MASTER THESIS

Title: Semi-parametric modeling of conditional skewness of financial returns

Student: Anton Petukhov

Advisor:
PhD in Economics, Tenured Professor at NES, Stanislav Anatolyev
PhD in Finance, Assistant Professor at NES, Stanislav Khrapov

Grade:

Signature:

Moscow 2013
Semiparametric modeling of conditional skewness of financial returns

Anton Petukhov
New Economic School
June 2013

Abstract

In this paper we propose a flexible method to model conditional skewness as a function of past standardized returns. This method adapts partially nonparametric model introduced by Engle & Ng (1993, The Journal of Finance 48, 1749–1778) to investigate news impact curve of volatility. The proposed model is estimated and analyzed on series of daily returns on major stock indices. We find that past returns may impact skewness in a way that significantly differs from those considered in the literature dealing with parametric conditional skewness modeling. We also conduct a Monte Carlo study to examine how well our estimation procedure identifies parameters of the model. Furthermore, using simulated data we show that direct models of skewness dynamics proposed in the literature can not capture skewness dynamics found in our empirical study.
1 Introduction

It is an established fact that stock returns exhibit nonzero skewness (see, for example Peiro, 1999; Harvey & Siddique, 1999; Franses & Van Dijk, 2000), and in several strands of literature, such as portfolio allocation, asset pricing, option valuation and risk-management skewness became an important object of research.

In the seminal work Kraus & Litzenberger (1976) develop a theoretical model, in which agents make their investment decisions not only on the mean-variance considerations but also taking into account skewness of returns. Since then many researchers investigated and underlined the role of skewness in portfolio allocation and asset pricing. Kane (1982), Simaan (1993), Athayde & Flóres (2004) among others consider theoretical issues of optimal portfolio choice taking into account the third moment.

Many papers provide empirical justification of the importance of skewness in asset pricing and portfolio allocation. Harvey & Siddique (2000) find that conditional skewness explains cross-sectional variation of expected returns, and that investors require a risk premium for systematic skewness. Patton (2004) shows empirically that taking into account conditional skewness and asymmetric dependence in portfolio selection may increase utility of a CRRA investor. Guidolin & Timmermann (2008) study effect of time-varying skewness and kurtosis on international portfolio allocation. Ghysels et al. (2011) using robust measures of skewness show that conditional skewness varies over time, find strong relation between conditional skewness and macroeconomic variables and discuss implications of these observations for international portfolio allocation.

Several papers investigate application of option implied measures of skewness in portfolio allocation. For example, Rehman & Vilkov (2012) find that option implied skewness is positively related to expected returns. DeMiguel et al. (2010) show how exploiting information on time-varying skewness implied by option prices can improve portfolio performance. Jha & Kalimiralli (2010) find that utilizing conditional skewness forecasts obtained using Hansen (1994) model in combination with option implied skewness may
yield significant gains in portfolio performance. Duffie & Pan (1997) emphasize the role of skewness in risk-management. There are many papers that propose models for value at risk taking into account time-varying skewness. Several examples are represented by Wilhelmsson (2009), Bali et al. (2008), Grigoletto & Lisi (2009) among others.

So, previous literature lead us to a conclusion that it is important to understand behavior of skewness and be able to model and forecast it. In the seminal work Hansen (1994) proposes an autoregressive method of conditional density modeling (ARCD), within which dynamics of parameters other than conditional mean and variance can be modeled. ARCD framework provides a very convenient way to model dynamics of higher order moments, and there is vast of literature that apply it to conditional skewness modeling. In section 2 we make a brief review of this literature. However, in all of these papers authors propose parametric forms of conditional skewness dynamics, but parametric models are subject to the risk of misspecification. In this paper we propose a semiparametric approach to conditional skewness modeling. This approach adapts so called partially nonparametric technique, which was introduced by Engle & Ng (1993) in the context of volatility modeling, to conditional skewness modeling. This approach is rather flexible and allows to capture patterns of behavior of skewness not allowed by parametric models proposed in the literature. In our empirical study we apply this method to modeling conditional skewness of returns on major stock indices: S&P 500, FTSE 100 and NIKKEI 225. Specifically, we assume that standardized returns follow the Skewed Generalized Error distribution, and model dynamics of asymmetry parameter of this distribution using this partially nonparametric technique. For S&P 500 index we find a functional connection between skewness and past returns that can not be revealed using simpler parametric models considered in the literature. However, we do not find the same connection for FTSE 100 and NIKKEI 225. We make several robustness checks of the results obtained for S&P 500 to modifications of our main model. Particularly we make a robustness check to the specification of the shape of the density of standardized returns by estimating the model that utilizes the Gram–Charlier density. We also implement a
simulation study, which allows us to evaluate, how well maximum likelihood procedure identifies parameters of our semiparametric model. We find that when there is rather weak connection between skewness and past returns, the maximum likelihood procedure may be not able to estimate parameters precisely enough. Due to this reason we may have not found for FTSE 100 and NIKKEI 225 results similar to those obtained for S&P 500. In addition, we examine how proposed in the literature direct parametric models of time-varying skewness fit dynamics found for S&P 500 in our empirical study. To do this we simulate the data from our flexible semiparametric model with parameters close to those obtained in empirical study. We find that GARCH-type models proposed by Harvey & Siddique (1999) and León et al. (2005) poorly capture dynamics of skewness revealed on S&P 500.

The rest of the paper is organized as follows. In section 2 a review of literature dealing with time-varying skewness modeling is presented. In section 3 we describe a new methodology of conditional skewness modeling. Section 4 describes the data used in the empirical study. In section 5 results of the empirical study are presented. In section 6 we conduct the simulation study. Section 7 concludes.

2 Existing specifications of skewness dynamics

In this section we make a brief review of developed in the literature parametric approaches to model time-varying conditional skewness in ARCD framework. Let \( \{r_t\} \) be a series of returns. Hansen (1994) ARCD framework assumes the following representation

\[
\begin{align*}
  r_t &= \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t, \\
  \mu_t &= \mathbb{E}[r_t|I_{t-1}], \\
  \sigma_t &= \mathbb{E}[(r_t - \mu_t)^2|I_{t-1}], \\
  I_t &= \text{information available at time } t, \text{ usually } I_t \text{ includes } r_t \text{ and its lags of all orders}. \\
  \mu_t \text{ can be specified using some parametric model, frequently it is modeled as a linear combination of lags of returns. In most cases volatility, } \sigma_t \text{, is assumed to follow GARCH-type dynamics. Variable } z_t \text{ has zero mean, unit}
\end{align*}
\]
variance and distribution that allows nonzero skewness. Sometimes researchers assume
distributions for $\varepsilon_t$ that have nonzero mean, in this case $\mu_t$ is not conditional expectation
of $r_t$ but just some function of variables known at date $t - 1$.

There are two approaches to model dynamics of skewness in this framework. The
first we call a direct approach. In the direct approach dynamics of skewness is defined
by an explicit equation for skewness $s_t$. This approach was proposed by Harvey &
Siddique (1999). In that paper authors consider Noncentral-$t$ distribution of standardized
returns and using analogy with the GARCH model for volatility develop autoregressive
conditional skewness model. In this model conditional skewness $s_t$ follows

$$s_t = \kappa_0 + \kappa_1 \varepsilon_{t-1}^3 + \kappa_2 s_{t-1}. \quad (2)$$

This approach was followed by León et al. (2005). In that paper authors utilize
modifications of the Gram–Charlier density for standardized returns and extend the
model of Harvey & Siddique (1999) by adding autoregressive dynamics of conditional
kurtosis $k_t$

$$s_t = \kappa_0 + \kappa_1 z_{t-1}^3 + \kappa_2 s_{t-1}, \quad (3)$$

$$k_t = \delta_0 + \delta_1 z_{t-1}^4 + \delta_2 k_{t-1}. \quad (4)$$

However, the direct approach has two major complications. Firstly, there are few
distributions that have skewness and kurtosis among parameters; usually skewness and/or
kurtosis depend on parameters of the distribution of $z_t$ through complex nonlinear map.
Due to this reason it becomes difficult to implement maximum likelihood procedure, as at
each iteration for each observation this nonlinear map should be inverted to obtain values
of parameters of the distribution from the current value of skewness and/or kurtosis.
Secondly, dynamics specified by equations (2), (3) and (4) does not restrict possible
values of skewness and kurtosis. However, there is a theoretical bound, in which all
possible values of the pair skewness and kurtosis of any distribution lie. For details see,
for example, Jondeau & Rockinger (2003). This means that for some values of $s_t$ and $k_t$ it will not be even possible to solve for parameters of the distribution of $z_t$. Harvey & Siddique (1999) solve the first complication directly (i.e. by inverting non linear map while computing log likelihood for each observation) and do not mention whether they encounter with the second complication during estimation. León et al. (2005) address the first complication by trying to utilize the Gram–Charlier distribution for $z_t$, which has two parameters: $s_t$ and $k_t$, which are exactly skewness and kurtosis. To overcome the second problem they decide to “modify” the Gram–Charlier density, so that it is defined for any values of $s_t$ and $k_t$. However, this solution annihilates the whole idea of conditional skewness and kurtosis modeling, because in the modified density parameters $s$ and $k$ have little relation to skewness and kurtosis. Due to mentioned above complications the direct approach is rarely used in the literature.

Another and more popular approach to model conditional skewness is to use some distribution of returns that incorporates a parameter reflecting skewness. Typically, this parameter is restricted between certain levels. Using logistic or exponential transformation this parameter is replaced with another parameter, which may take any value. After that, dynamics of the unbounded parameter is modeled.

Jondeau & Rockinger (2003) is one of the first papers that utilize this approach to investigate dynamics of conditional skewness and kurtosis. In that paper authors use Hansen Generalized-t (HGT) distribution, which has two parameters: asymmetry, $\lambda$, and degrees of freedom, $\eta$. They consider different parametric specifications for dynamics of $\lambda_t$ and $\eta_t$ and some of them were frequently used in subsequent literature. Hashmi & Tay (2007) extend one of these specifications to explore spillover effects on several Asian stock markets. Bali & Theodossiou (2008) exploit one of the specifications proposed by Jondeau & Rockinger (2003) to model value at risk and show that such model accurately

---

1 The formula defining the Gram–Charlier density can yield negative values when improper values of $s$ and $k$ are used. Detailed description of the Gram–Charlier density can be found in section 3.3.

2 To our knowledge Harvey & Siddique (1999) and León et al. (2005) are the only examples of implementing the direct method to model dynamics of conditional skewness. Brooks et al. (2005) use the direct approach to model conditional kurtosis.
estimates conditional value at risk.

Feunou et al. (2011) utilize this approach to model parameters of the HGT, Skewed Generalized Error (SGE) and Skewed Binormal distributions and find that the SGE distribution shows the best performance. Brännäs & Nordman (2003a) and Brännäs & Nordman (2003b) model dynamics of conditional skewness through parameters of the Log Generalized Gamma (LGG) distribution and Pearson type IV (PIV) distribution. The LGG distribution has one parameter, which authors connect with skewness. The PIV distribution has two parameters, one of which is closely connected to skewness. However, authors find that time-varying skewness in the model with the PIV distribution does not yield significant enhancement in comparison with the model with constant asymmetry parameter. Yan (2005) and Grigoletto & Lisi (2009) also exploit the PIV distribution to model dynamics of conditional skewness. In both of these papers authors do find strong evidence of time-varying conditional skewness in returns on stock indices.

We want to mark out here two specifications considered in the literature that are close to our simplest specification of semiparametric dynamics of skewness. It will be convenient to compare our results with the results obtained by the authors of these specifications. The first of them is the model considered in Feunou et al. (2011), in which standardized returns have the SGE distribution, that has two parameters. One of them, which reflects thickness of tails, is kept constant. The second, the asymmetry parameter, has the following dynamics

\[
\lambda_t = -1 + \frac{2}{1 + \exp(-\tilde{\lambda}_t)},
\]

\[
\tilde{\lambda}_t = \theta + \kappa_{0,+}z_{t-1}^+ + \kappa_{0,-}z_{t-1}^- + \rho\tilde{\lambda}_{t-1},
\]

where \(z_{t-1}^- = \min(0, z_t), z_{t-1}^+ = \max(0, z_t)\). Proposed in this paper semiparametric model nests this model as a simplest case. In our study we estimate this model and compare obtained results with the results of more flexible models.

The second is the model considered in Jondeau & Rockinger (2003). In this model
standardized residuals have the HGT distribution. This distribution also has two parameters: asymmetry parameter, $\lambda_t$, and thickness of tails parameter, $\eta$, which is kept constant. In this model $\lambda_t$ follows

$$\lambda_t = -1 + \frac{2}{1 + \exp(-\tilde{\lambda}_t)},$$

$$\tilde{\lambda}_t = \theta + \kappa \varepsilon_{t-1} + \rho \lambda_{t-1}.$$  \hspace{1cm} (6)

3 Semiparametric model for conditional skewness

In this section we describe the method proposed in this paper to model conditional skewness semiparametrically. Denote by $\{r_t\}$ a series of returns. We follow Hansen (1994) ARCD framework and continue to assume representation (1) for $r_t$. In our case $I_t$ includes $r_t$ and its lags of all orders. Volatility, $\sigma_t$, is assumed to follow GARCH-type dynamics, which will be specified in the next section. As before, variable $z_t$ has zero mean, unit variance and distribution that allows nonzero excess kurtosis and nonzero skewness. In our main specification we use the SGE distribution for $z_t$. The SGE distribution has two parameters. Both of these parameters influence skewness and kurtosis. One of them has greater impact on “thickness” of tails of the distribution and is assumed to be constant. Another controls asymmetry and dynamics of this parameter is modeled in the main specification. To check robustness of results to specification of the density we also consider a model which exploits the Gram–Charlier distribution for $z_t$.

3.1 Specification of volatility

To specify dynamics of $\sigma^2_t$ we use a GJR-GARCH model proposed in Glosten et al. (1993)

$$\sigma^2_t = \omega + \alpha \varepsilon^2_{t-1} + \gamma (\bar{\varepsilon}_{t-1})^2 + \beta \sigma^2_{t-1},$$ \hspace{1cm} (7)
where $\varepsilon_i^- = \min(0, \varepsilon_i)$. Engle & Ng (1993) find this specification to be the best in fitting the news impact curve of volatility. This model captures a number of stylized facts associated with volatility of financial returns. Particularly, it allows for asymmetric reaction of volatility to negative and positive past returns. It seems important to capture this effect in volatility dynamics, as asymmetry in volatility may interact with skewness dynamics and influence estimates of parameters in the model for skewness.

As potential misspecification of volatility dynamics may be crucial for estimates of skewness dynamics, as a robustness check we also estimate a specification, which exploits AGARCH model of Engle (1990)

$$\sigma_t^2 = \omega + \alpha(\varepsilon_{t-1} - \gamma)^2 + \beta \sigma_{t-1}^2.$$  \hspace{1cm} (8)

### 3.2 Skewed Generalized Error distribution

As has been mentioned, standardized residuals, $z_t$, in our main specification are assumed to have the SGE distribution. Depending on the values of parameters this distribution may have both fat and thin tails and nonzero skewness. This distribution is widely used in financial modeling literature. Anatolyev & Shakin (2007) use it to model intertrade durations of stock. Bali & Theodossiou (2008) apply this distribution to model value at risk. Feunou et al. (2011) model conditional skewness using this distribution, compare this model with models based on Skewed Generalized-t of Hansen (1994), and Binormal distribution of Feunou et al. (2013) and find that the model based on the SGE distribution performs the best.

Following notations of Feunou et al. (2011) the SGE density can be written as

$$f(z, \lambda, \eta) = C \exp \left( - \frac{|z + m|^{\eta}}{[1 + \text{sign}(z + m)\lambda]^{\theta}} \right), \hspace{1cm} (9)$$

$$C = \frac{\eta}{2\vartheta} \Gamma \left( \frac{1}{\eta} \right)^{-1}, \quad \theta = \Gamma \left( \frac{1}{\eta} \right)^{1/2} \Gamma \left( \frac{3}{\eta} \right)^{-1/2} S(\lambda)^{-1}, \quad m = 2\lambda AS(\lambda)^{-1},$$
where $\Gamma(\cdot)$ – Gamma function; $\eta$ and $\lambda$ – parameters of distribution, which are subject to restrictions $\eta > 0$, $-1 < \lambda < 1$. Parameters $\eta$ and $\lambda$ control tail thickness and asymmetry of the distribution correspondingly. On Figure 1 graphs of density of the SGE distribution for different values of parameters are presented. When $\lambda = 0$ the distribution is symmetric, when $\lambda > 0$ it is skewed to the right and when $\lambda < 0$ it is skewed to the left. When $\lambda = 0$ and $\eta = 2$ it coincides with the standard normal distribution.

Kurtosis and skewness of this distribution can be expressed in terms of parameters $\lambda$ and $\eta$ as

\[
Skew = A_3 - 3m - m^3, \\
Kurt = A_4 - 4A_3m + 6m^2 + 3m^4, \\
A_3 = 4\lambda(1 + \lambda^2)\Gamma\left(\frac{4}{\eta}\right)\Gamma\left(\frac{1}{\eta}\right)^{-1} \theta^3, \\
A_4 = (1 + 10\lambda^2 + 5\lambda^4)\Gamma\left(\frac{5}{\eta}\right)\Gamma\left(\frac{1}{\eta}\right)^{-1} \theta^4.
\]

On Figure 2 dependence of skewness on parameter $\lambda$ for different values of $\eta$ is presented. For reasonable values of $\eta$ skewness is an increasing function of $\lambda$, that is why it is convenient to model dynamics of skewness through dynamics of parameter $\lambda$.

### 3.3 The Gram–Charlier distribution

In this section we briefly describe the Gram–Charlier distribution. Due to some of its appealing properties the Gram–Charlier distribution gained popularity in financial modeling literature. Jondeau & Rockinger (2001) use it to model distribution of exchange rate returns, León et al. (2005) exploit its modification to model dynamics of skewness and kurtosis, Balaev (2011) applies this type of distribution to model multivariate densities of stock returns.

The Gram–Charlier distribution with zero mean and unit variance has the following
probability density function

\[ f(z, s, k) = \phi(z) \left( 1 + \frac{s}{3!}(z^3 - 3z) + \frac{k - 3}{4!}(z^4 - 6z^2 + 3) \right), \quad (11) \]

where \( \phi(z) \) – pdf of the standard normal distribution, \( s \) and \( k \) – parameters of the distribution. One important property of the Gram–Charlier density is that parameters \( s \) and \( k \) represent correspondingly skewness and kurtosis of the distribution

\[
\int_{-\infty}^{\infty} z^3 f(z, s, k) dz = s, \quad \int_{-\infty}^{\infty} z^4 f(z, s, k) dz = k.
\]

However, it should be noted that expression (11) represents a proper probability density function only for the limited set of values of parameters \( s \) and \( k \), for other values of parameters function \( f(\cdot, s, k) \) in formula (11) is not positive definite. Jondeau & Rockinger (2001) find the region for \( s \) and \( k \) within which expression (11) is positive for any real \( z \). This region is presented on Figure 3. This fact should be taken into account while modeling dynamics of skewness using the Gram-Charlier distribution.

### 3.4 Semiparametric specification of skewness dynamics

In this section we introduce semiparametric specification of skewness dynamics. Below we describe this specification for the case when standardized returns have the SGE distribution. In fact, this specification can be easily adapted for a model with another distribution of standardized returns, which has a parameter that determines asymmetry of the distribution. In the end of the section we show how this specification is adapted in the case of the Gram–Charlier distribution of standardized returns.

There are two ways in the literature to model dynamics of parameters of the distribution of standardized returns. In the first parameters are modeled as functions of standardized returns \( z_t \) (see, for example León et al., 2005; Feunou et al., 2011). Another approach is to model parameters as functions of \( \varepsilon_t = \sigma_t z_t \) (see, for example Harvey
We model dynamics of skewness as a function of $z_t$ and its lags, because skewness itself is a characteristic of the distribution of standardized returns. The second approach seems to us not justified enough.

In section 3.2 it was mentioned that skewness of the SGE distribution is determined by parameters $\eta$ and $\lambda$. For moderate values of $\eta$ skewness monotonically depends on asymmetry parameter $\lambda$ (see Figure 2). That is why it is convenient to model dynamics of skewness through dynamics of $\lambda$. It is also possible to allow the parameter $\eta$ to evolve dynamically, and even use similar semiparametric specification for it as we develop further for $\lambda$, but in order to maintain parsimony of the model we keep it constant.

So, in our main specification parameter $\eta$ is assumed to be constant, but parameter $\lambda$ through logistic transformation is modeled as a function of past standardized returns

$$\eta_t = \eta,$$

$$\lambda_t = \frac{2}{1 + \exp(-\tilde{\lambda}_t)},$$

$$\tilde{\lambda}_t = f(z_{t-1}, z_{t-2}, \ldots).$$ (12)

Logistic transformation of $\lambda_t$ ensures this parameter to lie in the interval $(-1, 1)$. To model function $f(z_{t-1}, z_{t-2}, \ldots)$ we use partially nonparametric technique introduced by Engle & Ng (1993). They proposed this method to estimate impact of news on volatility. Using this approach dependence of $\tilde{\lambda}_t$ on arguments $z_{t-2}, z_{t-3}, \ldots$ is expressed through dependence on $\tilde{\lambda}_{t-1}$ only. Dependence of $\tilde{\lambda}_t$ on $z_{t-1}$ is represented by a piecewise linear function.

Next, we describe this functional form in a way similar to Engle & Ng (1993). Let $m_+, m_-$ be some nonnegative integers; $\{\tau_i\}_{i=m_-}^{m_+}$ be a set of real numbers satisfying $\tau_{-m_-} < \tau_{-m_-+1} < \cdots < \tau_{m_+}$. Define dynamics of $\tilde{\lambda}_t$ by

$$\tilde{\lambda}_t = \theta + \rho \tilde{\lambda}_{t-1} + \sum_{i=0}^{m_+} \kappa_{i,+} (z_{t-1} - \tau_i)^+ + \sum_{i=0}^{m_-} \kappa_{i,-} (z_{t-1} - \tau_i)^-, \quad (13)$$
where $\rho$, $\theta$, $\kappa_{i,+}$ ($i = 0, 1, \ldots, m_+$), $\kappa_{i,-}$ ($i = 0, 1, \ldots, m_-$) are parameters to be estimated; $(z_{t-1} - \tau_i)^+ = \max(0, z_{t-1} - \tau_i)$, $(z_{t-1} - \tau_i)^- = \min(0, z_{t-1} - \tau_i)$. Inclusion of autoregressive term $\rho \tilde{\lambda}_{t-1}$ into dynamics of $\tilde{\lambda}_t$ is justified by the evidence of clusterization of skewness of stock indices presented in Jondeau & Rockinger (2003). When $m_+ = m_- = 0$, $\tau_0 = 0$ this specification coincides with (5). For fixed $\tilde{\lambda}_{t-1}$ functional form (13) defines $\tilde{\lambda}_t$ as a continuous piecewise linear function of $z_{t-1}$. For $z_{t-1} \in (\tau_0, \tau_1]$ this function has a slope $\kappa_{0,+}$; for $z_{t-1} \in (\tau_1, \tau_2]$ this function has a slope $\kappa_{0,+} + \kappa_{1,+}$; for $z_{t-1} \in (\tau_i, \tau_{i+1}]$ ($i \geq 0$) this function has a slope $\sum_{j=0}^i \kappa_{j,+}$; for $z_{t-1} \in (\tau_{-(i+1)}, \tau_{-i}]$ ($i \geq 0$) this function has a slope $\sum_{j=0}^i \kappa_{j,-}$.

Parameters $m_+, m_-, \{\tau_i\}_{i=-m_-}^{m_+}$ are usually chosen by a researcher, although some automated algorithms can be used. Higher values of $m_+$ and $m_-$ provide higher flexibility of the model, but at the same time this may lead to less precise estimation of parameters. So, while choosing the number of knots a compromise between parsimony and flexibility of the model should be found. In the case of volatility modeling Engle & Ng (1993) propose two simple methods to choose $\{\tau_i\}_{i=-m_-}^{m_+}$. The first method is to assign values of $\tau_i$ based on the order statistics of the explaining variable. The second is to take $\tau_i = i \cdot \sigma$ for $i \in \{-m_-, -(m_- - 1), \ldots, m_+\}$, where $\sigma$ is the unconditional standard deviation of the explaining variable. The first method is somewhat complicated in the case of conditional skewness modeling, because standardized returns $z_t$ are unobserved and, so, order statistics of $z_t$ are unavailable. In our empirical study we use the second approach. As $z_t$ have standard deviation equal to one, then in this case we take $\tau_i = i$ for $i \in \{-m_-,-(m_- - 1), \ldots, m_+\}$.

As has been mentioned earlier, we also conduct a robustness check to the specification of density using the Gram–Charlier distribution, which has two parameters: skewness $s$ and kurtosis $k$. In this case dynamics of skewness is modeled in the following way. We keep kurtosis constant and dynamics of the parameter $s$ is modeled through logistic transformation

\[
s_t = s_{\min}(k) + \frac{s_{\max}(k) - s_{\min}(k)}{1 + \exp(-\tilde{\lambda}_t)},
\]  

(14)
where $s_{\min}(k)$ and $s_{\max}(k)$ are minimal and maximal possible values of parameter $s_t$ for given $k$, for which the Gram–Charlier density (11) is defined (see Figure 3); $\tilde{\lambda}_t$ as before follows dynamics defined by (13).

4 Data

In our empirical study we use the data on stock indices taken from www.finance.yahoo.com. Although time-varying skewness may be also observed in returns on individual stock, in our study we use stock indices for the following reason. Our estimation sample covers about two decades, during which different factors within a particular company may change. Such factors may influence stock returns behavior and may be not captured by our model. For example, during two decades significant changes in the company’s capital structure may happen. Aggregate indices smooth such changes and it is more likely to reveal skewness dynamics in their returns rather than in individual stock returns for such sample sizes. So, we consider daily logarithmic returns on S&P 500, FTSE 100, NIKKEI 225. These indices are frequently studied in the empirical literature and it will be convenient to compare our results with predecessors. Log returns are calculated as

$$r_t = 100 \log(P_t/P_{t-1}),$$

where $P_t$ is an index close price at date $t$ adjusted for dividends and splits. The sample covers observations of returns from 2 January 1986 till 31 December 2005. In Table 1 we present summary statistics of the data.

It can be seen that all series have negative sample skewness and very large positive excess kurtosis. These observations demonstrate two established stylized facts: financial returns on stock and stock indices tend to have negative unconditional skewness and positive unconditional excess kurtosis (see, for example Harvey & Siddique, 1999; Peiro, 1999; Franses & Van Dijk, 2000; Premaratne & Bera, 2000).

On Figure 4 the graphs of dynamics of log returns are presented. All three series
exhibit volatility clustering and have a set of points that lie very far from return’s typical values. These points have a major contribution into high sample kurtosis.

Our dataset includes observation made on 19 October 1987, the day known as “Black Monday.” On this day S&P 500 experienced the largest in its history drop during one trading day: it lost more than 20%. At the same time our dataset does not include the period of extremely high volatility observed in 2008. As results of the estimation may be sensitive to such extreme events and periods, we make a robustness check by estimating our main specification on two additional samples. The first includes only years from 1988 till 2005 and the second includes years from 1986 till 2010.

5 Empirical study

5.1 Main results

In this section we present results of the estimation of different specifications of the model presented in section 3. Closest to our specifications for conditional skewness, considered in the literature, are models by Feunou et al. (2011) and Jondeau & Rockinger (2003), specified in equations (5) and (6) correspondingly. In the first mentioned paper authors estimate the model on S&P 500 data on different subsamples in the period from 1980 till 2009. In the second paper authors estimate the model on several indices including S&P 500, FTSE 100 and NIKKEI 225, their dataset covers years from 1971 till 1999. In this section we will compare our results with results, obtained in these papers.

Before going to estimation of the described above models we make AR(1) filtering of log returns\(^3\). We begin our empirical study with the estimation of the model, which has rather simple dynamics of skewness. In this model volatility follows GJR-GARCH dynamics, standardized returns have the SGE distribution with constant \(\eta\) and time-varying \(\lambda_t\), where \(\lambda_t\) is modeled using the partially nonparametric technique defined by (12) and (13) with \(m_- = m_+ = 0\) and \(\tau_0 = 0\). This dynamics of \(\lambda_t\) coincides with the

\(^3\)We also tried to make filtering using dummies for days of the week and got similar results.
one specified by formula (5). This model was considered by Feunou et al. (2011), except that in that work authors used NGARCH specification for volatility.

We use maximum likelihood procedure to estimate parameters of the model. Results of the estimation are presented in column “Specification 1” in Tables 2, 3 and 4 for S&P 500, FTSE 100 and NIKKEI 225 indices correspondingly. Estimated standard errors are presented in parentheses, robust standard errors – in brackets. Estimates of coefficients $\rho$, $\kappa_{0,-}$, $\kappa_{0,+}$ are of the main interest. For S&P 500 we got individually significant$^4$ coefficients $\kappa_{0,-}$ and $\kappa_{0,+}$, if usual standard errors are used and significant at 10% coefficient $\kappa_{0,-}$, if robust standard errors are used. Both of these coefficients have positive sign and are close to each other. The results for S&P 500 are similar to those obtained by Feunou et al. (2011) and Jondeau & Rockinger (2003). In the first work authors got significant positive estimates of coefficients $\kappa_{0,-}$ and $\kappa_{0,+}$ of the same order. In the second authors got significant positive estimate of the coefficient $\kappa$, see equation (6).

For FTSE 100 we got significant at 10% and positive estimate of $\kappa_{0,+}$, but $\kappa_{0,-}$ is insignificant. For NIKKEI 225 we got significant at 5% and positive estimate of $\kappa_{0,-}$ and insignificant $\kappa_{0,+}$. Jondeau & Rockinger (2003) for both of these indices obtained positive and significant estimates of coefficient $\kappa$. Our findings do not contradict results of Jondeau & Rockinger (2003): in our study significant kappas have the same sign as estimates of $\kappa$ in their work.

It can be concluded that discussed above specification reveals some connection between skewness and returns in the previous period, but this connection seems to be rather weak. Such results may be due to indeed weak connection or due to misspecification of the model. Using more flexible specification for dynamics of $\tilde{\lambda}_t$ below we show that for S&P 500 these observations are likely to be due to misspecification.

In the next model, which we call the main model, we add two more knots to the piecewise linear specification of $\tilde{\lambda}_t$. Now we take $m_+ = m_- = 1$ and place knots in points

---

$^4$We can not test joint significance of coefficients $\kappa_{0,-}$ and $\kappa_{0,+}$, because coefficients $\rho$ and $\theta$ are unidentified under the null of $\kappa_{0,-} = \kappa_{0,+} = 0$. Results of tests for individual significance should be interpreted with a caution: while testing significance of $\kappa_{0,-}$ ($\kappa_{0,+}$) it is assumed that $\tilde{\lambda}_t$ has nontrivial dependence on $z_t^+$ ($z_t^-$) for the parameters $\rho$ and $\theta$ to be identified under the nulls.
\( \tau_{-1} = -1, \tau_0 = 0, \tau_1 = 1. \) Results of the estimation of this model are presented in column “Specification 2” of Tables 2, 3 and 4. Estimates of the coefficients \( \kappa_{0,+}, \kappa_{0,-}, \kappa_{1,+} \) and \( \kappa_{1,-} \) are of the greatest interest. Looking at the estimates for S&P 500 one can see that estimates of \( \kappa_{0,+} \) and \( \kappa_{0,-} \) remained positive and increased dramatically in comparison with the previous specification. The estimates of \( \kappa_{1,+} \) and \( \kappa_{1,-} \) are negative. All four coefficients are individually significant. These estimates imply that positive standardized returns increase skewness (meaning make it more positive). Furthermore, up to some level of standardized return \( z_t \), higher \( z_t \) leads to higher skewness at time \( t + 1 \). However, after some level of \( z_t \) higher standardized returns have smaller impact on skewness. Negative returns, on the contrary, make skewness in the next period more negative, and the pattern of impact is similar: for moderately small in absolute value returns impact of negative returns rises with the absolute value of \( z_t \), but after some level the impact becomes smaller. On Figure 5 dependence of conditional skewness \( s_t \) on past standardized return \( z_{t-1} \) with 95% confidence bounds inferred from the estimated model is plotted for the case when \( \tilde{\lambda}_{t-1} = 0. \) This graph can be called the “news impact curve” (NIC) of skewness, by analogy with the NIC of volatility introduced by Engle & Ng (1993). It should be also noted that an estimate of the parameter of asymmetry in reaction of volatility to news, \( \gamma \), is high in absolute value and very significant. This observation does not support the conclusion of Harvey & Siddique (1999), made for S&P 500 on the sample of daily returns from 1969 till 1997, that evidence of asymmetry in volatility vanishes when skewness dynamics is taken into account.

However, in contrast with S&P 500 the estimates for FTSE 100 and NIKKEI 225 do not reveal the pattern of behavior of skewness observed for S&P 500: all of the coefficients \( \kappa_{0,+}, \kappa_{0,-}, \kappa_{1,+} \) and \( \kappa_{1,-} \) are individually insignificant. This fact may have several explanations. Firstly, once again misspecification may be the reason of such results: we may have chosen not very good knot points for these two indices, or the problem may lie in the shape of the density. The second possible explanation may be connected with the impossibility of precise estimation of the parameters under interest.
Below we estimate two additional specifications to check whether different position of knots and number of knots can help reveal stronger connection between skewness and standardized returns.

In column “Specification 3” of Tables 2, 3 and 4 we present results of the estimation of the model with semiparametric dynamics of skewness with knots in points $-2, 0, 2$. In column “Specification 4” results of the model with 5 knots in points $-2, -1, 0, 1, 2$ are presented. In this column we report only usual standard errors computed using estimated score vectors due to high sensitivity of numerical hessian to its estimation procedure. Results of the estimation of both specifications reveal the same shape of the NIC of skewness for S&P 500. In Specification 3 for FTSE 100 we obtained significant coefficients of $\kappa_{0,+}$ and $\kappa_{2,+}$ if robust standard errors are used. However, this result is not robust to number of knots: estimates of these coefficients are not significant in specification with 5 knots. Neither of these specifications revealed a clear NIC of skewness of NIKKEI 225: in both specifications all kappas are insignificant.

Before this moment it has been nothing said about the estimates of coefficient $\rho$. In all four specifications for S&P 500 the estimate of the coefficient is close to 0.7 and is very significant. This is an evidence of moderate persistence in conditional skewness of S&P 500 returns. This result coincides with those obtained by Jondeau & Rockinger (2003) and Feunou et al. (2011). Jondeau & Rockinger (2003) also found that conditional skewness of S&P 500 exhibit relatively high persistence. One of specifications considered by Feunou et al. (2011) coincides with our simplest specification of semiparametric skewness (results of estimation of this specification are presented in column “Specification 1” in Table 2). For this specification Feunou et al. (2011) obtained the estimates of $\rho$ close to 0.6 for S&P 500 for three subsamples, that cover periods: 1980–2009, 1980–1989, 1990–1999.

For FTSE 100 the obtained estimates of $\rho$ in all considered above specifications are near 0.7–0.8, for NIKKEI 225 they are about 0.3–0.4. The estimates of the coefficient for FTSE 100 are highly significant\(^5\) in all specifications; the estimates for NIKKEI 225

\(^5\)In order to conduct tests on significance of parameter $\rho$ one should be sure that $\hat{\lambda}_t$ nontrivially depends on the first lag of standardized return. In other case, as was pointed by Jondeau & Rockinger
are insignificant if usual standard errors are used, and two estimates become significant if robust standard errors are used. Jondeau & Rockinger (2003) obtained similar results for both FTSE 100 and NIKKEI 225. In their study they used another distribution, but in their model they still had similar parameter of persistence. Its estimate for FTSE 100 was 0.81 and highly significant, for NIKKEI 225 it was 0.39 and insignificant.

To sum up, in this subsection we have found that for two of three considered stock indices, namely S&P 500 and FTSE 100, skewness reveals clear persistent behavior. On the example of S&P 500 we have shown that the NIC of skewness may be highly nonlinear, however, for FTSE 100 and NIKKEI 225 our model has not been able to reveal a clear form of the NIC. In section 6 we conduct Monte Carlo experiments, which can explain the results obtained for FTSE 100 and NIKKEI 225. Using simulations we show that for small in absolute value kappas maximum likelihood procedure may yield too noisy estimates of these parameters. In the remainder of this section we check sensitivity of the NIC obtained for S&P 500 to modifications of the model and extreme events in the estimation sample.

5.2 Additional specifications

In this section we discuss results of the estimation of additional specifications. These specifications serve as robustness checks of results for S&P 500 to different alterations of the main model.

In column “Specification 1” of Table 5 we present results of the robustness check to specification of volatility for S&P 500. In this specification we use semiparametric model for skewness with 3 knots in points $-1, 0, 1$ and replace GJR-GARCH model of volatility with AGARCH. It can be seen that obtained estimates of parameter $\eta$ and parameters of skewness dynamics are very close to those obtained earlier indicating robustness of the results to specification of volatility.

(2003), tests may yield spurious results. We do not have solid evidence of individual significance of kappas for FTSE 100 and NIKKEI 225 and have no proper test to test their joint significance, so here we make an assumption that this dependence is nontrivial.
Next we conduct robustness check of our results to the specification of probability distribution of standardized returns. We do this by estimating an additional model, which utilizes the Gram–Charlier density for standardized returns on S&P 500 data. The complete specification of the model is the following. Volatility follows GJR-GARCH model as in (7), standardized returns have the Gram–Charlier distribution with constant kurtosis $k$ and dynamic skewness $s_t$. Dynamics of skewness is defined by equations (14) and (13). In dynamics of $\tilde{\lambda}_t$ we take $m_- = m_+ = 1$ and $\tau_0 = 0$, $\tau_{1,-} = -1$, $\tau_{1,+} = 1$. Results of the estimation are presented in Table 5. Although only two coefficients among $\kappa_{0,+}$, $\kappa_{0,-}$, $\kappa_{1,+}$ and $\kappa_{1,-}$ are statistically significant, all of the estimates of these parameters have the same signs as in specification with the SGE density. The fact that in this case the pattern of skewness dynamics seems to be less sharp than in the case of the SGE density may be connected with a narrower range of values that skewness can take due to restrictions imposed on parameters $s$ and $k$ of the Gram–Charlier density (see Figure 3). The estimate of $\rho$ in this specification is 0.6 and statistically significant indicating rather high persistence of conditional skewness. Thus, utilizing a different shape of density and similar model for dynamics of skewness revealed the same pattern in the behavior of conditional skewness.

In column “Specification 1” of Table 6 we present results of the robustness check of the shape of the NIC to exclusion of the autoregressive term from dynamics of $\tilde{\lambda}_t$. Here we estimate the model with the SGE density, GJR-GARCH volatility and semiparametric skewness with three knots and fix coefficient $\rho$ to be zero. The estimates of kappas are similar to those presented in column “Specification 2” of Table 2. However the estimate of $\kappa_{0,-}$ becomes significant only at 10% level and the estimate of $\kappa_{1,-}$ becomes insignificant, but still the shape of the NIC remains the same.

When introducing our semiparametric model in section 3 we made an assumption that $\tilde{\lambda}_t$ depends on lags of standardized returns of the order higher than one only through the autoregressive term $\rho \tilde{\lambda}_{t-1}$. To check this assumption we estimate a modification of our semiparametric model, which includes an additional term – lag of the standardized return.
of the order two

\[ \tilde{\lambda}_t = \theta + \rho \tilde{\lambda}_{t-1} + \kappa_{0,+} z_{t-1}^+ + \kappa_{0,-} z_{t-1}^- + \kappa_{1,+} (z_{t-1} - 1)^+ + \kappa_{1,-} (z_{t-1} + 1)^- + \nu z_{t-2}. \]

Results of the estimation are presented in column “Specification 2” of Table 6. The estimate of \( \nu \) is very small in absolute value and insignificant. The estimates of other parameters are very close to those obtained for the model without term \( \nu z_{t-2} \) (see column “Specification 2” in Table 2). So inclusion of lag of the order two into dynamics of skewness does not yield significant changes in obtained earlier dynamics.

5.3 Sensitivity to extreme events and periods

In section 4 it has been mentioned that our dataset includes the date of extreme event occurred in 1987 and does not include the period of high volatility in 2008. Here we check sensitivity of the results obtained for S&P 500 to these events. To do this we estimate the model with GJR-GARCH volatility and dynamic \( \tilde{\lambda}_t \) with 3 knots in points \(-1, 0, 1\) on two additional samples: the first covers daily returns from 1988 till 2005 inclusively, the second includes observations from 1986 till 2010 inclusively. Results of the estimation are presented in Table 7. It can be seen that exclusion of 1986 and 1987 years from estimation sample does not change estimates significantly. As before all 4 coefficients \( \kappa_{0,+}, \kappa_{0,-}, \kappa_{1,+}, \kappa_{1,-} \) are individually significant and have close values to those obtained earlier. For the sample starting in 1986 and ending in 2010 we see that estimates of parameters \( \kappa_{0,+}, \kappa_{0,-}, \kappa_{1,+}, \kappa_{1,-} \) became smaller in absolute value. Coefficient \( \kappa_{1,-} \) became insignificant, coefficients \( \kappa_{0,+} \) and \( \kappa_{0,-} \) are significant only at 10% level. Still, as before these coefficients have the same signs, and skewness exhibits the same pattern of dependence on past returns as was described above.
6 Simulation study

6.1 Quality of maximum likelihood estimates

In this section we analyze how maximum likelihood procedure identifies values of parameters in our main model, in which the data generating process (DGP) for returns $r_t$ is defined as follows. Volatility follows GJR-GARCH process (7), standardized returns have the SGE distribution with constant $\eta$ and dynamic $\lambda_t$, dynamics of $\lambda_t$ is defined by (12) and (13) with $m_+ = m_- = 1$, $\tau_0 = 0$, $\tau_{1,-} = -1$, $\tau_{1,+} = 1$.

In simulation study we consider 4 sets of parameters of the DGP. In all four sets we take $\omega = 0.013$, $\alpha = 0.02$, $\gamma = 0.09$, $\beta = 0.92$, $\theta = -0.02$, $\rho = 0.7$, these values are close to estimates obtained for S&P 500 (see Table 2). In two sets of parameters we take $\kappa_{0,+} = 0.3$, $\kappa_{0,-} = 0.3$, $\kappa_{1,+} = -0.3$, $\kappa_{1,-} = -0.4$. These values are relatively large in absolute value and also close to estimates obtained for S&P 500. In two other sets we take $\kappa_{0,+} = 0.1$, $\kappa_{0,-} = 0.1$, $\kappa_{1,+} = -0.05$, $\kappa_{1,-} = -0.05$. These values are chosen to be rather small in absolute value in comparison with two other sets and to exhibit similar shape of dependence of $\tilde{\lambda}_{t+1}$ on $z_t$. So, we have two pairs of sets of parameters and in each pair parameters of GARCH process and skewness dynamics are the same, but sets within each pair have different values of parameter $\eta$. This parameter is equal to either 1.25 (this value is relatively small and close to S&P 500 estimate) or 1.65 (this value is relatively large and close to FTSE 100 estimate).

For each set of parameters we generate $n = 100$ samples of size $N = 5000$ and estimate parameters using the same maximum likelihood procedure used in our empirical study. Then we calculate average estimates and their standard deviations across 100 sets of estimates. Furthermore, for each simulated sample we compute the series for conditional skewness $\{s_t\}_{t=1}^N$ using formula (10) and true values of parameters $\eta$ and $\lambda_t$ used in simulations. Then, we compute the series of skewness $\{s_t^{\text{inferred}}\}_{t=1}^N$ inferred by maximum likelihood estimates of parameters and estimates of standardized returns.
$\hat{z}_t = \frac{r_t}{\hat{\sigma}_t}$. Using these two series we calculate a measure of skewness fit

$$R_{Skew}^2 = 1 - \frac{\sum_{t=1}^{N} \left( s_t - \bar{s}_{\text{inferred}} \right)^2}{\sum_{t=1}^{N} (s_t - \bar{s})^2}.$$ 

This measure shows how well the procedure of estimation can recover the series of conditional skewness. Averages of $R_{Skew}^2$ and its standard deviations along with averages and standard deviations of estimates of parameters are presented in Table 8. It can be seen from this table that maximum likelihood procedure identifies rather large in absolute value parameters $\kappa_{0,+}, \kappa_{0,-}, \kappa_{1,+}, \kappa_{1,-}$ pretty well: standard deviation of estimates is not very large, and average values of estimates are close to true parameters, although there is obvious bias towards zero. However, for sets of parameters with relatively small in absolute value kappas identification is very poor: although there is no much evidence of bias, but high standard deviation of estimates relative to values of parameters wipes out any hope to get precise enough estimates from a single sample of size 5000 (samples in the empirical study had approximately this size). This finding means that even when parameters $\hat{\lambda}_{t+1}$ depend nontrivially (but not strongly) on $z_t$ our estimation procedure may not identify this dependence.

From Table 8 it can be also noted that lower value of $\eta$ yields more precise estimates of kappas: for $\eta = 1.25$ both bias and standard deviation of estimates are lower. This observation seems intuitive, because lower values of $\eta$ correspond to distribution with fatter tails and wider range of possible skewness (see Figure 2).

Estimated values of $R_{Skew}^2$ show that, when kappas are rather large, skewness recovered using maximum likelihood estimates fits true skewness pretty well: average value of $R_{Skew}^2$ equals 0.92 when $\eta = 1.25$ and 0.89 when $\eta = 1.65$ with standard deviations 0.05 and 0.07 correspondingly. To illustrate how well inferred skewness may reproduce the true skewness on Figure 6 we present 200 points from the series of true skewness and inferred skewness for a simulated sample of returns of size 5000. Values of parameters used in this
To sum up, from this simulation study several conclusions can be made. When values of kappas are large in absolute value, our estimates may have a downward bias and underestimate reaction of skewness to past returns, although obtained estimates are precise enough to investigate dynamics of skewness and recover the series for conditional skewness from the series of returns. When kappas are too small, maximum likelihood estimates of kappas are too noisy to provide a solid notion even about the signs of kappas. This fact can explain, why we may have not found significant reaction of skewness to past returns for FTSE 100 and NIKKEI 225 in our empirical study.

### 6.2 Analysis of existing direct models of skewness

In section 2 several parametric models for conditional skewness proposed in the literature were mentioned. In two papers, namely, Harvey & Siddique (1999) and León et al. (2005), the authors proposed direct methods to model conditional skewness dynamics. In both of these papers GARCH-type models were used, see formulas (2) and (3) for specifications of skewness dynamics proposed in Harvey & Siddique (1999) and León et al. (2005) correspondingly. The difference between these two specifications of skewness dynamics lies in the innovation term: in the first specification third power of return impacts skewness in the next period, in the second – third power of the standardized return.

In this section we analyze how these specifications capture skewness dynamics, when the DGP coincides with the model we estimated in section 5. As our semiparametric approach to modeling dynamics of skewness is rather flexible, this analysis can shed the light on how proposed earlier GARCH-type parametric models can fit skewness dynamics. More precisely, we assume that returns follow a model with GJR-GARCH
volatility, standardized residuals have SGE distribution with constant η and dynamic λt, and dynamics of λ is defined by expressions (12) and (13) with 3 knots at points −1, 0 and 1. Then we generate a series of returns rt and standardized returns zt from this process of length 5000. Values of parameters of the process are chosen close to estimates obtained for S&P 500. Parameters for volatility are ω = 0.013, α = 0.02, γ = 0.09, β = 0.92, parameters for dynamics of ˜λt are θ = −0.02, κ0,+ = 0.3, κ0,− = 0.3, κ1,+ = −0.3, κ1,− = −0.4, η = 1.25. For this generated series we exactly know the corresponding series of volatility σt, series of λt and, consequently, series of skewness st. Then using least squares we find parameters κ0, κ1 and κ2 in two specifications (we call them models 1 and 2):

\[
s_{t+1} = \kappa_0 + \kappa_1 s_t + \kappa_2 z_t^3 + u_t,
\]

\[
s_{t+1} = \kappa_0 + \kappa_1 s_t + \kappa_2 r_t^3 + u_t,
\]

where ut – is an error term. \( R^2 \) in these regressions will be an indicator of goodness of fit of the corresponding model. In addition, using nonlinear least squares we also fit the following model (model 3)

\[
s_{t+1} = \kappa_0 + \kappa_1 s_t + \kappa_2 |z_t|^{\psi_1} \sigma_t^{\psi_2} \text{sign}(z_t) + u_t.
\]

When \( \psi_1 = 3 \) and \( \psi_2 = 0 \) this model coincides with model 1, when \( \psi_1 = \psi_2 = 3 \) it coincides with model 2. Estimates of \( \psi_1 \) and \( \psi_2 \) will show how far from “reality” may be an assumption, that innovation terms should have power 3. \( R^2 \) in this regression will allow us to evaluate, how regression fit can be enhanced if we allow innovation term to have power different from 3.

In Table 9 results of the estimation of these models for skewness are presented. \( R^2 \) for both models 1 and 2 is rather small, it is below 0.5. However, \( R^2 \) for model 1 is higher by 0.02, this is an expected result, because in our data generating process \( s_{t+1} \) is defined by \( z_t \), so \( r_t \) has less in common with \( s_{t+1} \) than \( z_t \). \( R^2 \) for model 3 differs from that of models
1 and 2 dramatically, it is 0.95. Estimate of $\psi_2$ is very close to 0 and this result is also expected due to the reason described above. The estimate of $\psi_1$ is 0.444. This value is very far from 3.

Although in estimation of the models 1–3 we used the series of simulated $s_t$, this experiment shows that models of dynamics of skewness proposed by Harvey & Siddique (1999) and León et al. (2005) may be heavily misspecified. This may be the reason due to which Harvey & Siddique (1999) obtained the estimate of parameter $\kappa_2$ for S&P 500 very close to 0 in their empirical study. Results of this experiment may be useful if one wants to build a parsimonious direct parametric model for conditional skewness: including additional parameter, power of the innovation, may enhance the model significantly.

## 7 Conclusion

In this paper we proposed a flexible semiparametric method to conditional skewness modeling. It is based on partially nonparametric model introduced by Engle & Ng (1993) for volatility modeling. We estimated the proposed model on three major stock indices: S&P 500, FTSE 100 and NIKKEI 225. For S&P 500 we found strong and rather complex reaction of skewness to past returns that can not be captured by parametric models proposed in the literature. We checked robustness of the obtained results to specification of volatility process, shape of the density of standardized returns and different changes in specification of skewness dynamics. We also found that exclusion of the extreme events in October 1987 does not change our estimates significantly. Inclusion into estimation sample the period of high volatility observed in 2008 results in estimates yielding somewhat more fuzzy connection between skewness and past returns, but the shape of dependence remains the same.

We also investigated, whether our estimation procedure yields precise enough estimates of parameters of the semiparametric model. We found that for some values of parameters, that usually reflect weak reaction of skewness to past returns, the maximum
likelihood estimates are too noisy. Due to this reason for FTSE 100 and NIKKEI 225 we may have not found dependence observed for S&P 500. Using simulated data we also showed that direct parametric models of conditional skewness proposed in earlier literature are not able to capture skewness dynamics similar to that found in our empirical study.

8 Acknowledgments

The author is deeply grateful to his research advisors Stanislav Anatolyev and Stanislav Khrapov for responsive guidance and useful comments. I thank all participants of the research seminar under their supervision: Rostislav Berezovsky, Elena Buchatskaya, Alexander Firstenko, Grigory Franguridi, Nikita Kobotaev, Artem Tsalkovich and Matvey Vinokurov for active participation in discussions.
References


Appendix

Table 1: Summary statistics for log returns series

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500</th>
<th>FTSE 100</th>
<th>NIKKEI 225</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0352</td>
<td>0.0273</td>
<td>0.0042</td>
</tr>
<tr>
<td>Median</td>
<td>0.0551</td>
<td>0.0606</td>
<td>0.0288</td>
</tr>
<tr>
<td>Min</td>
<td>−22.8997</td>
<td>−13.0286</td>
<td>−16.1375</td>
</tr>
<tr>
<td>Max</td>
<td>8.7089</td>
<td>7.5970</td>
<td>12.4278</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.0874</td>
<td>1.0414</td>
<td>1.4252</td>
</tr>
<tr>
<td>Skewness</td>
<td>−2.0567</td>
<td>−0.5679</td>
<td>−0.1035</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>46.4866</td>
<td>11.3767</td>
<td>10.2043</td>
</tr>
<tr>
<td>Sample size</td>
<td>5047</td>
<td>5054</td>
<td>4923</td>
</tr>
</tbody>
</table>

Notes: Here are presented summary statistics for daily log returns series. Log returns are defined as $r_t = \log(P_t/P_{t-1})$, $P_t$ – closing level at date $t$. The data covers 20 years: from 1986 till 2005 inclusively.
Table 2: Results of estimation for S&P 500

<table>
<thead>
<tr>
<th></th>
<th>Specification 1</th>
<th>Specification 2</th>
<th>Specification 3</th>
<th>Specification 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega )</td>
<td>0.0125 (0.0029)</td>
<td>0.0125 (0.0020)</td>
<td>0.0130 (0.0094)</td>
<td>0.0130 (0.0023)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.0171 (0.0072)</td>
<td>0.0181 (0.0048)</td>
<td>0.0190 (0.0167)</td>
<td>0.0196 (0.0092)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.0930 (0.0148)</td>
<td>0.0938 (0.0099)</td>
<td>0.0932 (0.0070)</td>
<td>0.0943 (0.0124)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.9245 (0.0090)</td>
<td>0.9238 (0.0049)</td>
<td>0.9221 (0.0300)</td>
<td>0.9216 (0.0071)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>-0.0395 (0.0152)</td>
<td>-0.0219 (0.0084)</td>
<td>-0.0299 (0.0139)</td>
<td>-0.0092 (0.0367)</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.6702 (0.0872)</td>
<td>0.7008 (0.0271)</td>
<td>0.7191 (0.0482)</td>
<td>0.7121 (0.0631)</td>
</tr>
<tr>
<td>( \kappa_{0,+} )</td>
<td>0.1415 (0.0207)</td>
<td>0.2574 (0.0104)</td>
<td>0.1773 (0.0375)</td>
<td>0.2217 (0.0831)</td>
</tr>
<tr>
<td>( \kappa_{0,-} )</td>
<td>0.1207 (0.0417)</td>
<td>0.2753 (0.0262)</td>
<td>0.1603 (0.0270)</td>
<td>0.3084 (0.0839)</td>
</tr>
<tr>
<td>( \kappa_{1,+} )</td>
<td>-0.3980 (0.0246)</td>
<td>0.0807</td>
<td>-0.2380 (0.1933)</td>
<td></td>
</tr>
<tr>
<td>( \kappa_{1,-} )</td>
<td>-0.3230 (0.0642)</td>
<td>0.1561</td>
<td>-0.4220 (0.1847)</td>
<td></td>
</tr>
<tr>
<td>( \kappa_{2,+} )</td>
<td></td>
<td>-0.9598 (0.2755)</td>
<td>0.2916</td>
<td>-0.5391 (0.4055)</td>
</tr>
<tr>
<td>( \kappa_{2,-} )</td>
<td>-0.2473 (0.0852)</td>
<td>0.0869</td>
<td>0.0951 (0.1828)</td>
<td></td>
</tr>
<tr>
<td>( \eta )</td>
<td>1.2812 (0.0334)</td>
<td>1.2544 (0.0328)</td>
<td>1.2663 (0.0341)</td>
<td>1.2559 (0.0242)</td>
</tr>
</tbody>
</table>

Notes: Here are presented results of the estimation for S&P 500. Standard errors in parentheses, robust standard errors in brackets. For specification 4 robust standard errors are not presented due to high sensitivity of numerical hessian to estimation procedure. Estimated models: \( \varepsilon_t = \sigma_t \zeta_t \), \( \zeta_t \sim SGE(\lambda_t, \eta) \). \( \varepsilon_t \) – AR(1) filtered log returns. Volatility follows GJR-GARCH model: \( \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma (\varepsilon_{t-1})^2 + \beta \sigma_{t-1}^2 \). \( \lambda_t \) follows: \( \lambda_t = -1 + 2/(1 + \exp(-\tilde{\lambda}_t)) \), \( \tilde{\lambda}_t = \theta + \rho \tilde{\lambda}_{t-1} + \kappa_{0,+} \tilde{\zeta}_{t-1} + \kappa_{0,-} \tilde{\zeta}_{t-1} + \kappa_{1,+} (\tilde{\zeta}_{t-1} - 1)^+ + \kappa_{1,-} (\tilde{\zeta}_{t-1} - 1)^- + \kappa_{2,+} (\tilde{\zeta}_{t-1} - 2)^+ + \kappa_{2,-} (\tilde{\zeta}_{t-1} + 2)^- \). In specification 1 coefficients \( \kappa_{1,+}, \kappa_{1,-}, \kappa_{2,+} \) and \( \kappa_{2,-} \) are fixed to be zero. In specification 2 coefficients \( \kappa_{2,+} \) and \( \kappa_{2,-} \) are fixed to be zero. In specification 3 coefficients \( \kappa_{1,+} \) and \( \kappa_{1,-} \) are fixed to be zero.
Table 3: Results of estimation for FTSE 100

<table>
<thead>
<tr>
<th></th>
<th>Specification 1</th>
<th>Specification 2</th>
<th>Specification 3</th>
<th>Specification 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega )</td>
<td>0.0150 (0.0031) [0.0036]</td>
<td>0.0149 (0.0032) [0.0034]</td>
<td>0.0148 (0.0043) [0.0036]</td>
<td>0.0152 (0.0028)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.0363 (0.0082) [0.0081]</td>
<td>0.0364 (0.0083) [0.0080]</td>
<td>0.0361 (0.0080) [0.0080]</td>
<td>0.0364 (0.0085)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.0680 (0.0113) [0.0130]</td>
<td>0.0680 (0.0116) [0.0130]</td>
<td>0.0677 (0.0121) [0.0130]</td>
<td>0.0682 (0.0104)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.9129 (0.0097) [0.0115]</td>
<td>0.9131 (0.0099) [0.0106]</td>
<td>0.9136 (0.0125) [0.0114]</td>
<td>0.9126 (0.0086)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>-0.0300 (0.0263) [0.0251]</td>
<td>-0.0311 (0.0333) [0.0454]</td>
<td>-0.0430 (0.0297) [0.0283]</td>
<td>-0.0162 (0.0466)</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.7776 (0.1135) [0.1043]</td>
<td>0.7657 (0.1038) [0.1085]</td>
<td>0.7770 (0.1066) [0.0968]</td>
<td>0.7162 (0.1583)</td>
</tr>
<tr>
<td>( \kappa_{0,+} )</td>
<td>0.0683 (0.0393) [0.0390]</td>
<td>0.0955 (0.0683) [0.0792]</td>
<td>0.0972 (0.0451) [0.0437]</td>
<td>0.0523 (0.0933)</td>
</tr>
<tr>
<td>( \kappa_{0,-} )</td>
<td>0.0429 (0.0420) [0.0446]</td>
<td>0.0684 (0.0668) [0.1022]</td>
<td>0.0333 (0.0477) [0.0502]</td>
<td>0.1265 (0.0960)</td>
</tr>
<tr>
<td>( \kappa_{1,+} )</td>
<td>-0.0757 (0.1259) [0.1249]</td>
<td></td>
<td>0.0788 (0.2132)</td>
<td></td>
</tr>
<tr>
<td>( \kappa_{1,-} )</td>
<td>-0.0588 (0.1018) [0.1334]</td>
<td></td>
<td>-0.2286 (0.2122)</td>
<td></td>
</tr>
<tr>
<td>( \kappa_{2,+} )</td>
<td>-0.2447 (0.1735) [0.1176]</td>
<td></td>
<td>-0.3046 (0.4042)</td>
<td></td>
</tr>
<tr>
<td>( \kappa_{2,-} )</td>
<td>0.0132 (0.1366) [0.1143]</td>
<td></td>
<td>0.2097 (0.2742)</td>
<td></td>
</tr>
<tr>
<td>( \eta )</td>
<td>1.6520 (0.0468) [0.0690]</td>
<td>1.6524 (0.0466) [0.0685]</td>
<td>1.6514 (0.0468) [0.0688]</td>
<td>1.6549 (0.0348)</td>
</tr>
</tbody>
</table>

Notes: Here are presented results of the estimation for FTSE 100. Standard errors in parentheses, robust standard errors in brackets. For specification 4 robust standard errors are not presented due to high sensitivity of numerical hessian to estimation procedure. Estimated models: \( \varepsilon_t = \sigma_t z_t \), \( z_t \sim SGE(\lambda_t, \eta) \). \( \varepsilon_t \) - AR(1) filtered log returns. Volatility follows GJR-GARCH model: \( \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma (\varepsilon_{t-1})^2 + \beta \sigma_{t-1}^2 \). \( \lambda_t \) follows: \( \lambda_t = -1 + 2/(1 + \exp(-\tilde{\lambda}_t)) \), \( \tilde{\lambda}_t = \theta + \rho \tilde{\lambda}_{t-1} + \kappa_{0,+} z_{t-1} + \kappa_{0,-} z_{t-1} + \kappa_{1,+} (z_{t-1} - 1)^+ + \kappa_{1,-} (z_{t-1} - 1)^- + \kappa_{2,+} (z_{t-1} - 2)^+ + \kappa_{2,-} (z_{t-1} - 2)^- \). In specification 1 coefficients \( \kappa_{1,+}, \kappa_{1,-}, \kappa_{2,+} \) and \( \kappa_{2,-} \) are fixed to be zero. In specification 2 coefficients \( \kappa_{2,+} \) and \( \kappa_{2,-} \) are fixed to be zero. In specification 3 coefficients \( \kappa_{1,+} \) and \( \kappa_{1,-} \) are fixed to be zero.
Table 4: Results of estimation for NIKKEI 225

<table>
<thead>
<tr>
<th></th>
<th>Specification 1</th>
<th>Specification 2</th>
<th>Specification 3</th>
<th>Specification 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\omega)</td>
<td>0.0309 (0.0058)</td>
<td>0.0315 (0.0059)</td>
<td>0.0312 (0.0059)</td>
<td>0.0309 (0.0055)</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.0420 (0.0087)</td>
<td>0.0422 (0.0087)</td>
<td>0.0424 (0.0087)</td>
<td>0.0424 (0.0086)</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>0.1381 (0.0178)</td>
<td>0.1399 (0.0171)</td>
<td>0.1382 (0.0178)</td>
<td>0.1350 (0.0151)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.8815 (0.0103)</td>
<td>0.8803 (0.0104)</td>
<td>0.8810 (0.0103)</td>
<td>0.8827 (0.0087)</td>
</tr>
<tr>
<td>(\theta)</td>
<td>0.0202 (0.0390)</td>
<td>0.0169 (0.0324)</td>
<td>0.0076 (0.0486)</td>
<td>-0.0240 (0.0596)</td>
</tr>
<tr>
<td>(\rho)</td>
<td>0.3608 (0.2477)</td>
<td>0.3129 (0.1814)</td>
<td>0.3654 (0.2698)</td>
<td>0.3606 (0.3126)</td>
</tr>
<tr>
<td>(\kappa_{0,+})</td>
<td>-0.0347 (0.0492)</td>
<td>0.0134 (0.0661)</td>
<td>-0.0124 (0.0650)</td>
<td>0.0038 (0.1155)</td>
</tr>
<tr>
<td>(\kappa_{0,-})</td>
<td>0.1131 (0.0425)</td>
<td>0.0338 (0.0410)</td>
<td>0.0989 (0.0614)</td>
<td>-0.0292 (0.1166)</td>
</tr>
<tr>
<td>(\kappa_{1,+})</td>
<td>-0.0604 (0.1306)</td>
<td>0.0010 (0.1227)</td>
<td>0.0329 (0.2718)</td>
<td>0.4071 (0.2717)</td>
</tr>
<tr>
<td>(\kappa_{1,-})</td>
<td>0.1216 (0.0625)</td>
<td>0.0909 (0.0969)</td>
<td>0.4071 (0.2717)</td>
<td>-0.1811 (0.3624)</td>
</tr>
<tr>
<td>(\kappa_{2,+})</td>
<td>-0.1119 (0.1936)</td>
<td>0.1648</td>
<td>-0.1811 (0.3624)</td>
<td>-0.2950 (0.3120)</td>
</tr>
<tr>
<td>(\kappa_{2,-})</td>
<td>0.0234 (0.1094)</td>
<td>0.0898</td>
<td>-0.2950 (0.3120)</td>
<td>-0.2950 (0.3120)</td>
</tr>
<tr>
<td>(\eta)</td>
<td>1.3788 (0.0352)</td>
<td>1.3824 (0.0354)</td>
<td>1.3783 (0.0355)</td>
<td>1.3784 (0.0266)</td>
</tr>
</tbody>
</table>

Notes: Here are presented results of the estimation for NIKKEI 225. Standard errors in parentheses, robust standard errors in brackets. For specification 4 robust standard errors are not presented due to high sensitivity of numerical hessian to estimation procedure. Estimated models: 
\[ \varepsilon_t = \sigma_t z_t, \quad z_t \sim SGE(\lambda_t, \eta). \] 
\[ \varepsilon_t \sim AR(1) \text{ filtered log returns.} \] 
Volatility follows GJR-GARCH model: 
\[ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma (\varepsilon_{t-1})^2 + \beta \sigma_{t-1}^2 \]. 
\( \lambda_t \) follows: 
\[ \lambda_t = -1 + 2/(1 + \exp(-\tilde{\lambda}_t)), \quad \tilde{\lambda}_t = \theta + \rho \tilde{\lambda}_{t-1} + \kappa_{0,+} \tilde{z}_{t-1} + \kappa_{0,-} \tilde{z}_{t-1} + \kappa_{1,+} (z_{t-1} - 1)^+ + \kappa_{1,-} (z_{t-1} - 1)^- + \kappa_{2,+} (z_{t-1} - 2)^+ + \kappa_{2,-} (z_{t-1} - 2)^- \]. 
In specification 1 coefficients \( \kappa_{1,+}, \kappa_{1,-}, \kappa_{2,+} \) and \( \kappa_{2,-} \) are fixed to be zero. In specification 2 coefficients \( \kappa_{2,+} \) and \( \kappa_{2,-} \) are fixed to be zero. In specification 3 coefficients \( \kappa_{1,+} \) and \( \kappa_{1,-} \) are fixed to be zero.
<table>
<thead>
<tr>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega$</td>
<td>0.0000 (0.0024) [0.0033]</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0666 (0.0048) [0.0137]</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.4825 (0.0309) [0.0212]</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.9199 (0.0059) [0.0170]</td>
</tr>
<tr>
<td>$\theta$</td>
<td>-0.0242 (0.0105) [0.0234]</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.7008 (0.0277) [0.0513]</td>
</tr>
<tr>
<td>$\kappa_{0,+}$</td>
<td>0.2647 (0.0157) [0.0354]</td>
</tr>
<tr>
<td>$\kappa_{0,-}$</td>
<td>0.2743 (0.0256) [0.0514]</td>
</tr>
<tr>
<td>$\kappa_{1,+}$</td>
<td>-0.4014 (0.0293) [0.0362]</td>
</tr>
<tr>
<td>$\kappa_{1,-}$</td>
<td>-0.3295 (0.0476) [0.0493]</td>
</tr>
<tr>
<td>$\eta$</td>
<td>1.2520 (0.0326) [0.0488]</td>
</tr>
<tr>
<td>$k$</td>
<td>3.8499 (0.0934) [0.1090]</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses, robust standard errors in brackets. Specification 1: $\epsilon_t = \sigma_t z_t$, $z_t \sim SGE(\lambda_t, \eta)$, $\epsilon_t$ – AR(1) filtered log returns; volatility follows AGARCH model: $\sigma_t^2 = \omega + \alpha(\epsilon_{t-1} - \gamma)^2 + \beta\sigma_{t-1}^2$; $\lambda_t$ follows: $\lambda_t = -1 + 2/(1 + \exp(-\tilde{\lambda}_t))$, $\tilde{\lambda}_t = \theta + \rho\tilde{\lambda}_{t-1} + \kappa_{0,+} z_{t-1} + \kappa_{0,-} z_{t-1} + \kappa_{1,+} (z_{t-1} - 1) + \kappa_{1,-} (z_{t-1} + 1)^-$.

Specification 2: $\epsilon_t = \sigma_t z_t$, $z_t \sim GC(s_t, k)$, $\epsilon_t$ – AR(1) filtered log returns; volatility follows GJR-GARCH model: $\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \gamma(\epsilon_{t-1} - \beta\sigma_{t-2}^2)$; skewness $s_t$ follows: $s_t = -s_{min}(k) + (s_{max}(k) - s_{min}(k))/(1 + \exp(-\tilde{\lambda}_t))$, $\tilde{\lambda}_t = \theta + \rho\tilde{\lambda}_{t-1} + \kappa_{0,+} z_{t-1} + \kappa_{0,-} z_{t-1} + \kappa_{1,+} (z_{t-1} - 1) + \kappa_{1,-} (z_{t-1} + 1)^-$. 38
Table 6: Results of estimation of additional models for S&P 500

<table>
<thead>
<tr>
<th>Specification 1</th>
<th>Specification 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega )</td>
<td>0.0124 (0.0029) [0.0041]</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.0186 (0.0074) [0.0061]</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.0922 (0.0152) [0.0227]</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.9236 (0.0092) [0.0140]</td>
</tr>
<tr>
<td>( \theta )</td>
<td>-0.0939 (0.0588) [0.0586]</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.6959 (0.0299) [0.1078]</td>
</tr>
<tr>
<td>( \kappa_{0,+} )</td>
<td>0.2389 (0.1007) [0.0925]</td>
</tr>
<tr>
<td>( \kappa_{0,-} )</td>
<td>0.2331 (0.1121) [0.1293]</td>
</tr>
<tr>
<td>( \kappa_{1,+} )</td>
<td>-0.4523 (0.1844) [0.1599]</td>
</tr>
<tr>
<td>( \kappa_{1,-} )</td>
<td>-0.1637 (0.1383) [0.1480]</td>
</tr>
<tr>
<td>( \nu )</td>
<td>0.0043 (0.0158) [0.0278]</td>
</tr>
<tr>
<td>( \eta )</td>
<td>1.2809 (0.0342) [0.0556]</td>
</tr>
</tbody>
</table>

Notes: Here are presented results of estimation of two additional specifications for S&P 500. Standard errors in parentheses, robust standard errors in brackets. In both specifications: \( \varepsilon_t = \sigma_t z_t \), \( z_t \) \( \sim \) \( SGE(\lambda_t, \eta) \), \( \varepsilon_t \) \( \sim \) \( AR(1) \) filtered log returns; volatility follows GJR-GARCH model: \( \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma (\varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2) \); \( \lambda_t \) follows: \( \lambda_t = -1 + 2/(1 + \exp(-\hat{\lambda}_t)) \), \( \hat{\lambda}_t = \theta + \rho \lambda_{t-1} + \kappa_{0,+} z_{t-1}^+ + \kappa_{0,-} z_{t-1}^- + \kappa_{1,+} (z_{t-1} - 1)^+ + \kappa_{1,-} (z_{t-1} + 1)^- + \nu z_{t-2} \). In specification 1 parameters \( \rho \) and \( \nu \) are fixed to be zero.

Table 7: Results of estimation of the main model on different samples of S&P 500

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \omega )</td>
<td>0.0070 (0.0021) [0.0025]</td>
<td>0.0132 (0.0024) [0.0018]</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>0.0070 (0.0062) [0.0060]</td>
<td>0.0127 (0.0075) [0.0113]</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.0779 (0.0140) [0.0069]</td>
<td>0.1073 (0.0197) [0.0150]</td>
</tr>
<tr>
<td>( \beta )</td>
<td>0.9467 (0.0086) [0.0096]</td>
<td>0.9224 (0.0079) [0.0079]</td>
</tr>
<tr>
<td>( \theta )</td>
<td>-0.0032 (0.0394) [0.0080]</td>
<td>-0.0352 (0.0069) [0.0014]</td>
</tr>
<tr>
<td>( \rho )</td>
<td>0.6990 (0.0687) [0.0895]</td>
<td>0.7360 (0.0237) [0.0865]</td>
</tr>
<tr>
<td>( \kappa_{0,+} )</td>
<td>0.2157 (0.0650) [0.0453]</td>
<td>0.1672 (0.0114) [0.0911]</td>
</tr>
<tr>
<td>( \kappa_{0,-} )</td>
<td>0.2975 (0.0989) [0.0911]</td>
<td>0.1704 (0.0125) [0.0919]</td>
</tr>
<tr>
<td>( \kappa_{1,+} )</td>
<td>-0.3420 (0.1151) [0.1426]</td>
<td>-0.3205 (0.0250) [0.1181]</td>
</tr>
<tr>
<td>( \kappa_{1,-} )</td>
<td>-0.3686 (0.1340) [0.1295]</td>
<td>-0.2174 (0.0183) [0.1469]</td>
</tr>
<tr>
<td>( \eta )</td>
<td>1.2855 (0.0374) [0.0621]</td>
<td>1.2608 (0.0296) [0.0418]</td>
</tr>
</tbody>
</table>

Notes: Here are presented results of the estimation of the model for S&P 500 on samples from 1988 till 2005 and from 1986 till 2010 inclusively. Standard errors in parentheses, robust standard errors in brackets. Estimated model: \( \varepsilon_t = \sigma_t z_t \), \( z_t \) \( \sim \) \( SGE(\lambda_t, \eta) \), \( \varepsilon_t \) \( \sim \) \( AR(1) \) filtered log returns. Volatility follows GJR-GARCH model: \( \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma (\varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2) \); \( \lambda_t \) follows: \( \lambda_t = -1 + 2/(1 + \exp(-\hat{\lambda}_t)) \), \( \hat{\lambda}_t = \theta + \rho \lambda_{t-1} + \kappa_{0,+} z_{t-1}^+ + \kappa_{0,-} z_{t-1}^- + \kappa_{1,+} (z_{t-1} - 1)^+ + \kappa_{1,-} (z_{t-1} + 1)^- \).
Table 9: Results of estimation of different models for skewness dynamics on simulated data

<table>
<thead>
<tr>
<th>Model</th>
<th>$\kappa_0$</th>
<th>$\kappa_1$</th>
<th>$\kappa_2$</th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>-0.019</td>
<td>0.632</td>
<td>0.009</td>
<td></td>
<td></td>
<td>0.453</td>
</tr>
<tr>
<td>Model 2</td>
<td>-0.019</td>
<td>0.637</td>
<td>0.005</td>
<td></td>
<td></td>
<td>0.430</td>
</tr>
<tr>
<td>Model 3</td>
<td>-0.018</td>
<td>0.691</td>
<td>0.312</td>
<td>0.444</td>
<td>-0.017</td>
<td>0.951</td>
</tr>
</tbody>
</table>

Notes: Here are presented results of the estimation of 3 models of skewness dynamics on simulated data. Model 1: $s_{t+1} = \kappa_0 + \kappa_1 s_t + \kappa_2 s_t^3 + u_t$, Model 2: $s_{t+1} = \kappa_0 + \kappa_1 s_t + \kappa_2 s_t^3 + u_t$, Model 3: $s_{t+1} = \kappa_0 + \kappa_1 s_t + \kappa_2 |z_t|^{\psi_1} s_t^{\psi_2} \text{sign}(z_t) + u_t$. 

Table 8: Results of the simulation experiments

<table>
<thead>
<tr>
<th></th>
<th>$\theta$</th>
<th>$\rho$</th>
<th>$\kappa_{0,+}$</th>
<th>$\kappa_{0,-}$</th>
<th>$\kappa_{1,+}$</th>
<th>$\kappa_{1,-}$</th>
<th>$\eta$</th>
<th>$R^2_{\text{Skew}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGP parameters</td>
<td>-0.020</td>
<td>0.700</td>
<td>0.300</td>
<td>0.300</td>
<td>-0.300</td>
<td>-0.400</td>
<td></td>
<td>1.250</td>
</tr>
<tr>
<td>ML estimates</td>
<td>-0.016</td>
<td>0.699</td>
<td>0.260</td>
<td>0.262</td>
<td>-0.267</td>
<td>-0.348</td>
<td></td>
<td>1.236</td>
</tr>
<tr>
<td>STD of estimates</td>
<td>0.031</td>
<td>0.069</td>
<td>0.070</td>
<td>0.070</td>
<td>0.130</td>
<td>0.128</td>
<td>0.038</td>
<td>0.045</td>
</tr>
<tr>
<td>DGP parameters</td>
<td>-0.020</td>
<td>0.700</td>
<td>0.300</td>
<td>0.300</td>
<td>-0.300</td>
<td>-0.400</td>
<td></td>
<td>1.650</td>
</tr>
<tr>
<td>ML estimates</td>
<td>-0.017</td>
<td>0.701</td>
<td>0.231</td>
<td>0.220</td>
<td>-0.237</td>
<td>-0.294</td>
<td></td>
<td>1.662</td>
</tr>
<tr>
<td>STD of estimates</td>
<td>0.039</td>
<td>0.076</td>
<td>0.085</td>
<td>0.076</td>
<td>0.147</td>
<td>0.136</td>
<td>0.055</td>
<td>0.066</td>
</tr>
<tr>
<td>DGP parameters</td>
<td>-0.020</td>
<td>0.700</td>
<td>0.100</td>
<td>0.100</td>
<td>-0.050</td>
<td>-0.050</td>
<td></td>
<td>1.250</td>
</tr>
<tr>
<td>ML estimates</td>
<td>-0.024</td>
<td>0.609</td>
<td>0.102</td>
<td>0.098</td>
<td>-0.054</td>
<td>-0.061</td>
<td></td>
<td>1.242</td>
</tr>
<tr>
<td>STD of estimates</td>
<td>0.039</td>
<td>0.207</td>
<td>0.082</td>
<td>0.084</td>
<td>0.138</td>
<td>0.150</td>
<td>0.035</td>
<td>0.202</td>
</tr>
<tr>
<td>DGP parameters</td>
<td>-0.020</td>
<td>0.700</td>
<td>0.100</td>
<td>0.100</td>
<td>-0.050</td>
<td>-0.050</td>
<td></td>
<td>1.650</td>
</tr>
<tr>
<td>ML estimates</td>
<td>-0.019</td>
<td>0.649</td>
<td>0.084</td>
<td>0.081</td>
<td>-0.046</td>
<td>-0.051</td>
<td></td>
<td>1.654</td>
</tr>
<tr>
<td>STD of estimates</td>
<td>0.037</td>
<td>0.211</td>
<td>0.063</td>
<td>0.077</td>
<td>0.114</td>
<td>0.143</td>
<td>0.050</td>
<td>0.243</td>
</tr>
</tbody>
</table>

Notes: Here are presented results of the simulation experiments. For each set of parameters $n = 100$ samples of length $N = 5000$ were drawn from the following model: $r_t = \sigma_t z_t$, $z_t \sim SGE(\lambda_t, \eta)$. Volatility follows GJR-GARCH model: $\sigma_t^2 = \omega + \alpha z_{t-1}^2 + \gamma (\varepsilon_{t-1})^2 + \beta \sigma_{t-1}^2$. $\gamma = 0.09$, $\beta = 0.92$. “DGP parameters” – parameters used in simulations, “ML estimates” – mean of maximum likelihood estimates across 100 experiments, “STD of estimates” – standard deviation of maximum likelihood estimates across 100 experiments. Results for parameters of GARCH are not reported to save space. $R^2_{\text{Skew}}$ – measure of skewness fit computed as $R^2_{\text{Skew}} = 1 - \sum_{i=1}^{N}(s_t - s_t^{\text{inferred}})^2 / \sum_{i=1}^{N}s_t^2$, where $s_t$ – skewness of $z_t$, $s_t^{\text{inferred}}$ – skewness of $z_t$ inferred from generated sample $\{r_t\}$ and maximum likelihood estimates of parameters of the DGP.
Figure 1: The SGE density for different values of $\lambda$ and $\eta$
Figure 2: Skewness of the SGE distribution for different values of $\eta$ as a function of $\lambda$
Figure 3: Skewness and kurtosis boundary for the Gram–Charlier distribution

Notes: On this figure the boundary for skewness and kurtosis for the Gram–Charlier density is presented. For values of $s$ and $k$ inside the bounded region the formula $f(z,s,k) = \phi(z) \left(1 + \frac{s}{3!}(z^3 - 3z) + \frac{k}{4!}(z^4 - 6z^2 + 3)\right)$ defines a proper density, where $\phi(z)$ – pdf of the standard normal distribution.
Figure 4: Dynamics of log returns

Notes: Here is presented dynamics of daily log returns of three stock indexes, log returns are defined as $r_t = 100 \log \left( \frac{P_t}{P_{t-1}} \right)$, where $P_t$ an index close price at date $t$ adjusted for dividends and splits.
Figure 5: Dependence of skewness on past standardized returns for S&P 500

Notes: Solid line – inferred from the model conditional skewness $s_t$ of S&P 500 daily log returns depending on standardized residual $z_{t-1}$ conditional on $\tilde{\lambda}_{t-1} = 0$. Dashed lines – 95% confidence bounds.
Notes: On this figure the last 200 points of series of true and inferred skewness are presented. The sample of returns of length $N = 5000$ was drawn from the following model.

$$r_t = \sigma_t z_t, \quad z_t \sim SGE(\lambda_t, \eta).$$

Volatility follows GJR-GARCH model:

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \gamma (\varepsilon_{t-1})^2 + \beta \sigma_{t-1}^2.$$ 

$\lambda_t$ follows:

$$\lambda_t = -1 + \frac{2}{1 + \exp(-\tilde{\lambda}_t)}, \quad \tilde{\lambda}_t = \theta + \rho \lambda_{t-1} + \kappa_{0,+} z_{t-1}^+ + \kappa_{0,-} z_{t-1}^- + \kappa_{1,+} (z_{t-1} - 1)^+ + \kappa_{1,-} (z_{t-1} + 1)^-.$$ 

Values of parameters used in simulation are $\omega = 0.013$, $\alpha = 0.02$, $\gamma = 0.09$, $\beta = 0.92$, $\theta = -0.02$, $\rho = 0.7$, $\kappa_{0,+} = 0.3$, $\kappa_{0,-} = 0.3$, $\kappa_{1,+} = -0.3$, $\kappa_{1,-} = -0.4$, $\eta = 1.25$. Obtained maximum likelihood estimates are $\hat{\omega} = 0.011$, $\hat{\alpha} = 0.018$, $\hat{\gamma} = 0.089$, $\hat{\beta} = 0.928$, $\hat{\theta} = -0.047$, $\hat{\rho} = 0.618$, $\hat{\kappa}_{0,+} = 0.325$, $\hat{\kappa}_{0,-} = 0.247$, $\hat{\kappa}_{1,+} = -0.291$, $\hat{\kappa}_{1,-} = -0.352$, $\hat{\eta} = 1.21$. 

46