SFB 649 Discussion Paper 2006-033

Varying coefficient GARCH versus local constant volatility modeling. Comparison of the predictive power

Jörg Polzehl*
Vladimir Spokoiny*

* Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Varying coefficient GARCH versus local constant volatility modeling. Comparison of the predictive power*

Polzehl, Jörg
Weierstrass-Institute,
Mohrenstr. 39,
10117 Berlin, Germany
polzehl@wias-berlin.de

Spokoiny, Vladimir
Weierstrass-Institute,
Mohrenstr. 39,
10117 Berlin, Germany
spokoiny@wias-berlin.de

Abstract

GARCH models are widely used in financial econometrics. However, we show by mean of a simple simulation example that the GARCH approach may lead to a serious model misspecification if the assumption of stationarity is violated. In particular, the well known integrated GARCH effect can be explained by nonstationarity of the time series.
We then introduce a more general class of GARCH models with time varying coefficients and present an adaptive procedure which can estimate the GARCH coefficients as a function of time. We also discuss a simpler semiparametric model in which the β-parameter is fixed.
Finally we compare the performance of the parametric, time varying nonparametric and semiparametric GARCH(1,1) models and the locally constant model from Polzehl and Spokoiny (2002) by means of simulated and real data sets using different forecasting criteria. Our results indicate that the simple locally constant model outperforms the other models in almost all cases. The GARCH(1,1) model also demonstrates a relatively good forecasting performance as far as the short term forecasting horizon is considered. However, its application to long term forecasting seems questionable because of possible misspecification of the model parameters.

Keywords: varying coefficient GARCH, adaptive weights

JEL classification: C14, C22, C53.

*This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 *Economic Risk*.

1 Introduction

Autoregressive conditionally heteroscedastic (ARCH) and generalized autoregressive conditionally heteroscedastic (GARCH) models gained a lot of attention and are widely used in financial engineering since they were introduced by Engle (1982) and Bollerslev (1986). The simple GARCH(1,1) model is particularly popular. It models the observed log-returns R_t of the asset price process by the following two equations:

$$\begin{aligned} R_t &= \sigma_t \varepsilon_t \,, \\ \sigma_t^2 &= \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2 . \end{aligned}$$

Here ω, α, β are coefficients and σ_t^2 is the time varying volatility that is usually the target of analysis. The innovations ε_t are assumed zero mean and variance one conditioned on the σ-field $\mathcal{F}_{t-1}$ generated by the past observations. The GARCH(1,1) suggests a very natural and tractable model with only three parameters to be estimated. Moreover, this model allows to mimic many important stylized facts of financial time series like volatility clustering (alternating periods of small and large volatility) and persistent autocorrelation (slow decay of the autocovariance function of the absolute or squared returns). We cite from Engle (1995): *"The GARCH(1,1) is the leading generic model for almost all asset classes of returns. ... it is quite robust and does most of the work in almost all cases."*.

A simple parametric structure allows to directly apply the well developed parametric statistical methodology for estimation of the parameters and calibration of the model for real life applications and for studying the asymptotic properties of the estimates. The GARCH models are successfully applied to short term ahead forecasting of the volatility and particularly to Value-at-Risk problems, see McNeil and Frey (2000), Eberlein and Prause (2002).

However, a thorough analysis of the results delivered by the GARCH modeling raises some questions and indicates some problems.

For estimating the GARCH coefficients one usually applies a quasi likelihood approach. This means that the innovations ε_t are assumed i.i.d. standard normal and then the coefficients are obtained by maximizing the corresponding log-likelihood function. The resulting estimate is root-n consistent and asymptotically normal, see e.g., Berkes, Horvath and Kokoszka (2003) or Straumann and Mikosch (2003).

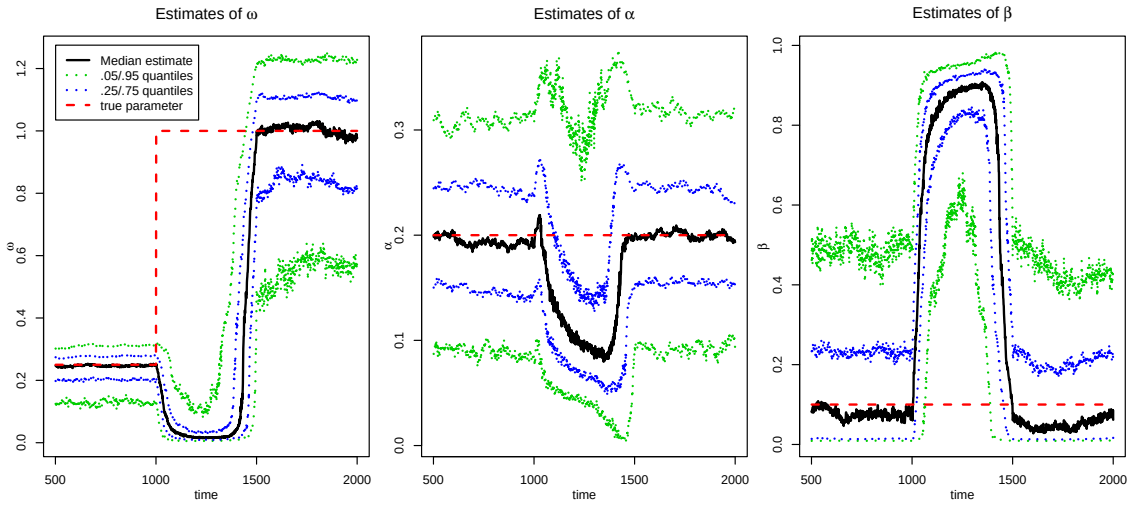


Figure 1: The true parameters (red) and the pointwise quantiles of the MLE's $\widehat{\omega}_t, \widehat{\alpha}_t, \widehat{\beta}_t$ obtained from the last 500 historical observations R_s for $s < t$.

However, for practical applications, the convergence is quite slow and one needs about 500 to 1000 data points to get a reasonable quality of estimation, especially for the coefficient β, see Mikosch and Stărică (2002, 2004). Giraitis and Robinson (2001) and Mikosch and Stărică (2004) discussed an alternative approach based on the Whittle estimator, for GARCH parameters, while Peng and Yao (2003) considered the LAD approach. However its performance is similar. Particularly, for 250 observations (corresponds to one year for daily data) the variability in the estimated β-parameter is quite high.

We also face a small identifiability problem. If $\alpha = 0$, then the parameters ω and β are not identifiable. Some additional boundary conditions on the process σ_t are necessary in this case. However, under the usual ergodicity condition, memory of boundary values is lost with the exponential speed. This yields some numerical problems for estimation of the parameters in the cases when α is near zero.

One more critical point is that GARCH modeling hardly extends to multiple time series, because of the overparametrization problem, see e.g. the BEKK model in Baba et al (1990) or Engle and Sheppard (2004).

However, it appears that the most crucial problem in the whole GARCH approach is that the GARCH models are not robust w.r.t. violation from the stationarity assumption. We illustrate this problem by a numerical experiment for an artificial change point model, see Figure 1. The observed data R_t for $t = 1, \ldots, 2000$

follow for $t < t_{cp} = 1000$ one GARCH(1,1) model with parameters $\omega_1 = 0.25$, $\alpha_1 = 0.2$ and $\beta_1 = 0.1$ and after $t = t_{cp}$ only the parameter ω jumps to $\omega_2 = 1$. We apply a scrolling window estimation procedure, that is, for every point t we estimate the parameters of the GARCH(1,1) model from the last 500 historical data R_s for $s \in [t-500, t-1]$. Therefore, for $t \leq t_{cp}$ we observe the performance of the GARCH estimator when the data generating process is indeed parametric GARCH(1,1). The resulting estimator is rather variable, however, it basically mimics the true model. For $t \in [t_{cp}+1, t_{cp}+500]$, the GARCH parameters are estimated from the subsample $R_{t-499}, \ldots, R_t$ which contains a jump in the ω-parameter at t_{cp}. We observe for such t that, even if most observations are from one model and only few of them come from the other model, the estimates are completely misspecified and in particular, the parameter β jumps to a value close to 1. Mikosch and Stărică (2004) and Stărică (2004) provide an explanation of this behavior: a GARCH(1,1) model, especially with a large value of the sum $\alpha + \beta$, is effectively very close to an exponential smoothing filter with memory parameter β. In other words, if the stationarity assumption is violated, GARCH modeling is essentially reduced to exponential smoothing of the latest observed squared returns. Mikosch and Stărică (2000, 2004) also argued that the other stylized facts of the financial time series like long range dependence, persistent autocorrelation and integration GARCH effect can be well explained by nonstationarity in the observed data.

In this paper we make an attempt to overcome this problem by considering the so called varying coefficients GARCH models. This means that the coefficients ω, α, β may vary with time and allows to model structural changes and external shocks in the considered time series. Varying coefficient models have been applied to model some financial time series in Fan, Jiang, Zhang and Zhou (2003) under the assumptions that the model parameters smoothly vary with time. We apply a more general approach that allows to include the case when the parameters spontaneously change. The estimation problem for such models is much more complicated than in the parametric case because we have to estimate three parameters which are possibly discontinuous functions of time. We also have to account that, even in a parametric case, a careful estimation of the GARCH-parameters from a small or moderate sample size is a hard task. To reduce the complexity of the model, apart

from the fully nonparametric model in which all three parameters are functions of time we consider a semiparametric model in which the parameter β is kept fixed and the two other parameters may vary with time. Additionally we consider the local constant volatility model where the coefficients α and β are zero and only the coefficient ω is a function of time. The latter model was considered in Polzehl and Spokoiny (2002) and Mercurio and Spokoiny (2004a, 2004b), see also Stărică and Granger (2004). Finally we compare these three models with the classical parametric GARCH(1,1) model.

For a comparison we use a number of simulated examples and look at different criteria like the prediction error, excess probability in Value-at-Risk (VaR) forecast and mean predictive VaR values.

We also apply the considered methods to real data including the DAX time series and the USD/GBP exchange rate series. For a comparison we look at the empirical counterparts of the criteria used in the simulations.

The results indicate that for both simulated and real data examples, the simple local constant model outperforms the other models including the more complicated non- and semiparametric models and delivers, in all cases, very reasonable results. At the same time, we observe that the fully nonparametric model has problems in identifying all the parameters as functions of time. A less variable semiparametric modeling delivers more stable results which also help to judge about statistical significance of the integrated GARCH effect.

The paper is organized as follows. Section 2 discusses the parameter estimation problem for the GARCH(1,1) model and indicates the related problems. Section 3 presents a varying coefficient GARCH model. The estimation problem for this model is discussed in Section 4. A modified procedure for the semiparametric GARCH model is briefly discussed in Section 4.4. Section 4.5 explains how the results of estimation can be used for out-of-sample forecasting of the volatility. Sections 5 and 6 illustrate the numerical performance of the methods by means of some simulated examples and applications to real data.

2 GARCH modeling and parameter estimation

Let the observed returns R_t obey the conditional heteroskedastic equation

$$R_t = \sigma_t \varepsilon_t \qquad t \geq t_0 \,,$$

where ε_t are "innovations" and σ_t is the volatility process. It is usually assumed that σ_t is measurable w.r.t. the σ-field $\mathcal{F}_{t-1}$ generated by the past observations R_s for $s < t$ and that the conditional distribution of the innovations given $\mathcal{F}_{t-1}$ fulfills $\boldsymbol{E}\,(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $\boldsymbol{E}\,(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$.

The GARCH(1,1) model specifies the volatility process σ_t^2 by the equation

$$\sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2 .$$

We denote $X_s = \sigma_s^2$ and $Y_s = R_s^2$ so that the process X_t obeys the linear autoregressive equation

$$X_s = \omega + \alpha Y_{s-1} + \beta X_{s-1} \,. \tag{2.1}$$

Usually all the coefficients are assumed nonnegative, that is, $\alpha \geq 0$, $\omega \geq 0$, $\beta \geq 0$. The condition $\alpha + \beta < 1$ ensures ergodicity of the process Y_t.

We denote by $\boldsymbol{\theta} = (\omega, \alpha, \beta)^\top$ the vector of parameters. Note that equation (2.1) does not uniquely determine the process $\{X_s\}$. Apart the vector $\boldsymbol{\theta}$, one has to specify the boundary (initial) value $\eta = X_{t_0}$ for some point t_0. However, the dependence on this parameter in the ergodic case is rather small, and we simply set $X_{t_0} = Y_{t_0} = R_{t_0}^2$. We therefore use the notation $X_s = X_s(\boldsymbol{\theta})$ to indicate the dependence of the volatility process on $\boldsymbol{\theta}$.

The structural linear equation can now be written as

$$X_s(\boldsymbol{\theta}) = \Psi_s(\boldsymbol{\theta})\boldsymbol{\theta} = \omega + \alpha Y_{s-1} + \beta X_{s-1}(\boldsymbol{\theta}), \tag{2.2}$$

with $\Psi_s(\boldsymbol{\theta}) = \bigl(1, Y_{s-1}, X_{s-1}(\boldsymbol{\theta})\bigr)$. Using this linear equation we can recursively compute the values $X_s(\boldsymbol{\theta})$, $s > t_0$, starting from the initial value $X_{t_0} = \eta$.

Similarly we obtain the derivatives $\nabla X_s(\boldsymbol{\theta}) = dX_s(\boldsymbol{\theta})/d\boldsymbol{\theta}$ and $\nabla^2 X_s(\boldsymbol{\theta}) = d^2 X_s(\boldsymbol{\theta})/d\boldsymbol{\theta}^2$. Namely it holds

$$\nabla X_s(\boldsymbol{\theta}) \;\; = \;\; \Psi_s^\top(\boldsymbol{\theta}) + \beta \nabla X_{s-1}(\boldsymbol{\theta}), \tag{2.3}$$

with the initial condition $\nabla X_t(\boldsymbol{\theta}) = 0$ for $t = t_0$. A similar recurrent formula applies for the matrix of second derivatives:

$$\nabla^2 X_s(\boldsymbol{\theta}) \;=\; \nabla\Psi_s(\boldsymbol{\theta}) + \nabla^\top \Psi_s(\boldsymbol{\theta}) + \beta\nabla^2 X_{s-1}(\boldsymbol{\theta}), \tag{2.4}$$

where $\nabla\Psi_s(\boldsymbol{\theta}) = (0, 0, \nabla X_{s-1}(\boldsymbol{\theta}))$ and $\nabla^2 X_s(\boldsymbol{\theta}) = 0$ for $s \le t_0$.

For estimating the parameter $\boldsymbol{\theta}$, one usually applies the quasi maximum likelihood approach assuming independent standard normal innovations $\{\varepsilon_s\}_{s\ge t_0}$. The log-likelihood for model (2.1) up to a constant term can be represented in the form

$$L(\boldsymbol{\theta}) = \frac{1}{2}\sum_{s\ge t_0} \ell(R_s, X_s(\boldsymbol{\theta}))$$

where $\ell(r, \sigma^2) = -(\log\sigma^2 + r^2/\sigma^2)/2$. We define the (quasi) maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$ by maximizing $L(\boldsymbol{\theta})$:

$$\widehat{\boldsymbol{\theta}} = \operatorname*{argsup}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \operatorname*{argsup}_{\boldsymbol{\theta}} \sum_{s\ge t_0} \ell(R_s, X_s(\boldsymbol{\theta}))). \tag{2.5}$$

The MLE $\widehat{\boldsymbol{\theta}}$ fulfills the estimating equation $dL(\boldsymbol{\theta})/d\boldsymbol{\theta} = 0$ leading to

$$\sum_{s\ge t_0} \big(Y_s - X_s(\boldsymbol{\theta})\big) |X_s(\boldsymbol{\theta})|^{-2}\, \nabla X_s(\boldsymbol{\theta}) = 0. \tag{2.6}$$

For solving this equation, one can apply an iterative Newton-Raphson procedure. Let some initial value $\boldsymbol{\theta}^{(0)}$ be fixed and let $\boldsymbol{\theta}^{(k-1)}$ be the estimated parameter vector after step $k-1$ for $k \ge 1$. One can compute the latent volatility process $X_s^{(k)} = X_s(\boldsymbol{\theta}^{(k-1)})$ by (2.2) and the derivatives $\nabla X_s^{(k)} = dX_s(\boldsymbol{\theta}^{(k-1)})/d\boldsymbol{\theta}$ and $\nabla^2 X_s^{(k)} = d^2 X_s(\boldsymbol{\theta}^{(k-1)})/d\boldsymbol{\theta}^2$ by (2.3) and (2.4) and define the update $\boldsymbol{\theta}^{(k)}$ as $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)} + (B^{(k)})^{-1} S^{(k)}$ with

$$\begin{aligned} S^{(k)} &= \sum_{s\ge t_0} \big|X_s^{(k)}\big|^{-2} \Big(U_s - X_s^{(k)}\Big) \nabla X_s^{(k)}, \\ B^{(k)} &= \sum_{s\ge t_0} \big|X_s^{(k)}\big|^{-2} \nabla X_s^{(k)} \left(\nabla X_s^{(k)}\right)^\top \\ &\quad + \sum_{s\ge t_0} \big|X_s^{(k)}\big|^{-2} \Big(U_s - X_s^{(k)}\Big) \left(\frac{2}{X_s^{(k)}} \nabla X_s^{(k)} \left(\nabla X_s^{(k)}\right)^\top - \nabla^2 X_s^{(k)}\right). \end{aligned} \tag{2.7}$$

The update $\boldsymbol{\theta}^{(k)}$ can be interpreted as gradient decent in direction of the estimated gradient of the log-likelihood. It is recommended to check that the this

update really improves the likelihood, that is, $L(\boldsymbol{\theta}^{(k)}) < L(\boldsymbol{\theta}^{(k-1)})$. If this inequality does not hold, the step in the gradient direction should be taken smaller, $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)} + \rho(B^{(k)})^{-1}S^{(k)}$ for some $\rho < 1$, e.g. $\rho = 1/2$ and checked again.

The constrains $\alpha \geq 0$, $\omega \geq 0$, $\beta \geq 0$ and $\alpha + \beta < 1$ can be naturally incorporated in the Newton-Raphson procedure using a barrier function. We omit the details.

3 Varying coefficient GARCH

Having the problems mentioned in the introduction in mind, we aim to extend the GARCH approach by including a possibility for structural changes. This can be done using the notion of a varying coefficient model. Namely, we assume that the GARCH parameters may depend on time t. We denote them as $\boldsymbol{\vartheta}_t = (\omega_t, \alpha_t, \beta_t)^\top$. Two special cases are usually considered in the literature. For change point models, the parameters change spontaneously at some time points and remain constant between them, see e.g. Chu (1995) and Mikosch and Stărică (2002). Smooth transition models assume that the parameters vary slowly and smoothly in time, cf. Fan, Jiang, Zhang and Zhou (2003). We do not assume any special dependence of the GARCH-parameters on time, in particular, our modeling approach applies to both change point and smooth transition models. Moreover, our approach applies even if $\boldsymbol{\vartheta}_t$ is a predictable random process. The varying coefficient GARCH(1,1) reads as follows:

$$R_t \sim \phi(\cdot, X_t), \qquad X_t = \omega_t + \alpha_t R_{t-1}^2 + \beta_t X_{t-1} = \Psi_t \boldsymbol{\vartheta}_t \tag{3.1}$$

where $\Psi_t = (1, R_{t-1}^2, X_{t-1})$ and $\boldsymbol{\vartheta}_t$ is now the vector composed by the elements ω_t, α_t and β_t. Each of them may vary with time t.

The target of the analysis is the parameter process $\Theta = (\boldsymbol{\vartheta}_t)_{t \geq t_0}$. This process uniquely defines the process $X = X(\Theta)$ due to (3.1), and hence, the distribution of the process $(R_t)_{t \geq t_0}$. Similarly to the parametric case, we define the (quasi) maximum likelihood estimate of the process Θ by maximizing the corresponding log-likelihood expression

$$L(\Theta) = \sum_{s \geq t_0} \ell(R_s, X_s(\Theta)). \tag{3.2}$$

The maximization is done over the class of all "admissible" processes Θ. Two examples of such classes have been already mentioned: change point models assume that the process Θ is piecewise constant while smooth transition models are effectively based on the smoothness assumption of this process. Our approach is more general and it includes these two examples as special cases. The only assumption we make about the process Θ is *local time homogeneity*. This means that for every time point t the parameter vector $\boldsymbol{\vartheta}_s$ is nearly constant within some neighborhood of the point t. To state this assumption in a more formal way, we need to explain how a local neighborhood of a point t can be described. Similarly to Polzehl and Spokoiny (2000, 2002, 2003) we apply *localization by weights*. Let, for a fixed t, a nonnegative weight $w_{t,s} \in [0,1]$ be assigned to the observation Y_s. The collection of weights $W_t = (w_{t,s})_{s \geq t_0}$ describes a local model corresponding to the point t.

We mention two examples of choosing the weights $w_{t,s}$. *Localization by a bandwidth* is defined by weights of the form $w_{t,s} = K_{\mathrm{loc}}(\boldsymbol{l}_{t,s})$ with $\boldsymbol{l}_{t,s} = |(t-s)/h|^2$ where h is a bandwidth and K_{loc} is a *location kernel*. This method is applied e.g. in Fan, Jiang, Zhang and Zhou (2003). *Localization by a window* simply means that the parametric structure is assumed to hold within some subset (window) U_t containing t. In this case the weights are defined as $w_{t,s} = \mathbf{1}(s \in U_t)$. This approach suits well to change point models where the parameter $\boldsymbol{\vartheta}$ is a piecewise constant function of t.

Following to the *adaptive weights* idea from Polzehl and Spokoiny (2000), we do not assume any special structure for the weights $w_{t,s}$. The weights will be computed from the data in a data driven way.

We apply a local perturbation approach to maximize the log likelihood $L(\Theta)$ from (3.2). This means that we change the process Θ locally near every point t and obtain the local estimation equation by maximizing $L(\Theta)$ for such local perturbations. Before we discuss this method in detail, it is important to note that, even if the parameter process Θ is changed only locally around some point t, the corresponding process $X_s(\Theta)$ changes for all $s > t$. This requires to consider the global log-likelihood even if the parameters are only locally perturbed.

Suppose that a process $\Theta^\circ = (\boldsymbol{\vartheta}_t^\circ)$ is fixed. This process can be viewed as starting value or preliminary estimate of the true process $\Theta = (\boldsymbol{\vartheta}_t)$. Let now W_t be a collection of weights $(w_{t,s})_{s \geq t_0}$ describing a local model at a point t. We

define for every value $\boldsymbol{\theta}$ a locally perturbed process $\widetilde{\Theta} = (\widetilde{\boldsymbol{\vartheta}}_s)$ as

$$\widetilde{\boldsymbol{\vartheta}}_s = w_{t,s}\boldsymbol{\theta} + (1 - w_{t,s})\boldsymbol{\vartheta}^\circ_s, \qquad \forall s \geq t_0.$$

The corresponding latent process denoted by $X_{t,s}(\boldsymbol{\theta}) = X_{t,s}(W_t, \boldsymbol{\theta}; \Theta^\circ)$, $s \geq t_0$, fulfills the equation

$$\begin{aligned} X_{t,s}(\boldsymbol{\theta}) &= \Psi_{t,s}(\boldsymbol{\theta})\,(w_{t,s}\boldsymbol{\theta} + (1 - w_{t,s})\boldsymbol{\vartheta}^\circ_s) \\ &= (w_{t,s}\omega + (1 - w_{t,s})\omega^\circ_s) + (w_{t,s}\alpha + (1 - w_{t,s})\alpha^\circ_s)\, Y_{s-1} \\ &\quad + (w_{t,s}\beta + (1 - w_{t,s})\beta^\circ_s)\, X_{t,s-1}(\boldsymbol{\vartheta}) \end{aligned} \tag{3.3}$$

where $\Psi_{t,s}(\boldsymbol{\theta}) = (1, Y_{s-1}, X_{t,s-1}(\boldsymbol{\theta}))$.

The updated value $\widehat{\boldsymbol{\vartheta}}_t$ of the process Θ at t is defined by maximizing the (quasi) likelihood expression corresponding to the process $X_{t,s}(\boldsymbol{\theta})$:

$$\widehat{\boldsymbol{\vartheta}}_t = \operatorname*{argsup}_{\boldsymbol{\theta}} L(W_t, \boldsymbol{\theta}, \Theta^\circ) = \operatorname*{argsup}_{\boldsymbol{\theta}} \sum_{s \geq t_0} \ell\big(Y_s, X_{t,s}(\boldsymbol{\theta})\big)\,. \tag{3.4}$$

As in the parametric case, the corresponding estimate $\widehat{\boldsymbol{\vartheta}}_t$ solves the equation

$$\sum_{s \geq t_0} \nabla X_{t,s}(\boldsymbol{\theta})\big(Y_s - X_{t,s}(\boldsymbol{\theta})\big)|X_{t,s}(\boldsymbol{\theta})|^{-2} = 0.$$

A numerical solution of this equation can be obtained by the Newton-Raphson procedure as described in Section 2. The definition of the process $X_{t,s}(\boldsymbol{\theta})$ in (3.3) leads to the following expression for the derivatives $\nabla X_{t,s}(\boldsymbol{\theta})$:

$$\begin{aligned} \nabla X_{t,s}(\boldsymbol{\theta}) &= w_{t,s}\Psi^\top_{t,s}(\boldsymbol{\theta}) + (w_{t,s}\boldsymbol{\theta} + (1 - w_{t,s})\boldsymbol{\vartheta}_s)\,\nabla\Psi_{t,s}(\boldsymbol{\theta}) \\ &= w_{t,s}\Psi^\top_{t,s}(\boldsymbol{\theta}) + (w_{t,s}\beta + (1 - w_{t,s})\beta^\circ_s)\,\nabla X_{t,s-1}(\boldsymbol{\theta}), \end{aligned} \tag{3.5}$$

with the starting conditions $\nabla X_{t,s}(\boldsymbol{\theta}) = 0$ for $s \leq t_0$. A similar recurrence formula applies for the matrix of second derivatives:

$$\nabla^2 X_{t,s}(\boldsymbol{\theta}) = w_{t,s}\nabla\Psi_s(\boldsymbol{\theta}) + w_{t,s}\nabla^\top\Psi_{t,s}(\boldsymbol{\theta}) + \big(w_{t,s}\beta + (1 - w_{t,s})\beta^\circ_s\big)\nabla^2 X_{t,s-1}(\boldsymbol{\theta}). \tag{3.6}$$

We can proceed exactly as in the parametric case described in Section 2.

The AWS procedure presented in the next section combines this method of estimating the process Θ with an approach for defining the weights $w_{t,s}$.

4 Adaptive weights smoothing

This section presents an estimation method for a varying coefficient GARCH given by (3.1). The underlying idea is to maximize the log likelihood $L(\Theta)$ from (3.2) in an iterative way. At every step we first describe in a data driven way a neighborhood of every point t in which the varying coefficient model (3.1) can be well approximated by a model with constant parameter values. We then apply the local perturbation approach to update the estimate of the process Θ as described in the previous section.

More precisely, we start defining at every point t a local model $W_t^{(0)}$ using the classical kernel weights with a very small bandwidth $h^{(0)}$. We then successively repeat two basis steps: for all $t \geq t_0$, we estimate the parameter $\boldsymbol{\vartheta}_t$ for the local model $W_t^{(k)} = (w_{t,s}^{(k)})_{s \geq t_0}$, and then, again for all $t \geq t_0$, we generate new larger local models $W_t^{(k+1)}$ using the obtained estimates $\widehat{\boldsymbol{\vartheta}}_t^{(k)}$, $k = 0, 1, 2 \ldots$.

4.1 Defining weights

The method for assigning weights $w_{t,s}^{(k)}$ which define the local model $W_t^{(k)}$ is the central point of the AWS procedure. As suggested in Polzehl and Spokoiny (2002, 2003), for every pair (t, s), the weight $w_{t,s}^{(k)}$ is defined using two different values: a *location penalty* $\boldsymbol{l}_{t,s}^{(k)}$ and a *statistical penalty* $\boldsymbol{s}_{t,s}^{(k)}$.

The location penalty $\boldsymbol{l}_{t,s}^{(k)} = (|t - s|/h^{(k)})^2$ is deterministic and depends only on the distance between t and s and on the bandwidth $h^{(k)}$ applied at step k. At the beginning of the iteration process, the bandwidth $h^{(0)}$ is taken very small leading to a strong localization. During iteration the bandwidth $h^{(k)}$ grows which relaxes the location penalty and allows to increase every local model. However, this increase is done in an adaptive (data-driven) way by use of the statistical penalty $\boldsymbol{s}_{t,s}^{(k)}$ which measures the difference in the parameter values for the local models $W_t^{(k-1)}$ and $W_s^{(k-1)}$. Following Polzehl and Spokoiny (2002, 2003), this penalty can be defined by the expressions

$$\begin{aligned} \boldsymbol{s}_{t,s}^{(k)} &= T_{t,s}^{(k)}/\lambda \\ T_{t,s}^{(k)} &= L(W_t^{(k-1)}, \widehat{\boldsymbol{\vartheta}}_t^{(k-1)}, \widehat{\Theta}^{(k-1)}) - L(W_t^{(k-1)}, \widehat{\boldsymbol{\vartheta}}_s^{(k-1)}, \widehat{\Theta}^{(k-1)}) \end{aligned}$$

where $\widehat{\Theta}^{(k-1)} = (\widehat{\boldsymbol{\vartheta}}_s^{(k-1)})$ is the estimate of the process $\Theta = (\boldsymbol{\vartheta}_s)$ obtained at the step $k-1$. The value $T_{t,s}^{(k)}$ can be interpreted as the test statistic for testing the

two sample hypothesis $\boldsymbol{\vartheta}_t = \boldsymbol{\vartheta}_s$: indeed, $L(W_t^{(k)}, \widehat{\boldsymbol{\vartheta}}_t^{(k-1)}, \widehat{\Theta}^{(k-1)})$ is the maximum of the log-likelihood $L(W_t^{(k)}, \boldsymbol{\theta}, \widehat{\Theta}^{(k-1)}))$ for the local model $W_t^{(k-1)}$ over all possible $\boldsymbol{\theta}$ and $T_{t,s}^{(k)}$ is defined as the discrepancy between this maximum and the particular value $L(W_t^{(k)}, \boldsymbol{\theta}, \widehat{\Theta}^{(k-1)})$ with $\boldsymbol{\theta} = \widehat{\boldsymbol{\vartheta}}_s^{(k-1)}$ coming from another local model $W_s^{(k-1)}$. The value λ can be treated as a critical value for this test. If the statistical penalty $\boldsymbol{s}_{ij}^{(k)}$ is large, then one can say that there is an empirical evidence that the GARCH parameters $\boldsymbol{\vartheta}$ are different at points s and t.

To reduce the computational effort of the procedure, one may also use the quadratic approximation of the log-likelihood:

$$\widetilde{T}_{t,s}^{(k)} = \bigl(\widehat{\boldsymbol{\vartheta}}_t^{(k-1)} - \widehat{\boldsymbol{\vartheta}}_s^{(k-1)}\bigr)^\top B_t^{(k-1)} \bigl(\widehat{\boldsymbol{\vartheta}}_t^{(k-1)} - \widehat{\boldsymbol{\vartheta}}_s^{(k-1)}\bigr)/2, \tag{4.1}$$

where $B_t^{(k-1)}$ is defined similarly to (2.7) using the weights $w_{t,s}^{(k-1)}$.

Suppose that for the pair (t, s), the penalties $\boldsymbol{l}_{t,s}^{(k)}$ and $\boldsymbol{s}_{t,s}^{(k)}$ have been computed. Polzehl and Spokoiny (2002) suggested to define the new weight $w_{t,s}^{(k)}$ such that the value $w_{t,s}^{(k)}$ is small if any of the penalties is large and that the different penalties act independently. This leads to a definition in form of a product:

$$w_{t,s}^{(k)} = K_{\mathrm{loc}}\bigl(\boldsymbol{l}_{t,s}^{(k)}\bigr) K_{\mathrm{st}}\bigl(\boldsymbol{s}_{t,s}^{(k)}\bigr),$$

where K_{loc} and K_{st} are two kernel functions on the positive semiaxis.

The choice of the initial estimates $\boldsymbol{\vartheta}_t^{(0)}$ is important. At the beginning we set the parameter $\beta_t^{(0)}$ to zero which reduces the GARCH(1,1)-model to ARCH(1). In such a case, the structural equation (3.1) reads $X_t = \omega_t + \alpha_t R_{t-1}^2$ and the value X_t is independent of the values ω_s, α_s for $s \neq t$. Therefore, one can define the starting values $\boldsymbol{\gamma}_t^{(0)} = (\omega_t^{(0)}, \alpha_t^{(0)})$ by optimization of the local log-likelihood

$$L(W_t^{(0)}, \boldsymbol{\gamma}) = \sum_{s \geq t_0} \ell(R_s, \omega + \alpha R_{s-1}^2) w_{t,s}^{(0)} \tag{4.2}$$

w.r.t. $\boldsymbol{\gamma} = (\omega, \alpha)$ where $w_{t,s}^{(0)} = K_{\mathrm{loc}}(|s-t|^2/h_0^2)$.

4.2 The procedure

We now present a formal description of the method. Important ingredients of the procedure are:

- the kernels K_{loc} and K_{st};

- the parameter λ;

- the initial bandwidth $h^{(0)}$, a factor $a > 1$ and the maximal bandwidth $h_{\max}$.

The choice of the parameters is discussed in Section 4.3. The procedure reads as follows:

1. Initialization: For every $t \geq t_0$, define the local model $W_t^{(0)}$ with weights $w_{t,s}^{(0)} = K_l(\boldsymbol{l}_{t,s}^{(0)})$ where $\boldsymbol{l}_{t,s}^{(0)} = \left(|t-s|/h^{(0)}\right)^2$ for all s. Next, set $\widehat{\beta}_t^{(0)} = 0$ and $(\widehat{\omega}_t^{(0)}, \widehat{\alpha}_t^{(0)}) = \operatorname{argmax}_{\boldsymbol{\gamma}=(\omega,\alpha)} L(W_t^{(0)}, \boldsymbol{\gamma})$, see (4.2). Set $k = 1$.

2. Iteration: for every $t = t_0, \ldots, T$

- **Calculate the adaptive weights:** For every point $s \geq t_0$ compute the penalties

$$\begin{aligned} \boldsymbol{l}_{t,s}^{(k)} &= \left(|t-s|/h^{(k)}\right)^2, \\ \boldsymbol{s}_{t,s}^{(k)} &= \lambda^{-1}\left\{L(W_t^{(k-1)}, \widehat{\boldsymbol{\vartheta}}_t^{(k-1)}, \widehat{\Theta}^{(k-1)}) - L(W_t^{(k-1)}, \widehat{\boldsymbol{\vartheta}}_s^{(k-1)}, \widehat{\Theta}^{(k-1)})\right\}. \end{aligned} \tag{4.3}$$

  where $L(W, \theta; \Theta)$ is given by (3.2) and (3.3). Define

$$w_{t,s}^{(k)} = K_{\mathrm{loc}}\left(\boldsymbol{l}_{t,s}^{(k)}\right) K_{\mathrm{st}}\left(\boldsymbol{s}_{t,s}^{(k)}\right)$$

  and $W_t^{(k)} = \left(w_{t,s}^{(k)}\right)_{s \geq t_0}$.

- **Estimate the parameter $\boldsymbol{\vartheta}_t$:** Define the local MLE $\widehat{\boldsymbol{\vartheta}}_t^{(k)}$ as

$$\widehat{\boldsymbol{\vartheta}}_t^{(k)} = \operatorname*{argsup}_{\boldsymbol{\theta} \in \Theta} L(W_t^{(k)}, \boldsymbol{\theta}, \widehat{\Theta}^{(k-1)}). \tag{4.4}$$

3. Stopping: Increase k by 1, set $h^{(k)} = ah^{(k-1)}$. If $h^{(k)} \leq h_{\max}$ continue with step 2. Otherwise terminate.

We denote the total number of iterations by k^*. The final estimates are obtained as $\widehat{\boldsymbol{\vartheta}}_t = \widehat{\boldsymbol{\vartheta}}_t^{(k^*)}$. The value $X_{t,t}^{(k^*)}$ can be naturally viewed as the estimate of the parameter σ_t^2 for the varying coefficient model (3.1).

4.3 Choice of parameters

The parameters of the procedure are selected similarly to Polzehl and Spokoiny (2002). We briefly discuss each of the parameters.

Kernels K_{st} and K_{loc}: The kernels K_{st} and K_{loc} must fulfill $K_{\mathrm{st}}(0) = K_{\mathrm{loc}}(0) = 1$, with K_{st} decreasing and K_{loc} non-increasing on the positive semi-axis. We recommend to take $K_{\mathrm{st}}(z) = e^{-z} I_{\{z \leq 6\}}$. We also recommend to apply a

compactly supported localization kernel K_{loc} to reduce the computational effort of the method. Similarly to Polzehl and Spokoiny (2002) we apply the triangle kernel $K_{\mathrm{loc}}(z) = (1 - z)_+$.

Initial bandwidth $h^{(0)}$, parameter a and maximal bandwidth $h_{\max}$: The starting bandwidth $h^{(0)}$ should be small. In general we select $h^{(0)}$ such that every initial local neighborhood $[t-h^{(0)}, t+h^{(0)}]$ contains sufficiently many design points to obtain an estimate of the parameter $\boldsymbol{\vartheta}_t$.

The parameter a controls the growth of the local neighborhoods. Our default choice is $a = 1.25$. The maximal bandwidth $h_{\max}$ may be very large, e.g. $h_{\max} = T$. However, this parameter can be used to bound the numerical complexity of the procedure. The exponential growth of the bandwidth $h^{(k)}$ ensures that the number of iterations k^* is at most logarithmic in the sample size.

Parameter λ: The most important parameter of the procedure is λ which scales the statistical penalty $\boldsymbol{s}_{t,s}$. Small values of λ lead to overpenalization which may result in unstable performance of the method in a homogeneous situation. Large values of λ result in a loss of adaptivity, i.e. less sensitivity to structural changes. A reasonable way to define the parameter λ for a specific application is based on the condition of free extension, which we also call "propagation condition". This means that in a homogeneous situation, i.e. when the process Θ is constant, the impact of the statistical penalty on the computed weights $w_{t,s}$ is negligible. This would result in a free extension of every local model. If the value $h_{\max}$ is sufficiently large, all the weights $w_{t,s}$ will be close to one at the end of iteration process and every local model will essentially coincide with the global one. Therefore, one can adjust the parameter λ using Monte-Carlo simulations. Simply select the minimal value of λ that still provides a prescribed probability to obtain the global model at the end of iteration process for the homogeneous (parametric) model $\boldsymbol{\vartheta}_t = \boldsymbol{\theta}$. The theoretical justification for such a choice is given in Polzehl and Spokoiny (2002).

Our default choice, obtained by this method, is $\lambda = q_\delta(\chi^2_3)$, that is, the δ-quantile of the χ^2 distribution with 3 degree of freedom, where $\delta = 0.99$.

4.4 Semiparametric modeling

In many situations a reasonable estimate of the parameter β requires a large sample size. This makes a local analysis relatively inefficient. A natural way to solve this problem is a semiparametric approach assuming that the parameter β is constant while the other parameters ω, α may vary with time. The AWS procedure can be easily adjusted to such models. Namely, at every iteration we locally estimate the varying coefficients $\boldsymbol{\gamma} = (\omega, \alpha)^\top$ while the value $\beta = \beta^{(k-1)}$ is kept fixed. Afterwards we update the parameter β. The basic AWS procedure reads exactly as described in Section 4.2. The only difference is that the parameter $\boldsymbol{\vartheta}$ should be replaced by $\boldsymbol{\gamma}$ and in the definition of the process $\Theta^{(k-1)}$ one should apply $\beta^{(k-1)}$ in place of $\beta_t^{(k-1)}$. For updating the parameter β, at the end of the iteration k, define for every vector β the process $X_s^{(k)}(\beta) = X_s^{(k)}(\beta, \Gamma^{(k)})$ with $\Gamma^{(k)} = (\boldsymbol{\gamma}_s^{(k)} = (\omega_s^{(k)}, \alpha_s^{(k)})^\top,\ s \geq 1)$ using the recurrence equation

$$X_s^{(k)}(\beta) = \omega_s^{(k)} + \alpha_s^{(k)} Y_{s-1} + \beta X_{s-1}^{(k)}(\beta).$$

The new estimate $\beta^{(k)}$ maximizes the log-likelihood $L(\beta) = \sum_{s \geq t_0} \ell\big(Y_s, X_s^{(k)}(\beta)\big)$ w.r.t. β. Again, the Newton-Raphson algorithm with the quadratic approximation (2.7) can be used.

4.5 Application to forecasting

The forecasting problem for the model (3.1) can be formulated as follows. Given the observations $R_1, \ldots, R_T$ estimate the value of the latent process X_t for some future point $t = T+j$ for $j \geq 1$, and predict the distribution of future observations R_t. A natural way of solving this problem (at least if the forecast horizon j is not too large) is to model the processes R_t and X_t for $t > T$ from the latest estimated model corresponding to $t = T$.

Let $\widehat{\boldsymbol{\vartheta}} = (\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})$ be $\widehat{\boldsymbol{\vartheta}}_T^{(k^*)}$ and $\widehat{X}_s = X_{T,s}^{(k^*)} = X_s(\widehat{\theta}_T)$ for $s = t_0, \ldots, T$. We then define X_{T+1} as

$$\widehat{X}_{T+1|T} = \widehat{\Psi}_{T+1} \widehat{\boldsymbol{\vartheta}} = \widehat{\omega} + \widehat{\alpha} R_T^2 + \widehat{\beta} \widehat{X}_T\,,$$

where $\widehat{\Psi}_{T+1} = (1, R_T^2, \widehat{X}_T)$. Using the estimate $\widehat{X}_{T+1|T}$ of X_{T+1} we can generate R_{T+1} from a GAussian distribution with variance $\widehat{X}_{T+1|T}$. These two steps, compute X_{T+j} and generate R_{T+j}, can be repeated for $t = T+2, T+3$. In general

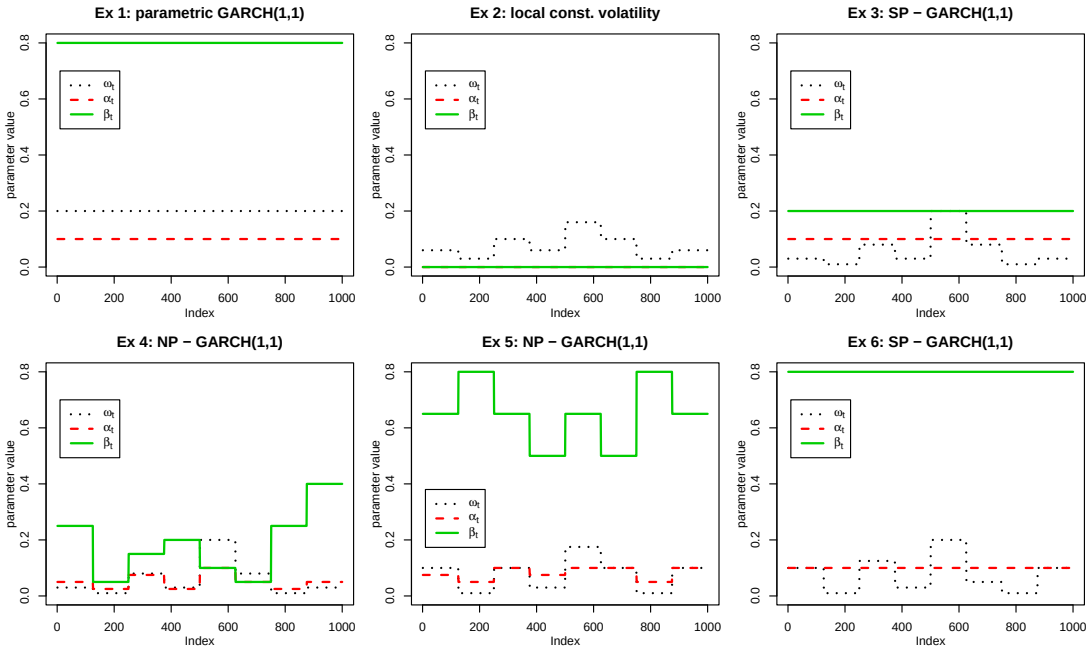


Figure 2: Parameters of simulated examples as functions of time.

there is no closed form expression for the distribution of the forecasted value R_{T+j}, but it can be numerically evaluated by Monte-Carlo simulations.

5 Simulated examples

The aim of this section is to illustrate the performance of the proposed models and compare them with the classical GARCH(1,1) model and the local constant AWS procedure for volatility estimation from Polzehl and Spokoiny (2002). The latter is a very particular and much simpler special case of the varying coefficient GARCH model with $\alpha = \beta = 0$ and only ω varying with time.

We especially focus on the "integrated GARCH" effect (value β close to one) and demonstrate that it can be artificially produced if the stationarity assumption is violated.

We use a set of six artificial examples to illustrate the predictive performance of parametric, non- and semiparametric GARCH(1,1) models. The sample size is set to $n = 1000$. Example 1 is a parametric GARCH(1,1) model with $\omega = 0.2$, $\alpha = 0.1$ and $\beta = 0.8$. Example 2 describes a local constant volatility model ($\alpha = \beta = 0$). Example 3 and 6 are generated as semiparametric GARCH(1,1) models with small and large values of β, respectively, while examples 4 and 5

Table 1: Simulation results for artificial examples 1-6. Simulation size 50. Mean estimated values of β, mean predictive likelihood, probability of exceeding the VaR and mean VaR obtained for the scrolling GARCH(1,1) estimate (from the last 250 observations), sequential AWS for nonparametric and semiparametric GARCH(1,1), and the sequential local constant volatility AWS procedure.

	Ex 1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6
Mean β	0.8	0.0	0.2	0.181	0.65	0.8
Mean $\widehat{\beta}$ GARCH	0.609	0.781	0.802	0.821	0.802	0.804
Mean $\widehat{\beta}$ NP-GARCH	0.558	0.551	0.491	0.520	0.566	0.622
Mean $\widehat{\beta}$ SP-GARCH	0.365	0.258	0.220	0.241	0.325	0.398
PL(10) GARCH	-1.732	1.450	1.313	1.395	0.073	-0.903
PL(10) NP-GARCH	-1.745	1.474	1.397	1.470	0.145	-0.816
PL(10) SP-GARCH	-1.735	1.481	1.449	1.517	0.213	-0.737
PL(10) Local Const	**-1.724**	**1.511**	**1.493**	**1.558**	**0.252**	**-0.710**
$100\widehat{P}_{\mathrm{EVaR}}(0.01, 10)$ GARCH	1.45	1.86	2.70	2.62	2.84	3.00
$100\widehat{P}_{\mathrm{EVaR}}(0.01, 10)$ NP-GARCH	1.54	1.81	2.30	2.27	2.62	2.70
$100\widehat{P}_{\mathrm{EVaR}}(0.01, 10)$ SP-GARCH	**1.31**	1.59	1.97	2.00	2.18	2.29
$100\widehat{P}_{\mathrm{EVaR}}(0.01, 10)$ Local Const	1.38	**1.48**	**1.82**	**1.82**	**2.11**	**2.25**
MVaR$(0.01, 10)$ GARCH	7.21	**2.07**	**2.12**	**2.02**	**3.75**	**6.04**
MVaR$(0.01, 10)$ NP-GARCH	**7.20**	2.10	2.18	2.08	3.82	6.17
MVaR$(0.01, 10)$ SP-GARCH	7.33	2.12	2.22	2.11	3.88	6.32
MVaR$(0.01, 10)$ Local Const	7.28	2.11	2.19	2.09	3.85	6.24

are entirely nonparametric GARCH(1,1) again with small and large values of β. Parameters are local constant and may change every 125 observations. Figure 2 illustrates the parameters used. The AWS estimates are computed sequentially based on all the observations from the past. For the parametric GARCH(1,1) model, a scrolling estimate from the last 250 observations is used.

We use the following criteria to compare the behavior of the estimates:

- Mean estimated value of β

$$\frac{1}{750}\sum_{t>250}\widehat{\beta}_t$$

- A predictive likelihood risk $\mathrm{PL}(k)$ with horizon $k = 10$

$$\mathrm{PL}(k) = -\frac{1}{(n-k-250)k}\sum_{t=251}^{n-k}\sum_{s=1}^{k}\left(\log\widehat{X}_{t+s|t} + \frac{X_{t+s}}{\widehat{X}_{t+s|t}}\right)$$

  where $\widehat{X}_{t+s|t}$ denotes the predicted volatility at time $t+s$ based on the estimated process using observations up to time t.

- Let the Value at Risk (VaR) at level δ and time horizon k be defined as

$$\mathrm{VaR}_t(\delta, k) = -q_\delta \sum_{s=1}^{k} \widehat{X}_{t+s|t} \tag{5.1}$$

with q_δ denoting the δ-quantile of the standard Gaussian distribution. We report an estimate of the mean probability $P_{\mathrm{EVaR}}(\delta, k)$ of exceeding VaR at level δ and time horizon k

$$P_{\mathrm{EVaR}}(\delta, k) = \frac{1}{n-k-250} \sum_{t=251}^{n-k} \boldsymbol{P}\left(\sum_{s=t+1}^{t+k} R_s < -\mathrm{VaR}_t(\delta, k) \right) \tag{5.2}$$

obtained from the simulations. This value should be possibly close to the nominal level δ.

- Finally we provide a mean VaR at level δ and time horizon k as

$$\mathrm{MVaR}(\delta, k) = \frac{1}{n-k-250} \sum_{t=251}^{n-k} \mathrm{VaR}_t(\delta, k) \tag{5.3}$$

again obtained from our simulations, cf. Fan and Gu (2003). This value characterizes the cost required to secure the asset.

Results of the simulations are summarized in Table 1. The results lead to the following conclusions:

- The GARCH model applied to data following a change point GARCH model leads to a misspecification with a large value of the estimated parameter $\widehat{\beta}$.

- The fully nonparametric GARCH model did not succeed to get a reasonable estimate of the varying parameter β. Again, the estimated $\widehat{\beta}_t$ is in mean much larger than the true value in Examples 2 to 4, while the semiparametric GARCH model seems to be much more successful in handling the change point models considered in our examples.

- The local constant model provides the best prediction quality for the 10 days forecasting horizon for all examples. The GARCH model leads to the worst results in almost all examples, while the semiparametric model is typically at the second place.

- The excess probability for the predicted VaR-quantiles is again optimized by the local constant estimate while for examples 5 and 6 the semiparametric model shows slightly better results. However, all the models provide a reasonable fit of the 1%-quantile.

- The averaged value of the VaR-quantile is in most cases minimized by the GARCH-model. In combination with the excess probability results one can judge that the GARCH-model tends to underestimate the VaR. This probably explains why GARCH models are so popular in risk management.

6 Applications to financial time series

We now apply our methodology to two time series, the German DAX index (August 1991 to July 2003) and the USD/GBP exchange rate (January 1990 to December 2000). Similarly to the simulation study, we compare four methods: the parametric GARCH(1,1), the non- and semiparametric GARCH(1,1) models and the local constant volatility model from Polzehl and Spokoiny (2003). We show up to which extend the four methods can explain phenomena observed for financial time series like heavy tails and long range dependence.

We investigate the predictive performance of the methods by estimating the predictive empirical likelihood risk $\mathrm{PEL}(k)$ at different time horizons k ranging from 2 weeks to half a year:

$$\mathrm{PEL}(k) = -\frac{1}{(n-k-500)k} \sum_{t=501}^{n-k} \sum_{s=1}^{k} \left(\log \widehat{X}_{t+s|t} + \frac{R_{t+s}^2}{\widehat{X}_{t+s|t}} \right) \tag{6.1}$$

where $\widehat{X}_{t+s|t}$ denotes the predicted volatility at time $t+s$ based on the estimated process using observations up to time t. We also provide estimates for the excess probability (5.2) of VaR and the mean VaR (5.3).

The top of Figure 3 shows the logarithmic returns of the DAX series, emphasizing strong variations in volatility. Additionally global and sequential estimates of the square root of the volatility obtained by the four methods under consideration are provided. Note that in principle all methods capture the same volatility structure over time. Similar results are observed for the USD/GBP exchange rate series.

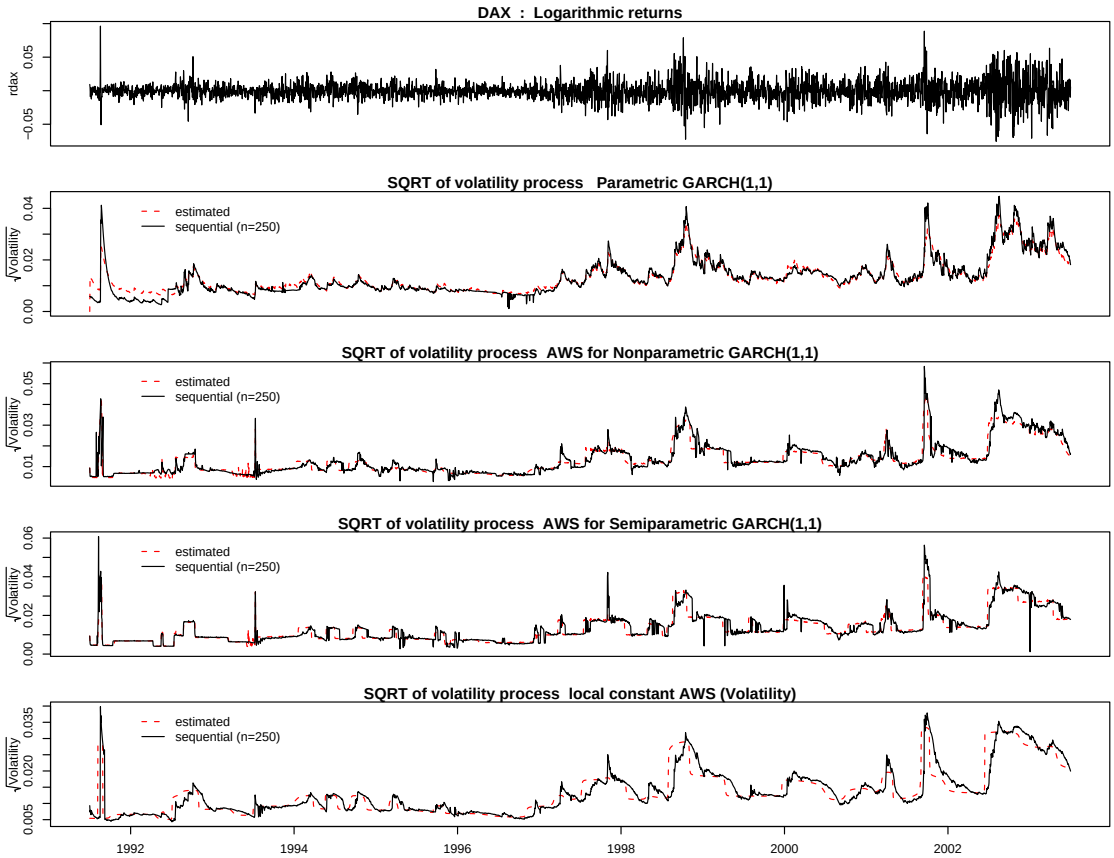


Figure 3: DAX: Logarithmic returns (top) and estimated volatility processes. Given are global estimates (dashed line) and sequential estimates (obtained from the last 500 observations, solid line) by parametric GARCH(1,1), AWS for nonparametric GARCH(1,1), AWS for semiparametric GARCH(1,1) and the local constant volatility model (from top to bottom).

Table 2: DAX and USD/GBP: Mean values for the nonlinear parameter.

	GARCH	NP-GARCH	SP-GARCH
DAX	0.862	0.609	**0.250**
USD/GBP	0.777	0.411	**0.227**

Cointegration in DAX and USD/GBP: fact or artifact?

In Table 2 we provide the mean estimate of the parameter β obtained using the parametric GARCH(1,1) model and its non- and semiparametric generalizations. Exactly as in our simulation study, for both time series, the estimated value of parameter β for the scrolling parametric GARCH(1,1) is close to one, while the results for the semiparametric model (given in boldface) indicate that this IGARCH effect can be artifact of nonstationarity of the time series.

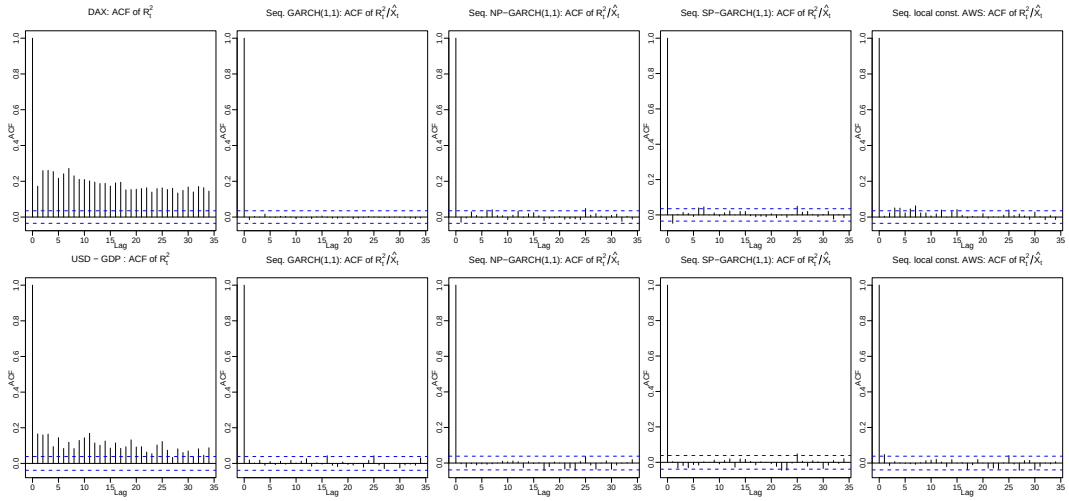


Figure 4: DAX and USD/GBP: ACF of squared log returns and squared standardized residuals (using sequential estimates) obtained for the four methods for DAX (top) and USD/GBP (bottom) volatility estimates, respectively.

Table 3: DAX and USD/GBP: Tail index of absolute logarithmic returns and standardized residuals (using sequential estimates). Critical values for Gaussian distributions with same sample size: 0.193 (.95), 0.202 (.99).

	log returns	residuals GARCH	residuals NP-GARCH	residuals SP-GARCH	residuals Local Const
DAX	0.324	0.225	0.195	0.190	0.188
USD/GBP	0.310	0.232	0.166	0.148	0.171

DAX and USD/GBP: Persistent ACF and Long Range Dependence Phenomenon

The autocorrelation function (ACF) of squared log returns R_t^2 and of squared standardized residuals $\widehat{\varepsilon}_t^2 = R_t^2/\widehat{\sigma}_t^2$ obtained for the four estimates are provided in Figure 4. The ACF of the log returns clearly indicates persistency, however, all four models under consideration, despite their quite different structure, allow to successfully explain the dependence structure. Hence, the long range dependence phenomenon in financial returns can be easily explained by nonstationarity of the financial market.

DAX and USD/GBP: Tail index behavior of the returns

To investigate the phenomenon of heavy tails we estimate the tail index of logarithmic returns R_t and standardized residuals $\widehat{\varepsilon}_t$ obtained by the four methods.

Table 4: DAX and USD/GBP: Mean predictive empirical likelihood risk for different forecast horizons. The best result for each time horizon in boldface.

Method	two weeks		one month		three months		six months	
	DAX	USD/GBP	DAX	USD/GBP	DAX	USD/GBP	DAX	USD/GBP
GARCH	7.54	9.44	7.42	9.40	7.02	9.31	6.73	9.22
NP-GARCH	7.54	9.31	7.47	9.25	7.28	8.57	7.15	8.47
SP-GARCH	7.53	8.46	7.49	7.65	7.35	7.68	7.26	7.70
Local Const	**7.56**	**9.46**	**7.52**	**9.45**	**7.39**	**9.40**	**7.3**	**9.35**

Table 5: DAX and USD/GBP: Probability to exceed the Value at Risk at 10 trading days. The best result in boldface.

Level	GARCH		NP-GARCH		SP-GARCH		Local Const	
	DAX	USD/GBP	DAX	USD/GBP	DAX	USD/GBP	DAX	USD/GBP
0.01	**0.0118**	0.0173	0.0133	0.0168	0.0129	0.0230	0.0137	**0.0149**
0.05	0.0556	0.0542	0.0551	0.0561	0.0594	0.0571	**0.0480**	**0.0538**

We use the AWS tail index estimate proposed in Polzehl and Spokoiny (2003). Results are provided in Table 3. Note that the estimated parameter for the standard normal random sample of the same size should be below 0.193 with probability 0.95 and below 0.202 with probability 0.99.

The logarithmic returns clearly show heavy tails. The estimated tail index for the standardized residuals is smaller for all methods. Note that the use of the parametric GARCH(1,1) model only partly explains the heavy tail effect while the other methods succeeded to eliminate the heavy tails in the standardized returns.

DAX and USD/GBP: Out-of-sample performance

Table 4 provides estimates of the predictive empirical likelihood risk (6.1) for four different time horizons ranging from two weeks to half a year. We observe, with respect to this criterion, that the local constant forecast significantly improves on the other three methods.

DAX and USD/GBP: Value-at-Risk performance

In Table 5 we provide estimates of the probability to exceed the VaR (5.1), defined at a 1% and 5% level using quantiles of a standard Gaussian distribution. The time horizon is two weeks. One can see that all the methods succeeded in forecasting the VaR-quantiles with, in most cases, best results for the local constant model.

Table 6: DAX and USD/GBP: Value at Risk at 10 trading days. The best result in boldface.

Level	GARCH		NP-GARCH		SP-GARCH		Local Const	
	DAX	USD/GBP	DAX	USD/GBP	DAX	USD/GBP	DAX	USD/GBP
0.01	**0.1021**	0.0402	0.1057	0.0405	0.1070	0.0401	0.1056	**0.0400**
0.05	**0.0722**	0.0284	0.0748	0.0286	0.0757	0.0283	0.0747	**0.0283**

Table 6 provides the mean (over time) VaR (5.3) assigned by the four methods. This value characterizes the cost required to secure the asset. Here all four methods demonstrate a similar performance with a small benefit of using the parametric GARCH(1,1) model for the DAX series and of the local constant modeling for the USD/GBP series.

DAX and USD/GBP: Conclusion

Overall we see an advantage in using the local constant volatility model. It seems preferable with respect to risk management and also provides a better explanation for heavy tails, long range dependence and many other stylized facts of the financial time series.

7 Conclusion and Outlooks

The paper shows that the parametric GARCH(1,1) modeling has serious problems if the assumption of stationarity is violated. In particular, the IGARCH effect in the GARCH(1,1) model seems to be an artifact of nonstationarity. An integrated GARCH performs essentially as an exponential smoothing filter. This yields a very good short term ahead forecasting performance. However, an application of the estimated model to long term prediction is questionable because of possible model misspecification. More arguments and a similar conclusion can be found in Stărică (2004).

Two new procedures are suggested which allow to model the nonstationarity in the observed financial time series via varying coefficient GARCH modeling. The method of estimation of time varying GARCH-models suggested in this paper as an extension of the Adaptive Weights idea from Polzehl and Spokoiny (2003) is very general in nature and can be easily extended to GARCH (p, q), or to EGARCH (p, q) and TGARCH (p, q) models. The both methods demon-

strate a reasonable performance, compared to the parametric GARCH(1,1) model. Especially the semiparametric model can be useful for the analysis of the integrated GARCH effect. However, the simulated results and applications to real data demonstrated that a more simple local constant model delivers better results in term of short time forecasting and applications to risk management.

We do not investigate the asymptotic properties and the rate of estimation delivered by the two proposed procedures. Although some properties can be established similarly to Polzehl and Spokoiny (2002), particularly, the important propagation condition. The main reason is that the obtained numerical results are mostly discouraging and do not motivate a rigorous theoretical study.

The general approach proposed in this paper and based on the adaptive weights idea seems to be applicable to many other models like hidden Markov chains, and can be very powerful in that area. This can be viewed as a topic of further research.

References

[1] Baba, Y., Engle, R.F., Kraft, D.F., and Kroner, K.F. (1990). Multivariate Simultaneous Generalized ARCH". *mimeo*, Department of Economics, University of California,

[2] I. Berkes, L. Horvath and P. Kokoszka (2003). Estimation of the maximal moment exponent of a GARCH(1,1) sequence, *Econometric Theory* **19**, 565–586,

[3] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, **31**, 307-327.

[4] Chu, C.-S.J. (1995). Detecting Parameter Shift in GARCH Models. *Econometric Reviews*, **14**, 241-266.

[5] Eberlein, E., Prause, K. (2002). The generalized hyperbolic model: financial derivatives and risk measures. In Mathematical Finance-Bachelier Congress 2000, H. Geman, D. Madan, S. Pliska, T. Vorst (Eds.), Springer Verlag, 245–267.

[6] Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. *Econometrica*, **50**, 987-1008.

[7] Engle, R. F. (ed.) (1995). *ARCH, selected readings*, Oxford University Press, Oxford.

[8] Engle, R.F. and Sheppard, K. (2004). Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. Unpublished manuscript.

[9] Fan, J., Jiang, J., Zhang, C. and Zhou, Z. (2003). Time-dependent diffusion models for term structure dynamics and the stock price volatility. *Statistica Sinica*, **13**, 965-992.

[10] Fan, J., Gu, J. (2003). Semiparametric estimation of value at risk, *Econometrics Journal*, **6**, 260–289.

[11] Giraitis, L. and Robinson, P.M. (2001). Whittle estimation of ARCH models. *Econometric Theory*, **17**, 608-631.

[12] McNeil, A. J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *J. Empirical Finance* **7**: 271–300.

[13] Mercurio, D. and Spokoiny, V. (2004a). Statistical inference for time-inhomogeneous volatility models. *Ann. Statist.*, **32** (2004), no.2, 577-602.

[14] Mercurio, D. and Spokoiny, V. (2004b). Estimation of time dependent volatility via local change point analysis with applications to Value-at-Risk, WIAS-Preprint No. 904, 2004.

[15] Mikosch, T. and Stărică, C. (2000). Is it really long memory we see in financial returns? Extremes and Integrated Risk Management, Ed. P. Embrechts, Risk Books, (2000).

[16] Mikosch, T. and Stărică, C. (2002). Changes of structure in financial time series and the GARCH model. Unpublished manuscript. `http://www.math.chalmers.se/ starica/publi1.html`

[17] Mikosch, T. and Stărică, C. (2004). Non-stationarities in financial time series, the long range dependence and the IGARCH effects. *The Review of Economics and Statistics* **86** 378-390.

[18] Yao, Q., and Peng, L. (2003). Least Absolute Deviations Estimation for ARCH and GARCH Models. Biometrika **90**, no. 4 (2003), pp. 967-975.

[19] Polzehl, J. and Spokoiny, V. (2000). Adaptive weights smoothing with applications to image segmentation. J. of Royal Stat. Soc., 62, Series B, 335–354.

[20] Polzehl, J. and Spokoiny, V. (2002). Local likelihood modeling by adaptive weights smoothing. Preprint 787. WIAS 2002. `http://www.wias-berlin.de/publications/preprints/787`.

[21] Polzehl, J. and Spokoiny, V. (2003). Varying coefficient regression modeling by adaptive weights smoothing. Preprint 818, WIAS, 2003. `http://www.wias-berlin.de/publications/preprints/818`

[22] Stărică, C. (2004). Is Garch(1,1) as good a model as the accolades of the Nobel prize would imply? Preprint. `http://www.math.chalmers.se/ starica/papersoft.pdf`

[23] Stărică, C., and Granger, C. (2004) Non-stationarities in stock returns. Preprint. `http://www.math.chalmers.se/ starica/0903.pdf`. To appear in *Review of Economics and Statistics*.

[24] Straumann, D. and Mikosch, T. (2003) Quasi-MLE in heteroscedastic times series: a stochastic recurrence equations approach. Preprint. `http://www.math.ku.dk/ mikosch/preprint.html`

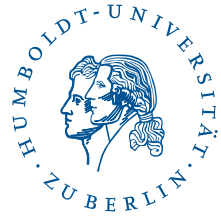
SFB 649 Discussion Paper Series 2006

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Calibration Risk for Exotic Options" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
002 "Calibration Design of Implied Volatility Surfaces" by Kai Detlefsen and Wolfgang K. Härdle, January 2006.
003 "On the Appropriateness of Inappropriate VaR Models" by Wolfgang Härdle, Zdeněk Hlávka and Gerhard Stahl, January 2006.
004 "Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification" by Harald Uhlig, January/February 2006.
005 "British Interest Rate Convergence between the US and Europe: A Recursive Cointegration Analysis" by Enzo Weber, January 2006.
006 "A Combined Approach for Segment-Specific Analysis of Market Basket Data" by Yasemin Boztuğ and Thomas Reutterer, January 2006.
007 "Robust utility maximization in a stochastic factor model" by Daniel Hernández–Hernández and Alexander Schied, January 2006.
008 "Economic Growth of Agglomerations and Geographic Concentration of Industries - Evidence for Germany" by Kurt Geppert, Martin Gornig and Axel Werwatz, January 2006.
009 "Institutions, Bargaining Power and Labor Shares" by Benjamin Bental and Dominique Demougin, January 2006.
010 "Common Functional Principal Components" by Michal Benko, Wolfgang Härdle and Alois Kneip, Jauary 2006.
011 "VAR Modeling for Dynamic Semiparametric Factors of Volatility Strings" by Ralf Brüggemann, Wolfgang Härdle, Julius Mungo and Carsten Trenkler, February 2006.
012 "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms" by Carsten Trenkler, February 2006.
013 "Penalties and Optimality in Financial Contracts: Taking Stock" by Michel A. Robe, Eva-Maria Steiger and Pierre-Armand Michel, February 2006.
014 "Core Labour Standards and FDI: Friends or Foes? The Case of Child Labour" by Sebastian Braun, February 2006.
015 "Graphical Data Representation in Bankruptcy Analysis" by Wolfgang Härdle, Rouslan Moro and Dorothea Schäfer, February 2006.
016 "Fiscal Policy Effects in the European Union" by Andreas Thams, February 2006.
017 "Estimation with the Nested Logit Model: Specifications and Software Particularities" by Nadja Silberhorn, Yasemin Boztuğ and Lutz Hildebrandt, March 2006.
018 "The Bologna Process: How student mobility affects multi-cultural skills and educational quality" by Lydia Mechtenberg and Roland Strausz, March 2006.
019 "Cheap Talk in the Classroom" by Lydia Mechtenberg, March 2006.
020 "Time Dependent Relative Risk Aversion" by Enzo Giacomini, Michael Handel and Wolfgang Härdle, March 2006.
021 "Finite Sample Properties of Impulse Response Intervals in SVECMs with Long-Run Identifying Restrictions" by Ralf Brüggemann, March 2006.
022 "Barrier Option Hedging under Constraints: A Viscosity Approach" by Imen Bentahar and Bruno Bouchard, March 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".



023 "How Far Are We From The Slippery Slope? The Laffer Curve Revisited" by Mathias Trabandt and Harald Uhlig, April 2006.
024 "e-Learning Statistics – A Selective Review" by Wolfgang Härdle, Sigbert Klinke and Uwe Ziegenhagen, April 2006.
025 "Macroeconomic Regime Switches and Speculative Attacks" by Bartosz Maćkowiak, April 2006.
026 "External Shocks, U.S. Monetary Policy and Macroeconomic Fluctuations in Emerging Markets" by Bartosz Maćkowiak, April 2006.
027 "Institutional Competition, Political Process and Holdup" by Bruno Deffains and Dominique Demougin, April 2006.
028 "Technological Choice under Organizational Diseconomies of Scale" by Dominique Demougin and Anja Schöttner, April 2006.
029 "Tail Conditional Expectation for vector-valued Risks" by Imen Bentahar, April 2006.
030 "Approximate Solutions to Dynamic Models – Linear Methods" by Harald Uhlig, April 2006.
031 "Exploratory Graphics of a Financial Dataset" by Antony Unwin, Martin Theus and Wolfgang Härdle, April 2006.
032 "When did the 2001 recession *really* start?" by Jörg Polzehl, Vladimir Spokoiny and Cătălin Stărică, April 2006.
033 "Varying coefficient GARCH versus local constant volatility modeling. Comparison of the predictive power" by Jörg Polzehl and Vladimir Spokoiny, April 2006.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".