Time-varying asymmetry and tail thickness in long series of daily financial returns

Błażej Mazur, Mateusz Pipień

PRELIMINARY DRAFT: 25th of November, 2016

Abstract

We develop a univariate parametric time-series framework that is capable of capturing low-frequency fluctuations in features of conditional distribution that include tail thickness and asymmetry therefore extending the usual scale/location modelling. Potential applications include analysis of long series of daily financial returns so we generalize upon an AR(1)-GJR-GARCH-In-Mean specification. The conditional distribution is assumed to be of the asymmetric generalized Student-$t$ form of Zhu and Galbraith (2010) transformed in order to obtain its zero-mean unit-variance counterpart. It has three shape parameters so is capable of capturing skewness being distinct from tail asymmetry. We make the shape parameters time-varying with low-frequency behavior approximated using a deterministic concept of Flexible Fourier Form of Gallant (1981) so that low-frequency fluctuations can be focused on. Consequently, changes in risk measures might arise from a number of distinct sources including also low-frequency changes in variance and in tail thickness (the latter considered separately for the left and the right tail). In-sample results for 60 years of daily S&P500 returns strongly support long-run time inhomogeneity not only in variance but also in tail behavior. Tail asymmetry is strongest from the beginning of the Great Moderation till the end of the century. In certain sub-periods it is driven by fluctuations in thickness of the left tail that make existence of finite fourth conditional moment extremely unlikely. In general, after tail asymmetry is accounted for, we find no evidence in favor of skewness. We also conduct out-of-sample predictive experiment covering two expanding sequences (with more than 3000 forecasts in total). Both in-sample and out-of sample results reveal differences between the 1987 and the 2008 crashes in terms of tail asymmetry. We also evaluate quality of out-of-sample density forecasts using CRPS and LPS measures and show consequences of long-run changes in asymmetry for inference on risk characteristics (VaR and ES).

Keywords: GARCH models, time varying parameters, Bayesian inference, generalized asymmetric Student-$t$, Flexible Fourier Form, Expected Shortfall, density forecasting, daily asset returns

JEL classification: C11, C58, G10
1 Introduction

One of the topics that have received a lot of interest is the analysis of relationships linking fluctuations in financial and real part of economy. There has been a lot of attention on the real part, business cycle fluctuations and low-frequency effects, as well as long-run swings in volatility. On the other hand, models describing financial markets are less often intended to take into account possible time inhomogeneity associated with low-frequency changes. Moreover, some of the models are impractical when considering dozens of thousands of observations (corresponding to e.g. fifty years of daily returns). We believe that analysis of low-frequency dynamics in characteristics of long series of financial returns is an important preliminary step for investigation of interdependencies linking business and financial fluctuations.

Models for financial time series are most often constructed with focus on volatility and tail thickness. There is also a growing body of evidence indicating that asymmetry is an important feature of such series. Its analysis is often restricted to the use of distributions with one-parameter skewing (like skew-$t$) or simple mixtures. However, simple skewed distributions or mixtures of symmetric distributions are unable to capture differences in tail behavior between positive and negative returns. Some recent papers highlight the importance of separating tail-asymmetry (or shape asymmetry in general) from skewness, see e.g. Zhu and Galbraith, Harvey and Lange. If a model is flexible enough to allow for more than one source of asymmetry, tail asymmetry is not necessarily affected by asymmetry around mode of the distribution (and vice versa). This is likely to allow for more adequate inference on tail thickness resulting in more reliable risk evaluation.

There is a number of studies investigating long-run changes in variance of long daily financial returns series. However, we are interested in tail behavior and skewness as well. In the paper we augment a standard univariate model to make it capable of taking into account long-run (low-frequency) fluctuations in various characteristics of conditional distribution going beyond just location and scale.

The reasons for considering such a model are twofold. Firstly, it is possible that financial fluctuations that are important for tracing linkages with the real part of economy are not necessarily restricted to the first two moments of conditional distribution of observables. For instance it is not unlikely that business cycle phases coincide with changes in risk patterns that are reflected not only by the evolution of volatility but also time-varying tail thickness, tail asymmetry or skewness. In-sample identification of such low-frequency fluctuations is an important preliminary step for more advanced analyses. Secondly, there exist a practical question whether taking into account low-frequency changes in the characteristics can result in improved accuracy of out-of-sample density forecast and superior risk evaluation.

In what follows we begin with a time-series specification with changes in conditional variance decomposed into a long-term and a short-term component, similar to models considered by Engle and Rangel (2008) as well as Amado and Teräsvirta (2009). However, we add two important extensions. Firstly, the conditional distribution is assumed to belong to two-piece skewed class with shape asymmetry. More specifically,
we make use of the generalized asymmetric $t$ of Zhu and Galbraith. The distribution has three shape parameters controlling skewness and thickness of left and right tail, therefore allowing for tail asymmetry being modelled separately from skewness. Secondly, we extend the idea of long-run changes in variance, as similar fluctuations are introduced into skewness, left tail thickness and right tail thickness of conditional distribution.

The resulting model class encompasses a considerable number of interesting nested special cases that are potentially useful in applied work. In particular it contains models displaying different deviations from stationarity (including specifications that are covariance stationary but non-stationary in the strict sense). Moreover, it can be interpreted as belonging to the strand of literature generalizing upon the stationarity assumption by introducing time-variation in model parameters.

In order to trace the long-run dynamics we make use of Flexible Fourier Form of Gallant (1984). It is important that the pattern of time-variation considered here is deterministic (being a direct alternative to e.g. spline-based models or models with sets of logistic functions). We refer to the pattern as long-run, since parametric restrictions are imposed in order to rule out high-frequency changes.

The Flexible Fourier Fourier form has some obvious limitations but is well-suited to capture cyclical behavior and implies non-explosive out-of-sample extrapolation. The latter might be problematic for splines or polynomials. Another reason for restricting time-variation in tail behavior to low frequency movements arises from the idea that it corresponds to the occurrence of rare events. Therefore it seems that abrupt or short-run changes might be difficult to identify, possibly resulting in an overfitted specification with limited predictive power.

A fundamentally different approach is based on latent stochastic processes with e.g Markov switching or change-point processes, latent AR processes etc. However, such models are much more difficult to estimate (especially when facing serious deviations from normality that are of interest here). Moreover, the degree of computational complexity for unobserved component models is quite high, which might be restrictive when long series are analyzed. We assume that the Flexible Fourier Form might provide a relatively simple and numerically feasible approximation to a pattern generated by some long-memory latent process. The resulting model can be useful as it casts some light on low-frequency time-inhomogeneity of more subtle stochastic characteristics of the process under consideration such as skewness, tail thickness and tail asymmetry of conditional distribution of daily asset returns.

For the sake of inference we make use of Bayesian techniques as their validity does not rely on existence of moments of conditional distribution. Throughout the paper we assume that its variance is finite, but the assumption is not crucial estimation-wise and could be relaxed. This might be important as our empirical results indicate non-existence of finite fourth conditional moment in certain sub-periods of the sample. We focus on investigating changes in rather subtle features of conditional distribution so the assessment of estimation uncertainty is crucial. Its reliable evaluation might be non-trivial within the
standard Maximum Likelihood approach. Moreover, as the model is highly non-linear and non-Gaussian, the usual large sample approximation might be inadequate. However, within the Bayesian setup the estimation uncertainty is fully accounted for and does not rely on additional assumptions as to existence of moments of conditional distribution. This is true for in-sample inference as well as out-of-sample density forecasting (including risk evaluation).

The rest of the paper is organized as follows. Firstly, we briefly review some of the existing results on modeling long-term time inhomogeneity in financial time-series. A simplified model with long-run and short-run variation in volatility is recalled in order to discuss connections with the existing literature. Secondly, we introduce a fully general model with long-run variation in various features of conditional distribution. We provide a parametric example (stating explicitly some short run dynamics) and discuss its special cases. Thirdly, we present an empirical application of the approach involving analysis of 15624 daily returns on S&P500 covering a time-span of sixty years.

Empirical model comparison based on full sample is conducted in order to identify the features that display long-run time-inhomogeneity. It turns out that there is clear evidence in favor of low-frequency variation in volatility and tail thickness (the latter is such that implies low-frequency changes in tail asymmetry). However, there is no evidence in favor of skewness (neither time-varying nor time-invariant, though we do not consider short-term variation in skewness). Interestingly, the appearance of strong tail asymmetry coincides with the beginning of the Great Moderation. The effect is driven by increasing thickness of the left tail (implying that the existence of finite fourth conditional moment is extremely unlikely for certain sub-periods). Tail asymmetry decreases towards the end of the sample, so we find serious differences between the crashes of 1987 and of 2008 (the former is characterized by strong tail asymmetry whereas the latter has occurred during a period of approximate tail symmetry). Throughout the sample tail asymmetry displays a quasi-cyclical pattern with reoccurrence of periods of approximate tail symmetry. It is possible that the pattern of long-term variation described here is to some extent influenced by properties of the Flexible Fourier Form. However, it seems that the evidence in favor of some form of long-term variation in tail thickness and tail asymmetry in the sample we consider is indeed quite strong.

In order to verify the above in-sample results we conduct an out-of-sample recursive forecasting experiment covering two sequences of expanding sub-samples. We pick one verification period with strong changes in tail asymmetry and one with a relatively stable pattern according to the full-sample results. Moreover, the periods include two largest negative returns in the sample. We compare density forecasting performance of a model allowing for low-frequency changes in features to a benchmark with no long-term variation using CRPS and LPS measures. Results of the comparison are in line with the full-sample analysis indicating clear performance gains for the more general model during the period with pronounced shifts in tail thickness and tail asymmetry. Moreover, the evidence does not rely on a limited
number of outliers but builds up in a rather systematic way. In particular predictive evidence in favor of a model with long-term time variation in tail asymmetry (against time-invariant benchmark) starts to build up about a year before the 1987 crash. We also investigate the implications of time-variation in volatility and shape for risk assessment analyzing differences in estimates of value-at-risk and expected shortfall.

2 Non-stationary extensions of a standard GARCH model

Lamoreux and Lastrapes (1990) and Engle and Mustafa (1992) suggested that parameters of GARCH-type models might be unstable over time. Consequently, the constancy of parameters initially imposed in GARCH-type processes was subject to criticism that prompted new studies concerning various generalisations. In particular Mikosh and Stărică (2004) state that the IGARCH effect is often spuriously supported by data, because in the case of long time series variability of parameters is natural; see also Hillebrand (2005).

One can indicate two basic approaches applied in this respect, one relating to variability governed by a random process and one relying on a deterministic framework. As to the first approach, Hamilton and Susmel (1994) assumed that stock returns are characterised by different ARCH processes at different points in time, with the shifts between processes mediated by a Markov chain. The study has opened new topics in financial econometrics based on the application of Markov switching mechanisms in volatility modelling. The second approach is recalled below in Section 2.1.

Another strand of the literature focuses on nonparametric methods using very general conditions concerning the regularity of parameters treated as functions of time, see Härdle et al. (2003), Mercurio and Spokoiny (2004), Spokoiny and Chen (2007). Čižek and Spokoiny (2009) present a review of literature concluding that relaxing time homogeneity of the process is a promising approach but causes serious problems with proper estimation methods. For instance, when some or all model parameters will vary over time, a more subtle treatment of testing structural breaks in financial returns may be obtained; see Fan and Zhang (1999), Cai et al. (2000), Fan et al. (2003). In general it is difficult to point out a practical approach that allows for nonparametric analysis of low-frequency changes in features concerning asymmetry or tail thickness of conditional distribution in long series of daily financial returns.

2.1 The basic case: long-run changes in volatility

Suppose that we observe logarithmic returns on a financial instrument with price $x_t$ at time $t$ given by the form: $y_t = 100 \ln \frac{x_t}{x_{t-1}}$. Denote by $y = (y_1, \ldots, y_T)$ the vector of modelled observations. We focus on
properties of the error process $\varepsilon_t$ in the observation equation

$$y_t - E(y_t|\Psi_{t-1}) = \varepsilon_t, \quad t = 1, \ldots, T,$$

having the following general compound structure:

$$\varepsilon_t = \sqrt{\tau_t} \sqrt{h_t(\omega, \Psi_{t-1})} z_t.$$

In (2) $h_t(\omega, \Psi_{t-1})$ describes short run fluctuations of volatility as a parametric GARCH-type function of the information set $\Psi_{t-1} = (\ldots, \varepsilon_{t-2}, \varepsilon_{t-1})$, i.e. the history of the process with parameters $\omega$. We assume that $\tau_t$ describes long term fluctuations of the volatility. The process $z_t$ represents a sequence of iid random variables with zero mean and unit variance. We implicitly condition on some fixed initial values $y_0$.

Importantly, the above formulation can be re-interpreted in terms of time-variation in parameters of the short-term volatility equation represented by $h_t$, so that $\sqrt{\tau_t} \sqrt{h_t(\omega, \Psi_{t-1})}$ is replaced by short-run component with time-varying parameters $\sqrt{h_t(\omega_t, \Psi_{t-1})}$ where time-evolution of $\omega_t$ is described by $\tau_t$, see APPENDIX.

Within the general framework (2) many alternative specifications have been analysed. In particular Engle and Rangel (2008) assume $h_t(\omega, \Psi_{t-1})$ to be of the GARCH(1,1) form whereas the low frequency changes in the volatility are described by the following specification, denoted here by $\tau_t^{ER}$:

$$\tau_t^{ER} = c \exp \left( \omega_0 t + \sum_{i=1}^{k} \omega_i (t - t_{i-1})^2 \right).$$

The original spline-GARCH model augments the above spline function with a linear function of some state variables $v_t/\beta$. Since we are interested in forecasting, such a block of exogenous variables would potentially require some extension for the sake of prediction. All the specifications considered here could in principle be extended in such manner. However, at the stage we exclude any exogenous variables from the long-run volatility equation and therefore we limit the above long-term component to the spline part.

Amado and Teräsvirta (2014) have developed a smooth transition approach to define $\tau_t$ using a positive-valued linear combination of bounded transition functions:

$$\tau_t^{AT} = \delta_0 + \sum_{l=1}^{r} \delta_l G_l(\frac{t}{T}),$$

where each $G_l(\frac{t}{T})$ is a generalised logistic transition function with parameter $k \in \mathbb{N}$ determining the number of exponential elements. In this approach the transition functions allow the unconditional variance of the process $\varepsilon_t$ to change smoothly with calendar time $\frac{t}{T}$. 
Engle and Rangel (2008) introduce time variability of unconditional variance as they incorporate into $\tau_{ER}^t$ a slow-moving component, represented by an exponential spline. Since $\tau_{ER}^t$ is built on the basis of a purely non-parametric construct, the pattern of long term changes in the volatility is not necessarily regular. This irregular character, dependent on the number of spines - $k$, is confirmed by empirical analyses conducted by Engle and Rangel (2008). The logistic function, applied by Amado and Teräsvirta (2014) in $\tau_{AT}^t$ also involves a slowly changing volatility component corresponding to variability of parameters according to a smooth transition mechanism. Just like in the previous case, irregular pattern of unconditional variance is obtained as a result of non-parametric construct included in the volatility equation.

Analogously to modelling economic activity of the real sector, it might be interesting to check whether the volatility of financial time series also evolves over time following phases similar to the usual expansion and contraction pattern of macroeconomic data. In order to verify the existence of cyclical features in long term changes of volatility in a more direct way, a process with an approximately periodic structure of unconditional moments was considered by Mazur and Pipień (2012). In the general setting they apply Flexible Fourier Form (FFF), see Galant (1981):

$$\tau_{MP}^t = \exp[f(t, \gamma)],$$

where

$$f(t; \gamma) = \sum_{i=1}^{F} (\gamma_{si} \sin(\phi_i t) + \gamma_{ci} \cos(\phi_i t)),$$

with $\gamma = (\gamma_{s1}, ..., \gamma_{sF}, \gamma_{c1}, ..., \gamma_{cF}, \phi_1, ..., \phi_F)$. Function $f(t, \gamma)$ is defined as a sum of periodic functions where parameters $\phi_i$ determine frequencies, while $\gamma_{si}$ and $\gamma_{ci}$ control amplitudes and phase shifts. Since the infinite Fourier series is replaced by its finite substitute, formula (4) yields an approximation of finite order $F$ to Almost Periodic (AP) function describing variability of the unconditional variance (see references quoted by Mazur and Pipień, 2012). As compared to $\tau_{AT}^t$ and $\tau_{ER}^t$, the slowly changing volatility component can be interpreted in terms of existence of cycles.

2.2 The general case: long-term variation in features of the conditional distribution

In order to generalize the above setup we begin with the following formulation for time-varying conditional location parameter:

$$\mu_t = \delta + \lambda_1 \sqrt{h_t(\omega, \Psi_{t-1}) \sqrt{\exp[f_c(t)]}} + f_m(t)$$

which includes a GARCH-In-Mean-type structure initially proposed by Engle, Lilien and Robins (1987). The long run volatility component $\tau_1$ in (2) is here denoted by $\exp[f_c(t)]$ while $h_t(\omega, \Psi_{t-1})$ represents its
short-term counterpart. Conditional volatility is implied by the following structure of error terms:

$$
\varepsilon_t = \sqrt{\exp[f_m(t)]} \sqrt{h_t(\omega, \Psi_{t-1})} z_t,
$$

where $$z_t$$'s have unit variance. Functions $$f_m(t)$$ and $$f_v(t)$$ are intended to generate long-run fluctuations in mean and variance of the observable process.

A full parametric model requires specification of short-run dynamics, though at the stage we do not formulate specific assumptions on the short-term volatility component $$h_t(\omega, \Psi_{t-1})$$ and the linkage between the observable process $$y_t$$, the location $$\mu_t$$ and the error term $$\varepsilon_t$$.

The generalization considered here relies on the fact that a more complicated distribution of $$z_t$$ is assumed and long-term variation is imposed on its shape parameters. Here $$z_t$$'s are independent D(0,1) variables that are no longer identically distributed. This adds a number of additional sources of non-stationarity in the strict sense.

We assume that $$z_t$$'s follow the generalized asymmetric Student-t of Zhu and Galbraith (2010), hereafter labelled AST. The distribution is a member of two-piece skew family. The skewed extension of the Student-t distribution proposed by Hansen (1994) results from an application of so-called two-piece method. Zhu and Galbraith (2010) generalise this approach and discuss a distribution obtained by means of mechanism incorporating inverse scale factor of Hansen (1994), see also Fernández and Steel (1998). The crucial feature of AST distribution is its shape asymmetry: the left and the right tail of the density decay at different polynomial rates.

An AST-distributed variable might be considered a mixture of two re-scaled half-t distributions with mode at zero (one negatively oriented) having different degrees of freedom. The skewing effect comes from the differences in weights attributed to the two pieces. They are also re-normalized in order to make the resulting probability density function continuous at the mode. The mixture is then shifted to a new location and re-scaled, so the resulting probability density function is:

$$
f_{AST}(z|m, s, \alpha, \nu_1, \nu_2) = \begin{cases} 
\frac{1}{s} \left[ 1 + \frac{1}{\nu_1} \left( \frac{z - m}{2asK(\nu_1)} \right)^2 \right]^{-\frac{\nu_1+1}{2}}, & z \leq m \\
\frac{1}{s} \left[ 1 + \frac{1}{\nu_2} \left( \frac{z - m}{2(1-\alpha)sK(\nu_2)} \right)^2 \right]^{-\frac{\nu_2+1}{2}}, & z > m
\end{cases},
$$

where $$s > 0$$ controls overall scale of the distribution and $$m$$ denotes its mode. Moreover, $$0 < \alpha < 1$$ corresponds to the weight of the left part in the mixture (being the probability mass to the left of the mode $$m$$) thus reflecting the two-piece skewing mechanism. Most importantly, the degrees of freedom $$\nu_1 > 0$$, $$\nu_2 > 0$$ are different in both parts of the distribution which implies tail asymmetry as $$\nu_1 \neq \nu_2$$.

Obviously, with $$\nu_1 = \nu_2$$, $$f_{AST}$$ defines the skewed Student-t distribution considered by Hansen (1994) and Fernández and Steel (1998). We hereafter separate shape asymmetry from skewness (see e.g. Harvey and
Lange, 2015) and hence any further reference to skewness is equivalent to the situation where $\alpha \neq 0.5$. This does not assume existence of the finite third moment of the distribution; moreover, the distribution that is not skewed (in the sense assumed here so with $\alpha = 0.5$) might nevertheless be asymmetric (due to the fact that $\nu_1 \neq \nu_2$).

In what follows we assume that the AST distributed variables $z_t$ have finite variance which is equivalent to $\nu_i > 2$ for $i = 1, 2$. Moreover, contrary to the original approach of Zhu and Galbraith (2010), the above distribution is here standardized in the sense that $m$ and $s$ in (7) are functions of shape parameters (i.e. of $\alpha$, $\nu_1$ and $\nu_2$) such that the resulting variable has zero mean and unit variance. This has consequences for estimation as the implied likelihood kernel is more complicated.

We emphasize that the restriction $\nu_i > 2$ might easily be relaxed in a modified model based on the original, untransformed distribution $\textup{AST}(\nu_1, \nu_2)$. The approach would focus on direct modelling of location $m$ and scale $s$ (or precision) instead of mean and variance. The treatment would be somewhat more general in terms of tail thickness by not excluding the cases with $0 < \nu_i < 2$ (therefore allowing for e.g. Cauchy tails). However, it would also imply interferences between dynamics of the shape parameters and the first two conditional moments (once finite). Here we prefer to parameterize the model explicitly in terms of mean and variance of errors in order to preserve certain connections with the existing literature. We generalize well-established models for dynamics in variance separating the variation in shape from the variation in the first two moments of the conditional distribution. However, we do not require third conditional moment to be finite, so we prefer to avoid references to e.g. time-varying conditional kurtosis for the sake of generality.

In the paper we investigate potential time-varying asymmetries of conditional distribution for daily asset returns. The above discussion suggests that one might prefer to utilize such asymmetry measures that do not require existence of higher-order moments.

The above setup implies that the sequence of independent $\textup{D}(0,1)$ variables $z_t$ following zero-mean unit-variance AST distribution is characterized by three time-varying shape parameters $\alpha_t$, $\nu_{1,t}$ and $\nu_{2,t}$ (corresponding to $\alpha$, $\nu_1$ and $\nu_2$ in (7)). Furthermore we let:

$$\alpha_t = 0.5 + \frac{1}{\pi} \arctan(\alpha_{0,0} + f_{\alpha}(t)),$$

$$\nu_{1,t} = \nu_{min} + \nu_{1,0} \exp(f_{lt}(t))$$

and

$$\nu_{2,t} = \nu_{min} + \nu_{2,0} \exp(f_{rt}(t)).$$

To summarize, our general formulation allows for long-run time variation in mean, variance, skewness and thickness of the left and the right tail, controlled by $f_m(t)$, $f_{\alpha}(t)$, $f_{lt}(t)$, $f_{rt}(t)$ and $f_{\nu}(t)$ being real-valued.
functions of time. Of course in the presence of the In-Mean effect (nonzero $\lambda_1$ in (5)), the mean-variance relationship is more involved. If $f_\alpha(t)$, $f_\xi(t)$ and $f_\gamma(t)$ describe fluctuations around zero, the average level of the skewing parameter $\alpha_t$, the left-tail degrees of freedom $\nu_{l,t}$ and the right tail degrees of freedom $\nu_{r,t}$ are controlled by $\alpha_{0,0}$, $\nu_{l,0}$ and $\nu_{r,0}$, respectively and no identification issues arise. As mentioned above, we set $\nu_{\min} = 2$ (implying $\nu_{l,t} > 2$ and $\nu_{r,t} > 2$) in order to ensure that variance of $z_t$ exists. The existence of higher-order moments of conditional distribution can be verified empirically.

2.3 A statistical model for daily returns

Practical use of the above general setup requires specific assumptions on short-run variability in conditional mean and variance. Thought we emphasize that the framework described in the previous section allows for a more complicated specification, for the sake of empirical analysis we make use of an AR(1) process:

\[
y_t = \mu_t + \rho(y_{t-1} - \mu_{t-1}) + \varepsilon_t,
\]

which is often used in analysis of daily financial returns, see for example Bauwens, Lubrano and Richard (1999) for a univariate case or Osiewalski and Pipień (2004) for a multivariate setting. The error term $\varepsilon_t$ in (8) has the compound scaling structure (4). Following Amado and Teräsvirta (2009, 2014) for the short-run volatility component $h_t(\omega, \Psi_{-1})$ we assume GJR(1,1) formulation proposed by Glosten, Jagannathan, Runkle (1993):

\[
h_t = \alpha_0 + \alpha_1^+ \xi_{t-1}^2 I(\xi_{t-1} \geq 0) + \alpha_1^- \xi_{t-1}^2 I(\xi_{t-1} < 0) + \beta_1 h_{t-1},
\]

with $\xi_t = \sqrt{h_t(\omega, \Psi_{-1})} z_t$, $\alpha_0 > 0$, $\alpha_1^+ \geq 0$, $\alpha_1^- \geq 0$, $\beta_1 \geq 0$. The news impact curve is therefore asymmetric as $\alpha_1^+ \neq \alpha_1^-$. Moreover, we assume that all the functions governing the long-run evolution of parameters, namely $f_m(t)$, $f_\alpha(t)$, $f_\xi(t)$, $f_\gamma(t)$ and $f_\gamma(t)$ represent Flexible Fourier Forms analogous to (4). As the number of terms in the Fourier expansion (denoted by $F$ in (4)) is now function-specific, we label it $F_m$, $F_\omega$, $F_\alpha$, $F_\gamma$ and $F_{\gamma\gamma}$, respectively. The parameters of individual functions are distinguished using similar superscripts, e.g. $\gamma^{(\omega)} = (\gamma_1^{(\omega)}, ..., \gamma_{sF}^{(\omega)}, \gamma_{sF}^{(\omega)}, ..., \gamma_{cF}^{(\omega)}, \phi_1^{(\omega)}, ..., \phi_F^{(\omega)})$ denote parameters of $f_\omega(t)$. We assume that $M_{F_m, F_\omega, F_\alpha, F_\gamma, F_{\gamma\gamma}}$ corresponds to the general, unconstrained model framework. As for instance $F_\omega = 0$ implies $f_\omega(t) = 0$, the model $M_{F_m, F_\omega, F_\alpha, F_{\gamma\gamma}}$ exhibits long-run variability in conditional mean, variance and skewness, but the right and the left tail of the conditional distribution display thickness that is constant over time. $M_{0, F_\omega, F_\alpha, F_{\gamma\gamma}}$ refers to a model with long-term variability in conditional variance only, illustrating the case considered in section 2.1. The model with no long-run variation (i.e. a GARCH-type model with constant parameters) is denoted by $M_{0,0,0,0,0}$, which here corresponds to AR(1)-GJR(1,1)-In-Mean process with AST disturbances. It allows for time-invariant skewness and tail asymmetry with dynamics.
in conditional mean and variance restricted to short-run fluctuations.

Moreover, we consider a number of additional special cases representing equality restrictions imposed on parameters of the Flexible Fourier Forms mentioned above. In particular $M^*$ denotes the case with $f_{lt} \equiv f_{rt}$ whereas $M^{**}$ indicates that $f_{lt}$ and $f_{rt}$ share the same frequency parameters ($\phi_{lt} = \phi_{rt}$) but have different amplitude/phase shift parameters (so $\gamma_{lt} \neq \gamma_{rt}$), with $F_{lt} = F_{rt} > 0$ (see [4]). However, even in the case of $f_{lt} \equiv f_{rt}$, tails of the conditional distribution are not necessarily the same, as long as $\nu_{1,0}$ and $\nu_{2,0}$ are different. Consequently, $M_{0,2,0,3,3}^*$ denotes a specification with long-term variation in variance and tail thickness only, with two Fourier components describing the evolution of volatility and three components for tail thickness. Moreover, dynamics of the left and the right tail degrees of freedom is based on the same frequency parameters though involve different amplitudes and phase shifts. Such a model allows for time-varying tail thickness and tail asymmetry though it rules out time-varying skewness.

The resulting model class encompasses quite a broad spectrum of possible deviations from stationarity, as for example $M_{0,0,2,1,1}$ corresponds to a model that is covariance stationary being non-stationary in the strict sense. The model class provides a useful tool for investigation of patterns of long frequency dynamics in features of the conditional distribution of daily returns.

Full likelihood of the unconstrained model $MF_{lm,F,r,F,u,F,t}$ is implied by the form of conditional density of observations. The latter is in turn implied by equation (EQ), together with (EQ) and the conditional density of the error terms $\varepsilon_t$:

$$p(\varepsilon_t|\Psi_{t-1}) =$$

$$= \begin{cases} 
\frac{B_d(\alpha_t, \nu_{1,t}, \nu_{2,t})}{\sqrt{\nu_{t,1}, \exp(f_{lt}(t))}} \left[ 1 + \frac{1}{\nu_{1,t}(2\alpha_{lt})^2} \left( \varepsilon_t d(\alpha_t, \nu_{1,t}, \nu_{2,t}) + c(\alpha_t, \nu_{1,t}, \nu_{2,t}) \right)^2 \right]^{\frac{1+\nu_{1,t}}{2}}, & \varepsilon_t \leq m_{lt,} \\
\frac{B_d(\alpha_t, \nu_{1,t}, \nu_{2,t})}{\sqrt{\nu_{t,1}, \exp(f_{lt}(t))}} \left[ 1 + \frac{1}{\nu_{2,t}(2(1-\alpha_{lt})^2) \left( \varepsilon_t d(\alpha_t, \nu_{1,t}, \nu_{2,t}) + c(\alpha_t, \nu_{1,t}, \nu_{2,t}) \right)^2 \right]^{\frac{1+\nu_{2,t}}{2}}, & \varepsilon_t > m_{lt,} 
\end{cases}$$

derivation of the above formula together with expressions for the relevant functions are given in Appendix A.

The full vector of all the model parameters (denoted by $\theta$) contains the following elements:

$$\theta = (\delta, \lambda_1, \rho, \alpha_0, \alpha_1^+, \alpha_1^-, \beta_1, \alpha_{0,0}, \nu_{1,0}, \nu_{2,0}, \gamma^{(m)}, \gamma^{(v)}, \gamma^{(\alpha)}, \gamma^{(lt)}, \gamma^{(rt)})$$

For the sake of Bayesian inference the likelihood function has to be augmented with priors for all the model parameters. For simplicity we assume that all the priors are proper and the joint prior is a product of priors for individual parameters:

$$p(\theta) = p(\delta)p(\lambda_1)p(\rho)p(\alpha_0)p(\alpha_1^+)p(\alpha_1^-)p(\beta_1)p(\alpha_{0,0})p(\nu_{1,0})p(\nu_{2,0})p(\gamma^{(m)})p(\gamma^{(v)})p(\gamma^{(\alpha)})p(\gamma^{(lt)})p(\gamma^{(rt)})$$
Additionally we assume that $p(\delta)$ and $p(\lambda_1)$ is the standard (i.e. zero-location unit-precision) Student-t with 2 degrees of freedom, $p(\alpha_0,0)$ is the standard Student-t with 3 degrees of freedom, $p(\rho)$ is uniform over (-1,1), $p(\alpha_0)$ represents an Exp(200) distribution. Moreover, for parameters of all the Flexible Fourier Forms we assume that $p(\gamma_{ci})$ and $p(\gamma_{si})$ are products of independent univariate Cauchy densities with unit precision, $p(\phi_{F})$ is multivariate uniform over the set obtained by imposing identification restrictions that eliminate the related label-switching effect. Individual elements of $\phi(\cdot)$ are assumed to satisfy $L < \phi(\cdot) < \ldots < \phi_F(\cdot) < U$, for appropriately chosen constants $0 < L < U < \pi$. The ordering provides identification (eliminating the redundant likelihood modes). As to the choice of $L$ and $U$, we decide to pick values that rule out cycle lengths shorter than a year and longer two-thirds of the time interval covering the observed time series.

In the case of GJR(1,1) parameters we assume $p(\alpha_{1}^{-})p(\alpha_{1}^{+})p(\beta_{1})$ is uniform over $[0,1]^3$. As for $\nu_{1,0}$ and $\nu_{2,0}$ we take exponential priors implying that once one assumes $f_{lt}(t) = 0$ and $f_{rt}(t) = 0$, the implied prior for $\nu_{1,t}$ and $\nu_{2,t}$ is Exp(10) truncated at 2. Under $f_{\alpha}(t) = 0$ the implied prior for $\alpha_t$ is approximately uniform over the (0.1, 0.9) range, making the values closer to 0 or 1 less likely. We assume that a maximum number of $F(\cdot)$ is fixed a priori, so the above setup defines a finite number of fully parametric specifications. Alternatively, one might want to impose priors on the numbers of Fourier components therefore pursuing a semi-parametric path. However, we believe that in order to avoid overfitting the appropriate values of $F(\cdot)$ should not be large.

As the number of model parameters is not very large, Bayesian inference can be carried out by means of standard MCMC methods and a Metropolis-Hastings algorithm (either random-walk or an independent proposal version) can be applied. For that purpose it is practical to transform constrained parameters into an unconstrained space.

The restrictions ensuring stationarity of the GJR-GARCH equation are not imposed here a priori, since a potential violation a posteriori might indicate some mis-specification in the long-run part of the model. For the sake of numerical stability one might choose to replace $\nu_{min} = 2$ with $\nu_{min} = 2 + v$ for some positive yet close to zero constant $v$. Setting e.g. $\nu_{min} = 100$ and $F_{lt} = F_{rt} = 0$ and $\nu_{1,0} = \nu_{2,0} = 0$ makes the model practically equivalent to a conditionally (skew) Gaussian specification. As to skewness of the conditional distribution, setting $\alpha_t = 0.5$ (which can be obtained by setting $\alpha_0,0 = 0$ and $f_{\alpha}(t) = 0$) implies absence of the skewing effect, whereas extreme values of $\alpha_t$ that (being very close to zero or one) imply that the conditional distribution practically becomes half-t.

To summarize, the general model proposed here encompasses a large number of nested special cases. The property makes model comparison easier therefore providing a flexible tool for empirical specification search. Within the approach it is crucial to consider a reasonable pattern of long-term time variation for the parameters under consideration. One obvious choice would be the use of latent stochastic processes. However, such a solution is not easy to implement in practice, as we consider time variation in several
parameters, potentially over dozens of thousands of observations in a non-Gaussian model. Moreover, if there is a linkage between risk-attitude (expressed as varying tail thickness) and business cycle, similar memory modeling in time variation of shape parameters would be necessary (allowing for low-frequency changes in the financial part). Business fluctuations with a period of say four years would correspond to a thousand of daily observations for the financial returns. A standard, autoregressive latent processes would be most likely pushed very close to the unit root non-stationarity, which is not necessarily relevant for out-of-sample projections.

As for \( f_\alpha(t), f_{lt}(t) \) and \( f_{rt}(t) \) we assume the general Fourier form \([1]\), though we impose restrictions on the frequency parameters that rule out very long or very short cycles. Certain advantages of FFF can be seen when considering out-of-sample inference. The crucial feature of the form is that it is bounded and less likely to imply unreasonable out-of-sample behavior (which is typical for polynomials or splines). Some alternative approaches assume sample decomposition into a sequence of homogenous sub-periods. The results are though sensitive to the partitioning criteria that are often rather arbitrary. The approach taken here requires no such preliminary steps. Moreover, models with latent random walks have been criticized for implying unreasonable out-of-sample properties, though maintaining good in-sample fit. Our choice of the Flexible Fourier Form is not free from disadvantages, but we take it as a compromise between flexibility and numerical feasibility.

Our interpretation of the Fourier-type time variation in parameters is that it approximates possible behavior of a long-memory latent process. As we practically rule out abrupt or very frequent changes, our models imply that in the forecasting horizons used in financial applications the parameters are practically constant, fixed at the levels that characterize the last part of the sample. However, we must emphasize that considering a constant parameter model on a short data series is not equivalent from a predictive point of view, one reason being that inference on tail behavior can be very sensitive to the number of observations used (see Zhu and Galbraith (2010)).

One has to bear in mind that finding a reasonable dynamic formulation for tail features is not trivial, since changes in tail behavior are statistically identified by occurrence of rare events. Consequently, application of a typical latent process allowing for frequent, abrupt changes might lead to practical identification issues or imply unreasonable out-of-sample behavior.

Non-parametric methods clearly allow for more flexibility in modelling shape of conditional distribution, as compared to the parametric approach pursued here. However, within non-parametric (or mixture) setup it is not easy to define shape features that can be meaningfully attributed with some pattern of time evolution, which is of interest here. Moreover, as we have indicated above, neither model construction nor estimation procedures should rely on assumptions that require existence of higher-order moments of the conditional distribution.
3 Data, in-sample model estimation and comparison

An empirical analysis making use of the general model framework discussed above is presented in this section. The dataset of interest consists of $T = 15624$ observations of daily logarithmic returns on S&P500 index in the postwar era (5th of January, 1950 - 7th of February, 2012). Our goal is to verify empirical relevance of long-run (low-frequency) variability in parameters that control features such as tail thickness and asymmetry of the conditional distribution. Moreover, the analysis aims at investigating practical consequences of allowing for such time-inhomogeneity in terms of general predictive performance and - particularly - risk assessment.

Initially we compare selected special cases of the general model, taking into account goodness of fit and estimation results based on full sample. We focus our attention on tail behavior and asymmetry. One specification allowing for the low-frequency variation is then chosen for a more detailed analysis. The model is compared to a benchmark specification with no long-run dynamics within a recursive out-of-sample forecasting framework assuming full reestimation with each observation added. Two verification windows are considered, covering 2035 and 1016 data points respectively. The windows are chosen to represent periods that differ with respect to variability of shape parameters. The point and density forecasts (for horizons ranging from one to five trading days ahead) are compared in terms of overall accuracy and results of risk assessment.

The benchmark model considered here, denoted by $\mathcal{M}_{0,0,0,0,0}$, is a AR(1)-GJR(1,1)-In-Mean process with AST disturbances. Its conditional distribution has time-invariant shape but allows for two-piece skewing and shape asymmetry (with different degrees of freedom parameter in each tail).

The alternative models under consideration augment the benchmark formulation by introducing additional long-run variability in mean, variance, skewing mechanism, left and right tail thickness, represented by $f_m(t)$, $f_v(t)$, $f_\alpha(t)$, $f_{lt}(t)$ and $f_{rt}(t)$ respectively. The dynamic behavior is driven by Flexible Fourier Forms [4] that are assumed to reflect low-frequency changes. Consequently, the frequency parameters are restricted so that sub-cycles of length shorter than one year are ruled out. The models being compared differ by the number of Fourier components (denoted by $F_\lambda$) in $f_m(t)$, $f_v(t)$, $f_\alpha(t)$, $f_{lt}(t)$ and $f_{rt}(t)$ being of the form [4].

3.1 Model fit

Table 1 presents the results of model comparison based on estimation using full sample. We report results of Bayesian estimation (decimal logarithms of marginal data density values with implied Bayes factors) and ML-based BIC score values, comparing the alternatives under consideration to the benchmark $\mathcal{M}_{0,0,0,0,0}$.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Model & $f_m(t)$ & $f_v(t)$ & $f_\alpha(t)$ & $f_{lt}(t)$ & $f_{rt}(t)$ \\
\hline
$\mathcal{M}_{0,0,0,0,0}$ & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
$\mathcal{M}_{1,0,0,0,0}$ & 0.01 & 0.01 & 0.01 & 0.01 & 0.01 \\
$\mathcal{M}_{0,1,0,0,0}$ & 0.02 & 0.02 & 0.02 & 0.02 & 0.02 \\
$\mathcal{M}_{0,0,1,0,0}$ & 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \\
$\mathcal{M}_{0,0,0,1,0}$ & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 \\
\hline
\end{tabular}
\caption{Model comparison results.}
\end{table}

1The reported Bayes factors are based on marginal data density approximated on the basis of the harmonic mean estimator of Newton and Raftery (1994). However, the models are nested so we assume that the model ranking is reflected rather correctly, even though numerical properties of the estimator are not superior.
Table 1: Goodness of fit comparison (full sample): differences in marginal data density and BIC score values against the benchmark model

<table>
<thead>
<tr>
<th>F</th>
<th>$M_{F,0.0,0.0}$</th>
<th>$M_{0,0.0,0.0}$</th>
<th>$M_{0,0.0,F,0}$</th>
<th>$M_{0.0,0,F,0}$</th>
<th>$M_{0.0,0,F,F}$</th>
<th>$M_{0,F,0,F,F}$</th>
<th>$M_{0,F,0,F,F}^*$</th>
<th>$M_{0,F,0,F,F}^{**}$</th>
<th>$M_{0,F,0,F,F}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

number of parameters $k$ (for $M_{0,0,0,0} k = 11$)

| $\log_{10}$ Bayes factor vs. $M_{0,0,0,0}$ | $\log_{10} p(y|M) = -7200.89$ |
|-----------------------------------------|---------------------------------|
| 1                                       | 0.49                            |
| 2                                       | 1.14                            |
| 3                                       | 0.29                            |
| 4                                       | 0.42                            |

$\text{BIC differences vs. } M_{0,0,0,0,0}$ (for $M_{0,0,0,0,0} \text{ BIC } = 33243.70$)

<table>
<thead>
<tr>
<th>F</th>
<th>16.75</th>
<th>-8.18</th>
<th>15.79</th>
<th>-13.04</th>
<th>11.61</th>
<th>-22.49</th>
<th>-4.94</th>
<th>-36.23</th>
<th>-35.41</th>
<th>-29.27</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>36.80</td>
<td>-5.78</td>
<td>33.44</td>
<td>-5.18</td>
<td>22.28</td>
<td>-26.75</td>
<td>5.96</td>
<td>-50.03</td>
<td>-26.70</td>
<td>-17.34</td>
</tr>
<tr>
<td>3</td>
<td>56.47</td>
<td>6.37</td>
<td>56.57</td>
<td>-14.03</td>
<td>38.54</td>
<td>-16.48</td>
<td>23.69</td>
<td>-41.66</td>
<td>7.65</td>
<td>13.53</td>
</tr>
<tr>
<td>4</td>
<td>87.03</td>
<td>16.09</td>
<td>73.17</td>
<td>-2.65</td>
<td>56.23</td>
<td>-4.48</td>
<td>36.88</td>
<td>-9.30</td>
<td>39.03</td>
<td>47.72</td>
</tr>
</tbody>
</table>

We begin by considering specifications with long-term variation present in just one model feature. The results that are included in the first five columns of Table 1 indicate that the low-frequency changes are empirically relevant for variance and tails of the conditional distribution of daily returns, with much less evidence in favor of such an effect in skewness parameter and no support for the variability in location parameter. As thickness of the right tail is considered, differences between models are more moderate. The model fit clearly improves with $F_v > 0$ or $F_t > 0$ according to the two criteria under consideration, though BIC-based choice promotes parsimony (i.e. smaller $F$).

In order to explore the problem further, we consider more complicated models with long-term variation present in tails and variance of the conditional distribution. Moreover, for simplicity we assume that $F_{(t)}$, denoting the number of terms in the Fourier expansion, is the same for all the functions in a given model. This is reflected in the notation as e.g. $M_{0,F,0,F,F}$. Additionally we assume that $F < 5$. In order to promote parsimony we include cases where $f_{lt}(t)$ and $f_{rt}(t)$ are the same, denoted with ($M^*$), or just share the same frequency parameters (having different amplitude parameters), denoted by ($M^{**}$). The results of model comparison for specifications with long-run FFF-type dynamics in more than just one feature are reported in subsequent columns of Table 1. Based on the above results we do not check all of the numerous model specifications, but focus on time-varying variance and tail thickness.

In terms of the Bayes factor all the other specifications under consideration are outperformed by $M_{0.4,0.4,4}$, having 47 parameters, though BIC indicates $M_{0.4,0.4,4,4}$, which has 23 parameters. The latter model assumes that $f_{lt} \equiv f_{rt}$, i.e. the same Flexible Fourier Form influences time-variation in the left and the right degrees of freedom (though the pattern of time variation is not identical, since we do not impose $\nu_{1,0} \equiv \nu_{2,0}$). However, one might think that the first model is overparametrized whereas the second is too restrictive. We therefore consider $M_{0.3,0.3,3}$ as well, which seems to be a reasonable compromise between parsimony and flexibility in describing time heterogeneity of the process. This is confirmed by the fact...
that if prior model probabilities are not equal, but proportional to \(2^{-k}\) (where \(k\) is the number of model parameters), as in Osiewalski and Steel (1993), \(M_{0,3,0,3,3}^{**}\) gains the highest posterior probability. The specification assumes time variability of tail thickness (with the same frequency parameters but different amplitudes and phase shifts in \(f_{it}\) and \(f_{rt}\)) and a separate pattern for \(f_c\) controlling long-run time evolution in variance. Imposing \(f_{it} \equiv f_{rt}\) in \(M_{0,3,0,3,3}^{**}\) (which leads to \(M_{0,3,0,3,3}^{*}\)) results in marginal data density decrease by four orders of magnitude, the model though still outperforms \(M_{0,0,0,F,F}^{0}\) and \(M_{0,0,0,F,F}^{**}\) characterized by time-varying tail behavior and constant long-run variance \((f_c = 0)\). Eventually, a detailed comparison of posterior inference on parameters of interest is conducted for the three following specifications: a model with no time variation \((M_{0,0,0,0,0})\), the BIC-preferred specification \((M_{0,2,0,2,2})\) and the model that is preferred according to Bayesian model comparison when prior model probabilities promote parsimony \((M_{0,3,0,3,3}^{**})\). The three models are nested, have 11, 23 and 35 free parameters, respectively, and ML estimation results in maximized log-likelihood value of -16568.7, -16485.8 and -16456.7.

Bayesian estimates of basic parameters common to all the specifications are presented in Table 2. We report marginal posterior median with quantiles of order 0.025 and 0.975\(^2\). Marginal posterior distributions of \(\rho\) and \(\delta\) are quite similar across the models, thought estimates of \(\delta\) move closer to zero in more complicated specifications. Estimation results for parameters \(\alpha\) — and \(\alpha\) strongly support asymmetric reaction of short-run conditional variance to the news. In the model with constant parameters, the effect of bad news on volatility is five times greater compared to that of good news. The asymmetry effect is strengthened in case of models with time varying parameters, as the ratio of \(\alpha_{-}\) to \(\alpha_{+}\), calculated at posterior medians, increases from 5.12 in model \(M_{0,0,0,0,0}\) to 10.04 and 11.48 in case of models \(M_{0,2,0,2,2}\) and \(M_{0,3,0,3,3}\) respectively (which is associated with both decrease in \(\alpha_{+}\) and an increase in \(\alpha_{-}\)). Marginal posterior distribution of \(\lambda_{1}\) is located in the region of positive values, indicating positive GARCH-In-Mean effect, thought it is dispersed so that some of the probability mass is allocated to negative values.

Amado and Teräsvirta (2014) discuss impact of time variability of parameters on persistence of the short-run GARCH-type volatility and find that allowing for dynamics of smooth transition type reduces the persistence effect quite strongly. We report posterior analysis of the persistence effect in the GARCH part of the process, described by the functions \(0.5(\alpha_{-} + \alpha_{+}) + \beta_{1}\) and \(\frac{\alpha_{0}}{1-0.5(\alpha_{-} + \alpha_{+})} - \beta_{1}\) (with the latter defining unconditional variance of the error term in model \(M_{0,0,0,0,0}\)). As low-frequency time-variation is introduced, persistence decreases significantly. This clearly indicates the empirical importance of long-run variability of parameters in reducing the persistence effect, though the persistence still remains relatively high. In the case of \(M_{0,3,0,3,3}^{**}\) the posterior dispersion of the persistence parameter does not increase, leaving very little posterior odds in favor of instability of the GARCH equation.

\(^2\)In order to facilitate comparison with the baseline model under the labels of \(\alpha_{0,0}, \nu_{1}\) and \(\nu_{2}\) we report estimates of \(0.5 + \frac{1}{\pi} \arctan(\alpha_{0,0}), \nu_{min} + \nu_{1,0}\) and \(\nu_{min} + \nu_{2,0}\), respectively, assuming \(\nu_{min} = 2.001\).
### Table 2: Posterior inference about parameters of interest in case of $M_{0,0,0,0,0}$, $M_{0,2,0,2,2}$ and $M_{0,3,0,3,3}$

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{0.025}$</th>
<th>$\theta_{0.500}$</th>
<th>$\theta_{0.975}$</th>
<th>$\theta_{0.025}$</th>
<th>$\theta_{0.500}$</th>
<th>$\theta_{0.975}$</th>
<th>$\theta_{0.025}$</th>
<th>$\theta_{0.500}$</th>
<th>$\theta_{0.975}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>-0.019</td>
<td>0.017</td>
<td>0.051</td>
<td>-0.026</td>
<td>0.008</td>
<td>0.041</td>
<td>-0.031</td>
<td>0.002</td>
<td>0.037</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.145</td>
<td>0.161</td>
<td>0.177</td>
<td>0.143</td>
<td>0.159</td>
<td>0.174</td>
<td>0.142</td>
<td>0.158</td>
<td>0.173</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.004</td>
<td>0.005</td>
<td>0.007</td>
<td>0.007</td>
<td>0.010</td>
<td>0.012</td>
<td>0.009</td>
<td>0.012</td>
<td>0.015</td>
</tr>
<tr>
<td>$\alpha_1^+$</td>
<td>0.086</td>
<td>0.099</td>
<td>0.114</td>
<td>0.090</td>
<td>0.104</td>
<td>0.119</td>
<td>0.094</td>
<td>0.108</td>
<td>0.124</td>
</tr>
<tr>
<td>$\alpha_1^-$</td>
<td>0.012</td>
<td>0.019</td>
<td>0.027</td>
<td>0.004</td>
<td>0.012</td>
<td>0.019</td>
<td>0.001</td>
<td>0.009</td>
<td>0.017</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.925</td>
<td>0.935</td>
<td>0.943</td>
<td>0.919</td>
<td>0.930</td>
<td>0.940</td>
<td>0.915</td>
<td>0.926</td>
<td>0.936</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>-0.032</td>
<td>0.022</td>
<td>0.079</td>
<td>-0.020</td>
<td>0.031</td>
<td>0.084</td>
<td>-0.015</td>
<td>0.039</td>
<td>0.090</td>
</tr>
<tr>
<td>$\nu_1$</td>
<td>4.776</td>
<td>5.464</td>
<td>6.304</td>
<td>4.588</td>
<td>5.395</td>
<td>6.397</td>
<td>4.814</td>
<td>5.781</td>
<td>7.106</td>
</tr>
<tr>
<td>$\alpha_{0,0}$</td>
<td>0.484</td>
<td>0.501</td>
<td>0.517</td>
<td>0.480</td>
<td>0.496</td>
<td>0.512</td>
<td>0.482</td>
<td>0.500</td>
<td>0.517</td>
</tr>
<tr>
<td>$0.5(\alpha_1^+ + \alpha_1^- + \beta_1)$</td>
<td>0.990</td>
<td>0.994</td>
<td>0.997</td>
<td>0.983</td>
<td>0.988</td>
<td>0.993</td>
<td>0.979</td>
<td>0.985</td>
<td>0.990</td>
</tr>
<tr>
<td>$1 - 0.5(\alpha_1^- + \alpha_1^+) - \beta_1$</td>
<td>0.574</td>
<td>0.817</td>
<td>1.685</td>
<td>0.632</td>
<td>0.792</td>
<td>1.114</td>
<td>0.636</td>
<td>0.770</td>
<td>1.031</td>
</tr>
</tbody>
</table>

Long-run changes in conditional variance (with $F_\nu > 0$) that are present in $M_{0,2,0,2,2}$ and $M_{0,3,0,3,3}$ imply that the unconditional variance of the processes is also time-varying (as it displays long-run 'cyclical trend' induced by $f_\nu$), making $\varepsilon_t$ in $\{1\}$ covariance nonstationary. Figure 2 presents characteristics of posterior distributions of time-varying unconditional standard deviation of error terms for the three models analysed here. The dynamic behavior of the posterior quantiles is rather similar for models analysed here. The dynamic behavior of the posterior quantiles is rather similar for models $M_{0,2,0,2,2}$ and $M_{0,3,0,3,3}$ despite differences in the number of parameters. The long-term volatility component displays peaks at the beginning of the sample and in mid-80’s, showing minima in mid 60’s, early 90’s and prior to the global financial crisis, with volatility increasing after 2006.

### 3.2 Long-term variability in shape of conditional distribution

Posterior inference about $\nu_1$ and $\nu_2$ confirms empirical adequacy of tail asymmetry inherent in the AST distribution of Zhu and Galbraith (2010). In the case of model $M_{0,0,0,0,0}$ the 95% credible sets calculated for those parameters have empty intersection and the left tail is heavier - the posterior median for $\nu_1$ in is equal to 5.5, compared to 7.6 for $\nu_2$.

Figure 2 presents plots of the posterior characteristics of the time-varying degrees of freedom parameters $\nu_{1,t}$ and $\nu_{2,t}$ compared to time-invariant parameters $\nu_1$ and $\nu_2$ for $M_{0,0,0,0,0}$. The pattern of evolution of tail thickness of conditional distribution is somewhat different in case of models $M_{0,2,0,2,2}$ and $M_{0,3,0,3,3}$. However, both models indicate periods where confidence sets for $\nu_{1,t}$ lie below confidence sets for $\nu_1$ so the left tail is significantly heavier compared to the benchmark model. This happens in late 80’s, but also in early 50’s and early 90’s, with posterior probability of $\nu_{1,t} > 4$ practically equal to zero (indicating non-existence of finite fourth conditional moment). Moreover, swings in $\nu_{1,t}$ display higher frequency in $M_{0,3,0,3,3}$ which is even more evident for $\nu_{2,t}$. The above results clearly show inadequacy of various model reductions, conveying strong statistical evidence against replacing the AST distribution with skew-$t$, $t$ or Gaussian distribution.
A more detailed analysis of tail behavior is presented in Figure 3 depicting dynamics of ratios of posterior means of functions of model parameters that correspond to quantiles of the AST distribution (the conditional distribution of the standardised process $z_t$ in (1)) to their $N(0, 1)$ counterparts. The quantiles are not equivalent to the quantiles of the predictive distribution but convey some information on its implied shape. The dotted lines represent quantiles of the same order of $N(0, 1)$ distribution. Moreover, Figure 4 represents posterior means of ratios of the quantiles of order $\alpha$ and $1 - \alpha$, for various values of $\alpha$. It is clear that the shape asymmetry inherent in the AST distribution has different effect on quantiles of different order. Quantiles corresponding to $\alpha$ equal 0.001, 0.005 and 0.01 show large relative changes in time and clearly exceed the Gaussian counterparts in absolute value (especially for low values of $\nu_{1,t}$), which has consequences for risk analysis. On the other hand, quantiles of order 0.05 and 0.1 are closer to zero and are rather stable in time. Similar effects, though somewhat less pronounced, are visible for the right tail as well.

Figure 4 reveals that tail asymmetry (with the left tail being heavier) is not a permanent feature of the series. It seems that periods of approximate tail symmetry and clear tail asymmetry follow each other (with a hint of much weaker asymmetry in the opposite direction in some cases). Moreover, quantiles of order 0.1 and 0.9 are rather unaffected, displaying almost symmetric behavior throughout the sample.

Tail asymmetry appears with the beginning of the Great Moderation in mid-80’s and fades out closer to the end of the century. Therefore the Black Monday in October 1987 is different from the great financial crisis of 2008 (and the dot-com bubble burst in 2002) as the 1987 crash occurred during the period of strong tail asymmetry which is not true as to the two latter events.

Results regarding tail asymmetry that are established in the literature highlight the importance of allowing for differences in tail thickness. Otherwise the tail index is averaged across tails which results in underestimation of the left tail thickness and potential distortions in risk analysis. This is confirmed by estimation results coming from the benchmark model $M_{0,0,0,0,0}$. However, estimation results obtained within the general model proposed here shed some additional light on the issue.

Firstly, tail thickness seems not to be time-homogenous. If the effect is ignored, a model with time-invariant tail index would pick up the averaged value which might lead to underestimation or overestimation of risk for certain sub-periods. In the dataset analysed here such averaging would fail to capture the fact that the existence of the fourth conditional moment is rejected (and existence of the third conditional moment is problematic) for long sub-periods of the sample, despite that the conditional variance exists by assumption. This is important since some of non-Bayesian inference methods often require existence of higher order moments of the error term distribution.

Secondly, tail-asymmetry is not a time-invariant feature of the conditional distribution of daily asset returns. Instead, its strength evolves over time and there are periods of strong tail asymmetry, but approximate tail symmetry seems to appear as well. Again, this effect vanishes in a model with time-
invariant tail asymmetry, being averaged over time.

Thirdly, the asymmetry introduced by two-piece skewing of the conditional distribution seems to be empirically irrelevant (at least for the dataset at hand). The conclusion is still valid when time-inhomogeneity is allowed for. This is important, since absence of the skewness in time-homogenous models could be induced by the fact that temporary shifts in skewness could cancel out on average. The only possibility not covered here (so hypothetically relevant) is short-term variation in skewness.

Fourthly, allowing for long-run shifts in conditional variance reduces persistence of its short-run movements. Moreover, allowing for dynamic shape asymmetry seems to lead to a more pronounced asymmetry in the news impact curve in short-run dynamics of conditional variance. Also the In-Mean effect seems to be somewhat more evident.

Some of the results presented here might be affected by specific properties of the Flexible Fourier Form \( F \) resulting in identification of spurious patterns that are strongly affected by outliers. In order to verify the above results and to investigate predictive properties of the models an out-of-sample analysis is conducted in the section to follow.

4 Out-of-sample performance and risk assessment

Two specifications chosen for a more detailed investigation of out-of-sample performance are \( M^{**}_{0.3,0.3,3} \) and the benchmark model \( M_{0,0,0,0,0} \). As a preliminary step we investigate contribution of individual observations to the total difference in maximized log-likelihood value (at MLEs for the two models) based on full sample, which is depicted in Figure 5. The most important finding is that the difference cannot be explained by a small number of influential observations. Instead, \( M^{**}_{0.3,0.3,3} \) seems to gain support rather systematically throughout the sample.

Certain periods yield quite a limited contribution in favor of the more complicated model (e.g. from second half of 70’s till mid 80’s, and from 2003 onwards). Interestingly, \( M^{**}_{0.3,0.3,3} \) seems to be able to collect strong support in periods of stability (in mid-60’s) as well as in periods of very high volatility (second half of 80’s, but also early 70’s, the same is true about 90’s and 2001). The fact that the model gains support in periods with characteristics that diverge from the average behavior (being exceptionally calm or exceptionally volatile) seems to confirm its potential to exploit time-inhomogeneity in certain characteristics of the process.

In order to verify out-of-sample predictive performance of the models and potential usefulness for the purpose of risk assessment, a predictive experiment using two verification windows was conducted\(^3\). The limited number of models chosen for out-of-sample predictive analysis results from computational limitations. As we analyze over three thousands of forecasts the numerical burden involved is quite considerable. For the first model a full sequence of separate MCMC chains with MHRW algorithm was used. The benchmark models has fewer parameters, so it was possible to build a Metropolis-Hastings sampler with an independent proposal applicable for a large group of sub-samples, using the same computations in a parallel way. The same candidate draw was considered for all the sub-samples in a group, with only one evaluation of (sequentially computed) likelihood function being sufficient for sample-specific acceptance steps, so the computational cost was considerably smaller. However, the solution is more demanding as to memory utilization.
verification windows are shown in Figure 5. The recursive forecasting setup is based on expanding sample (all the samples start at $T_0 = 1$, Jan 5th, 1950). Therefore, for the first window (labeled Period A) the shortest sample ends at $T_1 = 7590$ (Apr 3rd, 1980) and the longest sample ends at $T_1 + N_1 = 9624$ (Apr 21st, 1988) covering 2035 data points and for the second one (labeled Period B) the shortest sample ends at $T_2 = 14596$ (Jan 9th, 2008) and the longest sample ends at $T = T_2 + N_2$ (Jan 19th, 2012) covering 1016 data points. The models were fully reestimated with every observation added. Since we are interested in quite subtle features of the conditional distribution, we generate density forecasts and the ex post forecasting performance is evaluated using scoring rules (see Gneiting and Raftery (2007) and references therein), like continuous ranked probability score (CRPS) and log-predictive score (LPS). Averaged CRPS can be considered a generalization of the mean absolute forecast error, whereas cumulated LPS values for one period ahead forecasts can be given a Bayesian interpretation in terms of contribution to the difference in the marginal data density.

4.1 Value-at-Risk and Expected shortfall

To illustrate practical consequences of differences in model assumptions for risk assessment, we present ratios of the estimated risk characteristics (taking Value-at-Risk, VaR and Expected Shortfall, ES) for 1%, 5% and 10% from the two models based on predictive distributions for $h = 1$, see Figure 6. The ratios are taken relative to $M_{0,3,0,3,3}^{**}$, so values greater than one indicate that the time-homogeneous benchmark model $M_{0,0,0,0,0}$ suggests higher risk. During the first part of Period A (covering about two-thirds of the period length) the ratio is almost the same for all the probability levels under consideration and displays fluctuations around 1. This suggests that the differences between the two models can be fully attributed to discrepancy in inference on conditional scale. However, during the remaining part of period A the ratios visibly diverge. The divergence is most evident at 1%, where the factor gets close to 0.5 at some point. This indicates that the two models differ considerably in terms of inference on shape parameters as well (in particular tail thickness), with $M_{0,3,0,3,3}^{**}$ pointing to higher tail risk. Importantly, the model with time-variation in parameters indicated increasing risk for almost a year before the Black Monday. The opposite is true about period B. The ratios diverge throughout whole window which indicates differences in inference on shape of the conditional distribution. The proportional fluctuations visible in all the ratios are coming from differences in scale. However, the ratios are greater than one for most of the time and again show the strongest discrepancy at 1%. However this time $M_{0,3,0,3,3}^{**}$ suggests lower tail risk. In general one can see that differences in estimates of 1% VaR or ES due to presence of low-frequency changes in variance and tail behavior can be as large as $+/- 50\%$.

Our notion of CRPS corresponds to CRPS*(.,.) in Gneiting and Raftery (2007) p. 367; as to LPS, we refer to the formula (19) on p. 365 therein (though calculated here using decimal logs).
Table 3: Results of the predictive experiment on the expanding sample: performance of density and point forecasts. CRPS (the lower the better) is averaged over forecasts, LPS (decimal logs, the higher the better) is cumulated. 'MAE' and 'RMSFE' refer to posterior median point forecasts, whereas 'std. dev.' corresponds to a standard deviation of the predictive distribution averaged over the forecasts.

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Period A</th>
<th></th>
<th>Period B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CRPS log-score RMSFE MAE std. dev.</td>
<td>CRPS LPS RMSFE MAE std. dev.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.491 -1108.918 0.972 0.667 0.884</td>
<td>0.637 -639.639 1.288 0.876 1.119</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.513 -1154.613 0.984 0.677 0.900</td>
<td>0.658 -653.658 1.299 0.887 1.133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.509 -1146.692 0.983 0.675 0.901</td>
<td>0.656 -654.037 1.297 0.886 1.133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.516 -1159.970 0.982 0.674 0.902</td>
<td>0.649 -650.623 1.297 0.886 1.135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.508 -1150.878 0.982 0.674 0.903</td>
<td>0.656 -654.311 1.297 0.886 1.134</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M_{0.0,0.0,0} (without low-frequency fluctuations)

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Period A</th>
<th></th>
<th>Period B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CRPS log-score RMSFE MAE std. dev.</td>
<td>CRPS LPS RMSFE MAE std. dev.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.490 -1101.326 0.973 0.668 0.907</td>
<td>0.635 -640.135 1.286 0.875 1.062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.499 -1121.583 0.986 0.677 0.926</td>
<td>0.644 -646.120 1.297 0.886 1.071</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.498 -1120.411 0.984 0.675 0.930</td>
<td>0.644 -648.392 1.295 0.884 1.068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.498 -1121.239 0.983 0.675 0.933</td>
<td>0.645 -648.788 1.295 0.885 1.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.498 -1121.801 0.983 0.674 0.937</td>
<td>0.646 -650.533 1.295 0.885 1.060</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M_{0.3,0.3,3} (low-frequency fluctuations in asymmetric tail thickness and variance)

4.2 Evaluation and comparison of density forecasts

Table 3 presents a summary of the out-of-sample predictive experiment for all the horizons under consideration. CRPS is averaged over observations, LPS (computed using decimal logs) is cumulated. Differences in cumulated LPS for $h = 1$ can be interpreted in terms of predictive Bayes Factors. We also report RMSFE and MAE for point forecast (predictive median) and averaged predictive standard deviation. The density-based criteria i.e. LPS and CRPS indicate that the model with time-varying parameters performs better (the only exception being LPS for one day ahead forecasts in period B, though the difference is not large). In general, it is clear that analysis of period A provides stronger evidence in favor of $M_{0.3,0.3,3}$. Moreover, forecasts for longer horizons (in both verification periods) deliver more support for $M_{0.3,0.3,3}$. The criteria used for point forecast evaluation unsurprisingly show very little difference between the models, as the specifications differ in aspects other than conditional location. The differences indicated by averaged CRPS are not very large either, though being somewhat more pronounced compared to those of MAE (the latter can be interpreted as a counterpart of CRPS for point forecasts). Differences in average predictive standard deviation obtained from $M_{0.3,0.3,3}$ (compared to the benchmark) visible in Table 3 can be attributed to low-frequency changes in unconditional standard deviation visible in Figure 1.

Evolution of cumulated differences in LPS throughout the verification windows for $h = 1$ and $h = 5$ is depicted in Figure 7. The pattern for $h = 1$ is similar to the one depicted in Figure 5 (though the latter uses natural logs) for both cases. Period A covers about eight years and $M_{0.3,0.3,3}$ gains much of the total support during the last two years. Interestingly, strong evidence in favor of the model with time-varying tail behavior has started to build up more than a year before the Black Monday, signalling increasing inadequacy of the benchmark specification. During that period negative returns that are large...
in absolute value cause visible shifts in the cumulated log-score supporting \( M_{0,3,0,3,3} \) indicating that the model describes tail behavior more adequately. For \( h = 5 \), the support in favor of \( M_{0,3,0,3,3} \) builds up more systematically throughout the period, though again the increase is greater during the two final years; large jumps in cumulated LPS account for relatively small part of the total difference.

As to the second verification window (period B) the differences between models are less pronounced. For \( h = 1 \) the cumulated log-difference moves up and down by about one point (corresponding to changes of one order of magnitude in terms of a predictive Bayes Factors), with some evidence against \( M_{0,3,0,3,3} \) showing up closer to the end of the period. Visible shifts in the cumulated difference in favor of the benchmark model seem to correspond to negative returns most of the times, but not necessarily the ones that are largest in absolute value. For \( h = 5 \) the last part of the sub-sample provides support for \( M_{0,3,0,3,3} \) that dominates the total difference and again large jumps are relatively less influential.

In order to verify calibration of the density forecasts we analyze histograms representing PITs coming from the two competing models (for \( h = 1 \) and \( h = 5 \)), see Figure 8. For the sake of risk assessment the most important feature is the left tail adequacy represented by the first bar in each histogram. The model \( M_{0,3,0,3,3} \) displays superior left tail performance compared to the benchmark during period A. The benchmark specification seriously underestimates risk for the range of \( 0 - 0.02 \), whereas \( M_{0,3,0,3,3} \) slightly overestimates it. As to the other tail, both models underestimate tail thickness, but again \( M_{0,3,0,3,3} \) performs better. The opposite is true about the performance in the second window: the benchmark specification visibly underestimates the left tail, but \( M_{0,3,0,3,3} \) is even less successful. We do not attempt to test whether the differences between the models are significant because not all parts of the distribution are equally important for the sake of risk assessment. Visual inspection of PITs provides important insights into performance of density forecasts coming from the competing specifications that is coherent with the remaining results.

To investigate predictive performance of the two competing models in more detail, observation-specific differences in LPS are plotted against the PIT values (representing order of the quantile of the density forecast - according to \( M_{0,3,0,3,3}^{**} \) - at the actual outturn). The results are presented in Figure 9. Positive numbers on the vertical axis indicate that the realized observation provides evidence supporting \( M_{0,3,0,3,3}^{**} \) against \( M_{0,0,0,0,0} \). In period B negative tail observations support the benchmark model only. The opposite effect (although a bit less clear-cut) is visible during period A, where left-tail observations generate support for \( M_{0,3,0,3,3}^{**} \) in most cases. Right tail behavior does not display such systematic differences. Moreover, in the first period for \( h = 1 \), \( M_{0,3,0,3,3}^{**} \) slightly outperforms the benchmark between quantiles of order 0.3 and 0.7. Although individual differences in LPS are not large, the effect is systematic for a considerable fraction of observations resulting in an influential contribution to the total difference. During the second verification period the inferior left-tail performance of \( M_{0,3,0,3,3}^{**} \) is to some extent neutralized by performance gains between quantiles of order 0.05 and 0.3. One might see that although
LPS is commonly considered to be very sensitive to tail observations, shape differences in other parts of
the target distribution might offset inferior tail performance. Relatively weak LPS-based evidence against
$M_{0,3,0,3,3}$ in period B does not fully reveal limitations of the model with respect to predicting negative
tail outcomes at the time.

In general the results of out-of-sample predictive experiments are consistent with full-sample analysis.
The model with long-term variation in shape parameters shows clearly superior predictive performance
in a period with pronounced shifts in tail thickness and tail asymmetry (period A). Moreover, the out-of-
sample results confirm that the appearance of strong tail asymmetry and increasing thickness of the left
tail visible since 1985 do not rely on a 'Black Monday' outlier only but reflect a more systematic process.
It seems that the beginning of the Great Moderation coincides with a shift in shape characteristics
of conditional distribution of daily S&P500 returns. The distribution evolves from an approximately
Gaussian one (evident in mid-70’s) towards a heavy-tailed case displaying strong tail asymmetry (clearly
visible in late 80’s and early 90’s). Afterwards the tail asymmetry again becomes less evident, so the
dot-com bubble and burst and in particular the great financial crisis appear during another period of
approximate tail symmetry with less-heavy tails. We also show that allowing for long-term variability of
the shape characteristics affects risk evaluation (so that e.g. differences in estimates of 1% VaR or ES
are as large as 50

5 Conclusions

We present a univariate parametric time-series framework that takes into account low-frequency variation
in shape-related features of conditional distribution. We therefore generalize upon existing approaches
that focus on including similar dynamics in conditional variance only (see e.g. Spline-GARCH of Engle
and Rangel, 2008). The resulting model allows for long-run (low-frequency) changes in location, scale,
skewness, tail thickness and tail asymmetry and is intended to describe long series of daily financial
returns. The framework used in the empirical part requires the first two moments of the conditional
distribution to be finite. However, the notion of skewness does not require existence of the finite
third conditional moment. We impose such assumptions since we emphasize connections with existing
specifications of long-term changes in conditional variance. However, the mechanism described here can
be easily transformed into a version that allows for e.g. left tail of the conditional distribution to be of
Cauchy type. This requires replacing the standardized AST distribution with its original version by Zhu
and Galbraith (2010) and re-writing of the short-term dynamics in terms of conditional scale instead of
volatility. Hence, in general neither the model construction nor the inference methods described here
rely on potentially restrictive assumptions as to existence of moments of conditional distribution, which
might be even more important for a time-varying setup.
The generalization builds upon an AR(p)-GJR-GARCH-In-Mean model. The conditional distribution used belongs to the AST class proposed by Zhu and Galbraith (2010) and allows for two-piece skewing as well as asymmetry in tail thickness. In order to capture long-term variation in the shape features we adopt the idea of Flexible Fourier Form of Gallant (1981). The form contains explicit frequency parameters so it can be adjusted to represent low-frequency changes. Moreover, the form is bounded, which precludes inadequate out-of-sample consequences (which is not necessarily true about e.g. spline-based models). Our approach provides a feasible and computationally convenient tool for tracing long-run changes in shape characteristics (including asymmetric tail thickness) in long series of asset returns. Its potential competitors would have to face two problems. Firstly, it is not obvious how to impose more meaningful (e.g. latent stochastic) dynamics upon parameters that represent tail thickness. Even in the static case the inference on tail parameters is very sensitive to the number of observations. Therefore, from theoretical point of view it is unclear whether abrupt or frequent changes in tail behavior should be allowed for. Secondly, potential alternatives based on e.g. non-Gaussian models with latent long-memory processes driving the shape features are likely to be much more computationally demanding especially when facing long series or recursive predictive experiments.

In principle the mechanism of time-variation in shape parameters considered here (i.e. the Flexible Fourier Form) might be augmented by introducing exogenous variables similarly to the formulation of Engle and Rangel (2008). The option seems feasible though we do not pursue it here since out-of-sample prediction in such a setup generates a number of additional issues that are outside the scope of the paper.

We analyze over 60 years of daily S&P500 returns and find strong evidence in favor of long-run changes not only in variance but also in tail behavior. However, there is no evidence supporting long-term fluctuations in skewness or location. We also find that allowing for long-term variation in conditional volatility and tail thickness affects inference on other model parameters, as the In-Mean effect becomes more pronounced. Moreover, the short-term volatility equation (of the GJR-GARCH form) shows reduced persistence and more evident asymmetry in the news impact curve.

Moreover, we find that in early 50’s and during most of the 1985-1995 period the left tail degrees of freedom parameter is almost certainly below 4, so the existence of finite conditional fourth moment is extremely unlikely. Tail asymmetry (with the left tail being heavier) is therefore clearly present in the S&P500 dataset. However, the periods of tail asymmetry seem to be separated by periods where approximate tail symmetry prevails. Around mid-70’s the conditional distribution is approximately Gaussian. However, the beginning of the Great Moderation period in mid-80’s coincides with appearance of strong tail asymmetry and increasing thickness of the left tail. The effect vanishes towards the end of the century and after 2005 the conditional distribution is again approximately symmetric with less-heavy tails. Consequently, according to our results the global financial crisis and the crash of 2008 took place during a period of tail symmetry with relatively light tails. This is in contrast to the Black Monday crash
in 1987, where we find strong tail asymmetry with the left tail being heavier.

The above conclusions (which are based on in-sample inference) are also supported by out-of-sample predictive experiments. We investigate density forecasting performance of alternative models using LPS and CRPS criteria. We find that the model with long-term variation in volatility and tail behavior displays obvious performance gains (in terms of LPS) during a period with shifts in (asymmetric) tail thickness (i.e. after mid 80’s). Moreover, the results are not fully determined by few influential outliers.

The predictive evidence in favor of the model with time-varying tail asymmetry starts to build up for about a year before the Black Monday crash. We also illustrate differences in risk assessment (compared to a benchmark model with no long-term variation). The estimates of VaR and ES differ systematically due to discrepancies in inference on volatility and shape parameters between the time-varying and the time-invariant specifications considered here.

Appendix A

Appendix B

For a bounded function $g(\cdot,\gamma)$ we have the following equivalences:

1. For each $n \in N$, $E(\varepsilon_n^p)$ exists and $E(\varepsilon_n^p) = g(t,\gamma)\bar{Z}E(\xi_n^p)$ if and only if $E(\xi_n^p)$ exists

2. For each $n \in N$, $E(\varepsilon_n^p|\Psi_{t-1})$ exists and $E(\varepsilon_n^p|\Psi_{t-1}) = g(t,\gamma)\bar{Z}h_t(\theta,\Psi_{t-1})\bar{Z}E(\xi_n^p)$ if and only if $E(\xi_n^p)$ exists.

For a GARCH(1,1) process with $h_t$ of the form:

$$ h_t = \alpha_0 + \alpha_1 \xi_{t-1}^2 + \beta_1 h_{t-1}, \quad (10) $$

for $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, the process $\{\varepsilon_t, t \in Z\}$ defined by (2), generated by the GARCH(1,1) process $\{\xi_t, t \in Z\}$ has the following properties:

1. $E(\varepsilon_t|\Psi_{t-1}) = 0$

2. $V(\varepsilon_t|\Psi_{t-1}) = g(t,\gamma)h_t$

3. $E(\varepsilon_t) = 0$

4. $V(\varepsilon_t) = E(\varepsilon_t^2) = g(t,\gamma)\frac{\alpha_0}{1-\alpha_1-\beta_1}$, if additionally $\alpha_1 + \beta_1 < 1$

It is clear that process $\{\varepsilon_t, t \in Z\}$ is nonstationary in the strict sense and also covariance nonstationary, as function $g(\cdot,\gamma)$ introduces variability of the unconditional variance. Consequently, variability over time of the conditional variance of $y_t$ is decomposed into GARCH(1,1) effect and a deterministic component
that changes dispersion of the conditional distribution according to the form of function $g$. In this case, if we rewrite the equation for conditional variance in the GARCH-type form we have:

$$E(\varepsilon_t^2|\Psi_{t-1}) = g(t, \gamma) h_t = \alpha_{0,t} + \alpha_{1,t} \xi_{t-1}^2 + \beta_{1,t} h_{t-1},$$

(11)

where $\alpha_{0,t} = g(t, \gamma) \alpha_0$, $\alpha_{1,t} = g(t, \gamma) \alpha_1$ and $\beta_{1,t} = g(t, \gamma) \beta_1$. Hence, the process $\{\varepsilon_t, t \in \mathbb{Z}\}$ can be also interpreted as a GARCH(1,1) model with time varying parameters.

Equation (11) involves very similar idea to the construct proposed by Baillie and Morana (2009). However in (11) we focus on a simpler GARCH-type process and do not generalise equation for $h_t$ to the fractionally integrated GARCH, considered in Baillie and Morana (2009). But we allow time variability of each parameter in equation for conditional variance of the process $\{\varepsilon_t, t \in \mathbb{Z}\}$.

Just like for the GARCH(1,1) process, one may consider properties of the process $\{\varepsilon_t, t \in \mathbb{Z}\}$, when different functional forms of $h_t$ are assumed. In the empirical part of the paper we consider Asymmetric-GARCH specification, proposed by Glosten, Jagannathan, Runkle (1993), with $h_t$ of the following form:

$$h_t = \alpha_0 + \alpha_1^+ \xi_{t-1}^2 I(\xi_{t-1} \geq 0) + \alpha_1^- \xi_{t-1}^2 I(\xi_{t-1} < 0) + \beta_1 h_{t-1},$$

(12)

with $\alpha_0 > 0, \alpha_1^+ \geq 0, \alpha_1^- \geq 0, \beta_1 \geq 0$. We will denote this specification by GJR(1,1). Analogously to the case with GARCH(1,1), the conditional second moment of $\varepsilon_t$ is given by the form:

$$E(\varepsilon_t^2|\Psi_{t-1}) = \alpha_{0,t} + \alpha_{1,t}^+ \xi_{t-1}^2 I(\xi_{t-1} \geq 0) + \alpha_{1,t}^- \xi_{t-1}^2 I(\xi_{t-1} < 0) + \beta_{1,t} h_{t-1}$$

(13)

where $\alpha_{0,t} = g(t, \gamma) \alpha_0$, $\alpha_{1,t}^+ = g(t, \gamma) \alpha_1^+$, $\alpha_{1,t}^- = g(t, \gamma) \alpha_1^-$ and $\beta_{1,t} = g(t, \gamma) \beta_1$. This leads us to the Asymmetric-GARCH model with time varying parameters.

Explicit formulae of conditional and unconditional moments in case of GARCH($p,q$), for $p > 1$ and $q > 1$ and also in case of GJR model can be found in Bollerslev (1986), Chiangli (1997) and Chiangli and Teräsvirta (1999).
References


Figure 1: In-sample results: long-run dynamics of variance - posterior inference on unconditional standard deviation of the error term $\varepsilon_t$ in (1). Solid lines denote posterior median with 50%, 75% and 95% credible sets based on quantiles of the marginal posterior distribution.
Figure 2: In-sample results: time-varying tail behavior. posterior inference on degrees of freedom in AST distribution of the error term $\varepsilon_t$ in (1). Solid lines denote posterior median with 50%, 75% and 95% credible sets (obtained from quantiles of the posterior distribution of $\nu_{1,t}$ and $\nu_{2,t}$).
Figure 3: In-sample results: time-varying shape of the conditional distribution. Solid lines represent posterior means of functions of parameters corresponding to ratios of quantiles of the $d(0,1)$ AST distribution to its Gaussian counterparts.

<table>
<thead>
<tr>
<th>Left tail</th>
<th>$M_{0.0,0.0}$</th>
<th>Right tail</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td><img src="image3.png" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Graph" /></td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Graph" /></td>
<td><img src="image8.png" alt="Graph" /></td>
<td><img src="image9.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
Figure 4: In-sample results: time-varying tail asymmetry. Posterior means for parameter functions representing ratios of tail quantiles of the conditional distribution (belonging to AST class 7) in $M_{0.3,0,3,3}^{**}$.

Figure 5: In-sample results: cumulated differences in log-likelihood (at MLEs for the whole sample) between $M_{0.3,0,3,3}^{**}$ and $M_{0,0,0,0,0}$ decomposed into contribution of individual observations.
Figure 6: Out-of-sample predictive performance: differences in risk analysis. Ratios of risk characteristics (VaR and ES) obtained from $M_{0,0,0,0,0}$ and $M_{0,3,0,3,3}^{**}$ at $h = 1$, with results from $M_{0,3,0,3,3}^{**}$ in denominator.

<table>
<thead>
<tr>
<th>Period A</th>
<th>Period B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expected Shortfall</strong></td>
<td><strong>Value at Risk</strong></td>
</tr>
<tr>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Period A Expected Shortfall</td>
<td>Period B Value at Risk</td>
</tr>
</tbody>
</table>

Figure 7: Out-of-sample predictive performance: cumulated difference in LPS between $M_{0.3,0.3,3}$ and $M_{0.0,0.0,0}$ (calculated with decimal logs) - positive numbers are in favor of $M_{0.3,0.3,3}$.
Figure 8: Out-of-sample predictive performance: Histograms of PITs of the realized forecasts for one- to five- trading days ahead.

<table>
<thead>
<tr>
<th>$M_{0.0,0.0,0.0}$</th>
<th>$M_{0.3,0.3,3.3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Period A, $h = 1$</strong></td>
<td><strong>Period A, $h = 5$</strong></td>
</tr>
<tr>
<td><img src="image" alt="Histogram" /></td>
<td><img src="image" alt="Histogram" /></td>
</tr>
<tr>
<td><strong>Period B, $h = 1$</strong></td>
<td><strong>Period B, $h = 5$</strong></td>
</tr>
<tr>
<td><img src="image" alt="Histogram" /></td>
<td><img src="image" alt="Histogram" /></td>
</tr>
</tbody>
</table>
Figure 9: Out-of-sample predictive performance: differences in LPS/CRPS corresponding to every realized observation (between $M_{0,3,0,3,3}^{**}$ and $M_{0,0,0,0,0}$) plotted vs PIT of the realized observation (according to $M_{0,3,0,3,3}^{**}$). Positive values characterize realizations supporting $M_{0,3,0,3,3}^{**}$ against $M_{0,0,0,0,0}$.