A general class of SemiGARCH models based on the Box-Cox Transformation

Xuehai Zhang, Yuanhua Feng and Christian Peitz

April 2017
A general class of SemiGARCH models based on the Box-Cox transformation

Xuehai Zhang, Yuanhua Feng and Christian Peitz
Faculty of Business Administration and Economics, Paderborn University, Germany

Abstract

The paper proposes a wide class of semiparametric GARCH models by introducing a scale function into a GARCH class model for featuring long-run volatility dynamics, which can be thought of as an MEM (multiplicative error model) with a varying scale function. Our focus is to estimate the scale function under suitable weak moment conditions by means of the Box-Cox transformation of the absolute returns. The estimation of the scale function is independent of any GARCH specification. To overcome the drawbacks of the kernel and the local linear approaches, a non-negatively constrained local linear estimator of the scale function, which is then proposed to fit a suitable parametric GARCH model to the standardized residuals, is considered. Asymptotic properties of the proposed nonparametric and parametric estimators are studied in detail and iterative plug-in algorithms are developed for selecting the bandwidth and transformation parameters, which are selected by MLE and JB statistic. The algorithms are also carried out independently without any parametric specification in the stationary part. Application to real data sets show that the proposals work very well in practice.

1 Introduction

Despite the success of the ARCH (autoregressive conditional heteroskedasticity, Engle, 1982) and GARCH (generalized ARCH, Bollerslev, 1986) models for modeling conditional (short-run) volatility dynamics in stock market returns, their implications for long-run volatility are restrictive, in the sense that these models imply a constant unconditional long-run volatility, i.e.
it implies that the stock market returns are stationary. However, in recent years it was realized that, this feature does not seem to be consistent with the time series behavior of volatilities of stock returns. Different extensions of the standard GARCH model are hence proposed for capturing the long-run volatility patterns observed in the data. For instance, Feng (2004) introduced a SemiGARCH (semiparametric GARCH) model by employing a smooth volatility trend (also called the scale function) into the standard GARCH model and proposed to estimate it using a data-driven kernel regression. Van Bellegem and von Sachs (2004) discussed the forecasting of financial time series under the time varying unconditional variance. A general time varying ARCH process was introduced by Dahlhaus and Rao (2006). Engle and Rangel (2008) put forward a Spline-GARCH model with a nonparametric volatility trend, which is defined as a function of the observation time, i.e. the location, and exogenous macroeconomic variables and is estimated by an exponential quadratic spline. Engle et al. (2008) extended this idea to a GARCH-MIDAS model, which combines the ideas of the Spline-GARCH model and of mixed data sampling (MIDAS), to investigate detailed macroeconomic sources of long-run dynamics of stock market volatility. Amado and Teräsvirta (2017) developed a specification technique for building multiplicative time-varying GARCH models by decomposing the variance into an conditional and a unconditional component, which is smooth over time.

In this paper, a general class of semiparametric GARCH models is introduced including the SemiGARCH model as a special case. Similar to the SemiGARCH model, the total volatility is defined as a product of a scale function and a conditional volatility component and the effect of exogenous variables is not considered. The key difference between the current proposal and the SemiGARCH model is that here the parametric part is not specified beforehand but to be chosen after estimating the scale function. Different specifications will lead to different models. By rewriting the GARCH formulations, it is shown that a semiparametric GARCH model is asymptotically equivalent to the GARCH model used in the parametric part with a time-varying scale parameter, while the other parameters remain constant. This provides us with a deep insight into the current proposal and indicates possible further extensions of it. To estimate the scale function, we propose the use of a constrained non-negative local linear regression, to ensure that the resulting scale function is (at least almost surely) positive. It is shown that the constrained local linear regression defined in this paper is asymptotically equivalent to the common local linear regression. Note that the data-driven algorithm proposed by Feng (2004) does not apply to the general framework considered in this paper because of the Box-Cox power transformation in the scale function. MLE and Jarque-Bera (JB) statistics are
applied in the selection of transformation parameter. Hence, the main focus of this paper is on
the development of a quick, stable data-driven algorithm for the practical implementation of
the general semiparametric GARCH approach. For this purpose, a fully data-driven iterative
plug-in bandwidth selector algorithm is proposed following Gasser et al. (1991), Herrmann et
al. (1992) and Beran and Feng (2002). The application to data examples shows that such
bandwidth and transformation parameter selection rules work well. Furthermore, a simple
test is introduced to determine, if a semiparametric GARCH or a parametric GARCH model
should be used. This test shows that the unconditional volatility during a financial crisis is
significantly higher than that in other sub-periods. It seems to be possible to develop a suitable
method for detecting the effect of a financial crisis by means of the proposal in this paper.
Further, the estimation and selection of a suitable parametric GARCH model based on the
standardized returns is also discussed. Some results in this paper can be easily adapted to the
Spline-GARCH or GARCH-MIDAS models. For instance, both of them are GARCH models
with a varying scale parameter determined by the time and other exogenous variables.

The paper is organized as follows. The model is introduced in Section 2. Section 3 discusses
the semiparametric estimation of the proposed model, the data-driven algorithm and the test
method. Data examples in Section 4 illustrate the practical usefulness of the proposal. Final
remarks in Section 5 conclude the paper. Sketched proofs of some results are given in the
appendix.

2 The models

Let \( y_t, t = 0, 1, ..., n \), denote the prices of some stock index and \( r_t \) their (log-)returns. To model
the slowly changing unconditional variance and conditional heteroskedasticity at the same time,
the following semiparametric GARCH class model (Feng, 2004) for the conditional distribution
of \( r_t \) is introduced:

\[
r_t = \mu(\tau) + s(\tau)\sqrt{h_t} \varepsilon_t,
\]

where \( \tau = t/n \) is the rescaled time, \( \mu(\cdot) \) stands for a smooth drift, \( s(\cdot) > 0 \) is a smooth scale
function and \( h_t \) is the conditional variance of the rescaled process \( \xi_t = r_t^*/s(\tau) = \sqrt{h_t} \varepsilon_t \), where
\( r_t^* = r_t - \mu(\tau) \). It is assumed that \( \xi_t \) also has unit variance so that the model is uniquely defined.
This implies that the unconditional mean of \( h_t \) is 1, i.e. \( E(h_t) = 1 \). Although our focus is on
the estimation of $s^2(\cdot)$ and $h_t$, a nonparametric drift function is included for modelling possible long-term deterministic changes in the mean of $y_t$. We will see that the asymptotic properties of $s^2(\cdot)$ will not be affected by the estimation errors in $\hat{\mu}(\cdot)$. Model (1) defines indeed a sequence of models. The process $r_t$ is non-stationary unless $\mu(\cdot)$ and $s(\cdot)$ are both constant. But $r_t$ is locally stationary following Dahlhaus (1997). In practice, returns may also have no nonparametric drift function. For simplicity, it will not be considered in the current paper, because our focus is on the estimation of $s(\cdot)$ and $\sigma_t$. Moreover, it is well known that under common regularity conditions the effect of the error in a nonparametric estimator of an unknown drift function on the estimation of $s(\tau)$ is asymptotically negligible. Then Model (1), without the drift function, is reduced to

$$r_t = s(\tau)\sqrt{h_t}\varepsilon_t.$$  \hspace{1cm} (2)

Model (2) is a general SemiMEM (semiparametric multiplicative error model) defined by introducing a smooth scale function into the MEM proposed by Engle (2002). Hence, all of the results given in this paper hold for a model with a nonparametric drift function, provided that the drift function is estimated by another well-developed data-driven algorithm.

The stationary process $\xi_t$ can be analyzed using any suitable GARCH class model and different parametric specifications on $h_t$ will lead to different semiparametric GARCH class models. If it is assumed that $\xi_t$ follows a standard GARCH model, we have

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \xi_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j},$$  \hspace{1cm} (3)

where $\alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p \geq 0$, $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ and $\alpha_0 = 1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j$. Due to the restriction $E(h_t) = 1$, $\alpha_0$ is no more a free parameter. Equations (2) and (3) together define the SemiGARCH model introduced by Feng (2004). See also Feng and McNeil (2008) for an extension of this model to high-frequency financial data.

2.1 The Box-Cox SemiGARCH models

Nonparametric estimation of variance functions is well known in the literature. Local polynomial estimation of variance functions in nonparametric regression with independent errors
is studied e.g. by Ruppert et al. (1997). Kernel and local linear estimators of conditional variance in nonlinear time series are proposed by Feng and Heiler (1998) and Fan and Yao (1998), respectively. A kernel estimator has the so-called boundary problem, which will not only affect the bias of the estimate at a boundary point but also reduce the convergence rate of the MISE (mean integrated squared error). A modified kernel estimator with second order bias at a boundary point is proposed by Hall and Presnell (1999). However, their proposal does not apply to the end points $x = 0$ or $x = 1$. A local linear estimator does not share this problem, however, it may be sometimes negative. In this paper we will hence propose the use of local linear estimators with a simple non-negative restriction. Moreover, a consistent estimator of $s^2(\tau)$ as a smoother of $\xi_t^2$ with a given bandwidth requires the existence of the fourth moments of $\xi_t$. But the selection of the bandwidth in this case requires the existence of the eighth moments of $\xi_t$. This will clearly affect the stability of the estimated scale function. This drawback hinders the application of such non-parametric variance estimators to financial data series, because the marginal distribution of a financial time series may have heavy tails. To solve this problem, we propose to estimate the scale function from the Box-Cox power transformation $|r_t|^\lambda$ with $0 < \lambda \leq 2$. If $\lambda \leq 1$ is used, the existence of the fourth moment of $\xi_t$ is sufficient for developing a convergent bandwidth selector. If the power transformation parameter is regarded as $\lambda = 0$, the Box-Cox transformation of $|r_t|^\lambda$ reduces to a logarithmic form, and obviously $\xi_t$ will follow a Log-GARCH process. The logarithmic transform can be applied to some financial variables, such as realized volatility, volatility index etc., so as to convert multiplicative models to additive models. For simplicity, $0 < \lambda \leq 2$ is applied to the Box-Cox transformation without the consideration of logarithmic transformation in the paper. Note that normally $\lambda = 2$ is used. Now, $s^2(\tau)$ is estimated first. $\hat{s}(\tau)$ is then obtained by taking the square root of $\hat{s}^2(\tau)$. Our proposal is to estimate the local mean of $|r_t|^\lambda$ first and then take the $\lambda$-th root of this local mean as an estimator of the scale function. The relationship between this estimate and the traditional one is as follows. Define $c_\lambda = E(|\xi_t|^\lambda)$, which is 1 for $\lambda = 2$ following the definition. For $\lambda \neq 2$, we have $c_\lambda \neq 1$ but its concrete value depends on the distribution of $\xi_t$ and will change from case to case. We will see that a nonparametric estimator based on $|r_t|^\lambda$ is indeed an estimator of $g(\tau, \lambda) = c_\lambda s_\lambda^\lambda(\tau)$ and not that of $s_\lambda^\lambda(\tau)$. Hence $[\hat{g}(\tau, \lambda)]^{1/\lambda} \approx c_\lambda^{1/\lambda} \hat{s}_\lambda(\tau) \neq \hat{s}(\tau)$, if $\lambda \neq 2$. To estimate the scale function from $r_t^2$ is the most natural method. However, note that the difference between $[g(\tau, \lambda)]^{1/\lambda}$ and $s_\lambda(\tau)$ is just a constant factor. Hence the use of $[g(\tau, \lambda)]^{1/\lambda}$ as an alternative scale function is equivalent to the use of $s_\lambda(\tau)$. Thus, for given $\lambda$,
Model (2) can be rewritten as
\[ r_t = s_{\lambda}(\tau) \xi_{\lambda,t}, \tag{4} \]
where \( s_{\lambda}(\tau) = c_1^{1/\lambda} s(\tau) = [g(\tau, \lambda)]^{1/\lambda} \) and \( \xi_{\lambda,t} = c_\lambda^{-1/\lambda} \xi_t \) is another stationary process with \( E(|\xi_{\lambda,t}|^\lambda) = 1 \). Obviously, \( \xi_{\lambda,t} \) and \( \xi_t \) share the same properties but with a different scale parameter. Hence the resulting estimator based on \(|r_t|^\lambda\) can be used to remove the effect of the slowly changing scale in \( r_t \). For \( \lambda = 2 \) we have \( s_2(\tau) = s(\tau) \). Otherwise, \( s_{\lambda}(\tau) \) and \( s(\tau) \) have different scale parameters. We see \( s_{\lambda}(\tau) \) can also be used as the scale function of the proposed model, which can be estimated consistently from \(|r_t|^\lambda\). There are different further transformations which can be used to estimate an equivalent scale function. The power transformations (or equivalently the Box-Cox transformations with non-negative power transformation parameter) are just the simplest examples. We refer the reader to Eagleson and Müller (1997) for more general description on this point.

It is clear that model (4) is an improved alternative of model (2) based on the Box-Cox transformation. We can propose Model (3) and Model (4) together as the Box-Cox SemiGARCH model, providing a new semiparametric methodology by introducing a power transformation parameter \( \lambda \) into the scale function. Similar to the time varying GARCH models, any kind of GARCH models can be selected as an extension in the parametric part of generalized Box-Cox SemiGARCH class models. If \( \lambda = 2 \) and the parametric part is a GARCH process, it is the standard SemiGARCH model proposed by Feng (2004). The Semi-APARCH model (Feng and Sun, 2013) is also another specification, applying the absolute returns and APARCH model in the parametric part.

3 The semiparametric estimation procedure

The generalized Box-Cox SemiGARCH class models introduced in the last section can be estimated using a semiparametric procedure. At first, \( s_{\lambda}(\tau) \) can be estimated by some non-parametric regression approach consistently without any parametric assumptions on \( \sigma_t \) and \( \varepsilon_t \). In the paper, local polynomial regression is applied. The slowly changing scale function can be estimated and removed under very weak moment condition \( E(|\xi_t|^\lambda) < \infty \) for any \( \lambda > 0 \) based on suitable power transformation of the data. A simple constrained local polynomial regression, which is approximately the same as the standard local polynomial regression, is proposed to
ensure that the resulting scale function is always positive. Then the conditional variance can be analyzed further using the GARCH class models based on the standardized returns.

3.1 Estimation of \( s(\tau) \)

Let \( \zeta_t = |\xi_{\lambda,t}|^\lambda - 1 \) with \( E[\zeta_t] = 0 \). Model (4) can be expressed as

\[
|r_t|^\lambda = g(\tau, \lambda) + g(\tau, \lambda)\zeta_t,
\]

which is a nonparametric regression with heteroskedastic time series errors and \( g(\tau, \lambda) \) is the trend function and the scale function at the same time. Let \( K(u) \) be a kernel function and \( b > 0 \) be the bandwidth. A local linear estimator of \( g(\tau, \lambda) \) at \( 0 \leq \tau \leq 1 \) is obtained by minimizing

\[
Q(\lambda, b) = \sum_{t=1}^{n} \left\{ |r_t|^\lambda - a_0 - a_1(\tau_t - \tau) \right\}^2 K\left( \frac{\tau_t - \tau}{b} \right).
\]

This results in \( \tilde{g}(\tau, \lambda) = \hat{a}_0 \). The advantage of a local linear estimator is that the bias of it is always of the order \( O(b^2) \). This is in particular important for application, because the forecasting of the trend is mainly carried out based on the estimation at the right end point. A problem is that \( \tilde{g}(\tau, \lambda) \) obtained above sometimes negative, in particular when the sample size is small and a small bandwidth is used. To ensure the non-negativity, we propose to use the final estimator \( \hat{g}(\tau, \lambda) = |\tilde{g}(\tau, \lambda)| \), which is almost surely positive. The use of \( \hat{g}(\tau, \lambda) \) instead of \( \tilde{g}(\tau, \lambda) \) is reasonable. Firstly, it can be shown that \( |\hat{g}(\tau, \lambda) - g(\tau, \lambda)|^2 \leq |\tilde{g}(\tau, \lambda) - g(\tau, \lambda)|^2 \). That is the performance of \( \hat{g}(\tau, \lambda) \) is not worse than that of \( \tilde{g}(\tau, \lambda) \) following the MSE (mean squared error). Moreover, negative values of \( \tilde{g}(\tau, \lambda) \) are just a limited sample problem, because the probability that \( |\tilde{g}(\tau, \lambda) - \tilde{g}(\tau, \lambda)| > \Delta \) for any \( \Delta > 0 \) tends to zero in an exponential rate. This is shown in the following lemma, where Assumptions A1 to A4 are described in the appendix.

**Lemma.** Suppose that a bandwidth of the order \( b = n^{-\lambda} \) with \( 0 < \lambda < 1 \) is used and \( \tilde{g}(\tau, \lambda) \) is consistent, asymptotically normal with bias \( B[\tilde{g}(\tau, \lambda)] = O(n^{-m}) \) and variance \( \text{Var}[(\tilde{g}(\tau, \lambda))] \approx \tilde{\sigma}^2 n^{-m_2} \), where \( \eta_1, \eta_2 > 0 \). If the assumptions A1 to A4 hold, then we have

\[
nP[\hat{g}(\tau, \lambda) \neq \tilde{g}(\tau, \lambda)] = nP[\tilde{g}(\tau, \lambda) < 0] \to 0, \text{ as } n \to \infty.
\]

The result of the lemma also holds if \( n \) is replaced by \( n^k \) for any \( k > 1 \), e.g. \( k = 2 \). Hence, when \( n \) and \( b \) are both large, then \( \tilde{g}(\tau, \lambda) < 0 \) will practically never happen, if \( b \) is large.
enough. Also, there is no difference between the asymptotic properties of \( \hat{g}(\tau, \lambda) \) and \( \tilde{g}(\tau, \lambda) \). Note that \( r_t \) are uncorrelated. The scale function \( \hat{g}(\cdot) \) defined above has the same asymptotic properties as those for a nonparametric regression estimator with independent errors and a non-constant scale function. For more theoretical discussions on these topics we refer the reader to Beran et al. (2015), where the estimation of the scale function in a semiparametric ACD (autoregressive conditional duration) model for daily average transaction durations is considered. The authors also obtained detailed asymptotic results of the constrained local linear estimator in that context. According to the similarity between the ACD and the GARCH models, asymptotic results of \( \hat{g}(\tau, \lambda) \) can be derived based on their results by replacing the average durations there with \( |r_t|^\lambda \).

The key idea behind our proposal is that although \( \hat{g}^{1/\lambda}(\tau, \lambda) \) is not a consistent estimator of \( s_\lambda(\tau) \), it can be directly used to remove the non-stationarity in returns, because \( \hat{\xi}_{\lambda,t} = r_t/\hat{g}^{1/\lambda}(\tau, \lambda) \) is also an approximately stationary process. Comparing the general formula (5) with the special case with \( \lambda = 2 \), we can see that, instead of the estimation of the scale function \( s(\tau) \) in \( r_t \), here the scale function in the process \( r_{\lambda,t} \) is indeed directly estimated, where \( r_{\lambda,t} = \text{sign}(r_t)|r_t|^{\lambda/2} \) with \( r_{2,t} = r_t \). As far as we know, there is still no study in the literature on the estimation of the scale function in a SemiGARCH framework based on the power transformation \( |r_t|^\lambda \). The main purpose is to develop a consistent data-driven estimator of the nonparametric scale function \( s_\lambda(\tau) \). If higher robustness is of interest, \( \lambda \leq 1 \) can be used and the assumptions that \( E(\xi_t^4) < \infty \) together with further regularity conditions is sufficient for developing a convergent bandwidth selector. This is the same moment condition required for estimating the GARCH parameters using conditional QLM (quasi maximum likelihood). In this paper, we will still consider the use of \( \lambda \leq 2 \) and in Fig. 5, we can see the selected \( \lambda \) are obviously smaller than 1, which means the stricter robustness requirement can be fulfilled.

Model (5) is an extension of Model (4) in Feng (2004), where only the special case with \( \lambda = 2 \) is considered. Asymptotic properties of \( \hat{g}(\tau, \lambda) \) can hence be proved analogously. The following summarizes and compares the asymptotic behavior of \( \hat{s}_\lambda(\tau) \) and \( \hat{g}(\tau, \lambda) \), where \( \text{MSE}[\hat{s}_\lambda(\tau),b] \) and \( \text{MSE}[\hat{g}(\tau, \lambda),b] \) denote the mean squared error of the two estimators obtained with the bandwidth \( b \). Assume now that \( E(\xi_t^4) < \infty \), a consistent estimate of \( s_\lambda(\tau) \) can be obtained as follows. Note that \( \hat{\xi}_{\lambda,t} \approx c_\lambda^{-1/\lambda} \hat{\xi}_t \) and that \( E(\xi_t^2) = 1 \). This leads to a consistent estimate of \( c_\lambda \)

\[
\hat{c}_\lambda = \left[ \frac{1}{n} \sum_{t=1}^{n} \hat{\xi}_{\lambda,t}^2 \right]^{-\lambda/2}
\]

(8)
and
\[
\hat{c}_\lambda^{-1/\lambda} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \hat{c}_{\lambda,t}^{2}}.
\]  
(9)

We obtain
\[
\hat{s}_\lambda(\tau) = \hat{c}_\lambda^{-1/\lambda} [\hat{g}(\tau, \lambda)]^{1/\lambda},
\]  
(10)

through rescaling the sample variance of the standardized returns to be one. Note that, so long as \( \hat{g}(\tau, \lambda) \) is consistent, the effect of the error in it on \( \hat{c}_\lambda \) is asymptotically negligible. Hence both of \( \hat{c}_\lambda \) and \( \hat{c}_\lambda^{-1/\lambda} \) are still \( \sqrt{n} \)-consistent. It leads to the conclusion that the MSE of \( \hat{s}_\lambda(\tau) \) in this way is approximately \( c_\lambda^{-2/\lambda} \text{MSE}[\hat{g}(\tau, \lambda)] \), which is still of the order \( O(n^{-4/5}) \), provided that \( \hat{g}(\tau, \lambda) \) is obtained by a suitable data-driven algorithm. However, \( \hat{s}_\lambda(\tau) \) is not an efficient estimate if \( \lambda \neq 2 \) is used, because the optimal bandwidth for estimating \( g(\tau, \lambda) \) is different to that for estimating \( s_\lambda(\tau) \). Moreover, we see that to obtain a consistent nonparametric estimate of \( s_\lambda(\tau) \), the condition \( E(\xi^4_t) < \infty \) is also necessary. To avoid possible confusion, we propose to estimate \( g(\tau, \lambda) \) for some chosen \( \lambda \) and to calculate \( \hat{\xi}_{\lambda,t} \) at first. Then we can obtain \( \hat{c}_\lambda \) and \( \hat{s}_\lambda(\tau) \) following (8) through (10). Finally, we will calculate the standardized returns \( \hat{\xi}_t = r_t / \hat{s}_\lambda(\tau) \) again, which are approximately independent of the choice of \( \lambda \) and will be used for further analysis.

Suppose that \( E(\xi^4_t) < \infty \) and the Assumptions A1 to A4 stated in the appendix hold. For any \( 0 < \lambda \leq 2 \) and \( 0 \leq \tau \leq 1 \), both \( \hat{s}_\lambda(\tau) \) and \( \hat{g}(\tau, \lambda) \) are consistent estimators of \( s_\lambda(\tau) \) and \( g(\tau, \lambda) \), respectively. Further, it also holds that \( \text{MSE}[\hat{s}_\lambda(\tau), b] \approx c_\lambda^{-2/\lambda} \text{MSE}[\hat{g}(\tau, \lambda), b] \), and \( \hat{s}_\lambda(\tau) \) and \( \hat{g}(\tau, \lambda) \) have the same asymptotically optimal bandwidth. The finding of particular interest is that \( \hat{s}_\lambda(\tau) \) and \( \hat{g}(\tau, \lambda) \) have the same asymptotically optimal bandwidth. Note that our aim is to estimate \( s_\lambda(\tau) \). However, it is straightforward to select the bandwidth for \( \hat{g}(\tau, \lambda) \). The bandwidth is just what we need for an optimal estimate of \( s_\lambda(\tau) \). So the problem is solved well. There is no need to develop a separate bandwidth selection procedure for estimating \( s_\lambda(\tau) \). Asymptotic properties of \( \hat{g}(\cdot) \) can be obtained following known results in nonparametric regression with dependent errors (see e.g. Altman, 1990 and Hart, 1991). In the sequel, some necessary results are summarized. For a kernel function \( K \), define \( R(K) = \int K^2(u)du \) and \( I(K) = \int u^2K(u)du \). Let \( S_{|\xi|} = \sum_{k=-\infty}^{\infty} \gamma_{|\xi|}(k), \) where \( \gamma_{|\xi|}(k) = \text{cov}(|\xi_0|^\lambda, |\xi_k|^\lambda) \), then the following holds.

**Theorem 1.** Under the Assumptions A1 through A5 stated in the appendix, the following holds,

i) The bias of \( \hat{g}(\tau, \lambda) \) is \( B[\hat{g}(\tau, \lambda)] = E[\hat{g}(\tau, \lambda)] - g(\tau, \lambda) = \frac{1}{2} b^2 [g(\tau, \lambda)]'' I(K) + o(b^2) \).
ii) The variance of \( \hat{g}(\tau, \lambda) \) is given by
\[
\text{Var}(\hat{g}(\tau, \lambda)) = S|\xi| \lambda g(\tau, \lambda)^2 R(K) nb + o\left(\frac{1}{nb}\right) = V nb + o\left(\frac{1}{nb}\right),
\] (11)
where \( V = S|\xi| \lambda g(\tau, \lambda)^2 R(K) \).

iii) If a bandwidth \( b = o(b_A) \) is used, the bias is asymptotically negligible and
\[
\sqrt{nb}[\hat{g}(\tau, \lambda) - g(\tau, \lambda)] \xrightarrow{p} N(0, V),
\] (12)
where \( V \) is as defined in (11).

In Theorem 1, the asymptotic bias and variance of \( \hat{g}(\tau, \lambda) \) are obviously similar to those in nonparametric regression with some specific GARCH or ACD class model, because of the similarity in calculating the sum of the autocovariance. The result of Theorem 1 iii) also indicates that \( \hat{g}(\tau, \lambda) \) is asymptotically unbiased and asymptotically normal with a bandwidth of a smaller order than \( b_A \).

**Theorem 2.** Suppose that Assumptions A1 to A4 hold, we have:

i) The mean squared error (MSE) of \( \hat{g}(\tau, d) \) is,
\[
\text{MISE}[\hat{g}(\tau, \lambda)] = \{B[g(\tau, \lambda)]\}^2 + V[g(\tau, \lambda)]
= S|\xi| \int \frac{g^2(\tau, \lambda) d\tau R(K)}{nb} + \max \left\{ o(b^4), o\left(\frac{1}{nb}\right) \right\},
\] (13)

at any point \( 0 < \tau < 1 \), the local asymptotically optimal bandwidth is given by
\[
b_A(\tau) = \left( S|\xi| \frac{R(K)}{I^2(K)} \left\{ \frac{[g(\tau, \lambda)]^2}{[g(\tau, \lambda)]^\prime} \right\}^2 \right)^{1/5} n^{-1/5}.
\] (14)

ii) The mean integrated squared error (MISE) of \( \hat{g}(\tau, \lambda) \) is,
\[
\text{MISE}[\hat{g}(\tau, \lambda)] = \int_0^1 \{\hat{g}(\tau, \lambda) - g(\tau, \lambda)\}^2 d\tau
= b^4 \int \left\{ \left[ [g(\tau, \lambda)]^\prime \right]^2 \right\}^2 d\tau I(K) + S|\xi| \int \frac{g^2(\tau, \lambda) d\tau R(K)}{nb} + \max \left\{ o(b^4), o\left(\frac{1}{nb}\right) \right\},
\] (15)

then the (global) asymptotically optimal bandwidth is given by
\[
b_A = \left( S|\xi| \frac{R(K)}{I^2(K)} \int \left\{ \frac{[g(\tau, \lambda)]^2}{[g(\tau, \lambda)]^\prime} \right\}^2 d\tau \right)^{1/5} n^{-1/5}.
\] (16)
Results in Theorem 2 are closely related to those for a local linear estimator of the mean function with heteroskedastic time series errors. Furthermore, note that the result in the first part of Theorem 2 does not hold at a boundary point, because the kernel constants in the asymptotic variance and asymptotic bias of \( \hat{g}(\tau, \lambda) \) at the boundary change from point to point. But this does not affect the asymptotic MISE, so that the global bandwidth can be calculated over \( \tau \in [0, 1] \).

### 3.2 Semiparametric estimation of a given model

The unknown parameters of chosen GARCH class models can be estimated from \( \hat{\xi}_t \) by approximate (conditional) QLM method proposed in the literature. A suitable model can also be selected using e.g. the BIC.*

Denote the true unknown parameter vector of a chosen GARCH model by \( \theta_0 \). Let \( \hat{\theta} \) be the estimate of \( \theta_0 \) obtained from \( \hat{\xi}_t \) and \( \hat{\theta} \) denotes the standard QMLE obtained under the assumption that \( \xi_t \) is observable. It is well known that under suitable regularity conditions \( \hat{\theta} \) is \( \sqrt{n} \)-consistent and asymptotically normal. The additional variance caused by the errors in \( \hat{\xi}_t \) is asymptotically negligible. The \( O(b^2) \) term in \( B_\theta \) is due to \( E[\hat{s}_\lambda(\tau) - s_\lambda(\tau)] \) and the \( O((nb)^{-1}) \) term due to Cov \( [\xi_t^2, \hat{s}_\lambda(\tau)] \). If a bandwidth \( O(n^{-1/2}) < b < O(n^{-1/4}) \) is used, \( B_\theta \) is asymptotically negligible. Now \( \hat{\theta} \) is also \( \sqrt{n} \) consistent and asymptotically normal. If the data-driven algorithm proposed in the next section is used, the bias term \( B_\theta \) will be of the order \( O(n^{-2/5}) \). We see that in the general SemiGARCH class models \( \sqrt{n} \)-consistent parametric estimation is no longer possible, if the scale function changes over time. In the special case, when \( r_t \) follow a stationary GARCH class model, a bandwidth of the order \( O_p(1) \) will be selected by the proposed the data-driven algorithm in the next section. Now, the parametric estimation is still \( \sqrt{n} \)-consistent but is inefficient. This means that some efficiency will be lost, if a generalized semiparametric GARCH class model is fitted to some stationary GARCH class process. In the next section a simple stationary test is proposed based on the selected bandwidth. If this test is significant, the proposed semiparametric model will be used. Otherwise, stationary generalized GARCH class models should be employed.

*In the paper, for fitting GARCH models, the R packages "fGarch" and "rugarch" are applied and the data-driven algorithm to be proposed in the next section is also carried out in R.
3.3 The bandwidth selection algorithm

Numerous criteria for selecting the bandwidth in nonparametric regression are proposed. One bandwidth selection rule which works well in different contexts, is the iterative plug-in (IPI) idea (Gasser et al., 1991). This approach will also be used in the current paper. Note that the estimation of \( g(\tau, \lambda) \) is just the estimation of the scale function in \(|r_i|^\lambda\).

The IPI algorithm is developed based on the formula of the asymptotically optimal bandwidth for estimating \( \hat{g}(\tau, \lambda) \), \( b_A \) say, which can be obtained by adapting those known results properly. In the sequel, this formula will be given without proof. For a kernel function \( K(u) \), define \( R(K) = \int K^2(u)du \) and \( I(K) = \int u^2 K(u)du \). Under regularity assumptions, in particular the assumption that \( \gamma_{|\xi|\lambda}(k) \) are absolutely summable, the asymptotically optimal bandwidth minimizing the dominating part of the MISE is given by

\[
b_A = \left( \frac{S_{|\xi|\lambda} R(K)}{T^2(K) \int \{[g(\tau, \lambda)]''\}^2 d\tau} \right)^{1/5} n^{-1/5}.
\]

(17)

To select \( b \), \( S_{|\xi|\lambda} \) has to be estimated. And the IPI idea is successfully applied to select bandwidth in different contexts (see e.g. Herrmann et al., 1992, Brockmann et al., 1993, Beran and Feng, 2002, and Ghosh and Draghicescu, 2002). An IPI bandwidth selector makes use of (17). The procedure is started with a bandwidth \( b_0 \). In Gasser et al. (1991), Herrmann et al. (1992) and Brockmann et al. (1993), the starting bandwidth \( b_0 = n^{-1} \) is used. Beran and Feng (2002) proposed to use \( b_0 = n^{-5/7} \) so that the starting bandwidth satisfies \( b_0 \to 0 \) and \( nb_0 \to \infty \). The bandwidth \( b_0 = 0.5n^{-1/5} \) is used by Feng (2004), which is of the optimal order \( O(n^{-1/5}) \). In this paper we proposed to select the starting bandwidth from a set of given bandwidths using the CV (cross-validation, Wahba and Wold, 1975) criterion, so that the algorithm is fully data-driven. It is well known that the choice of the starting bandwidth only has a clear effect on the required number of iterations but not on the finally selected bandwidth.

In an IPI algorithm, a bandwidth \( b_{\lambda,j} \) for estimating the second derivative \([g(\tau, \lambda)]''\) is calculated from \( \hat{b}_{j-1} \) using some inflation method. The choice of the inflation method is very important, because the rate of convergence of an IPI bandwidth selector depends on this choice. The original proposal of Gasser et al. (1991), applied a multiplicative inflation method (MIM), where \( b_{\lambda,j} = \hat{b}_{j-1} \cdot n^\alpha \) with \( \alpha = 1/10 \). Now, we have \( b_{\lambda,j} = O(n^{-1/10}) \), once convergence is reached. This ensures that the variance of \( \hat{b}/b_A \) has the fastest rate of convergence \( O(n^{-1/2}) \)
but the bias of \( \hat{b} \) is relatively large, where \( \hat{b} \) denotes the finally selected bandwidth. An exponential inflation method (EIM), \( b_{λ,j} = (\hat{b}_j - 1)^α \), was proposed by Beran and Feng (2002). The authors proposed to use the optimal choice \( α = 5/7 \), which minimizes the MSE of \( \int [g(τ, λ)]'' dτ \). Numerical experiments show that sometimes the MIM method does not work well, because the inflation factor \( n^{1/10} \) depends strongly on \( n \), and the range of the sample size considered in the current context is very large. The EIM method with \( α = 5/7 \) works well in different contexts and will be used in this paper.

Ghosh and Draghicescu (2002) proposed to estimate some unknown functions in bandwidth selection for quantile regression with time series errors directly from the data. Following their idea, it is proposed to estimate the unknown quantity \( S_{|ξ|^λ} \) non-parametrically by the sum of the sample autocovariances \( \hat{γ}_{|ξ|^λ}(k) \) of the residuals until some lag \( M \), where \( M \) satisfies \( M → ∞ \) and \( M/n → 0 \). Bühlmann (1996) proposed the optimal window selection of Bartlett window and \( C^2 \)-window with IPI. Bartlett window is selected as the lag window and in the following \( M = [3n^{1/5}] = O(n^{1/5}) \) will be used, where \([·]\) denotes the integer part. Under this choice, the effect of the error in \( \hat{S}_{|ξ|^λ} \) on the finally selected bandwidth is asymptotically negligible. Note that \( \hat{γ}_{|ξ|^λ}(k) \) tends to zero very fast. Hence, the finally selected bandwidth will not be changed clearly, if a larger \( M \), e.g. \( M = [4n^{1/5}] \), is used. Also note that bandwidths \( \hat{b}_j \) obtained in several iterations at the beginning are usually inconsistent. It is not good to use those bandwidths to estimate \( S_{|ξ|^λ} \). Following Herrmann et al. (1992), we select the bandwidth first by ignoring the correlation and scale change. In this stage a simple difference-based variance estimator \( \hat{γ}_{|ξ|^λ} = \frac{1}{2(n-1)} \sum_{t=2}^{n} (|ξ_t|^λ - |ξ_{t-1}|^λ)^2 \) will be used. The bandwidth selected at the end of this stage will be used as a new starting point for selecting the bandwidth under correlated errors with a smooth scale function. From now on \( \hat{S}_{|ξ|^λ} \) will be estimated and adapted in each iteration. The detailed bandwidth selection algorithm is discussed with the selection of power transformation parameter in the next subsection.

3.4 The power transformation parameter selection algorithm

Let \( b_0 \) denote the starting bandwidth, depending on an initial \( λ_0 \) value input. In the application, the starting input values \( λ_0 = 2, 1, 0.5 \) and \( 0.1 \) will be considered. In the algorithm, \( λ_0 = 1 \) is applied and the results remain with the other initial \( λ \) values.

The proposed data-driven algorithm is as follows:
1. Obtain $\hat{\mu}(\tau_t)$ using an IPI algorithm and let $\hat{r}_t^* = r_t - \hat{\mu}(\tau_t)$.

2. Select $\hat{b}_0$ from $b_{0,i} = c_{0,i}n^{-1/5}$ with $c_{0,i} = 0.05, 0.10, 0.15, 0.20, 0.25$ using the CV criterion and the starting power transformation parameter input $\lambda_0 = 1$. Then put $j = 1$.

3. Select a bandwidth by ignoring the correlation and scale change.

4. In the $m$-th iteration for $m > 1$,
   a) Let $\Delta \lambda = 0.001$ be the interval of $\lambda$ and $\lambda_n = n \cdot \Delta \lambda$, where $5 \leq n \leq 1000$.
   b) Determine the $\hat{\lambda}_{m-1} = \lambda_n$ by maximizing the MLE or minimizing the JB statistic.
   c) Increase $m$ by one and repeat b) until $\hat{\lambda}_m$ is convergence and let $\hat{\lambda} = \hat{\lambda}_m$.

5. Let $J_1$ be the number of iterations in the last stage. In the $j$-th iteration with $j > J_1$,
   a) Estimate $\hat{g}(\tau, \hat{\lambda})$ with $b_{j-1}$. Let $\hat{\xi}_t = \hat{r}_t / \hat{g}(\tau, \hat{\lambda})$ and $\hat{S}_{|\xi|^k} = \sum_{|k|<M} \hat{r}_t^{|k|}(k)$.
   b) Let $[\hat{g}(\tau, \hat{\lambda})]^n$ denote the estimate of $[g(\tau, \hat{\lambda})]^n$ obtained by using $b_{\lambda,j} = b_{j-1}^{5/7}$.
   c) Improve $b_{j-1}$ by
   \[
   b_j = \left( \frac{\hat{S}_{|\xi|^k} R(K)}{K^2 \int [\hat{g}(\tau, \hat{\lambda})]^n d\tau} \right)^{1/5} n^{-1/5}. \tag{18}
   
   d) Increase $j$ by one and repeatedly carry out b) and c) until convergence is reached or until a given maximal number of iterations has been done.

The finally selected power transformation parameter $\hat{\lambda}$ and bandwidth $\hat{b}_A$ are obtained in the last iteration of Step 4 and Step 5, respectively. In Step 1, the scale change is also ignored to save computing time. The condition $|b_j - b_{j-1}| < 1/n$ is used as a convergence criterion of $\hat{b}$, since such a difference is negligible. The bandwidth $\hat{b}_0$ used in Step 2 provides an object starting point of the algorithm. The maximal number of iterations, which indeed does not play any role in a common case, are 20 in Steps 1 and 3 and 30 in Step 5. The $\hat{\lambda}$, in Step 4, is a stable global power transformation parameter of the Box-Cox transformation. It means that now the scale function is estimated from the $\lambda$-th power of the absolute returns instead of the squared returns. Note that both the estimated scale function with the selected power transformation parameter $\lambda$ and the scale function applied during the descaled process is $\hat{g}(\tau, \hat{\lambda})$. Obviously, the convert parameter $\hat{c}_\lambda^{-1/\lambda}$ in Equation (9) can not be neglected.
3.5 A simple stationary test

The proposed semiparametric models should be used, only if the underlying process is non-stationary in the mean and/or non-stationary in the variance. In the sequel a simple method is proposed to test whether the variance of the process is constant. Similarly, a test of the stationarity in the mean can also be carried out, if this is of interest.

Assume that the acf of $|\xi_t|^\lambda$ is absolutely summable for any $0 < \lambda \leq 2$. Our null-hypothesis $(H_0)$ assumes that the process $r_t$ is stationary with constant standard deviation $s(\tau) \equiv s_0$ and finite fourth moments. Let $\hat{g}(\tau, \lambda)$ be the estimator of $g(\tau, \lambda)$ defined above with a bandwidth $b$ such that $b \to 0$ and $nb \to \infty$ as $n \to \infty$. Under $H_0$ it is clear that $g(\tau, \lambda)$ is also a constant $g_0(\lambda) \equiv c_\lambda s_0^\lambda$. Under $H_0$ and corresponding regularity assumptions, we have

$$\sqrt{n\hat{b}_{\lambda}}[\hat{g}(\tau, \lambda) - g_0(\lambda)] \to N[0, R(K)V_{|\xi|^\lambda}],$$

where $R(K) = \int K^2(u)du$ is the kernel constant in the asymptotic variance of $\hat{g}(\tau, \lambda)$ and $V_{|\xi|^\lambda} = S_{|\xi|^\lambda}^0 g_0^2(\lambda)$, where $S_{|\xi|^\lambda}^0$ is similar to $S_{|\xi|^\lambda}$ in Theorem 1 but defined for the process $\xi_t^\lambda = r_t^\lambda/g_0(\lambda)$. The overall variance $g_0(\lambda)$ can be estimated by the sample variance of $\{r_t^\lambda\}$, $\hat{g}_0(\lambda)$ say. The quantity $S_{|\xi|^\lambda}^0$ can be estimated from $\hat{S}_{|\xi|^\lambda}^0 = \hat{r}_t^\lambda/\hat{g}_0(\lambda)$ following the idea in the last subsection. Let $\hat{SD}_\sigma = [\hat{V}_{|\xi|^\lambda} R(K)/(n\hat{b}_{\lambda})]^{1/2}$ and $Z_{\alpha/2}$ be the $\alpha/2$ upper quantile of the standard normal distribution for given confidence level $\alpha$. Then $\hat{g}_0(\lambda) \pm Z_{\alpha/2}\hat{SD}_\sigma$ provide the approximate $(1-\alpha)*100\%$ confident bounds of $g_0(\lambda)$ under the stationary assumption on $\{r_t^\lambda\}$. If more than $\alpha * 100\%$ of the estimates $\hat{g}(\tau, \lambda)$ are clearly outside these confidence bounds, it indicates that $\{r_t^\lambda\}$ is non-stationary in the variance and a semiparametric model should be used. Otherwise, generalized parametric GARCH models will be preferable.

4 Applications

Several major stock market indexes are selected to carry out the algorithm. In the following empirical research, Standard & Poor’s 500 Index (S&P) and Deutscher Aktienindex 30 (DAX) from January 1996 to December 2015, covering 20 financial years, are applied as research samples. In the results, the general SemiGARCH class models are well fitted with the actual financial data.
4.1 The results of power transformation parameter $\lambda$ selection

An IPI algorithm is carried out to calculate the $\hat{\lambda}$, which will be used as the Box-Cox power transformation parameter in the scale function of the general SemiGARCH class models and it is discovered that the fixed $\lambda$ is extremely important in financial markets. In Fig. 1, we set different initial $\lambda$ inputs, which are 2, 1, 0.5 and 0.1, respectively and the fixed $\lambda$ searching processes of DAX and S&P are displayed.

During the $\lambda$ selection, a 6th-procedure IPI process is employed and it is proven that most of the data sets will reach their convergence values after the second IPI procedure, only a few need a third step (Beran and Feng, 2002). The $\hat{\lambda}$ seems to be independent with the initial $\lambda$ input, tending to a fixed value quickly. In the search process, the R packages ‘tseries’ and ‘MASS’ are used to select $\lambda$ of JB and MLE, respectively. When calculating the $\lambda$ value with JB, the formula of JB can also be directly applied and the input series is simply the standard returns. The difference between the $\hat{\lambda}$ of JB with both of the methods is so tiny that it can be neglected. To ensure the positivity of the input series, we have to remove the mean value and then obtain the descaled returns, because of the zero values in the returns series. Obviously, there is no significant difference between the fixed $\lambda$ and the optimal $\lambda$ we get until the second IPI procedure, e.g. the optimal $\lambda$ in the second IPI procedure of S&P with JB (starting $\lambda = 0.1$) is 0.300 and the $\hat{\lambda}$ after all the IPI procedures is also 0.3. Similarly in the example of DAX with MLE, $\lambda$ of the second IPI with starting $\lambda = 2$ is 0.321 and the $\hat{\lambda}$ is also 0.321. However, the $\lambda$ in the first step result of IPI may also coincide with the selected value, e.g. for DAX with JB, when the initial $\lambda = 1$, the $\lambda$ values are always 0.335 in the IPI process. In the figure, it is also discovered, if the input $\lambda$ is far above or below the final convergence value, $\hat{\lambda}$ will decrease or increase quickly to the final determined value. For returns series, it is obvious that the determined value is around 0.25. Ding and Granger (1996) also indicated that fourth root transformation is preferred to the square returns or the absolute ones.

In Fig. 2, the power parameters appear a U-shape curve and the optimal power $\lambda$ is where it minimizes the JB statistic or maximizes the MLE with fixed bandwidth $\hat{b}$. Please note that in this figure, the estimated trend is removed and the existence of the trend does not affect the U-shape curve. The 95% confidence intervals of MLE can be calculated based on the $\chi^2$ distribution, displayed as horizontal dashed line in the figure. We can observe that most of the minimum values marked by the dashed lines are around one quarter and they are actually
the $\lambda$ and their values of DAX and S&P are exactly 0.335, 0.300 and 0.321, 0.290 with JB and MLE, respectively. Obviously, the optimal $\hat{\lambda}$ with JB and MLE are quite close. However, the optimal $\lambda$ values selected by both methods are dramatically far away from that of $\lambda = 2$ or even $\lambda = 1$ and the $Q$ values also differ from those with absolute returns or the square returns. We understand in a financial market, if there are various extreme observations, a precise scale function with lower $\lambda$ is required to reveal the trend of the returns. In contrast the consideration of smoothness of the scale function leads to the relatively higher $\lambda$ value in this case. Further, if lower $\lambda$ is applied, the requirement of higher order moments does not exist. At the left boundary, it is found that the JB value increases at an extremely exponential speed, when $\lambda$ is close to zero and following the definition of Box-Cox transformation, there is no doubt that a logarithmic transform has to be considered. In this case, the descaled returns will follow a Log-GARCH process, which is a crucial connection between the additive and multiplicative models. For another, we can conclude if $\lambda = 2$ is selected, generally the JB values are far away from the optimal one. It has to be argued from the figure if a classic least square process with square returns is employed, although it is robust.

According to Fig. 3, we can observe the data sets before and after the Box-Cox transformation. Obviously, if the $\hat{\lambda}$ is applied, the distribution of the return series seems to be closer to a normal distribution. Further, we can see, the $\hat{\lambda}$ are far smaller than their values used before, even smaller than one half. It means, if the existence of second order moment is ensured, the scale function can be estimated under a weak moment condition.

### 4.2 The selection of the parametric models

If the power transformation parameter is fixed, the returns of the financial market indexes can be modeled with a general SemiGARCH process. The returns we considered in the scale function do not follow the square ($\lambda = 2$) or the absolute ($\lambda = 1$) patterns by manual, but rather that of the $\lambda$-th power, which is selected by means of the $\lambda$ selection algorithm.

The smoothing results are displayed in Fig. 4 and Fig. 5. The $\lambda$ used here is selected by MLE and similar results can also be obtained if JB is applied. Also, the stationarity test in the variance is based on the $\hat{\lambda}$ by means of MLE. The returns seem to be more stationary after removing the scale function, regarded as the long term component. In addition, clear GARCH cluster effects can still be observed, because the short term component displayed in a GARCH
class process is barely affected by removing the long term component. In other words, the financial returns can be divided into the long and short components, which can be described by the scale function using Box-Cox transformation and the descaled process using the GARCH class process.

In the paper, the GARCH, APARCH, EGARCH and csGARCH models of order (1,1), (1,2), (2,1) and (2,2) are chosen to analyse the conditional heteroscedasticity in the stationary standardized returns. It is also discovered that there is no significance with the mean function of the return series and it will not be considered in the model fitting. The innovations in the models are assumed to follow a normal- and t-distribution. In Table 1, the BIC of models with $\lambda = \hat{\lambda}$, 1, 2 are provided and it is discovered that the EGARCH(2, 1) model with t-distribution are always selected because of the minimum BIC values. From Table 2, the shape parameter, also known as degree of freedom of the innovation distribution in all cases are significantly greater than 8, which means that the eighth moment of $\xi_t$ exists and little heavy tails of the innovations distribution in the six research markets. Meanwhile, the degree of freedom of S&P is obviously lower than that of DAX, which means that the appearance possibility of extreme returns in US market is much higher than that in German market. For another, in EGARCH models, $\gamma_1$ indicates no longer the leverage effect but the size effect of the past returns on volatility, which is typically a cluster effect. Besides, the leverage effect is denoted as $\alpha_1$, being always negative to reveal the aggravation of past negative returns. In the cases of S&P and DAX, it is discovered that the leverage parameters are obviously determined by the negative sum of $\alpha_1$ and $\alpha_2$. It seems that the leverage effect in EGARCH models is weaker than that if other parametric models are applied, such as APARCH models.

5 Final remarks

In the paper, we put forward a wide class of SemiGARCH models with Box-Cox transformation. A data-driven algorithm is also carried out in the transformation parameter selection, which is a great improvement in the scale function estimation of SemiGARCH models. The parameter $\lambda$ we applied in the scale function estimation is obtained after several IPI procedures until it converges, also a supplement in displaying the behavior of long term component in SemiGARCH models. In the parametric part, general GARCH models can be selected to describe the performance of the returns after removing the long-run trend. In the paper, GARCH class models
are discussed as the cluster models to show the short run behaviours in some major financial markets of the world. It is found, if more extreme values are in a market, the transformation parameter $\lambda$ tends to be smaller, for example the $\hat{\lambda}$ of both DAX and S&P are only about a quarter, indicating the stability in the two stock markets. It is also proven from the distribution of innovations that the innovation of DAX follows a $\varepsilon_t \sim t(10.3433)$ distribution, exhibiting the existence of eighth moment and little heavy tails.

The framework of general SemiGARCH class models is set up in the paper, however, some open questions still have to be discussed further. e.g. the statistical properties of $\hat{\lambda}$ have not been fully explored yet. The optimal selection of the constant value at zero point of spectral density in the IPI procedures is also of great interest.
References


Appendix

Proof of Lemma. Let \( Z_n = n^{n/2} \{ \tilde{g}(\tau, \lambda) - B[\tilde{g}(\tau, \lambda) - g(\tau, \lambda)]/\tilde{\sigma} \} \). Under the assumptions of Lemma 1, \( Z_n \) is asymptotically standard normal. We have

\[
P(\tilde{g}(\tau, \lambda) < 0) \leq P[|\tilde{g}(\tau, \lambda) - g(\tau, \lambda)| > g(\tau, \lambda)]
\]

\[
= P[|\tilde{g}(\tau, \lambda) - B[\tilde{g}(\tau, \lambda)] - g(\tau, \lambda) + B[\tilde{g}(\tau, \lambda)]| > g(\tau, \lambda)] \\
\leq P[|\tilde{g}(\tau, \lambda) - B[\tilde{g}(\tau, \lambda)] - g(\tau, \lambda)| > m(x) - |B[\tilde{g}(\tau, \lambda)]|] \\
\leq P[|\tilde{g}(\tau, \lambda) - B[\tilde{g}(\tau, \lambda)] - g(\tau, \lambda)| > g(\tau, \lambda)/2]
\]

if \( n \) is large enough. Furthermore, we have

\[
P\{|\tilde{g}(\tau, \lambda) - B[\tilde{g}(\tau, \lambda)] - g(\tau, \lambda)| > g(\tau, \lambda)/2\} = P\{|Z_n| > n^{n/2}g(\tau, \lambda)/(2\tilde{\sigma})\}.
\]

Defining \( z_n^o = n^{n/2}g(\tau, \lambda)/(2\tilde{\sigma}) \), we have \( n = L_g(z_n^o)^{2/\eta_2} \), where \( L_g = [2\tilde{\sigma}/g(\tau, \lambda)]^{2/\eta_2} \). Furthermore let \( Z \sim N(0, 1) \). Then

\[
nP\{|Z_n| > z_n^o\} \approx nP\{|n| > z_n^o\} \\
= 2L_g(z_n^o)^{2/\eta_2} \int_{z=z_n^o}^{\infty} e^{-z^2/2} dz \\
= 2L_g \int_{z=z_n^o}^{\infty} (z_n^o)^{2/\eta_2} e^{-z^2/2} dz \to 0,
\]

as \( n \to \infty \), because all moments of \( Z \) are finite. Lemma is proved.

To prove the results of Theorem 1, the following assumptions are required.

A1. The scale function \( g(\tau, \lambda) \) is strictly positive, bounded, and at least twice continuously differentiable on \([0, 1]\).

A2. The kernel \( K(u) \) is a symmetric density with compact support \([-1, 1]\).

A3. The bandwidth \( b \) satisfies \( b \to 0 \) and \( nb \to \infty \) as \( n \to \infty \).

A4. \( \{ \zeta_t \} \) is a stationary process with unit mean and absolutely summable autocovariance.

A5. The stationary process \( \{ \zeta_t \} \) can be represented as \( \zeta_t = 1 + \sum_{i=0}^{\infty} \lambda_i \nu_{t-i} \), where \( \{ \nu_t \} \) is a sequence of uncorrelated zero-mean innovations with finite variance, \( \sum_{i=0}^{\infty} \lambda_i \neq 0 \) and \( \sum_{i=0}^{\infty} |\lambda_i| < \infty \).
Assumptions A1 to A3 are the regular nonparametric regression conditions. A4 is the requirement of the GARCH model. A5 is a sufficient regularity condition which ensures that the sample means of ζ_t and ξ_t are both asymptotically normal.

**Proof of Theorem 1.**

Following Lemma, we can conclude that \( B[\hat{g}(\tau, \lambda)] = B[\tilde{g}(\tau, \lambda)] + o_p(n^{-1/2}) \) and \( \text{Var}[\hat{g}(\tau, \lambda)] = \text{Var}[\tilde{g}(\tau, \lambda)] + o_p(n^{-1}) \). The proof will hence simply be given for the unrestricted local linear estimator \( \tilde{g}(\tau, \lambda) \).

i) Bias: Since \( \tilde{g}(\tau, \lambda) \) is a linear smoother, the bias \( B[\tilde{g}(\tau, \lambda)] = E[\tilde{g}(\tau, \lambda)] - g(\tau, \lambda) \) is the same as in the nonparametric regression with i.i.d. errors. This is the formula given i).

ii) Variance: The local linear estimator \( \tilde{g}(\tau, \lambda) \) is a linear estimator \( \tilde{g}(\tau, \lambda) = \sum_{i=1}^{T} w_i^\tau y_i \). It is well known that the weights \( w_i^\tau \) are asymptotically equivalent to those defined by the equivalent kernel, i.e.

\[
 w_i^\tau = \frac{K_\tau \left( \frac{\tau_i - \tau}{b_\tau} \right)}{\sum_{i=1}^{n} K_\tau \left( \frac{\tau_i - \tau}{b_\tau} \right)} \left[ 1 + o(1) \right] = \frac{1}{nb_\tau} K_\tau \left( \frac{\tau_i - \tau}{b_\tau} \right) \left[ 1 + o(1) \right]
\]

for \( |\tau_i - \tau| \leq b \) and zero otherwise.

Note that the autocovariances of \( \xi_t \) and \( \zeta_t \) are the same. Furthermore, let \( K \) be an integer such that \( K \to \infty \) and \( K/nb_\tau \to 0 \), as \( n \to \infty \). For instance, we may choose \( K = [\sqrt{n}b_\tau] \) where \( [\tau] \) denotes the integer part of \( \tau \). Defining \( b_K = K/n \) we have \( b_K/b_\tau \to 0 \) as \( n \to \infty \). The variance of \( \tilde{g}(\tau, \lambda) \) is given by

\[
 \text{Var}[\tilde{g}(\tau, \lambda)] = \sum_{|\tau_i - \tau| \leq b_\tau} \sum_{|\tau_j - \tau| \leq b_\tau} w_i^\tau w_j^\tau \text{Cov}[g(\tau_i, \lambda)\xi_i, g(\tau_j, \lambda)\xi_j]
\]

\[
 = \sum_{|\tau_i - \tau| \leq (b_\tau - b_K)} \sum_{|\tau_j - \tau| \leq b_\tau} w_i^\tau w_j^\tau \text{Cov}[g(\tau_i, \lambda)\xi_i, g(\tau_j, \lambda)\xi_j]
\]

\[
 + \sum_{|\tau_i - \tau| > (b_\tau - b_K)} \sum_{|\tau_j - \tau| \leq b_\tau} w_i^\tau w_j^\tau \text{Cov}[g(\tau_i, \lambda)\xi_i, g(\tau_j, \lambda)\xi_j]
\]

\[
 =: V_1 + V_2,
\]

where \( V_1 \) indicates the contribution of the observations in the middle part of the window and \( V_2 \) the contribution in the boundary of the window. The definition of \( K \) and \( b_K \) ensures that \( V_2 = o(V_1) \), i.e. \( \text{Var}[\tilde{g}(\tau, \lambda)] \approx V_1 \). Note that the condition \( |\tau_i - \tau| \leq b_\tau - b_K \) ensures that \( \tau_i \),
τ

j with |i − j| ≤ K are all within the window. This will simplify the analysis in the next part. Denote by \( V_{1i} \) the \( i \)th sum over \( \tau_j \) in \( V_1 \) for given \( \tau_i \). Then

\[
V_{1i} = \sum_{|\tau_j - \tau_i| \leq b_r} w^*_{i} w^*_{j} \text{Cov} \left[ g(\tau_i, \lambda) \xi_i, g(\tau_j, \lambda) \xi_j \right] \\
= \sum_{|i-j| \leq K} w^*_{i} w^*_{j} \text{Cov} \left[ g(\tau_i, \lambda) \xi_i, g(\tau_j, \lambda) \xi_j \right] \\
+ \sum_{|i-j| > K} w^*_{i} w^*_{j} \text{Cov} \left[ g(\tau_i, \lambda) \xi_i, g(\tau_j, \lambda) \xi_j \right] \\
=: V^C_{1i} + V^T_{1i},
\]

where \( V^C_{1i} \) denotes the contribution of the covariances in the central part with lags |k| ≤ K, whereas \( V^T_{1i} \) is the contribution of the covariances in the tail part. For the first term in (21) we have

\[
V^C_{1i} = \sum_{|i-j| \leq K} w^*_{i} w^*_{j} \text{Cov} \left[ g(\tau_i, \lambda) \xi_i, g(\tau_j, \lambda) \xi_j \right] \\
= \frac{1}{(nb_r)^2} K^2 \left( \frac{\tau_i - \tau}{b_r} \right) \left[ 1 + o(1) \right] \left[ g^2(\tau_i, \lambda) \right] \left[ 1 + O(b_K) \right] \sum_{|k| \leq K} \gamma(k) \\
\approx \frac{2\pi c_f}{(nb_r)^2} K^2 \left( \frac{\tau_i - \tau}{b_r} \right) g^2(\tau_i, \lambda). \tag{22}
\]

The second term in (21) is asymptotically negligible, because

\[
V^T_{1i} = \sum_{|i-j| > K} w^*_{i} w^*_{j} \text{Cov} \left[ g(\tau_i, \lambda) \xi_i, g(\tau_j, \lambda) \xi_j \right] \\
\leq \sum_{|i-j| > K} |w^*_{i} w^*_{j}| \text{Cov} \left[ g(\tau_i, \lambda) \xi_i, g(\tau_j, \lambda) \xi_j \right] \\
\leq \frac{C_i}{(nb_r)^2} \left[ 1 + o(1) \right] \sum_{|k| > K} |\gamma(k)| \\
= o \left( \frac{1}{(nb_r)^2} \right), \tag{23}
\]

where

\[
C_i = \sup_{|i-j| > K} \left| K \left( \frac{\tau_i - \tau}{b_r} \right) K \left( \frac{\tau_j - \tau}{b_r} \right) g(\tau_i, \lambda) g(\tau_j, \lambda) \right|.
\]
This leads to

\[
\text{Var} [\tilde{g}(\tau, \lambda)] = \left\{ \sum_{|\tau_i - \tau| \leq (b_{r} - b_K)} \sum_{|\tau_j - \tau| \leq b_{r}} w_{i}^{T} w_{j}^{T} \text{Cov} [g(\tau_i, \lambda) \xi_i, g(\tau_j, \lambda) \xi_j] \right\} [1 + o(1)]
\]

\[
= \left\{ \sum_{|\tau_i - \tau| \leq (b_{r} - b_K)} \sum_{|i-j| \geq K} V_{ii}^{C} \right\} [1 + o(1)]
\]

\[
= \frac{2\pi c_{f} \tilde{g}^{2}(\tau, \lambda)}{(nb_{r})} \left\{ \sum_{|\tau_i - \tau| \leq (b_{r} - b_K)} \frac{1}{(nb_{r})} K_{r}^{2} \left( \frac{\tau_i - \tau}{b_{r}} \right) \right\} [1 + o(1)]
\]

\[
= \frac{2\pi c_{f} \tilde{g}^{2}(\tau, \lambda) R(K_{r})}{(nb_{r})} [1 + o(1)]
\]

(24)

as given in Theorem 1 ii).

iii) Here a more general result \(\sqrt{n b_{r} [\tilde{g}(\tau, \lambda) - B[\tilde{g}(\tau, \lambda)] - g(\tau, \lambda)] \rightarrow N(0, V)\) can be proved, with \(V\) defined in (11). This leads to \(\sqrt{n b_{r} [\tilde{g}(\tau, \lambda) - g(\tau, \lambda)] \rightarrow N(0, V)\), when \(b_{r} = o(b_{A})\), because \(\sqrt{n b_{r} B[\tilde{g}(\tau, \lambda)] \rightarrow 0}\). Define \(\delta = \tilde{g}(\tau, \lambda) - B[\tilde{g}(\tau, \lambda)] - g(\tau, \lambda)\). Note that

\[
\delta(\tau) = \sum_{t=1}^{n} w_{t}^{*} [g(\tau, \lambda) \xi_t]
\]

\[
= \sum_{t=1}^{n} w_{t}^{*} \xi_t,
\]

(25)

where \(w_{t}^{*} = w_{t}^{r} g(\tau, \lambda)\). It is easy to check that the regularity conditions (4.2) and (4.3) in Beran and Feng (2002) are jointly fulfilled by \(w_{t}^{*}\) and \(\lambda_{t}\). Hence, following Theorem 1 in Beran and Feng (2002), \(\delta(\tau)\) is asymptotically normal, if the sample mean of \(\xi_t\) is. The latter is guaranteed by A1 and A5. Hence, Theorem 1 follows.

\(\diamond\)

**Proof of Theorem 2.**

i) The formula for the MSE of \(\tilde{g}(\tau, \lambda)\) is the sum of the square bias and the variance. The bias and variance follow i) and ii) in Theorem 1.

ii) The MISE can be calculated on the whole support \([0, 1]\), because the contribution of the estimated values in the boundary area is asymptotically negligible.

\(\diamond\)
Figure 1: The IPI process with JB and MLE

Figure 2: The $\hat{\lambda}$ with JB and MLE
Figure 3: The histogram of DAX and S&P with JB and MLE
Figure 4: The smoothing results of DAX Index from Jan 1996 to Dec 2015
(a) Closing price of the S&P Index from Jan 1996 to Dec 2015

(b) The log–returns of the S&P Index

(c) Scale functions for the selected $\lambda$, $\lambda=1$ and $\lambda=2$ with 95%–confidence bounds under constant scale

(d) Standardized returns calculated by means of the scale function with the selected $\lambda$ in (c)

Figure 5: The smoothing results of S&P Index from Jan 1996 to Dec 2015
<table>
<thead>
<tr>
<th>Index</th>
<th>Model</th>
<th>Order</th>
<th>$\lambda = \hat{\lambda}$</th>
<th>$\lambda = 1$</th>
<th>$\lambda = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>DAX</td>
<td>GARCH</td>
<td>(1,1)</td>
<td>2.7472 2.7323</td>
<td>2.7483 2.7340</td>
<td>2.6981 2.6556</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.7489 2.7340</td>
<td>2.7501 2.7358</td>
<td>2.7032 2.6596</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.7452 2.7291</td>
<td>2.7461 2.7308</td>
<td>2.6960 2.6525</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.7469 2.7308</td>
<td>2.7477 2.7324</td>
<td>2.6977 2.6542</td>
</tr>
<tr>
<td></td>
<td>APARCH</td>
<td>(1,1)</td>
<td>2.7220 2.7091</td>
<td>2.7221 2.7098</td>
<td>2.6699 2.6362</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.7236 2.7107</td>
<td>2.7238 2.7114</td>
<td>2.6749 2.6396</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.7229 2.7103</td>
<td>2.7231 2.7109</td>
<td>2.6703 2.6355</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.7246 2.7119</td>
<td>2.7248 2.7126</td>
<td>2.6720 2.6372</td>
</tr>
<tr>
<td></td>
<td>EGARCH</td>
<td>(1,1)</td>
<td>2.7214 2.7082</td>
<td>2.7210 2.7087</td>
<td>2.6767 2.6351</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.7229 2.7097</td>
<td>2.7225 2.7101</td>
<td>2.6803 2.6377</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.7158 2.6993</td>
<td>2.7152 2.6998</td>
<td>2.6674 2.6238</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.7168 2.7005</td>
<td>2.7162 2.7010</td>
<td>2.6686 2.6251</td>
</tr>
<tr>
<td></td>
<td>csGARCH</td>
<td>(1,1)</td>
<td>2.7500 2.7352</td>
<td>2.7507 2.7368</td>
<td>2.6772 2.6578</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.7516 2.7368</td>
<td>2.7523 2.7384</td>
<td>2.6778 2.6607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.7480 2.7321</td>
<td>2.7486 2.7336</td>
<td>2.6736 2.6545</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.7497 2.7338</td>
<td>2.7502 2.7353</td>
<td>2.6753 2.6562</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>GARCH</td>
<td>(1,1)</td>
<td>2.7243 2.6988</td>
<td>2.7310 2.7060</td>
<td>2.7002 2.6753</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.7259 2.7005</td>
<td>2.7326 2.7076</td>
<td>2.7018 2.6769</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.7207 2.6956</td>
<td>2.7278 2.7030</td>
<td>2.6975 2.6726</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.7208 2.6971</td>
<td>2.7278 2.7044</td>
<td>2.6976 2.6740</td>
</tr>
<tr>
<td></td>
<td>APARCH</td>
<td>(1,1)</td>
<td>2.6745 2.6590</td>
<td>2.6797 2.6651</td>
<td>2.6481 2.6339</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.6761 2.6606</td>
<td>2.6813 2.6667</td>
<td>2.6496 2.6355</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.6769 2.6623</td>
<td>2.6824 2.6684</td>
<td>2.6510 2.6368</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.6738 2.6605</td>
<td>2.6796 2.6669</td>
<td>2.6488 2.6362</td>
</tr>
<tr>
<td></td>
<td>EGARCH</td>
<td>(1,1)</td>
<td>2.6725 2.6565</td>
<td>2.6776 2.6626</td>
<td>2.6472 2.6325</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.6740 2.6581</td>
<td>2.6791 2.6642</td>
<td>2.6487 2.6341</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.6651 2.6487</td>
<td>2.6708 2.6551</td>
<td>2.6404 2.6248</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.6662 2.6499</td>
<td>2.6720 2.6563</td>
<td>2.6416 2.6261</td>
</tr>
<tr>
<td></td>
<td>csGARCH</td>
<td>(1,1)</td>
<td>2.7275 2.7021</td>
<td>2.7341 2.7092</td>
<td>2.7032 2.6782</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1,2)</td>
<td>2.7292 2.7038</td>
<td>2.7358 2.7109</td>
<td>2.7049 2.6798</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,1)</td>
<td>2.7237 2.6988</td>
<td>2.7308 2.7061</td>
<td>2.6997 2.6750</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2,2)</td>
<td>2.7238 2.7003</td>
<td>2.7308 2.7076</td>
<td>2.7004 2.6758</td>
</tr>
</tbody>
</table>
Table 2: Fitting results of the Semi-EGARCH(2, 1)-t models with selected $\lambda$

<table>
<thead>
<tr>
<th>Data</th>
<th>Stat.</th>
<th>$\mu$</th>
<th>$\omega$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\beta_1$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAX</td>
<td>coeff.</td>
<td>0.0378</td>
<td>-0.0083</td>
<td>-0.2189</td>
<td>0.0991</td>
<td>0.9454</td>
<td>-0.0865</td>
<td>0.2515</td>
<td>10.4867</td>
</tr>
<tr>
<td></td>
<td>s.e.</td>
<td>0.0122</td>
<td>0.0033</td>
<td>0.0225</td>
<td>0.0233</td>
<td>0.0081</td>
<td>0.0338</td>
<td>0.0344</td>
<td>1.3717</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>coeff.</td>
<td>0.0286</td>
<td>-0.0111</td>
<td>-0.2567</td>
<td>0.0858</td>
<td>0.9461</td>
<td>-0.1336</td>
<td>0.2591</td>
<td>9.6926</td>
</tr>
<tr>
<td></td>
<td>s.e.</td>
<td>0.0116</td>
<td>0.0034</td>
<td>0.0234</td>
<td>0.0243</td>
<td>0.0068</td>
<td>0.0346</td>
<td>0.0356</td>
<td>1.2381</td>
</tr>
<tr>
<td>Year</td>
<td>Authors</td>
<td>Title</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-06</td>
<td>Xuehai Zhang, Yuanhua Feng, Christian Peitz</td>
<td>A general class of SemiGARCH models based on the Box-Cox transformation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-05</td>
<td>Juliane Proelss, Denis Schweizer, Volker Seiler</td>
<td>Do Announcements of WTO Dispute Resolution Cases Matter? Evidence from the Rare Earth Elements Market</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-04</td>
<td>Yuanhua Feng, Thomas Gries</td>
<td>Data-driven local polynomial for the trend and its derivatives in economic time series</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-03</td>
<td>Sonja Brangewitz, Behnud Mir Djawadi, Angelika Endres, Britta Hoyer</td>
<td>Network Formation and Disruption - An Experiment Are efficient networks too complex?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-02</td>
<td>Volker Seiler, Evgenia Taach</td>
<td>Effectiveness of Official Development Assistance – Further Evidence from DAC Countries</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017-01</td>
<td>Nico Stöckmann</td>
<td>“The Impact of Refugees on Per Capita Income” A Gravity Model Approach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016-06</td>
<td>Jochen Manegold</td>
<td>Stackelberg Competition among Intermediaries in a Differentiated Duopoly with Product Innovation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016-05</td>
<td>Republished as CIE Working Paper 2017-05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016-04</td>
<td>Marlon Fritz, Thomas Gries, Yuanhua Feng</td>
<td>Growth Trends and Systematic Patterns of Booms and Busts – Testing 200 Years of Business Cycle Dynamics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016-03</td>
<td>Thomas Gries, Claus-Jochen Haake</td>
<td>An Economic Theory of ‘Destabilization War’</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016-02</td>
<td>Daniel Kaimann, Nadja Maraun, Joe Cox</td>
<td>Identifying preferences and heterogeneity of consumers in multiplayer video games</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016-01</td>
<td>Republished as CIE Working Paper 2016-05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-05</td>
<td>Sonja Brangewitz, Claus-Jochen Haake, Philipp Möhlmeier</td>
<td>Strategic Formation of Customer Relationship Networks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-04</td>
<td>Sonja Brangewitz, Jochen Manegold</td>
<td>Competition and Product Innovation of Intermediaries on a Differentiated Duopoly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-03</td>
<td>Achim Voß, Mark Schopf</td>
<td>Lobbying over Exhaustible-Resource Extraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-02</td>
<td>Maximilian Müller, Denis Schweizer, Volker Seiler</td>
<td>Wealth Effects of Rare Earth Prices and China’s Rare Earth Elements Policy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015-01</td>
<td>Yuanhua Feng, Chen Zhou</td>
<td>An iterative plug-in algorithm for realized kernels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014-13</td>
<td>Republished as CIE Working Paper 2015-02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014-12</td>
<td>Sonja Brangewitz, Behnud Djawadi, René Fahr, Claus-Jochen Haake</td>
<td>Quality Choices and Reputation Systems in Online Markets – An Experimental Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014-11</td>
<td>Daniel Kaimann, Joe Cox</td>
<td>The Interaction of Signals: A Fuzzy set Analysis of the Video Game Industry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>