, the updating mechanism for the TVP under Student-t innovations is such that the observation is downplayed when it is perceived as an outlier. These differences are most visible in the last part of the sample, where the underlying trend inflation from the model with Gaussian innovations is often revised after large, but one-off, releases of inflation.[17] 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 Gaussian model, whereas they have a smaller effect under a Student-t distribution. However, the variation in the time-varying intercept is offset by the variation in the autoregressive coefficients, and the

[16] Adding more lags has very little impact on the estimates of long-run mean (see Appendix D). Therefore, the choice of the lag length has an impact only on the shape of the dynamics toward the long-run level, i.e. on short to medium horizon forecasts.

[17] Aastveit et. al (2014) have recently reported evidence of instability in standard VARs since the financial crisis. Even though our results are not directly comparable to theirs, who use a parameter driven model and investigate the issue in a multivariate setting, it is worth noting that in our setting specifications of the model that take into account for the presence of outliers are less likely to show pronounced instabilities.

model ends up delivering rather smooth long-run forecasts.[18]

Imposing an 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.[19]

[Insert Figure 2]

Figure 3 reports measures of changes in volatility for the AR(1) specification with Gaussian and Student-t innovations.[20] For both 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 (2015). Yet, there are some interesting differences between the two models. First, the model based on the Student-t distribution is more robust to single outliers; under Gaussian distribution the volatility seems to be disproportionately affected by very few observations in the last part of the sample.[21] 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 also 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 reported and emphasized the importance of the low frequency variation in the volatility.

[Insert Figure 3]

[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. 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] It is worth noting that the pattern in the long-run trend is quite similar to the one found by Chan et al. (2013), despite the fact that they use a different model specification and different estimation techniques.

[20] Small differences can be appreciated when comparing the trend-only model to the AR(p) specifications (see Appendix D).

[21] 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 with the Student-t distribution. Therefore, with a Gaussian model one would retrieve a misleading picture of inflation uncertainty, which surpasses by far the level reached in the 70s. This result maps into a severely biased estimate of the densities that are going to be significantly fatter as a result of a single observation.

4.2 Forecast evaluation

In this section we assess the forecasting performance of the model. The various specification are evaluated against the SW model that is usually considered to be a good benchmark for inflation forecasting.[22] The models are 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.

4.2.1 Point Forecast

Table 2 reports the results for the point forecast. Despite the well-known performance of the benchmark model, many of the alternative specifications we consider tend to have lower RMSEs and MAEs. The differences tend to disappear at longer horizons. The superior forecasting performance is also statistically significant for many specifications.[23] For instance, the AR(4) model reduces the loss by roughly 15%, both for the one quarter and one year ahead forecast. Imposing bounds on the long-run mean does not seem to improve the performance of the various specifications.[24] Most importantly, a comparison between the Gaussian and Student-t models reveals little differences in terms of point forecasts.

[Insert Table 2]

Looking at the relative performance in different sub-samples reveals that the score-driven models are superior at the beginning and at the end of sample, while the SW model is slightly better in the low volatile period (from mid-80s to early 2000). None of these differences are significant using the fluctuation tests of Giacomini and Rossi (2010), highlighting a relatively high volatility of the forecast errors.

4.2.2 Density Forecast

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 (2015) highlight the relevance of deflation risk, and the prediction of the latter requires an estimation of the overall density. Table 3 reports the results from the

[22]The SW model is estimated by Bayesian MCMC methods, and the Gibbs Sampling algorithm is broken into the following steps: (i) sampling of the variance of the noise component using the independent Metropolis Hastings as in Jacquier et al. (2002); (ii) sampling of the variance of trend component as in (i); (iii) sampling of the trend component using the algorithm developed by Carter and Kohn (1994).

[23]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.

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

density forecast exercise. As highlighted by Diebold et al. (1998), a correctly conditionally calibrated density forecast produces probability integral transforms (PITs) that are uniformly distributed.[25] For the one step-ahead we evaluate the calibration of the densities looking at the results of Berkowitz's (2001) LR test and the nonparametric test of Rossi and Sekhposyan (2014).[26] For horizons beyond the one step-ahead we report the test of Knüppel (2015), which is robust to the presence of serial correlation of the PITs. The results suggest that, at all horizons, the density forecasts of specifications with Gaussian innovations, as well as for the SW model, are not well calibrated. In order to understand why this is the case Figure 4 plots the empirical distribution function (p.d.f.) of the PITs for the AR(1) specification.[27] From both plots it is evident that models with Gaussian innovations tend to produce densities in which too many realizations fall in the middle of the distribution relative to what we would expect if the data really were Normally distributed.[28] The one step-ahead density forecasts are well calibrated for all the specifications with Student-t distribution. Two key features are important to deliver correctly calibrated forecasts. First, the volatility is not affected by observations in the tail of distribution, thus varying in a way that better captures changes in the dispersion of the central part of the density. Second, the distribution by nature has a slower decay in the tail. As such, it allows for higher probability of extreme events. According to Knüppel's (2015) test, the multistep density forecasts are well calibrated when we allow for Student-t innovations and also we impose bounds to the long-run mean. This restriction effectively imposes a tight anchor to far ahead forecasts, which turns out to be important in order to avoid that too many realizations fall in the tails of the distribution.

[Insert Table 3]

[Insert Figure 4]

Table 3 reports measures of the relative performance of density forecasts for various models, evaluating the relative improvements with respect to the SW model. The average logarithm score (ALS) suggests that 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 at all forecasting horizons. We also evaluate density forecasts looking at the average Continuous Ranked Probability Score (CRPS). The latter provides a metric for the evaluation of the density that is more resilient in the presence of outliers (Gneiting and Ranjan, 2007). The models under Student-t innovation continue to dominate the SW benchmark. Yet, since the CRPS is less sensitive to the realizations falling at the tails of the forecast density, the performance from the Student-t model are roughly in

[25] The PITs are also iid for one step ahead forecasts.

[26] The latter is still valid also in the presence of parameter estimation errors.

[27] Appendix D reports a similar plot for the other specifications. These figures confirms that only the PITs of the densities from the adaptive models with Student-t innovations resemble a uniform distribution.

[28] The histogram of the PITs for the SW model is similar to the one obtained for the score-driven models with Gaussian distribution, meaning that it produces densities that are overall too wide relative to the optimal density.

line with the ones produced by the model under Gaussian innovations under this alternative scoring metric. The model with Student-t innovations provides a better characterization of the density at the tail of the distribution. This superiority is clearly highlighted when looking at the ALS.[29]

4.2.3 Robustness

Forecasting performance over time. It is possible that the superiority of the heavy tail specification is not stable over the entire forecasting sample. For instance, the robustness of the model under a Student-t distribution can potentially delay the updating of the parameters in presence of a marked, isolated, structural break. In this case, a Gaussian model adapts more quickly to the new regime, while the Student-t model initially downweighs the observations occurring after the break. In order to understand whether the improvements under Student-t are driven by particular episodes, or specific sample periods, as opposed to be stable over time, Figure 5 reports: (a) the fluctuation test of Giacomini and Rossi (2010) applied to the ALS (left panel), and (b) the cumulative sum of the log-score (CLS) over time (right panel). We consider the trend-only specification with Student-t versus the Gaussian case and the SW model, focussing on the one quarter ahead forecast.[30] The value of the Giacomini and Rossi (2010) statistics is always positive and the CLS is rising. This suggests that the densities produced by the heavy tails model deliver a consistently higher log-score throughout the sample. The differences between Gaussian and Student-t innovations are, however, not statistically significant in the 90s and the CLS shows a slight decline over this period. This is not surprising since in the 90s inflation has been quite stable and as such we would not expect considerably different densities produced by the two models.

[Insert Figure 5]

Evaluating the importance of local stationarity restrictions. In the empirical analysis we have always imposed local stationarity, as a result inflation forecasts are forced to revert to the time-varying long-run mean. Instead, when local stationarity restrictions are not enforced, forecasts are in principle allowed to diverge over time. Although one can justify the imposition of stationarity restrictions on theoretical grounds, a natural question to ask is how relevant those restrictions are in practice. Table 4 highlights that the density forecasts produced without imposing stationarity restrictions are roughly in line with the restricted counterpart only for short-run forecast, i.e. one quarter ahead. However, at the 1 and 2 years horizon the density forecasts produced by a model that does not impose stationarity restrictions delivers

[29] The adaptive model developed in this paper delivers a model-consistent way to deal with time-variation in presence of heavy tailed distribution. Appendix C explores the importance of using an updating mechanism for the parameters, which is consistent with the score-driven approach as opposed to some ad-hoc specifications. Specifically, allowing for low degrees of freedom as well as for an updating mechanism that downplays the importance of outliers are both important ingredients to achieve well calibrated density forecasts.

[30] The results are qualitatively similar when other autoregressive specifications and longer forecast horizons are considered. Those results are not reported, but are available upon request.

significantly worse ALS and CRPS than the restricted counterpart.[31]

[Insert Table 4]

4.2.4 Comparison with a TVP-VAR with stochastic volatility

In this section we compare the forecasting performance of our adaptive model with that of an alternative benchmark, the TVP-VAR with stochastic volatility. This model is often considered the state of the art technology when it comes to modelling macroeconomic time series facing structural changes. In particular, D'Agostino et al. (2013) have highlighted that the TVP-VAR produces forecasts that are superior to a number of alternative specifications (both with fixed coefficients and more parsimonious specifications with TVP). Barnett et al. (2014) highlight that TVP-VAR leads to large improvements in forecast of UK macroeconomic time series. Clark and Ravazzolo (2015) compare different volatility specifications, including specifications with fat tails, and conclude that none dominates the standard stochastic volatility specification considered as a benchmark in this section.

We consider a VAR with 2 lags and three variables: unemployment, the short term interest rate, and inflation.[32,33] The TVP-VAR produces long-run forecasts that are potentially more precise than the one produced by a univariate model. In fact, the multivariate setting allows the TVP-VAR to exploit the information contained in the level of unemployment and the interest rate in predicting long-run inflation and the associated uncertainty around the forecast. In order to check for that in this section we consider forecasts of inflation up to 4 years ahead. In order to save space we focus on a comparison with the AR(1) specification of the adaptive model. The adaptive model specifications and the TVP-VAR produce very similar point forecasts (see Table 5). Whereas the Student-t specifications are typically marginally better both in terms of RMSE and MAE, the difference with respect to the forecasts produced by the TVP-VAR are never statistically significant according to the Giacomini and White (2006) test.[34]

[Insert Table 5]

Table 6 turns to the comparison of the models in terms of density forecast. Density forecasts produced by the TVP-VAR are well calibrated at the short horizons but fail to correctly

[31]The densities for the one quarter ahead are correctly calibrated only for the models with Student-t innovations. At longer forecast horizons are typically not correctly calibrated.

[32]The choice of variables and lag order is in line with the literature (e.g. Primiceri, 2005, and D'Agostino et al., 2013). The results with a single lag are qualitatively in line with the ones reported in this section and are available from the authors upon request.

[33]The TVP-VAR model with stochastic volatility model is estimated by Bayesian MCMC methods. The Gibbs Sampling algorithm is broken into the following steps: (i) sampling of the time varying coefficients of the VARs using the algorithm developed by Carter and Kohn (1994); (ii) sampling of the diagonal elements of the VAR covariance matrix model using the Metropolis Hastings as in Jacquier et al. (2002); (iii) sampling of the off-diagonal elements of the VAR covariance matrix using the algorithm developed by Carter and Kohn (1994). The priors of the model are chosen as in Cogley and Sargent (2005).

[34]Table D.1 in Appendix D reports that there is no statistically significant difference in point forecasts for any of the autoregressive specifications of the adaptive model.

characterize the distribution of inflation for longer horizons.[35] Moreover, the TVP-VAR produces CRPS which are broadly in line with the one produced by the adaptive AR(1) model. However, there are large differences in terms of average log-score (ALS). The ALS associated with the density produced by the TVP-VAR model is almost always superior to the one produced by the specification of the model with Gaussian innovations, but it is substantially lower than most specifications which allow for heavy tails. Those differences are large and always statistically significant according to the Amisano and Giacomini (2007) test. These results hold at all forecasting horizons, including forecasts of inflation 4 years ahead.[36]

[Insert Table 6]

[Insert Figure 6]

Figure 6 reports the cumulative logarithm score of the models for different forecasting horizons of the AR(1) specification against the TVP-VAR. The log-score differences between the adaptive model with heavy tail and the TVP-VAR are almost always positive for forecasts horizons longer than one quarter. For short-run forecasts there are two periods, in the late 70s and the mid-90s, when the TVP-VAR produces higher log-scores. The gains in these two periods are however overshadowed by the superior performance of the adaptive heavy tail model for the remaining part of the sample. The differences between the two models seem to be larger in the last decade of the sample, with inflation showing a more volatile and erratic behaviour. Overall, the gains in terms of log-score are larger for more distant horizons. Summarizing, in this comparison a clear ranking emerges: when forecasting inflation the adaptive AR model with Gaussian errors is dominated by a state-of-the art multivariate TVP benchmark but the adaptive model with heavy tails proves to be superior in density forecast to both approaches. Although our univariate model does not allow us to generalize this conclusion to other variables, it strongly points to potential advantages in combining heavy tails with TVP.

4.3 Additional empirical evidences

4.3.1 Alternative inflation indicators

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 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 forecasts of US CPI inflation. However, CPI

[35] Figure D.10 in Appendix D shows the PITs associated to the forecast and highlights that the long-horizons forecasts of TVP-VAR tend to be too wide so that far too many of the inflation turnout are scored in the middle part of the distribution.

[36] Comparing the results in Table 6 to the ones in Table 3, we note that the forecasts from the TVP-VAR model with stochastic volatility are typically superior to the one produced by the SW model. Moreover, the superiority of the adaptive model with Student-t innovations holds trough for all the autoregressive alternative of the model considered in the previous section.

inflation is notoriously more volatile than other inflation indicators, such as the PCE deflator and GDP deflator.[37] In order to assess whether the improvements of the heavy tails model carry through for other measures of inflation, we repeat the forecasting exercise using these two additional 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 Appendix D.[38] Table 7 shows that for these two indicators the one quarter ahead densities forecast are well calibrated only under the Student-t specifications. For 1 and 2 years ahead forecast most of the specifications tend to be poorly calibrated. An inspection of the PITs reveals that the problems are less severe when we allow for the Student-t distribution. Moreover, the models under Student-t tend to outperform the SW benchmark, as well as all the Gaussian specifications, and by a considerable margin, when looking at the ALS. The CRPS tend to be similar between the Gaussian and Student-t specifications and, in all cases, are lower than the SW benchmark. Interestingly, since those two indicators are smoother than CPI inflation, adding lags of inflation can in some cases deliver significant improvements in the density forecast.

[Insert Table 7]

4.3.2 International evidence

Cecchetti et al. (2007) highlight the presence of similarities in inflation dynamics across countries. Therefore, we investigate the performance of our model for CPI inflation of the remaining G7 countries. Table 8 reports a summary of the density evaluation focussing on the trend-only model.[39] The results are in line with those reported for the US. Specifically, the model with heavy tails results in large gains in terms of ALS and similar CRPS. Moreover, for one quarter ahead forecasts the densities are well calibrated only when the Student-t distribution is allowed. At longer horizons most specifications tend to be not well calibrated. An inspection of the PITs reveals that the problems are more severe for the model with Normal innovations.

[Insert Table 8]

5 Conclusion

In this paper we derive an adaptive algorithm for time-varying autoregressive models in presence of heavy tails. Following Creal et al. (2013) and Harvey (2013), the score of the

[37] SW report that their model is better suited for those smoothed series.

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

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

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. In the presence of Student-t innovations the updating mechanism for the TVP downplays the signal when the observation is perceived to be in the tail of the distribution. A given prediction error is more likely to be categorized as part of the tails in periods of low volatility than in periods of high volatility. 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. As such, it entails a clear computational advantage, especially when restrictions on the parameters are imposed.

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 inflation dynamics. The specifications with 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 well documented low-frequency variations.

The proposed model produces forecasts that are comparable in terms of point forecasts and far superior in terms of density forecasts than the model of Stock and Watson (2007) and a TVP-VAR with stochastic volatility. Allowing for heavy tails is found to be a key ingredient to obtain well calibrated density forecasts; only in this case one has a correct characterization of the density at the tail of the distribution. Imposing restrictions to the TVP of the model is relevant in order to produce better forecasts. For long-run horizons, imposing local stationarity improves substantially the performance in terms of density forecasts. Moreover, imposing the bounds on the long-run mean of the model improves the density forecasts for 1 and 2 years ahead.

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, 2013) to the multivariate case, where the dimensions of the model might be so large that the use of computationally intensive simulation methods is infeasible, and imposing restrictions is typically problematic.

References

Aastveit, K., A. Carriero, T. Clark and M. Marcellino (2014). Have Standard VARs Remained Stable Since the Crisis? Cleveland Fed Working paper 14-11.

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

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

Barnett, A., Haroon, M. and K. Theodoridis (2014). Forecasting UK GDP growth and inflation under structural change. A comparison of models with time-varying parameters. International Journal of Forecasting, vol. 30(1), 129-143.

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

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.

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

Carter, C. K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. Biometrika, 81(3), 541-553.

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

Cecchetti, S. G., Hooper, P., Kasman, B. C., Schoenholtz, K. L. and M.W. Watson (2007). Understanding the Evolving Inflation Process. U.S. Monetary Policy Forum 2007.

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.

Clark, T.E. and T. Doh (2014). A Bayesian Evaluation of Alternative Models of Trend Inflation. International Journal of Forecasting, vol. 30(3), 426-448.

Clark, T. E. and F. Ravazzolo (2015). The macroeconomic forecasting performance of autoregressive models with alternative specifications of time-varying volatility. Journal of Applied Econometrics, 30(4), 551-575.

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 (2015). Measuring Price-Level Uncertainty and Instability in the U.S., 1850-2012. Review of Economics and Statistics, 97(4), 827838.

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.

Cúrdia, V., Del Negro, M. and D.L. Greenwald (2014). Rare shocks, Great Recessions. Journal of Applied Econometrics, 29, 1031-1052.

D'Agostino, A., Gambetti, L. and D. Giannone (2013). Macroeconomic forecasting and structural change. Journal of Applied Econometrics, 28(1), 82-101.

Dahlhaus, R. (2012). Locally stationary processes. Handbook of Statistics, 30.

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.

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 (2001). Learning and Expectations in Macroeconomics. Princeton University Press.

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

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 (2014). Inference on stochastic time-varying coefficient models. Journal of Econometrics, 179(1), 46-65.

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

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

Gneiting T. and A. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association 102, 359-378.

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, 109(507), 1112-1122.

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

Jacquier, E., Polson, N. G., and Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics, 20(1), 69-87.

Knüppel, M. (2015). Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments. Journal of Business and Economic Statistics, 33(2), 270-281.

Koop, G. and D. Korobilis (2013). Large Time Varying Parameter VARs. Journal of Econometrics, 177(2), 185-198.

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

Koopman S.J., A. Lucas and M. Scharth (2016). Predicting time-varying parameters with parameter-driven and observation-driven models. Review of Economics and Statistics, 98, 97-110.

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.

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.

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.

Pesaran, M.H. and A. Timmermann (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics, 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, 177(2), 134-152.

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. Economics Working Papers 1416, Department of Economics and Business, Universitat Pompeu Fabra.

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

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

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), 1-54.

Figures and Tables

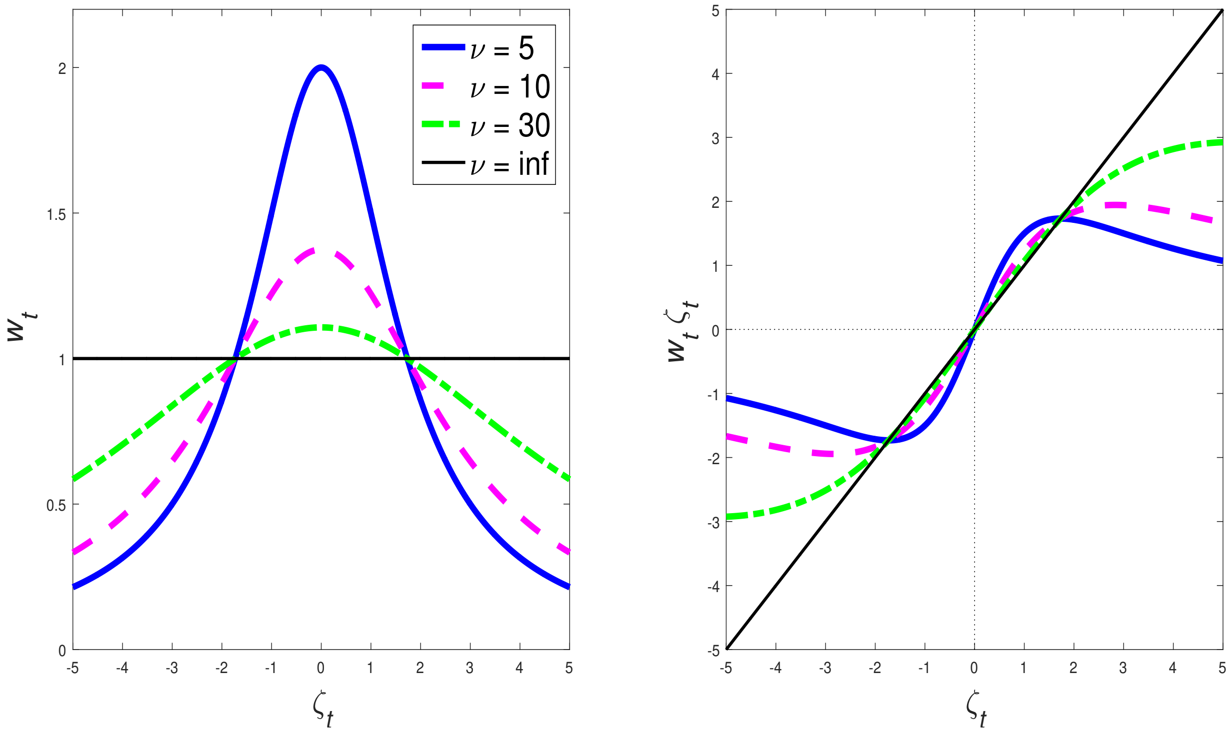

Figure 1: The left panel plots of the weights, w_t, against the standardized errors, $\zeta_t = \varepsilon_t/\sigma_t$, for different values of the degrees of freedom υ. The right panel plots the weighted standardized errors, $w_t\zeta_t$, against the standardized errors, ζ_t. The latter is also known as the influence function.

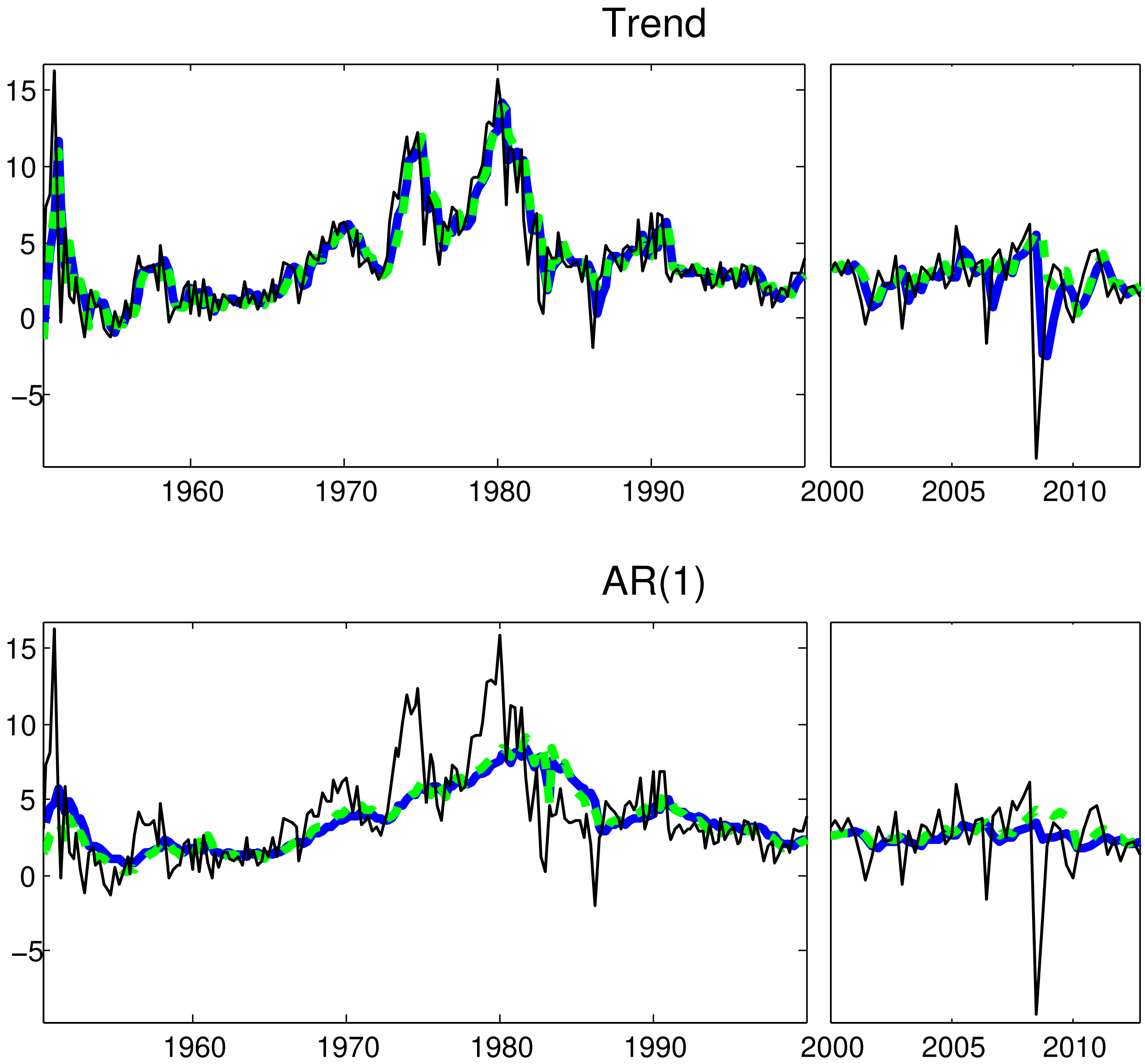

Figure 2: Implied "long-run" inflation, $\mu_t = \phi_{0,t}/(1 - \sum_{j=1}^{p} \phi_{j,t})$, together with the realized inflation. Trend-only specification (upper panel) and AR(1) specification (lower panel) for the model with Gaussian (continuous line) and Student-t innovations (dashed line).

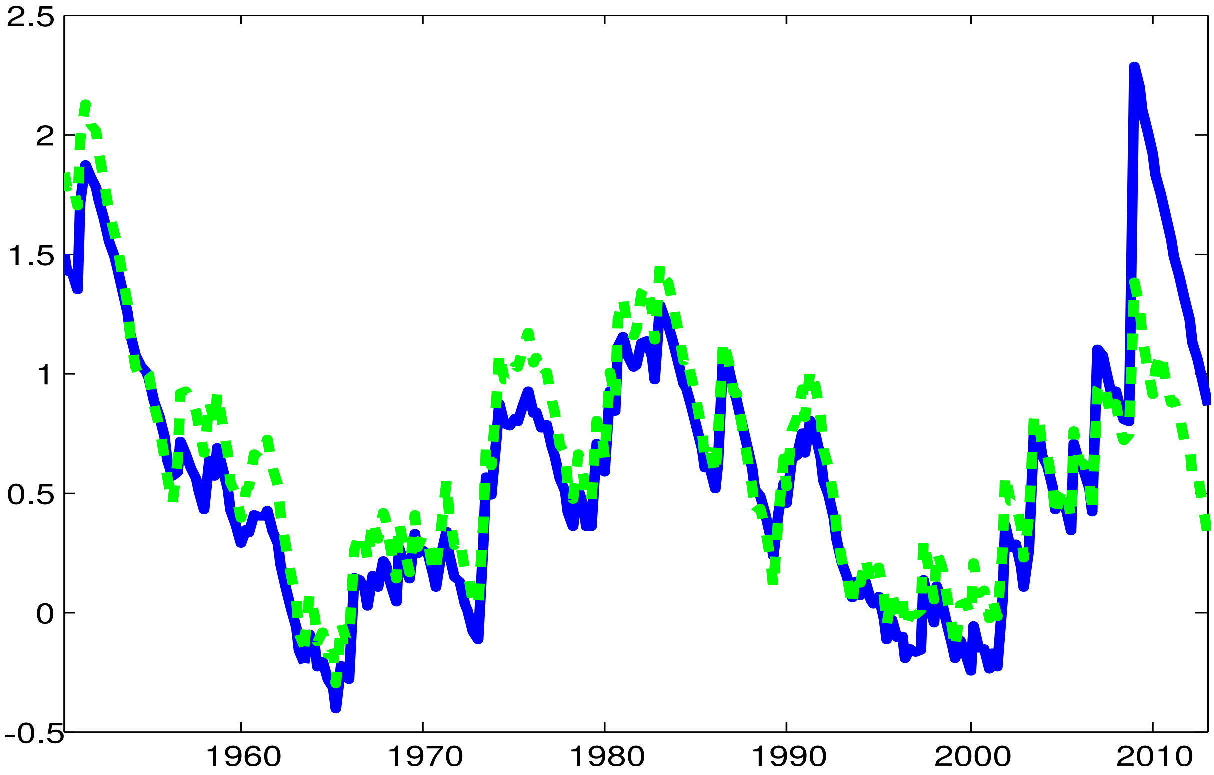

Figure 3: Implied volatility, $\log \sigma_t$, for the AR(1) specification of the model with Gaussian (continuous line) and Student-t innovations (dashed line).

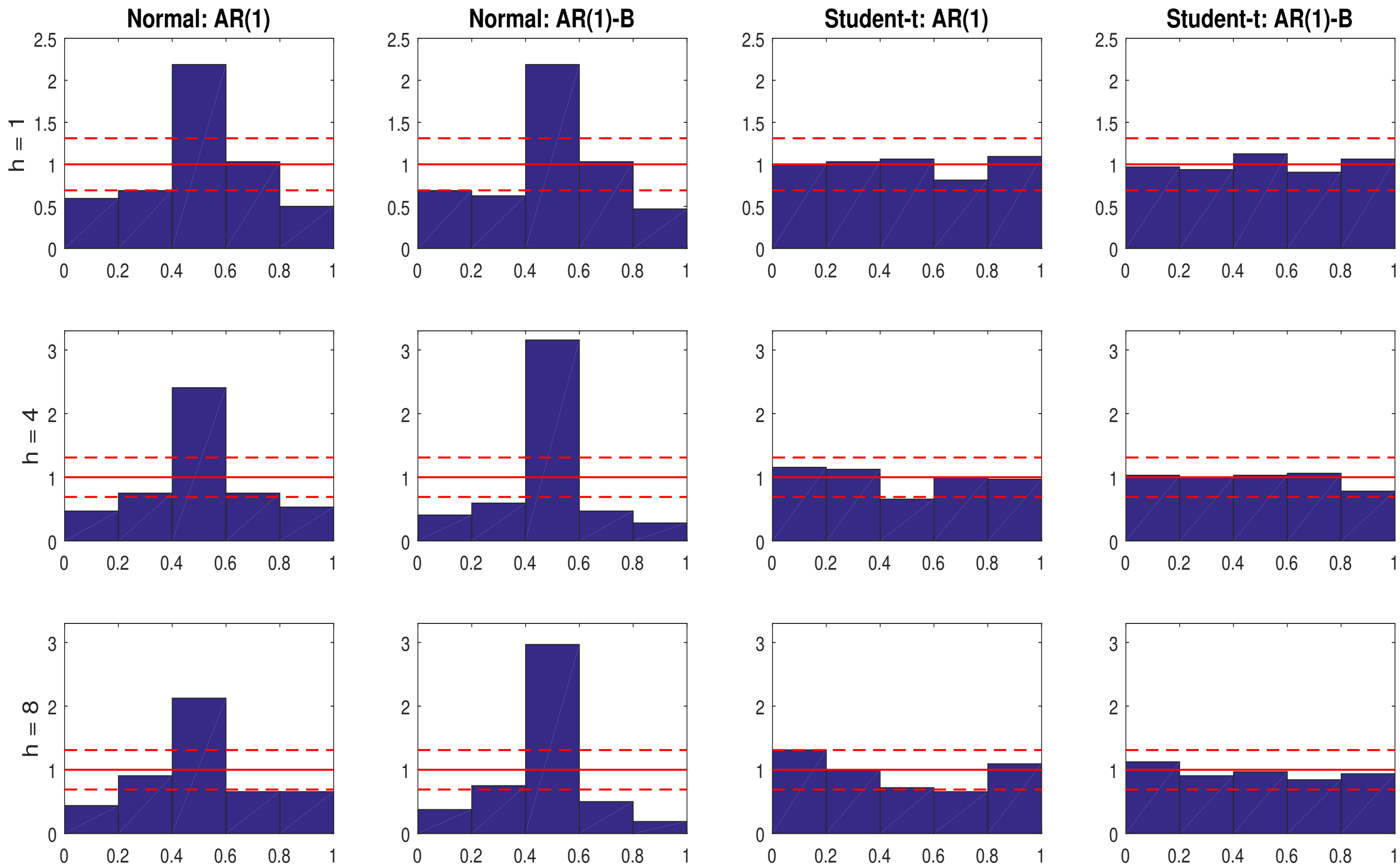

Figure 4: Probability density function of the PITs (normalized) for various AR(1) specifications of the model for one quarter (upper panels), one year (middle panel) and two years (lower panel) ahead forecasts. The dashed lines denote the 95% confidence interval constructed using a Normal approximation to a binomial distribution, as in Diebold et al. (1998).

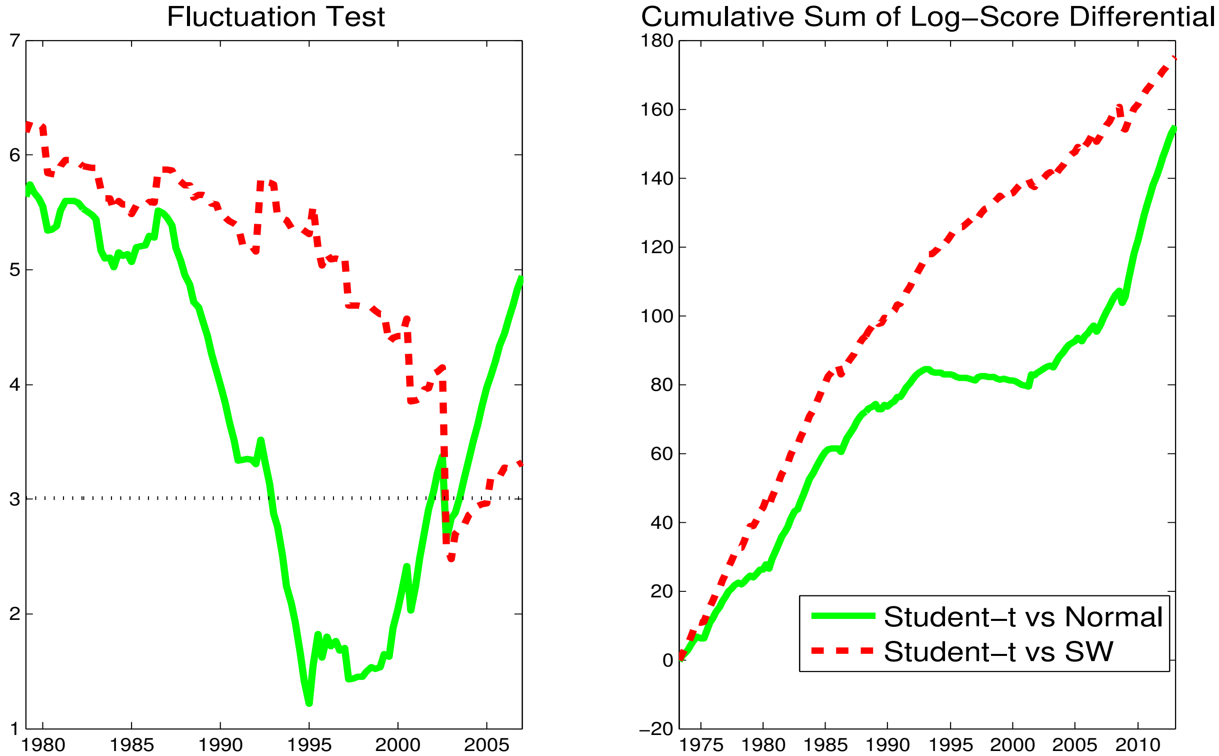

Figure 5: Left panel: fluctuation test statistics (Giacomini and Rossi, 2010) and the 5% critical value of the two sided test. The test compares the average log-score of two models over a window of 4 years, the dates on the x-axis correspond to the mid-point of the window. Positive values of the fluctuation statistic imply that the Student-t model outperforms the alternative. Right panel: cumulative sum of the logarithm score over time. A rising line suggests that the Student-t model outperforms the alternative. Student-t vs. SW model (dashed line) and Student-t vs. Normal model (continuous line). The Gaussian and Student-t model are for the trend-only specification. In all cases the statistics are computed for the one quarter ahead forecast horizon.

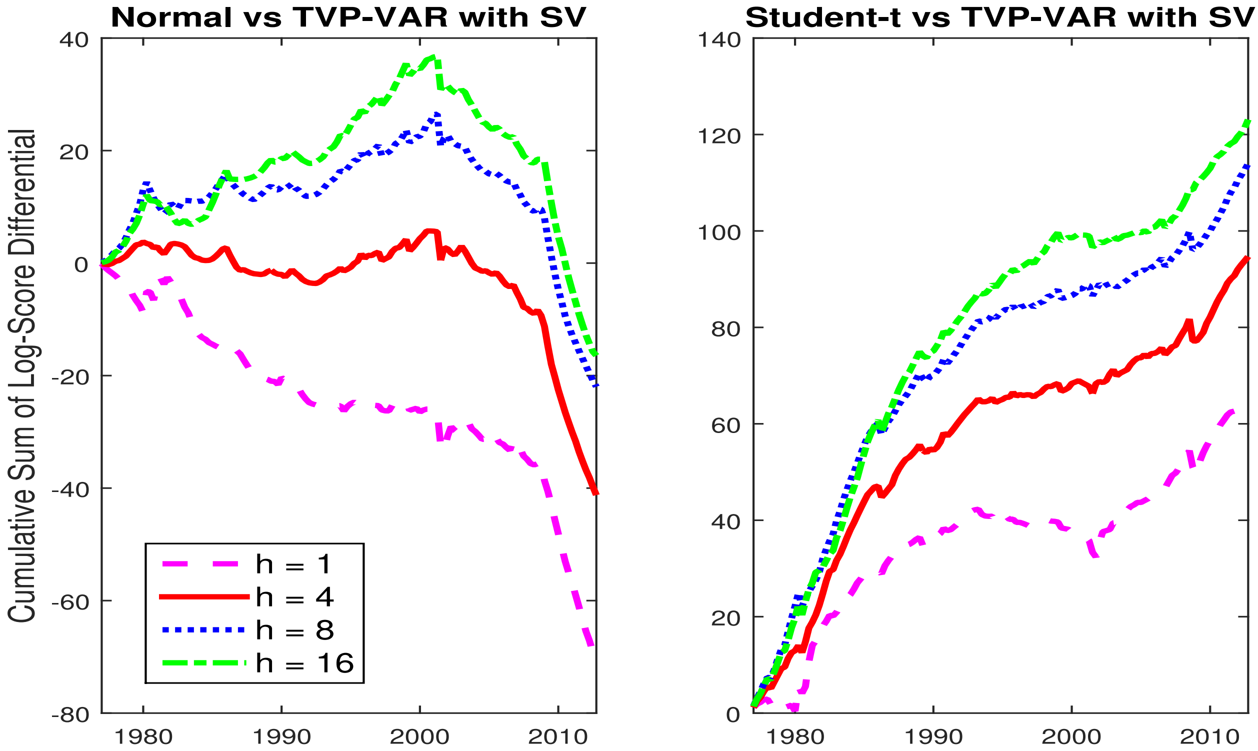

Figure 6: Cumulative sum of the logarithm score over time: comparison with a TVP-VAR with stochastic volatility. A rising line suggests that the adaptive AR(1) model with Gaussian (left panel) and Student-t innovations (right panel) outperforms the TVP-VAR with stochastic volatility.

	Normal					Student-t					
	κ_ϕ	κ_σ	LL	AIC	BIC	κ_ϕ	κ_σ	υ	LL	AIC	BIC
Trend	0.531	0.147	-549.114	1102.228	1109.349	0.471	0.212	5.331	-523.183	1052.365	1063.046
	(0.021)	(0.022)				(0.070)	(0.085)	(1.211)			
AR(1)	0.148	0.183	-541.147	1086.294	1093.415	0.170	0.170	5.137	-519.597	1045.195	1055.877
	(0.023)	(0.023)				(0.001)	(0.001)	(0.002)			
AR(2)	0.117	0.166	-551.3741	1106.748	1113.869	0.098	0.217	5.738	-526.518	1059.036	1069.718
	(0.022)	(0.023)				(0.001)	(0.001)	(0.001)			
AR(4)	0.136	0.214	-544.280	1092.556	1099.681	0.1381	0.260	6.207	-520.511	1047.023	1057.705
	(0.022)	(0.023)				(0.001)	(0.001)	(0.001)			
Trend-B	0.205	0.487	-604.327	1212.654	1219.775	0.820	0.246	5.875	-561.747	1129.495	1140.177
	(0.022)	(0.021)				(0.165)	(0.114)	(1.406)			
AR(1)-B	0.0861	0.283	-535.919	1075.838	1082.959	0.073	0.262	4.208	-520.567	1047.134	1057.816
	(0.020)	(0.023)				(0.000)	(0.000)	(0.001)			
AR(2)-B	0.098	0.267	-535.412	1074.825	1081.946	0.068	0.253	4.743	-520.794	1047.588	1058.270
	(0.021)	(0.023)				(0.000)	(0.000)	(0.000)			
AR(4)-B	0.096	0.273	-545.230	1094.460	1101.582	0.087	0.295	5.639	-521.315	1048.630	1059.312
	(0.021)	(0.023)				(0.000)	(0.000)	(0.001)			

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 π_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, AR(p) denotes the specification with lags order equal to p. Normal corresponds to $\varepsilon_t \sim N(0, \sigma_t^2)$, while Student-t to $\varepsilon_t \sim t_\upsilon(0, \sigma_t^2)$. The estimated static parameters are κ_ϕ, κ_σ and υ (std. error in brackets). LL is the Log-likelihood, AIC and BIC are the information criteria.

	RMSFE			MAE		
	h=1	h=4	h=8	h=1	h=4	h=8
SW	2.403	2.989	3.192	1.617	2.076	2.299
Normal						
Trend	0.908	0.938	1.038	0.898	0.978	1.042
	(0.029)	(0.246)	(0.501)	**(0.037)**	(0.677)	(0.354)
Trend-B	1.306	1.074	0.943	1.436	1.160	1.019
	(0.000)	(0.445)	(0.651)	**(0.000)**	(0.139)	(0.884)
AR(1)	0.877	0.886	0.899	0.896	0.929	0.903
	(0.004)	**(0.062)**	(0.262)	**(0.036)**	(0.287)	(0.284)
AR(1)-B	0.947	0.902	0.917	0.981	0.928	0.923
	(0.305)	(0.177)	(0.478)	(0.757)	(0.346)	(0.488)
AR(2)	0.889	0.870	0.951	0.895	0.925	0.950
	(0.007)	**(0.032)**	(0.534)	**(0.024)**	(0.240)	(0.537)
AR(2)-B	0.932	0.879	0.910	0.952	0.914	0.895
	(0.109)	**(0.060)**	(0.375)	(0.359)	(0.195)	(0.258)
AR(4)	0.862	0.850	0.923	0.871	0.909	0.919
	(0.004)	**(0.016)**	(0.255)	**(0.016)**	(0.132)	(0.202)
AR(4)-B	1.043	0.902	0.926	0.973	0.906	0.896
	(0.741)	(0.140)	(0.537)	(0.700)	(0.141)	(0.315)
Student-t						
Trend	0.926	0.927	1.028	0.886	0.958	1.015
	(0.143)	(0.213)	(0.708)	**(0.028)**	(0.443)	(0.811)
Trend-B	1.273	1.042	0.914	1.323	1.070	0.963
	(0.001)	(0.662)	(0.491)	**(0.000)**	(0.481)	(0.758)
AR(1)	0.878	0.901	0.918	0.888	0.949	0.911
	(0.003)	(0.064)	(0.324)	**(0.026)**	(0.393)	(0.291)
AR(1)-B	0.869	0.882	0.898	0.891	0.882	0.855
	(0.004)	**(0.100)**	(0.404)	**(0.040)**	(0.114)	(0.198)
AR(2)	0.918	0.897	0.990	0.915	0.949	0.976
	(0.042)	**(0.068)**	(0.901)	**(0.060)**	(0.413)	(0.768)
AR(2)-B	0.922	0.898	0.948	0.960	0.914	0.908
	(0.056)	(0.114)	(0.608)	(0.391)	(0.188)	(0.340)
AR(4)	0.869	0.852	0.919	0.873	0.904	0.906
	(0.005)	**(0.012)**	(0.262)	**(0.017)**	(0.115)	(0.215)
AR(4)-B	0.929	0.902	0.965	0.945	0.923	0.936
	(0.215)	(0.121)	(0.746)	(0.318)	(0.204)	(0.499)

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

	h=1						h=4					h=8				
	BK	RS	ALS		CRPS		KN	ALS		CRPS		KN	ALS		CRPS	
SW	0.000	9.801	-2.684	-	1.681	-	0.070	-2.855	-	1.910	-	0.169	-3.006	-	2.165	-
Normal																
Trend	0.145	4.422	-2.559	0.128	1.147	0.000	0.000	-2.914	0.724	1.534	0.000	0.000	-3.098	0.649	1.756	0.001
AR(1)	0.004	3.782	-2.486	0.003	1.123	0.000	0.004	-2.837	0.827	1.427	0.000	0.094	-2.900	0.299	1.482	0.000
AR(2)	0.006	4.692	-2.536	0.034	1.135	0.000	0.001	-2.801	0.529	1.412	0.000	0.014	-2.897	0.322	1.556	0.000
AR(4)	0.146	3.481	-2.466	0.004	1.096	0.000	0.003	-2.792	0.394	1.392	0.000	0.011	-2.848	0.066	1.513	0.000
Trend-B	0.058	7.832	-3.107	0.000	1.667	0.850	0.000	-3.672	0.008	1.784	0.200	0.000	-3.725	0.009	1.731	0.000
AR(1)-B	0.005	3.192	-2.454	0.000	1.168	0.000	0.000	-3.040	0.000	1.471	0.000	0.000	-3.172	0.002	1.560	0.000
AR(2)-B	0.457	3.721	-2.627	0.681	1.181	0.000	0.000	-3.051	0.030	1.451	0.000	0.000	-3.230	0.011	1.560	0.000
AR(4)-B	0.978	4.096	-2.602	0.402	1.198	0.000	0.000	-2.985	0.121	1.494	0.000	0.000	-3.195	0.021	1.605	0.000
Student-t																
Trend	0.700	0.702	-1.590	0.000	1.160	0.000	0.000	-2.000	0.000	1.535	0.000	0.000	-2.327	0.000	1.752	0.001
AR(1)	0.998	0.132	-1.633	0.000	1.139	0.000	0.179	-1.763	0.000	1.438	0.000	0.044	-1.809	0.000	1.489	0.000
AR(2)	0.501	0.552	-1.625	0.000	1.151	0.000	0.011	-1.704	0.000	1.423	0.000	0.000	-1.782	0.000	1.558	0.000
AR(4)	0.791	0.256	-1.598	0.000	1.108	0.000	0.073	-1.670	0.000	1.402	0.000	0.012	-1.699	0.000	1.515	0.000
Trend-B	0.005	0.272	-1.569	0.000	1.674	0.926	0.000	-1.830	0.000	1.791	0.217	0.000	-1.981	0.000	1.737	0.000
AR(1)-B	0.995	0.042	-1.677	0.000	1.180	0.000	0.571	-1.746	0.000	1.497	0.000	0.372	-1.782	0.000	1.588	0.000
AR(2)-B	0.759	0.361	-1.667	0.000	1.196	0.000	0.597	-1.649	0.000	1.474	0.000	0.532	-1.653	0.000	1.585	0.000
AR(4)-B	0.755	1.332	-1.627	0.000	1.212	0.000	0.297	-1.610	0.000	1.515	0.000	0.080	-1.590	0.000	1.631	0.000

Table 3: Density Forecasts for the US-CPI inflation 1973Q1-2012Q4. 'h' denotes the forecast horizon. 'BK' denotes the p-value of the test proposed by Berkowitz (2001), 'RS' is the test proposed by Rossi and Sekhposyan (2014), with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%). 'ALS' denotes the Average Log Score and 'CRPS' denotes the Continuous Ranked Probability Score. For both statistics we report the associated p-values of Amisano and Giacomini's (2007) test with respect to the SW model. 'KN' denotes the p-value of Knüppel's (2015) test.

	h=1						h=4					h=8				
	BK	RS	ALS Diff		CRPS Ratio		KN	ALS Diff		CRPS Ratio		KN	ALS Diff		CRPS Ratio	
Normal																
AR(1)	0.039	3.721	-0.001	0.939	0.994	0.706	0.005	-0.345	0.000	1.272	0.000	0.083	-0.431	0.000	1.396	0.000
AR(2)	0.049	5.041	-0.040	0.031	1.002	0.862	0.009	-0.320	0.000	1.323	0.000	0.057	-0.457	0.000	1.584	0.000
AR(4)	0.142	4.032	-0.056	0.059	1.004	0.836	0.067	-0.343	0.000	1.349	0.000	0.085	-0.410	0.000	1.661	0.000
Student-t																
AR(1)	0.619	0.462	0.095	0.000	0.990	0.537	0.095	0.016	0.791	1.264	0.000	0.007	-0.049	0.528	1.384	0.000
AR(2)	0.252	0.812	0.039	0.088	0.999	0.955	0.000	-0.088	0.197	1.313	0.000	0.000	-0.224	0.035	1.563	0.000
AR(4)	0.286	0.256	-0.009	0.806	1.005	0.802	0.000	-0.215	0.009	1.341	0.000	0.000	-0.413	0.001	1.643	0.000

Table 4: Density Forecasts for the US-CPI inflation 1973Q1-2012Q4 - Stationary Restrictions. 'h' denotes the forecast horizon. 'BK' denotes the p-value of the test proposed by Berkowitz (2001). 'RS' is the test proposed by Rossi and Sekhposyan (2014), with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%). 'ALS Diff' and 'CRPS Ratio' denote the difference in the ALS and the ratio of CRPS with respect their counterpart with stationary restrictions. Both statistics have the associated p-values of Amisano and Giacomini's (2007) test. 'KN' denotes the p-value of Knüppel's (2015) test.

	RMSFE				MAE			
	h=1	h=4	h=8	h=16	h=1	h=4	h=8	h=16
TVP-VAR	2.131	2.786	3.143	3.196	1.477	1.927	2.004	2.167
Normal								
AR(1)	0.989	0.950	0.913	0.976	0.981	1.001	1.036	1.056
	(0.857)	(0.658)	(0.453)	(0.743)	(0.759)	(0.991)	(0.696)	(0.504)
AR(1)-B	1.068	0.968	0.931	1.025	1.074	1.000	1.059	1.097
	(0.335)	(0.796)	(0.632)	(0.847)	(0.290)	(0.998)	(0.579)	(0.535)
Student-t								
AR(1)	0.990	0.966	0.932	1.010	0.973	1.023	1.045	1.090
	(0.886)	(0.755)	(0.528)	(0.901)	(0.661)	(0.789)	(0.630)	(0.337)
AR(1)-B	0.980	0.946	0.912	0.985	0.975	0.950	0.981	0.980
	(0.757)	(0.670)	(0.545)	(0.908)	(0.701)	(0.612)	(0.852)	(0.889)

Table 5: Point forecasts for US CPI inflation 1973Q1–2012Q4: Comparison with a TVP-VAR with stochastic volatility. The Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are expressed in relative terms with respect to the TVP-VAR with stochastic volatility. The forecast horizon is 'h'. In brackets are the p-values of Giacomini and White's (2006) test. Values in bold denote a significance at the 10% level.

	h=1						h=4				
	Berk	RS	ALS		CRPS		KN	ALS		CRPS	
TVP-VAR	0.290	0.781	-2.068	–	1.095	–	0.234	-2.290	–	1.411	–
Normal											
AR(1)	0.004	3.782	-2.486	0.000	1.123	0.623	0.004	-2.837	0.000	1.427	0.704
AR(1)-B	0.005	3.192	-2.454	0.000	1.168	0.242	0.000	-3.040	0.000	1.471	0.363
Student-t											
AR(1)	0.998	0.132	-1.633	0.000	1.139	0.450	0.179	-1.763	0.000	1.438	0.600
AR(1)-B	0.995	0.042	-1.677	0.000	1.180	0.166	0.571	-1.746	0.000	1.497	0.205

	h=8					h=16				
	KN	ALS		CRPS		KN	ALS		CRPS	
TVP-VAR	0.025	-2.375	–	1.544	–	0.001	-2.384	–	1.644	–
Normal										
AR(1)	0.094	-2.900	0.000	1.482	0.763	0.075	-2.807	0.000	1.640	0.929
AR(1)-B	0.000	-3.172	0.000	1.560	0.474	0.000	-3.199	0.000	1.795	0.054
Student-t										
AR(1)	0.044	-1.809	0.000	1.489	0.823	0.001	-1.872	0.000	1.632	0.814
AR(1)-B	0.372	-1.782	0.000	1.588	0.273	0.370	-1.770	0.000	1.814	0.021

Table 6: Density Forecasts for the US-CPI inflation 1973Q1-2012Q4: Comparison with a TVP-VAR with stochastic volatility. 'h' denotes the forecast horizon. 'BK' denotes the p-value of the test proposed by Berkowitz (2001), 'RS' is the test proposed by Rossi and Sekhposyan (2014), with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%). 'ALS' denotes the Average Log Score and 'CRPS' denotes the Continuous Ranked Probability Score. For both statistics we report the associated p-values of Amisano and Giacomini's (2007) test with respect to the TVP-VAR with stochastic volatility model. 'KN' denotes the p-value of Knüppel's (2015) test.

		h=1						h=4					h=8				
		BK	RS	ALS		CRPS		KN	ALS		CRPS		KN	ALS		CRPS	
PCE Deflator	SW	0.000	11.705	-2.453	-	1.326	-	0.196	-2.604	-	1.474	-	0.313	-2.743	-	1.648	-
	Normal																
	Trend	0.023	2.844	-2.249	0.105	0.805	0.000	0.000	-2.783	0.444	1.102	0.000	0.000	-3.011	0.209	1.256	0.000
	AR(1)	0.278	2.596	-2.165	0.001	0.830	0.000	0.000	-2.607	0.914	1.198	0.000	0.000	-2.727	0.961	1.264	0.000
	AR(2)	0.234	2.776	-2.199	0.018	0.819	0.000	0.000	-2.550	0.500	1.154	0.000	0.002	-2.690	0.600	1.224	0.000
	AR(4)	0.134	3.659	-2.237	0.076	0.866	0.000	0.000	-2.760	0.008	1.332	0.054	0.000	-2.892	0.020	1.443	0.028
	Trend-B	0.067	3.198	-2.607	0.192	1.099	0.000	0.000	-3.114	0.104	1.228	0.001	0.000	-3.365	0.048	1.200	0.000
	AR(1)-B	0.253	2.776	-2.126	0.000	0.818	0.000	0.000	-2.646	0.406	1.173	0.000	0.000	-2.801	0.216	1.263	0.000
	AR(2)-B	0.605	3.322	-2.113	0.002	0.798	0.000	0.000	-2.533	0.428	1.095	0.000	0.000	-2.639	0.097	1.177	0.000
	AR(4)-B	0.001	2.116	-2.286	0.237	0.822	0.000	0.001	-2.554	0.550	1.092	0.000	0.000	-2.852	0.177	1.227	0.000
	Student-t																
	Trend	0.430	0.375	-1.976	0.262	0.813	0.000	0.000	-1.943	0.000	1.101	0.000	0.000	-2.217	0.000	1.251	0.000
	AR(1)	0.783	0.060	-1.627	0.000	0.842	0.000	0.364	-1.657	0.000	1.213	0.000	0.029	-1.739	0.000	1.275	0.000
	AR(2)	0.981	0.219	-1.640	0.000	0.832	0.000	0.293	-1.695	0.000	1.167	0.000	0.020	-1.773	0.000	1.234	0.000
	AR(4)	0.977	0.164	-1.697	0.000	0.883	0.000	0.131	-1.783	0.000	1.349	0.072	0.002	-1.887	0.000	1.453	0.027
	Trend-B	0.033	0.477	-1.539	0.000	1.101	0.000	0.000	-1.773	0.000	1.229	0.001	0.000	-1.901	0.000	1.201	0.000
	AR(1)-B	0.270	0.366	-1.630	0.000	0.830	0.000	0.175	-1.575	0.000	1.189	0.000	0.015	-1.635	0.000	1.278	0.000
	AR(2)-B	0.819	0.692	-1.885	0.000	0.831	0.000	0.011	-1.886	0.000	1.136	0.000	0.000	-1.896	0.000	1.214	0.000
	AR(4)-B	0.444	0.180	-1.618	0.000	0.830	0.000	0.449	-1.588	0.000	1.102	0.000	0.077	-1.571	0.000	1.240	0.000
GDP Deflator	SW	0.000	13.368	-2.316	-	1.135	-	0.344	-2.423	-	1.231	-	0.381	-2.535	-	1.356	-
	Normal																
	Trend	0.000	0.426	-2.252	0.859	0.564	0.000	0.000	-2.763	0.339	0.796	0.000	0.000	-5.570	0.092	0.964	0.000
	AR(1)	0.000	3.749	-1.854	0.002	0.621	0.000	0.000	-2.203	0.010	1.031	0.004	0.000	-2.458	0.699	1.110	0.001
	AR(2)	0.000	2.815	-1.879	0.014	0.604	0.000	0.000	-2.204	0.041	0.965	0.000	0.000	-2.545	0.856	1.069	0.000
	AR(4)	0.000	1.016	-3.116	0.335	0.563	0.000	0.000	-2.656	0.395	0.774	0.000	0.000	-3.779	0.096	0.923	0.000
	Trend-B	0.000	0.666	-3.201	0.249	0.905	0.000	0.000	-2.766	0.108	0.999	0.000	0.000	-5.297	0.094	1.007	0.000
	AR(1)-B	0.000	5.113	-1.740	0.000	0.604	0.000	0.000	-2.222	0.001	1.030	0.003	0.000	-2.496	0.855	1.234	0.118
	AR(2)-B	0.000	0.745	-1.810	0.005	0.561	0.000	0.000	-1.883	0.000	0.769	0.000	0.000	-2.263	0.009	0.899	0.000
	AR(4)-B	0.000	5.474	-1.733	0.000	0.635	0.000	0.000	-2.221	0.001	1.013	0.001	0.000	-2.559	0.513	1.298	0.465
	Student-t																
	Trend	0.726	0.168	-1.567	0.000	0.566	0.000	0.000	-1.930	0.000	0.785	0.000	0.000	-2.304	0.077	0.946	0.000
	AR(1)	0.190	2.894	-1.447	0.000	0.622	0.000	0.000	-1.637	0.000	1.028	0.003	0.000	-1.793	0.000	1.107	0.001
	AR(2)	0.193	1.959	-1.477	0.000	0.604	0.000	0.000	-1.674	0.000	0.961	0.000	0.000	-1.842	0.000	1.063	0.000
	AR(4)	0.463	0.726	-1.684	0.000	0.574	0.000	0.265	-1.880	0.000	0.783	0.000	0.134	-1.968	0.000	0.924	0.000
	Trend-B	0.120	0.245	-1.565	0.000	0.906	0.000	0.078	-1.715	0.000	0.996	0.000	0.000	-1.886	0.000	1.004	0.000
	AR(1)-B	0.612	0.700	-1.470	0.000	0.605	0.000	0.293	-1.513	0.000	1.028	0.002	0.016	-1.650	0.000	1.229	0.103
	AR(2)-B	0.234	1.206	-1.485	0.000	0.564	0.000	0.009	-1.483	0.000	0.774	0.000	0.000	-1.493	0.000	0.903	0.000
	AR(4)-B	0.008	0.647	-1.512	0.000	0.644	0.000	0.001	-1.498	0.000	1.014	0.000	0.000	-1.471	0.000	1.279	0.307

Table 7: Density Forecasts for the US PCE Deflator and US GDP Deflator 1973Q1-2012Q4. 'h' denotes the forecast horizon. 'BK' denotes the p-value of the test proposed by Berkowitz (2001), 'RS' is the test proposed by Rossi and Sekhposyan (2014) with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%). 'ALS' denotes the Average Log Score and 'CRPS' denotes the Continuous Ranked Probability Score. For both statistics we report the associated p-values of Amisano and Giacomini's (2007) test with respect to the SW model. 'KN' denotes the p-value of Knüppel's (2015) test.

	h=1						h=4					h=8				
	BK	RS	ALS		CRPS		KN	ALS		CRPS		KN	ALS		CRPS	
Normal																
CA	0.000	5.726	-2.457	-	1.053	-	0.000	-2.519	-	1.269	-	0.087	-3.284	-	1.496	-
FR	0.000	1.646	-2.398	-	0.761	-	0.000	-2.953	-	1.018	-	0.000	-3.156	-	1.214	-
DE	0.061	1.742	-1.910	-	0.762	-	0.051	-2.189	-	0.982	-	0.000	-2.391	-	1.230	-
JP	0.000	2.312	-2.587	-	0.894	-	0.000	-3.221	-	1.292	-	0.000	-4.492	-	1.559	-
IT	0.005	5.127	-2.599	-	1.011	-	0.671	-2.676	-	1.208	-	0.349	-2.658	-	1.369	-
UK	0.034	7.595	-2.737	-	1.285	-	0.293	-3.005	-	1.579	-	0.000	-3.402	-	1.847	-
Student-t																
CA	0.527	0.357	-1.446	0.000	1.052	0.270	0.001	-1.663	0.000	1.242	0.125	0.000	-1.943	0.042	1.425	0.115
FR	0.483	0.843	-1.430	0.011	0.762	0.184	0.000	-1.788	0.021	0.997	0.357	0.000	-1.902	0.003	1.157	0.837
DE	0.062	0.441	-1.457	0.000	0.770	0.020	0.856	-1.664	0.000	0.958	0.211	0.005	-1.916	0.000	1.154	0.000
JP	0.161	1.252	-1.518	0.000	0.901	0.002	0.000	-2.011	0.000	1.262	0.157	0.000	-2.303	0.003	1.471	0.000
IT	0.431	1.001	-1.521	0.000	1.029	0.000	0.466	-1.687	0.000	1.191	0.102	0.014	-1.835	0.000	1.302	0.539
UK	0.886	0.384	-1.632	0.000	1.311	0.000	0.018	-1.872	0.000	1.558	0.053	0.000	-2.057	0.000	1.758	0.707

Table 8: Density Forecast for the G7 countries CPI inflation 1973Q1-2012Q4 - Trend specification. 'h' denotes the forecast horizon. 'BK' denotes the p-value of the test proposed by Berkowitz (2001), 'RS' is the test proposed by Rossi and Sekhposyan (2014) with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%), 'ALS' denotes the Average Log Score, 'CRPS' denotes the Continuous Ranked Probability Score, for both statistics we report the associated p-values of Amisano and Giacomini's (2007) test with respect to the Normal distribution. 'KN' denotes the p-value of Knüppel's (2015) test.

A Derivations and proofs

A.1 The score vector in Section 2.1

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

$$\ell_t = c(\eta) + d_t + g_t,$$

with

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

where $\zeta_t = \varepsilon_t/\sigma_t$ and $\Gamma(\cdot)$ is the Euler's gamma function. Let $\nabla_t = \partial\ell_t/\partial f_t$ denotes the gradient function, partitioning 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

$$\frac{\partial g_t}{\partial \zeta_t^2} = -\frac{\eta+1}{2(1-2\eta+\eta\zeta_t^2)}$$

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

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

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

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

We compute the information matrix as $\mathcal{I}_t = -\mathbb{E}(H_t)$, 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

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

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

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

and

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

The cross-derivative $H_{\phi\sigma,t} = -x_t \frac{\varepsilon_t}{\sigma_t^4}$ and therefore $\mathcal{I}_{\phi\sigma,t} = 0$. 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 scaled score vector is

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

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

A.2 Example in Section 2.2

Considering (1) with time varying mean:

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

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

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

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

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

The same weighting pattern is obtained when regressors are included. Since the weights across time, γ_t, are affected by the cross-sectional weights, w_t, we cannot obtained the robust filter for μ_{t+1} as solution of a re-weighted quadratic criterion function as Ljung and Soderestrom (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 it is straightforward to obtain the expression for the variance its implied weighting pattern.

A.3 Proof of Theorem 1 in Section 3.1

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 (16), 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 (16) 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

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

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.

B 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). Civilian Unemployment Rate (UNRATE). These data are seasonally adjusted at the origin. 3-Month Treasury Bill: Secondary Market Rate (TB3MS).

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

C Robustness: the importance of the score-driven updating mechanism

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. Someone could argue that a score-driven law of motion may be not a crucial ingredient and that standard TVP model would deliver comparable results by simply assuming the heavy tails distribution. Here, we want to assess the importance of using the law of motion of the parameters consistent with the score-driven model in presence of heavy-tails. We compare the density forecast of the two alternative "misspecified" cases. Firstly, we consider the case where the dynamic of the parameters is driven by the law of motion under Normal distribution (11)-(12) but the appropriate density is the Student-t. This is in spirit of the t-GARCH model of Bollerslev (1987) and it is labeled "Miss1". Secondly, we use the estimated the model under Gaussian distribution and we 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 labeled "Miss2". Table C.1 reports differences in the ALS and in the CPRS ratio with respect to the benchmark score-driven model with Student-t. Miss1 delivers ALS that are comparable with the benchmark model. However, it is strongly outperformmed by the benchmark in terms of the CRPS. Conversely, Miss2 performs rather poorly compared to the benchmark model when one looks at the ALS, even if it produces similar CRPS. Those results suggest that the low degree of freedom and the score-driven law of motion are both important to achieve well calibrated density forecasts.

	h=1						h=4					h=8				
	BK	RS	ALS Diff		CRPS Ratio		KN	ALS Diff		CRPS Ratio		KN	ALS Diff		CRPS Ratio	
Miss-1																
Trend	0.923	1.521	-0.065	0.053	1.466	0.001	0.000	-0.051	0.189	1.304	0.004	0.000	0.010	0.868	1.226	0.013
AR(1)	0.935	0.992	0.061	0.018	1.321	0.000	0.083	0.070	0.044	1.240	0.001	0.002	0.025	0.549	1.243	0.001
AR(2)	0.803	0.992	0.100	0.000	1.313	0.000	0.077	0.092	0.000	1.246	0.000	0.002	0.092	0.001	1.184	0.002
AR(4)	0.898	0.841	0.071	0.001	1.276	0.010	0.149	0.059	0.017	1.224	0.013	0.013	0.044	0.140	1.206	0.017
Trend-B	0.000	2.116	0.093	0.024	1.289	0.040	0.000	0.135	0.055	1.268	0.061	0.000	0.000	0.997	1.283	0.036
AR(1)-B	0.582	0.992	0.141	0.000	1.470	0.051	0.641	0.256	0.000	1.424	0.084	0.401	0.256	0.000	1.398	0.093
AR(2)-B	0.604	0.702	0.077	0.038	1.223	0.031	0.570	0.121	0.004	1.198	0.046	0.208	0.061	0.203	1.192	0.052
AR(4)-B	0.884	2.401	0.082	0.005	2.922	0.001	0.337	0.054	0.023	2.939	0.002	0.225	0.058	0.044	2.780	0.005
Miss-2																
Trend	0.481	0.992	-0.144	0.000	1.005	0.027	0.000	-0.194	0.000	0.994	0.014	0.000	-0.141	0.035	0.991	0.000
AR(1)	0.653	0.272	-0.093	0.000	1.000	0.575	0.030	-0.080	0.000	1.000	0.641	0.003	-0.099	0.000	1.000	0.668
AR(2)	0.601	0.256	-0.092	0.000	1.002	0.269	0.007	-0.146	0.000	0.999	0.501	0.000	-0.157	0.000	0.999	0.562
AR(4)	0.516	0.462	-0.138	0.000	1.004	0.002	0.020	-0.177	0.000	1.002	0.164	0.001	-0.210	0.000	1.001	0.368
Trend-B	0.000	2.862	-0.279	0.000	1.004	0.066	0.000	-0.229	0.000	1.001	0.651	0.000	-0.186	0.000	1.002	0.336
AR(1)-B	0.173	0.702	-0.113	0.001	0.998	0.110	0.479	0.031	0.420	0.996	0.003	0.310	0.070	0.077	0.996	0.000
AR(2)-B	0.191	0.576	-0.108	0.001	1.000	0.843	0.232	-0.063	0.096	0.999	0.199	0.075	-0.086	0.010	1.000	0.793
AR(4)-B	0.218	1.521	-0.143	0.000	1.002	0.078	0.079	-0.138	0.000	1.002	0.170	0.054	-0.140	0.000	1.002	0.046

Table C.1: Density Forecast for the US-CPI inflation 1973Q1-2012Q4 - Importance of the score driven updating rule. ‘h’ denotes the forecast horizon. ‘BK’ denotes the p-value of the test proposed by Berkowitz (2001), ‘RS’ is the test proposed by Rossi and Sekhposyan (2014) with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%). ‘ALS Diff’ and ‘CRPS Ratio’ denote the difference in the ALS and the ratio of CRPS with respect to the correctly specified model. For both statistics we also report the associated p-values of Amisano and Giacomini’s (2007). ‘KN’ denotes the p-value of Knüppel’s (2015) test.

D Additional Figures and Tables

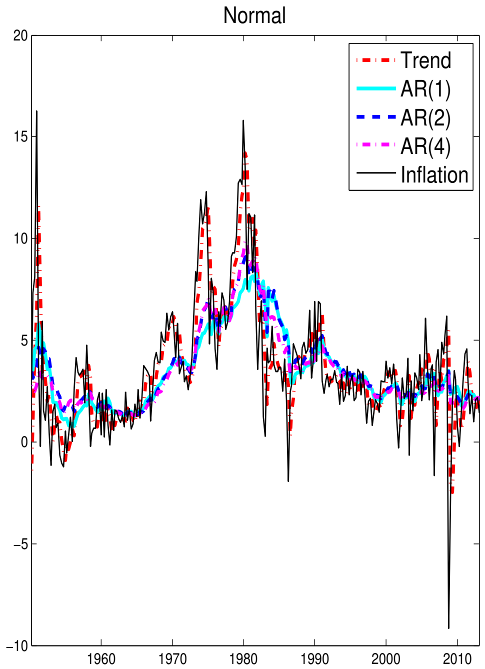

Figure D.1: Implied "long-run" inflation, $\mu_t = \phi_{0,t}/(1 - \sum_{j=1}^{p} \phi_{j,t})$, together with the realized inflation: left panel Gaussian models, right panel Student-t models.

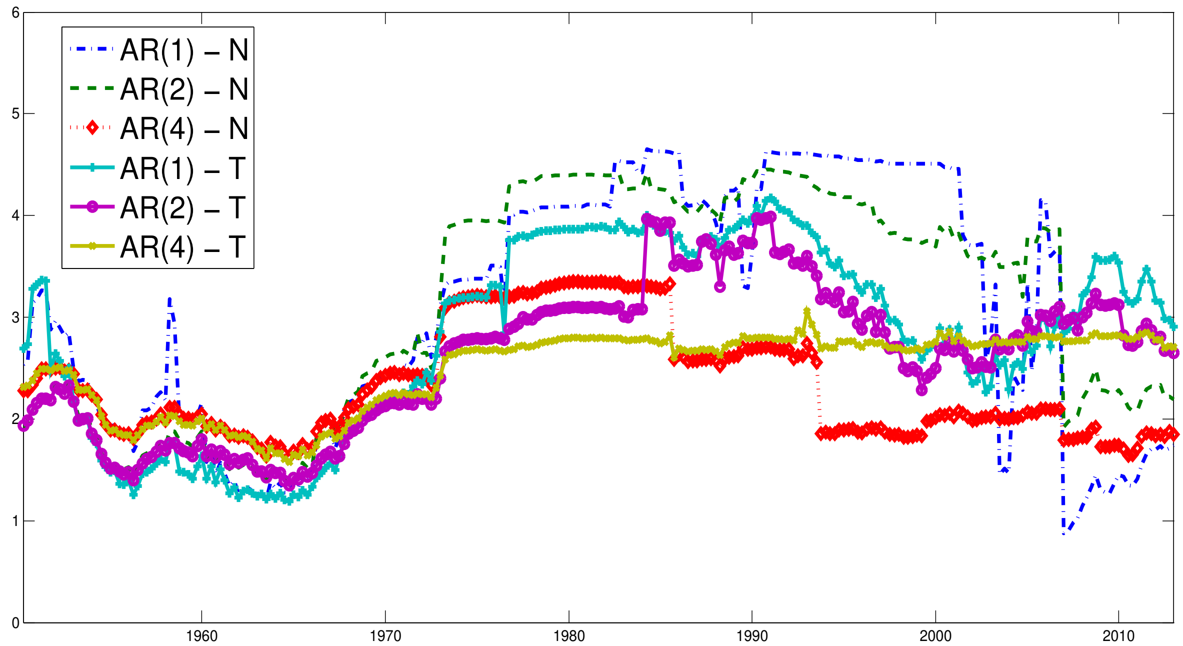

Figure D.2: The implied "restricted" long-run trend for various specifications, "N" denotes Gaussian distribution and "T" for Student-t distribution.

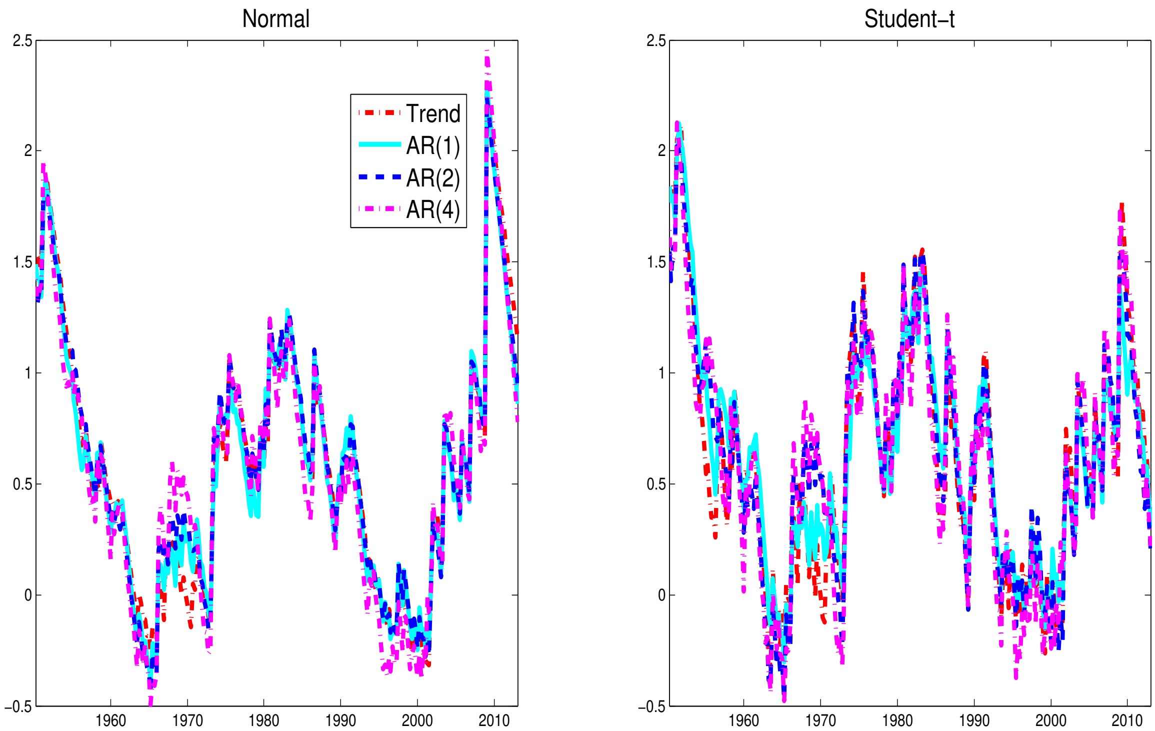

Figure D.3: Implied volatility, $\log \sigma_t$, for different specifications: left panel Gaussian models, right panel Student-t models.

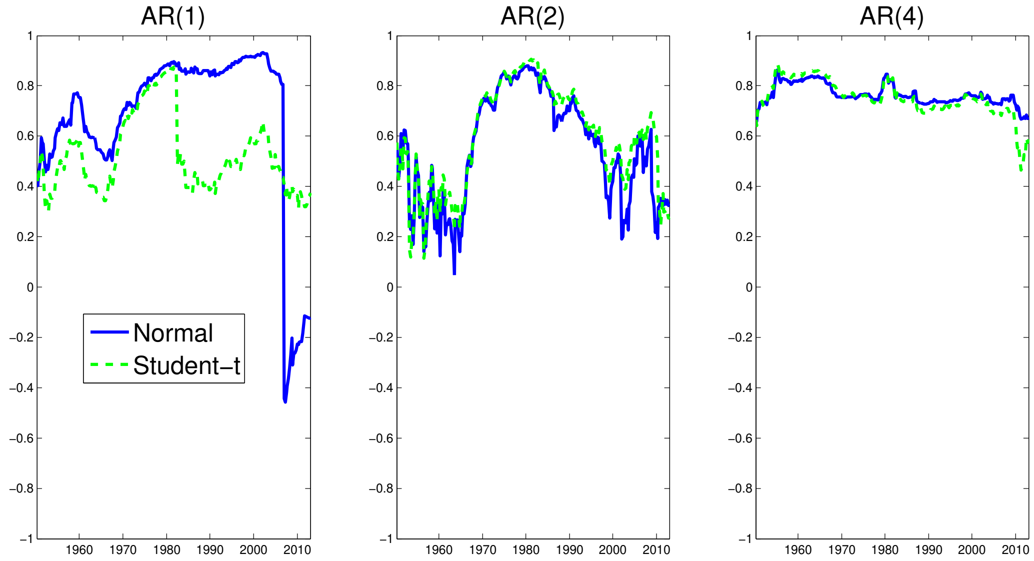

Figure D.4: Largest eigenvalue for various specifications of the model with Student-t innovations (dashed line) and Gaussian innovations (continuous line).

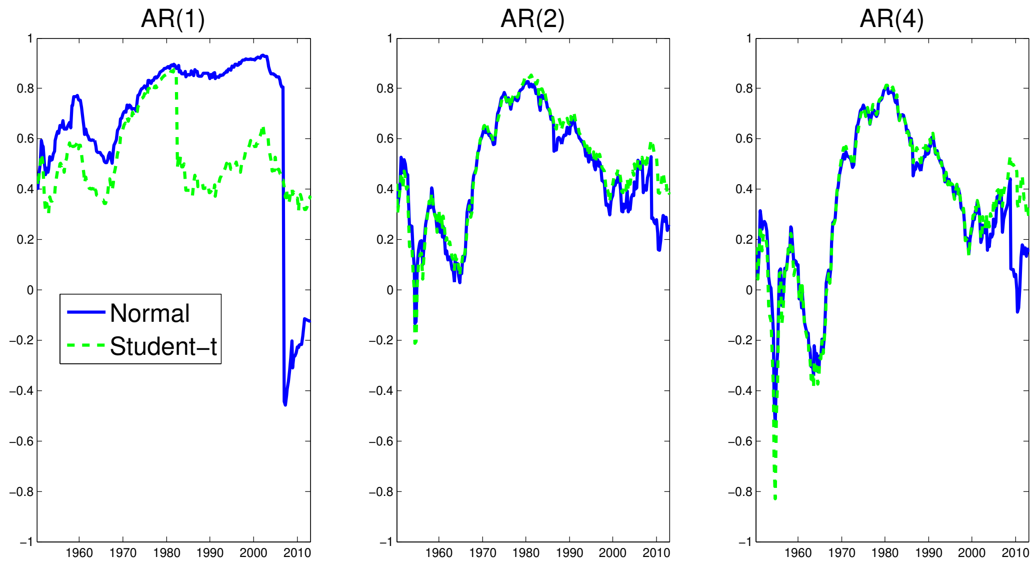

Figure D.5: Sum of the ARs coefficients for various specifications of the model with Student-t innovations (dashed line) and Gaussian innovations (continuous line).

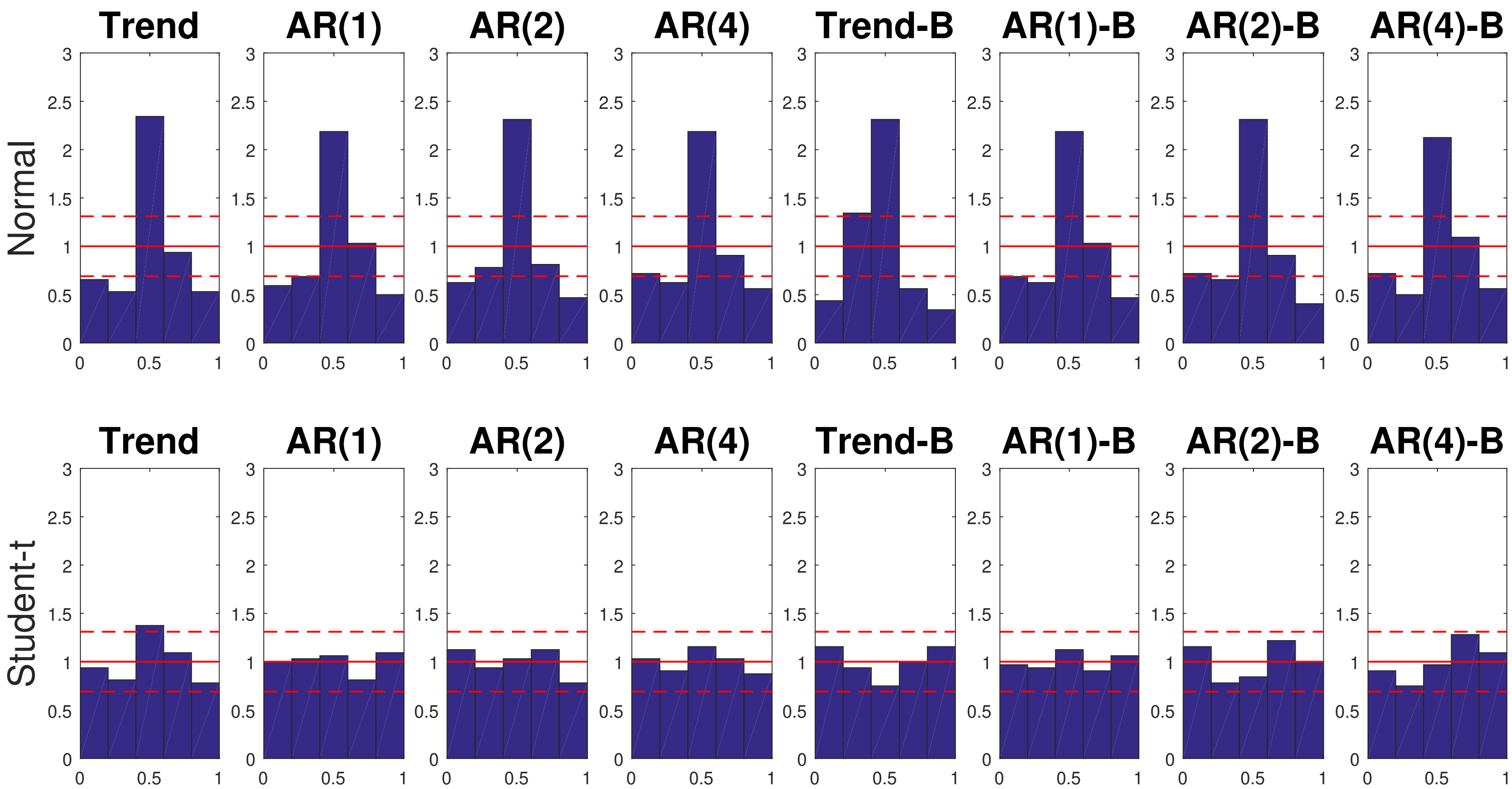

Figure D.6: Probability density function of the PITs (normalized) for various specifications of the model with 95% confidence interval (dashed lines) constructed using a Normal approximation to a binomial distribution as in Diebold et al. (1998). Forecasting horizon: one quarter.

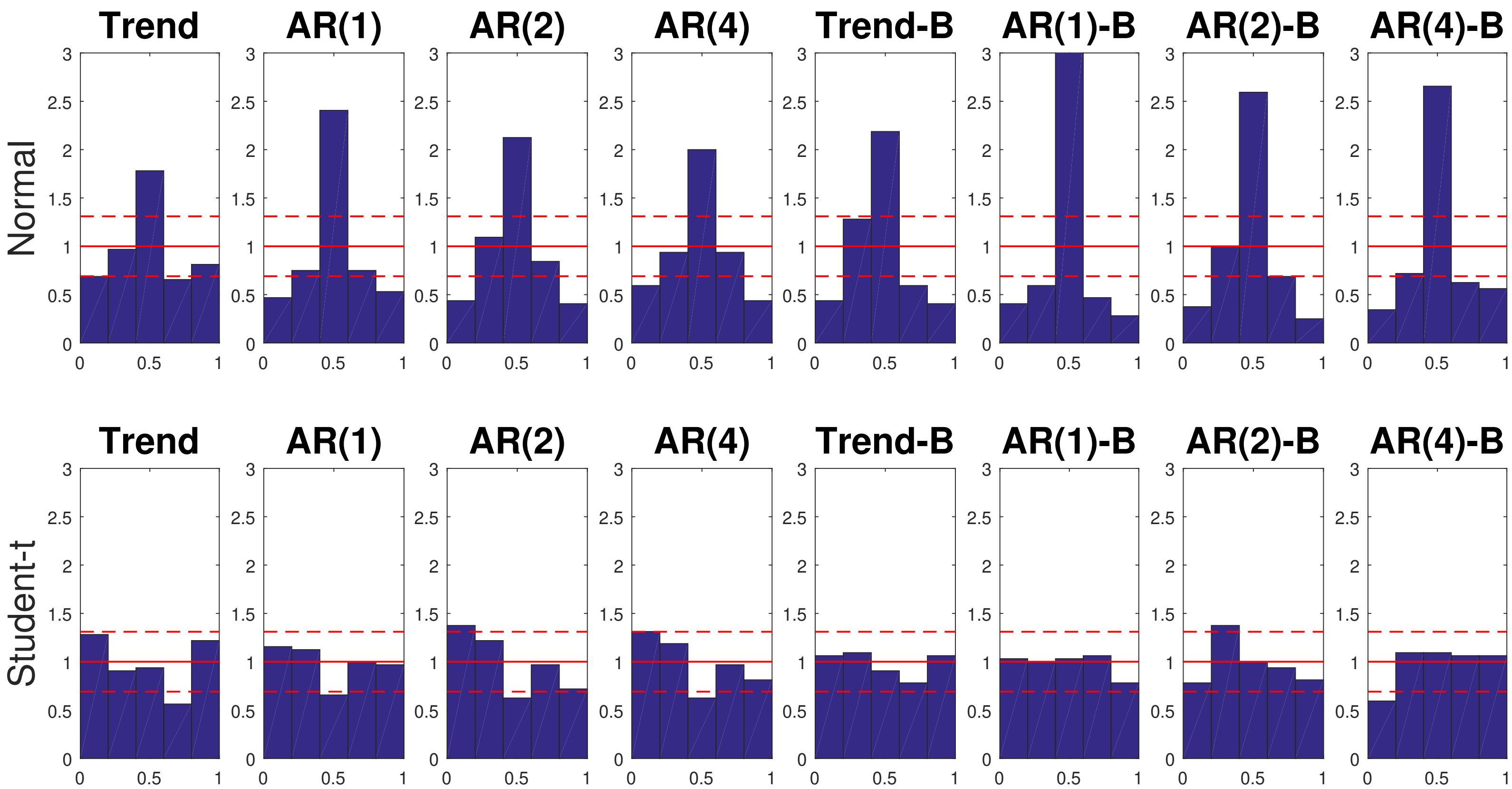

Figure D.7: Probability density function of the PITs (normalized) for various specifications of the model with 95% confidence interval (dashed lines) constructed using a Normal approximation to a binomial distribution as in Diebold et al. (1998). Forecasting horizon: one year.

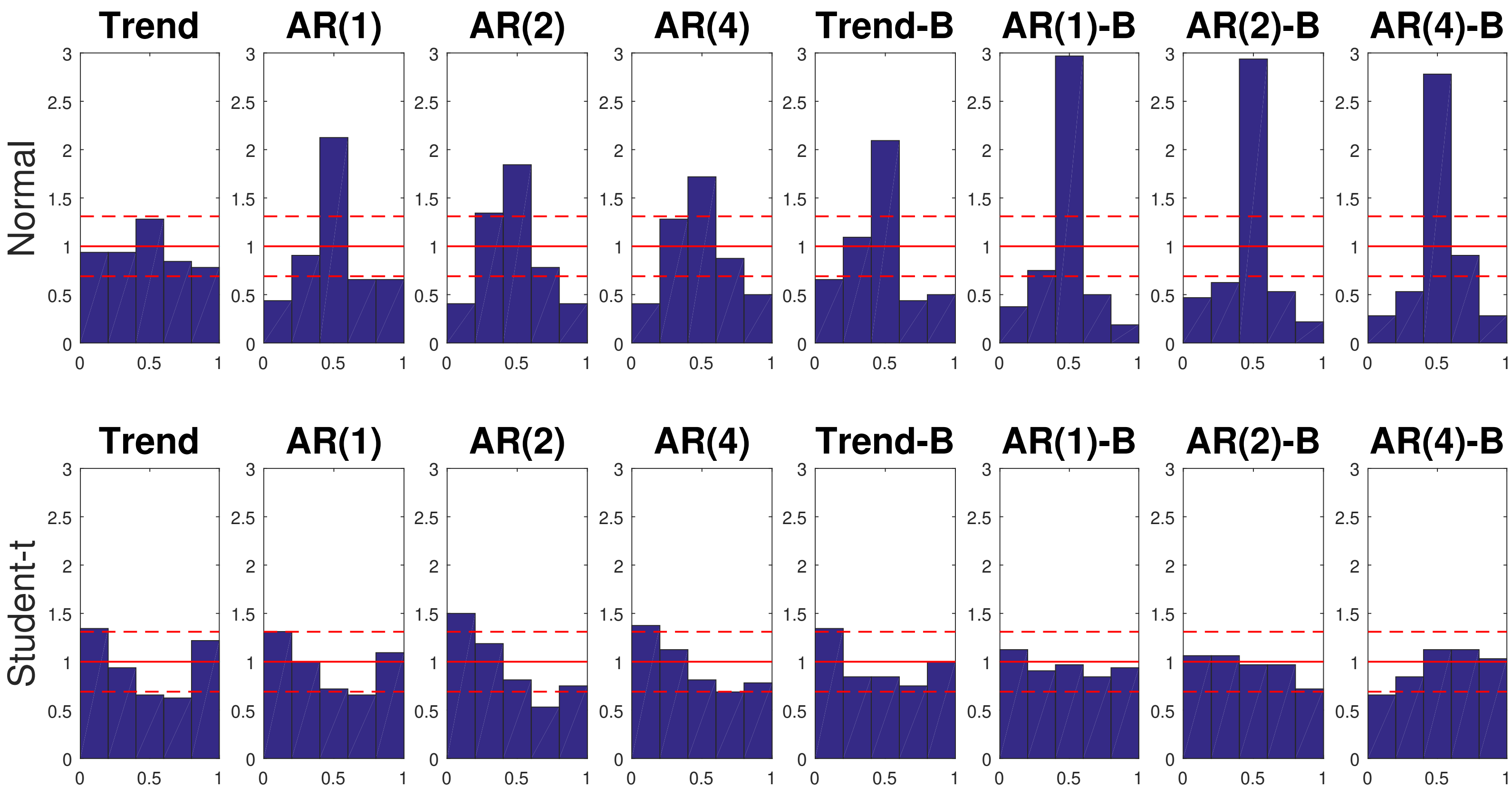

Figure D.8: Probability density function of the PITs (normalized) for various specifications of the model with 95% confidence interval (dashed lines) constructed using a Normal approximation to a binomial distribution as in Diebold et al. (1998). Forecasting horizon: two years.

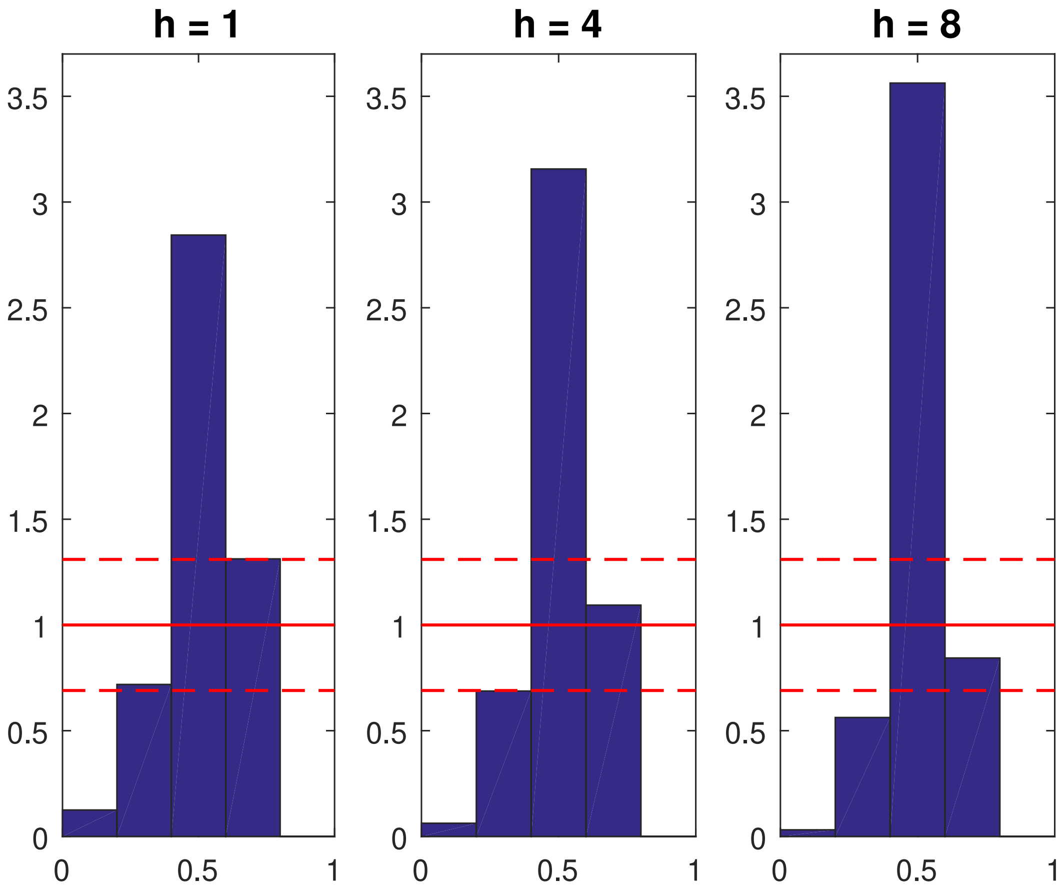

Figure D.9: Probability density function of the PITs (normalized) for one quarter (left panel), one year (middle panel) and two years (right panel) density forecasts produced under the SW model. The dashed lines denote the 95% confidence interval obtained using a Normal approximation to a binomial distribution as in Diebold et al. (1998).

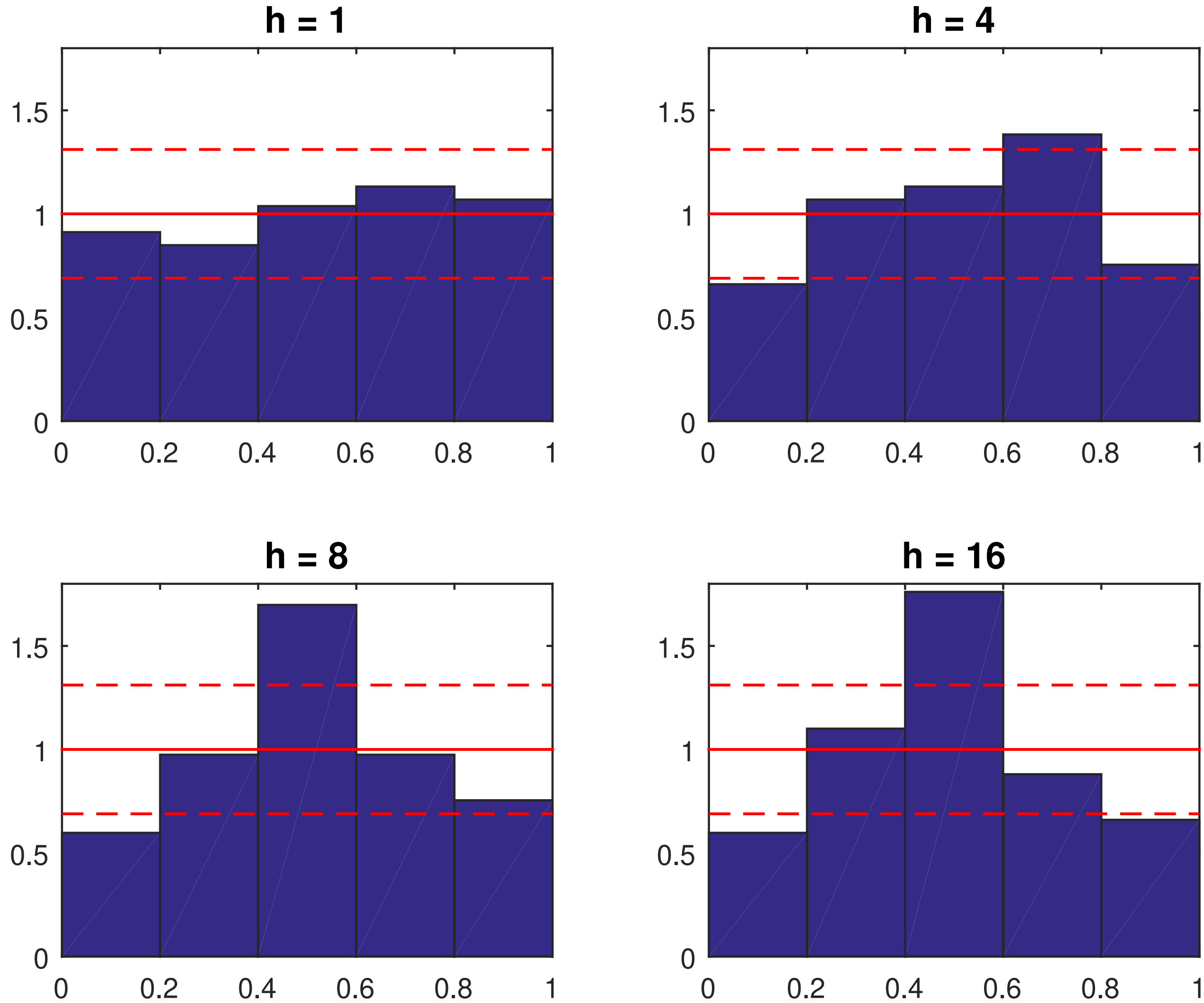

Figure D.10: Probability density function of the PITs (normalized) for one quarter (left panel), one year (middle panel) and two years (right panel) density forecasts produced under the TVP-VAR with stochastic volatility model. The dashed lines denote the 95% confidence interval obtained using a Normal approximation to a binomial distribution as in Diebold et al. (1998).

	RMSFE				MAE			
	h=1	h=4	h=8	h=16	h=1	h=4	h=8	h=16
TVP-VAR	2.131	2.786	3.143	3.196	1.477	1.927	2.004	2.167
Normal								
Trend	1.024	1.007	1.055	1.145	0.984	1.054	1.195	1.208
	(0.717)	(0.941)	(0.372)	(0.180)	(0.761)	(0.440)	**(0.006)**	**(0.066)**
Trend-B	1.473	1.152	0.958	0.990	1.572	1.249	1.169	1.144
	(0.000)	(0.290)	(0.795)	(0.936)	**(0.000)**	**(0.046)**	(0.252)	(0.410)
AR(1)	0.989	0.950	0.913	0.976	0.981	1.001	1.036	1.056
	(0.857)	(0.658)	(0.453)	(0.743)	(0.759)	(0.991)	(0.696)	(0.504)
AR(1)-B	1.068	0.968	0.931	1.025	1.074	1.000	1.059	1.097
	(0.335)	(0.796)	(0.632)	(0.847)	(0.290)	(0.998)	(0.579)	(0.535)
AR(2)	1.003	0.933	0.966	1.004	0.979	0.996	1.090	1.081
	(0.969)	(0.507)	(0.700)	(0.957)	(0.728)	(0.958)	(0.250)	(0.407)
AR(2)-B	1.051	0.943	0.924	0.980	1.042	0.984	1.027	1.012
	(0.457)	(0.613)	(0.530)	(0.823)	(0.488)	(0.854)	(0.748)	(0.908)
AR(4)	0.972	0.912	0.938	1.001	0.953	0.979	1.054	1.033
	(0.666)	(0.390)	(0.430)	(0.994)	(0.451)	(0.764)	(0.395)	(0.701)
AR(4)-B	1.176	0.968	0.940	1.020	1.065	0.976	1.028	0.998
	(0.292)	(0.801)	(0.677)	(0.889)	(0.483)	(0.801)	(0.764)	(0.989)
Student-t								
Trend	1.044	0.994	1.044	1.131	0.969	1.032	1.164	1.172
	(0.605)	(0.950)	(0.374)	(0.254)	(0.665)	(0.649)	**(0.005)**	(0.129)
Trend-B	1.436	1.118	0.928	0.952	1.449	1.152	1.104	1.043
	(0.000)	(0.419)	(0.653)	(0.677)	**(0.000)**	(0.205)	(0.444)	(0.782)
AR(1)	0.990	0.966	0.932	1.010	0.973	1.023	1.045	1.090
	(0.886)	(0.755)	(0.528)	(0.901)	(0.661)	(0.789)	(0.630)	(0.337)
AR(1)-B	0.980	0.946	0.912	0.985	0.975	0.950	0.981	0.980
	(0.757)	(0.670)	(0.545)	(0.908)	(0.701)	(0.612)	(0.852)	(0.889)
AR(2)	1.035	0.963	1.006	1.105	1.002	1.023	1.119	1.182
	(0.620)	(0.672)	(0.925)	(0.410)	(0.974)	(0.737)	(0.141)	(0.229)
AR(2)-B	1.040	0.963	0.962	1.047	1.051	0.985	1.041	1.086
	(0.549)	(0.740)	(0.726)	(0.636)	(0.386)	(0.855)	(0.576)	(0.383)
AR(4)	0.979	0.914	0.933	1.013	0.955	0.974	1.040	1.061
	(0.754)	(0.394)	(0.413)	(0.892)	(0.482)	(0.719)	(0.608)	(0.587)
AR(4)-B	1.048	0.967	0.980	1.041	1.035	0.994	1.074	1.014
	(0.534)	(0.785)	(0.884)	(0.781)	(0.609)	(0.947)	(0.447)	(0.923)

Table D.1: Point forecasts for US CPI inflation 1973Q1–2012Q4: Comparison with TVP-VAR with stochastic volatility. The Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are expressed in relative terms with respect to the SW model. The forecast horizon is ‘h’. In brackets are the p-values of Giacomini and White’s (2006) test. Values in bold denote a significance at the 10% level.

	PCE Deflator						GDP Deflator					
	RMSFE			MAE			RMSFE			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	2.169	2.204	1.188	1.572	1.683	1.267	1.628	1.804	0.899	1.201	1.359
Normal												
Trend	0.879	0.966	1.056	0.872	0.938	1.035	0.858	0.969	1.021	0.880	0.928	1.003
	(0.007)	(0.489)	(0.426)	**(0.008)**	(0.262)	(0.581)	**(0.014)**	(0.578)	(0.739)	**(0.005)**	(0.180)	(0.967)
Trend-B	1.247	1.055	0.951	1.252	1.025	0.964	1.463	1.183	1.022	1.465	1.194	1.067
	(0.001)	(0.511)	(0.680)	**(0.001)**	(0.782)	(0.760)	**(0.000)**	0.141	0.890	**(0.000)**	(0.109)	(0.649)
AR(1)	0.892	1.040	1.074	0.898	1.011	1.015	0.931	1.244	1.182	0.960	1.239	1.216
	(0.025)	(0.656)	(0.676)	**(0.080)**	(0.914)	(0.930)	(0.325)	(0.116)	(0.423)	(0.558)	(0.113)	(0.317)
AR(1)-B	0.881	0.985	1.029	0.885	0.985	1.025	0.875	1.158	1.204	0.938	1.250	1.366
	(0.015)	(0.839)	(0.849)	**(0.052)**	(0.875)	(0.874)	**(0.080)**	(0.167)	(0.256)	(0.341)	**(0.054)**	**(0.066)**
AR(2)	0.886	1.011	1.048	0.885	0.970	0.977	0.913	1.175	1.148	0.941	1.145	1.160
	(0.011)	(0.902)	(0.784)	**(0.034)**	(0.763)	(0.888)	(0.191)	(0.231)	(0.512)	(0.345)	(0.314)	(0.450)
AR(2)-B	0.878	0.940	0.981	0.871	0.939	0.985	0.840	0.924	0.948	0.875	0.928	1.002
	(0.005)	(0.319)	(0.881)	**(0.012)**	(0.416)	(0.907)	**(0.015)**	(0.287)	(0.659)	**(0.014)**	(0.273)	(0.986)
AR(4)	0.934	1.123	1.214	0.958	1.150	1.190	0.834	0.941	1.016	0.868	0.920	1.004
	(0.259)	(0.193)	(0.276)	(0.525)	(0.165)	(0.305)	**(0.019)**	(0.345)	(0.763)	**(0.011)**	(0.168)	(0.930)
AR(4)-B	0.878	0.927	1.015	0.879	0.897	0.986	0.925	1.124	1.246	0.989	1.230	1.434
	(0.035)	(0.273)	(0.875)	**(0.047)**	(0.127)	(0.881)	(0.302)	(0.249)	(0.172)	(0.871)	**(0.069)**	**(0.034)**
Student-t												
Trend	0.905	0.949	1.029	0.883	0.908	1.003	0.823	0.954	1.029	0.858	0.935	1.008
	(0.097)	(0.335)	(0.696)	**(0.031)**	(0.122)	(0.970)	**(0.008)**	(0.481)	(0.680)	**(0.003)**	(0.263)	(0.906)
Trend-B	1.271	1.075	0.982	1.313	1.064	1.012	1.434	1.148	0.984	1.420	1.144	1.016
	(0.000)	(0.372)	(0.880)	**(0.000)**	(0.498)	(0.924)	**(0.000)**	(0.235)	(0.920)	**(0.000)**	(0.223)	(0.908)
AR(1)	0.895	1.048	1.079	0.901	1.013	1.015	0.935	1.250	1.187	0.962	1.239	1.214
	(0.028)	(0.605)	(0.659)	**(0.088)**	(0.902)	(0.930)	(0.348)	(0.112)	(0.417)	(0.572)	(0.116)	(0.324)
AR(1)-B	0.873	0.955	0.993	0.877	0.952	0.988	0.865	0.974	1.003	0.886	0.931	1.021
	(0.011)	(0.504)	(0.961)	**(0.040)**	(0.584)	(0.934)	**(0.035)**	(0.775)	(0.975)	**(0.039)**	(0.353)	(0.837)
AR(2)	0.889	1.008	1.047	0.889	0.969	0.979	0.916	1.179	1.150	0.940	1.141	1.153
	(0.011)	(0.921)	(0.788)	**(0.036)**	(0.754)	(0.898)	(0.205)	(0.227)	(0.514)	(0.339)	(0.329)	(0.474)
AR(2)-B	0.900	0.929	0.980	0.901	0.946	1.000	0.817	0.917	0.935	0.864	0.922	0.997
	(0.024)	(0.179)	(0.831)	**(0.063)**	(0.391)	(0.998)	**(0.007)**	(0.299)	(0.648)	**(0.007)**	(0.298)	(0.981)
AR(4)	0.882	0.891	0.974	0.900	0.882	0.968	0.982	1.317	1.237	0.957	1.188	1.123
	(0.012)	**(0.049)**	(0.669)	**(0.045)**	**(0.045)**	(0.595)	(0.806)	**(0.090)**	(0.376)	(0.533)	(0.227)	(0.590)
AR(4)-B	0.880	0.888	0.975	0.889	0.884	0.966	0.927	0.981	1.052	0.974	0.982	1.059
	(0.037)	**(0.070)**	(0.831)	**(0.062)**	**(0.098)**	(0.772)	(0.315)	(0.772)	(0.526)	(0.666)	(0.763)	(0.470)

Table D.2: Point forecast of the US PCE Deflator 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. 'h' is the forecast horizon, in brackets the p-values of the Giacomini and White (2006) test (star when it is significant at 10% level).

			h=1						h=4					h=8				
			BK	RS	ALS		CRPS		KN	ALS		CRPS		KN	ALS		CRPS	
Normal	CA	Trend	0.000	5.726	-2.457	-	1.053	-	0.000	-2.519	-	1.269	-	0.087	-3.284	-	1.496	-
		AR(1)	0.000	6.125	-2.370	0.102	1.005	0.132	0.000	-2.641	0.005	1.250	0.705	0.000	-2.715	0.387	1.455	0.553
		AR(2)	0.000	5.341	-2.359	0.057	1.014	0.093	0.000	-2.548	0.433	1.203	0.045	0.000	-2.618	0.316	1.409	0.027
		AR(4)	0.000	7.604	-2.504	0.367	1.099	0.245	0.000	-2.801	0.000	1.323	0.416	0.000	-2.887	0.557	1.550	0.590
	FR	Trend	0.000	1.646	-2.398	-	0.761	-	0.000	-2.953	-	1.018	-	0.000	-3.156	-	1.214	-
		AR(1)	0.000	1.154	-2.203	0.541	0.780	0.495	0.000	-2.519	0.276	1.062	0.426	0.000	-2.942	0.451	1.296	0.247
		AR(2)	0.000	18.541	-2.552	0.724	1.197	0.000	0.000	-3.862	0.123	3.441	0.000	0.000	-3.982	0.089	4.415	0.000
		AR(4)	0.000	18.541	-2.730	0.449	1.334	0.000	0.000	-4.275	0.027	4.344	0.000	0.000	-4.512	0.006	6.214	0.000
	DE	Trend	0.061	1.742	-1.910	-	0.762	-	0.051	-2.189	-	0.982	-	0.000	-2.391	-	1.230	-
		AR(1)	0.002	3.281	-2.003	0.001	0.821	0.029	0.051	-2.270	0.245	1.039	0.056	0.839	-2.392	0.991	1.215	0.710
		AR(2)	0.000	3.446	-2.019	0.003	0.824	0.028	0.048	-2.216	0.695	1.000	0.573	0.695	-2.318	0.352	1.152	0.069
		AR(4)	0.000	3.584	-1.963	0.216	0.811	0.103	0.000	-2.194	0.949	1.049	0.042	0.000	-2.289	0.296	1.203	0.490
	JP	Trend	0.000	2.312	-2.587	-	0.894	-	0.000	-3.221	-	1.292	-	0.000	-4.492	-	1.559	-
		AR(1)	0.000	5.904	-2.584	0.980	0.958	0.137	0.000	-3.031	0.391	1.498	0.011	0.000	-4.116	0.233	1.823	0.020
		AR(2)	0.000	6.085	-2.522	0.537	0.932	0.287	0.000	-2.810	0.094	1.378	0.248	0.000	-3.662	0.059	1.648	0.350
		AR(4)	0.000	12.202	-2.621	0.917	1.024	0.007	0.000	-3.248	0.935	1.433	0.016	0.000	-3.444	0.204	1.657	0.038
	IT	Trend	0.005	5.127	-2.599	-	1.011	-	0.671	-2.676	-	1.208	-	0.349	-2.658	-	1.370	-
		AR(1)	0.000	11.901	-2.588	0.952	1.164	0.004	0.000	-2.811	0.129	1.541	0.000	0.000	-2.834	0.006	1.682	0.002
		AR(2)	0.000	19.253	-2.865	0.158	1.363	0.000	0.000	-3.082	0.000	1.830	0.000	0.000	-3.098	0.000	1.992	0.000
		AR(4)	0.000	16.485	-2.810	0.253	1.244	0.000	0.000	-2.986	0.006	1.596	0.000	0.000	-3.045	0.000	1.712	0.000
	UK	Trend	0.034	7.595	-2.737	-	1.285	-	0.293	-3.005	-	1.579	-	0.000	-3.402	-	1.848	-
		AR(1)	0.018	6.529	-2.855	0.082	1.329	0.509	0.029	-3.283	0.009	1.761	0.003	0.137	-3.681	0.068	1.962	0.152
		AR(2)	0.000	13.657	-2.828	0.151	1.334	0.414	0.000	-3.165	0.283	1.893	0.003	0.000	-3.317	0.702	2.165	0.037
		AR(4)	0.000	15.181	-3.102	0.000	1.603	0.000	0.000	-3.752	0.000	2.155	0.000	0.000	-4.012	0.012	2.562	0.000
Student-t	CA	Trend	0.527	0.357	-1.446	0.000	1.052	0.270	0.001	-1.663	0.000	1.242	0.125	0.000	-1.943	0.042	1.426	0.350
		AR(1)	0.362	0.897	-1.403	0.000	1.005	0.133	0.067	-1.428	0.000	1.224	0.699	0.006	-1.530	0.009	1.387	0.548
		AR(2)	0.365	1.076	-1.447	0.000	1.016	0.111	0.039	-1.516	0.000	1.179	0.045	0.000	-1.646	0.012	1.341	0.025
		AR(4)	0.370	0.928	-1.489	0.000	1.102	0.207	0.011	-1.607	0.000	1.298	0.397	0.000	-1.795	0.022	1.477	0.591
	FR	Trend	0.483	0.843	-1.430	0.011	0.762	0.184	0.000	-1.788	0.021	0.997	0.357	0.000	-1.902	0.003	1.157	0.362
		AR(1)	0.010	2.837	-1.627	0.061	0.786	0.366	0.000	-1.882	0.055	1.041	0.418	0.000	-2.086	0.018	1.225	0.298
		AR(2)	0.327	1.863	-1.561	0.038	1.200	0.000	0.014	-1.577	0.012	3.335	0.000	0.000	-1.749	0.001	4.142	0.000
		AR(4)	0.000	2.093	-1.341	0.016	1.344	0.000	0.000	-1.375	0.007	4.173	0.000	0.000	-1.459	0.001	5.734	0.000
	DE	Trend	0.062	0.441	-1.457	0.000	0.770	0.020	0.856	-1.664	0.000	0.958	0.211	0.005	-1.916	0.000	1.154	0.000
		AR(1)	0.064	0.720	-1.535	0.000	0.832	0.007	0.234	-1.652	0.000	1.023	0.042	0.093	-1.826	0.000	1.152	0.600
		AR(2)	0.032	0.984	-1.536	0.000	0.835	0.007	0.356	-1.620	0.000	0.983	0.504	0.208	-1.783	0.000	1.092	0.047
		AR(4)	0.146	1.154	-1.433	0.000	0.813	0.084	0.044	-1.577	0.000	1.028	0.041	0.001	-1.710	0.000	1.145	0.465
	JP	Trend	0.161	1.252	-1.518	0.000	0.901	0.002	0.000	-2.011	0.000	1.262	0.157	0.000	-2.303	0.003	1.471	0.000
		AR(1)	0.000	5.191	-1.589	0.002	0.970	0.067	0.000	-2.021	0.000	1.477	0.006	0.000	-2.252	0.007	1.739	0.016
		AR(2)	0.000	4.131	-1.554	0.001	0.942	0.151	0.000	-1.962	0.000	1.360	0.180	0.000	-2.178	0.006	1.574	0.317
		AR(4)	0.034	2.226	-1.643	0.004	1.110	0.000	0.014	-1.760	0.000	1.527	0.000	0.001	-1.838	0.001	1.713	0.000
	IT	Trend	0.431	1.001	-1.521	0.000	1.029	0.000	0.466	-1.687	0.000	1.191	0.102	0.014	-1.835	0.000	1.303	0.571
		AR(1)	0.000	5.904	-1.344	0.000	1.187	0.001	0.000	-1.387	0.000	1.514	0.000	0.000	-1.435	0.000	1.596	0.001
		AR(2)	0.000	5.687	-1.402	0.000	1.385	0.000	0.000	-1.436	0.000	1.785	0.000	0.000	-1.479	0.000	1.882	0.000
		AR(4)	0.000	10.450	-1.436	0.000	1.267	0.000	0.000	-1.569	0.000	1.573	0.000	0.000	-1.566	0.000	1.636	0.000
	UK	Trend	0.886	0.384	-1.632	0.000	1.311	0.000	0.018	-1.872	0.000	1.558	0.053	0.000	-2.057	0.000	1.760	0.824
		AR(1)	0.595	0.420	-1.673	0.000	1.367	0.199	0.948	-1.682	0.000	1.776	0.001	0.645	-1.778	0.000	1.916	0.052
		AR(2)	0.824	0.881	-1.551	0.000	1.355	0.241	0.728	-1.540	0.000	1.861	0.002	0.448	-1.646	0.000	2.058	0.036
		AR(4)	0.015	5.303	-1.397	0.000	1.616	0.000	0.000	-1.294	0.000	2.136	0.000	0.000	-1.282	0.000	2.467	0.000

Table D.3: Density Forecast for the G7 countries CPI inflation 1973Q1-2012Q4 - All specifications. 'h' denotes the forecast horizon. 'BK' denotes the p-value of the test proposed by Berkowitz (2001), 'RS' is the test proposed by Rossi and Sekhposyan (2014) with critical values 2.25 (1%), 1.51 (5%), 1.1 (10%), 'ALS' denotes the Average Log Score, 'CRPS' denotes the Continuous Ranked Probability Score, for both statistics we report the associated p-values of Amisano and Giacomini's (2007) test with respect to the Trend specification with Normal distribution. 'KN' denotes the p-value of Knüppel's (2015) test.