CHAPTER 8

Estimating and Forecasting GARCH Models in the Presence of Structural Breaks and Regime Switches

Eric Hillebrand\textsuperscript{a} and Marcelo C. Medeiros\textsuperscript{b}

\textsuperscript{a}Department of Economics, Louisiana State University, Baton Rouge, LA 70803, USA
E-mail address: erhil@lsu.edu

\textsuperscript{b}Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil
E-mail address: mcm@econ.puc-rio.br

Abstract

In this chapter, we outline the statistical consequences of neglecting structural breaks and regime switches in autoregressive and GARCH models and propose two strategies to approach the problem. The first strategy is to identify regimes of constant unconditional volatility using a change point detector and estimate a separate GARCH model on the resulting segments. The second approach is to use a multiple-regime GARCH model, such as the Flexible Coefficient GARCH (FCGARCH) specification, where the regime-switches are governed by an observable variable. We apply both alternatives to an array of financial time series and compare their forecast performance.

Keywords: Volatility, GARCH models, multiple regimes, structural breaks, change-point detection, finance, asymmetry, forecasting

JEL classifications: C22, C53

1. Introduction

Upon casual observation of stock market indices like the Dow Jones or S&P 500, it seems that since around early 2004, the stock market has entered a tranquil phase again. This is after a long period of high volatility that lasted from the onset of the Asian crisis in 1997 until 2003. Financial econometrics discusses three main routes to account for this phenomenon of volatility clustering. One model class comprises highly persistent but stationary models, like the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Engle, 1982; Bollerslev, 1986) or stochastic volatility models with high persistence (Taylor, 1986). Another class consists of models that are somewhat hybrid with regard...
to their stationarity properties, like the Integrated GARCH (IGARCH) process
(Engle and Bollerslev, 1986; Nelson, 1990), which is strictly but not weakly sta-
tionary, or long memory models (Granger and Joyeux, 1980; Ding et al., 1993;
Baulie et al., 1996). The third approach is the development of models with time-
varying parameters. Within each segment of constant parameters (or regime),
the local process is stationary, but the global model is non-stationary due
to the changing regimes. These regimes introduce apparent high persistence
(or long-range dependence) if models are estimated on the full sample with-
out accounting for the parameter changes (Diebold, 1986; Hillebrand, 2005;
Mikosch and Stărică, 2004). The most popular models for switching volatility
regimes have abrupt regime transitions that are governed by a Markov
chain (Hamilton and Susmel, 1994; Cai, 1994; Gray, 1996), hard thresholds
(Rabemananjara and Zakoian, 1993; Li and Li, 1996; Liu et al., 1997), or display
smooth transitions (Hagerud, 1997; Gonzalez-Rivera, 1998).

Recently, Hyung et al. (2008) (in Chapter 9) and Davidson and Sibbert-
tsen (2005) discuss the possible sources of long memory in financial volatility,
mainly structural breaks and aggregation. Hyung et al. (2008) show that a myr-
iad of nonlinear models that display short memory locally, especially models
with infrequent breaks, can generate data with long memory behavior. Examples
of such nonlinear models include the break model of Granger and Hyung
(2004), the volatility component model of Engle and Lee (1999), and the regime
switching model proposed by Hamilton and Susmel (1994), and discussed fur-
ther in Diebold and Inoue (2001).

In this study, we will outline the statistical consequences of neglecting struc-
tural breaks in autoregressive and GARCH models. A popular remedy of the
problem is to identify regimes of constant unconditional volatility using a
change point detector such as the statistic proposed by Kokoszka and Leipus
(1999, ?), and estimate a GARCH process on the resulting segments. A model
that circumvents this two-step procedure is the Flexible Coefficient GARCH
(FCGARCH) model of Medeiros and Veiga (2004), where the regime-switches
are governed by an observable variable. Contrary to a locally stationary ap-
proach and novel to this model, it allows for local non-stationary regimes while
the time series remains globally stationary. We will apply this model to an ar-
ray of financial time series and compare its forecast performance to that of the
segment-wise GARCH estimation.

This chapter is organized as follows. In Section 2, we describe the statisti-
cal consequences of neglecting parameter changes in autoregressive (AR) and
GARCH models. We show that changes in the unconditional mean introduce a
bias in the estimation of AR models and, even more severely, GARCH models.
In Section 3, the FCGARCH(1, 1) model is presented as a model that allows
for changing parameters and that circumvents the problem of the bias due to ne-
glected changes. Its main properties are summarized. A forecasting experiment
that compares the FCGARCH(1, 1) model with a segmented GARCH(1, 1)
model is considered in Section 4. Section 5 concludes the chapter.
2. The statistical consequences of neglecting parameter changes

In this section, we will describe the geometry of autoregressive model estimation in the presence of parameter changes that are not accounted for. Autoregressive models, and in particular GARCH, exhibit a distinct bias in this case. This makes it paramount to formulate a model that can accommodate parameter changes, to which we will turn in the next section.

Consider the autoregressive model of order $p$:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

where $\epsilon_t$ is some error process; for example, zero-mean white noise with variance $\sigma^2$. It is well known that these models exhibit spurious high persistence when they are estimated on data that were generated by a process that displays changing means plus some random zero-mean error (Chen and Tiao, 1990; Perron, 1989). That is, the sum $\sum_{j=1}^{p} \hat{\phi}_j$ of the estimated autoregressive parameters, a common measure for persistence (Andrews and Chen, 1994), is close to one in this case, regardless of the data-generating values of the parameters $\phi_j$.

Hillebrand (2005, 2006) shows that this is a general phenomenon in all autoregressive models, including GARCH, and that it occurs for changes in all parameters that change the unconditional mean. We will briefly give some intuition for this result.

Figure 1 shows the situation for an AR(1) model that is estimated on data that contain a switch in the constant from $c_1 = 1$ to $c_2 = 15$ halfway through the sample of 500 points. The slope stays constant at $\phi = 0.20$ and the standard deviation is 3. The change in the unconditional mean from $\mu_1 = c_1/(1-\phi) = 1.25$ to $\mu_2 = c_2/(1-\phi) = 18.75$ determines the geometry of the estimation problem.

If there is no change point, the estimation can capture the data-generating slope. In the case of a parameter change, the alignment of the two different means $\mu_1$ and $\mu_2$ along the identity in the space $(y_{t-1}, y_t)$ causes the estimator of the slope parameter to pick up the slope of the identity. This readily extends to the case of several change points and higher-order autoregressive models. Note that the same change in the unconditional means could have been generated by a switch in the autoregressive parameter from $\phi_1 = 0.2$ to $\phi_2 = 0.9467$ and letting $c = 1$ on the entire sample.

Two main forces are at work: the absolute difference $|\mu_1 - \mu_2|$ in the means and the variance $\sigma^2$ of the error process. In the lower panel of Figure 1, we repeat the experiment using exactly the same data-generating parameters except that the standard deviation is now 9. The large variance of the error process blurs out the difference in the means and allows the estimation to get closer to the data-generating slope. If the difference in the means largely dominates the variance, the regression equation

$$y_t = X \hat{\beta} + \hat{\epsilon}_t,$$

where $X = [1 \ y_{t-1}], y_t = (y_2, \ldots, y_T)'$, $1 = (1, \ldots, 1)'$, $\epsilon_t = (\epsilon_2, \ldots, \epsilon_T)$, and $\beta = (c, \phi)'$, becomes in the limiting case as $T \to \infty$ (the break fraction is
Fig. 1. AR(1) model estimated on data generated with a change in the constant $c$ from 1 to 15. The data-generating slope is 0.20 throughout.

$(1) \quad \hat{\epsilon}_t = y_t - \hat{\mu} - \hat{\phi} y_{t-1}$

$(2) \quad y_t = X \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \hat{\epsilon}_t = y_{t-1} + \hat{\epsilon}_t.$

Obviously, in this case $\hat{\epsilon}_t$ cannot capture the data generating disturbance $\epsilon_t$. However, taking expectations, (2) holds on both segments and $\sum_{t=1}^T \hat{\epsilon}_t / T = 0$.

In order to show that in the case of a neglected change point (2) holds with a zero-mean error $\hat{\epsilon}_t$, we need to show that in this case

$y_t = X(X'X)^{-1}X'y_{t-1} + \hat{\epsilon}_t = y_{t-1} + \hat{\epsilon}_t.$

Denote the sizes of the two segments of constant parameters by $T_1$ and $T_2$, $T_1 + T_2 = T$. Using the representation

$y_t = \mu + \sum_{j=0}^{t-1} \phi^j \epsilon_{t-j} + \phi^i y_0 - c\phi^{i-1}/(1 - \phi)$

and ignoring initial terms, we can write $y_t = \mu + \eta_t$, where $\eta_t$ is an MA($t-1$) process in the error $\epsilon_t$. Thus, in the case of a change point,

$y_t = \begin{bmatrix} \mu_1 + \eta_{2,T_1} \\ \mu_2 + \eta_{T_1+1,T} \end{bmatrix}$, $\begin{bmatrix} \mu_1 + \eta_{1,T_1} \\ \mu_2 + \eta_{T_1+1,T-1} \end{bmatrix}$. 

$X = \begin{bmatrix} 1 & \mu_1 + \eta_{1,T_1} \\ 1 & \mu_2 + \eta_{T_1+1,T-1} \end{bmatrix}$.
where $\mu_i = (\mu_{i1}, \ldots, \mu_{ir})'$ and $\eta_{t1}, \ldots, \eta_{t2}$ are centered to zero. The two vectors on the right-hand side of $y_t$ are of size $T_1 - 1$ and $T_2$, respectively. The two block matrices $[1 \; \mu_i + \eta_i]$ on the right-hand side of $X$ are of size $T_1 \times 2$ and $(T_2 - 1) \times 2$, respectively. Then, tedious but straightforward calculation shows that the projection matrix is given by

$$
X(X'X)^{-1}X' = \frac{1}{\det X'X} \left[ \begin{array}{c}
T_2 (\mu_2^2 - \mu_1 \mu_2)' + T_2 (\mu_1 - \mu_2) \mu_1' + f_1(\eta) \\
T_1 (\mu_1^2 - \mu_1 \mu_2)' + T_1 (\mu_2 - \mu_1) \mu_2' + f_2(\eta)
\end{array} \right],
$$

where $\det X'X = (T - 1)(T_1 \mu_1^2 + T_2 \mu_2^2 + \sum \eta_t^2) - (T_1 \mu_1 + T_2 \mu_2 + \sum \eta_t)^2$, and that

$$
X(X'X)^{-1}X'y_{t-1} = X(X'X)^{-1}X' \left[ \begin{array}{c}
\mu_1 + \eta_{2,T1} \\
\mu_2 + \eta_{1,T+1,T}
\end{array} \right] = \left[ \begin{array}{c}
\mu_1 + g_1(\eta) \\
\mu_2 + g_2(\eta)
\end{array} \right].
$$

Here, the multiplications in $\mu_2^2$ and $\mu_1 \mu_2$ are element-wise. The $f_i(\eta)$ and $g_i(\eta)$ are functions in the MA process through which the variance of the error process influences the locus of the estimation line. They determine how far the estimation line is pushed away from the identity and towards the data-generating slope. The larger the difference in the means relative to the variance, the smaller the variance tends to zero, the $f_i$ and $g_i$ tend to zero and (2) holds exactly with an error of $\hat{\epsilon}_t = 0$. With a non-zero variance, $\hat{\epsilon}_t = (g_1(\eta) - \eta_{1,T-1}, g_2(\eta) - \eta_{T1,T-1})'$ contains squared terms in $\epsilon_t$. Taking expectations, these result in functions of $\sigma^2$. For a rigorous derivation of these functions see Hillebrand (2006).

Contrary to AR($p$) models for the first moment, the GARCH($p$, $q$) model of Engle (1982) and Bollerslev (1986) specifies an autoregressive structure for the second moment:

$$
\log S_t - \log S_{t-1} = f(t; \mathbf{b}) + \epsilon_t, \quad (3)
$$

$$
\epsilon_t | F_{t-1} \sim N(0, h_t), \quad (4)
$$

$$
h_t = \omega + \sum_{j=1}^q \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^p \beta_j h_{t-j}. \quad (5)
$$

Here $S_t$ is the price of a financial asset, $f(t; \mathbf{b})$ is some conditional mean function with parameter $\mathbf{b}$, $F_t$ is the information set, and $\sum_{j=1}^q \alpha_j + \sum_{j=1}^p \beta_j < 1$ is the stationarity condition. A GARCH model is usually estimated by maximum likelihood, so the projection matrix considerations above do not carry over directly, but the geometry of the situation is the same, as shown in Figure 2. Compared to AR models, the bias problem is exacerbated in GARCH models because the stochasticity of the conditional volatility equation (5) originates in the $\epsilon_{t-j}^2$. 


Fig. 2. The geometry of neglected change points in the GARCH(1, 1) model. The top left panel shows a simulated GARCH(1, 1) variance series with a change in the constant $\omega$. The top right panel shows the estimated GARCH(1, 1) variance series using the data from the top left panel and ignoring the change point. The bottom left panel shows a simulated GARCH(1, 1) series without change point. The bottom right panel shows the estimated GARCH(1, 1) variance series for the German stock index DAX.

term only. There is no contemporaneous error that is orthogonal to the regressors $\epsilon_{t-1}^2$ and $h_{t-1}$. Therefore, in GARCH processes we do not find the interplay of the distance in the unconditional local means $\mu_i = \omega_i / (1 - \sum_j \alpha_i,j - \sum_j \beta_i,j)$ with the variance of an orthogonal error process. As a result, the estimation of persistence in GARCH processes is much more sensitive to neglected change points than in simple autoregressive models.

The top left panel of Figure 2 shows synthetically generated GARCH(1, 1) conditional volatility data $(h_t)$ with a change in the unconditional mean in $(\epsilon_{t-1}^2, h_{t-1}, h_t)$ space. The two regimes are distinct. In both regimes, the sum $\alpha + \beta$ of the data-generating autoregressive parameters is 0.80. The top right panel shows the estimated conditional volatility $\hat{h}_t$ when a GARCH(1, 1) model is estimated on the segmented data but the change point is ignored. Again, the alignment of the two unconditional local means on the identity lets the single estimation hyperplane coincide with the identity. The estimated sum $\hat{\alpha} + \hat{\beta}$ of the autoregressive coefficients is 0.98, far from the data-generating value of 0.80.

For comparison, the bottom left panel shows synthetic GARCH(1, 1) data with $\alpha + \beta = 0.98$ and no change-points. At least visually, this case cannot
be distinguished from the case of neglected change points in the top right panel where the two regimes are projected onto a single hyperplane.

The bottom right panel shows the estimated GARCH(1, 1) conditional volatility for the time series of the German stock index DAX described in Section 4.1. It is impossible to tell which case, neglected change points or true high persistence, generated the real data, since the conditional volatility $h_t$ cannot be observed. This has immediate consequences for forecasting since the forecast generated from a locally stationary model with low persistence differs markedly from the forecast of a high persistence model.

3. The flexible coefficient GARCH (1, 1) model

3.1. Overview

As we saw in the previous section, the GARCH model is highly sensitive to changes in the unconditional mean. Therefore, it is desirable to have a model that treats the GARCH parameters as a function of time and allows for regime changes. In this section, we consider the flexible coefficient GARCH model of Medeiros and Veiga (2004). The model is defined as follows.

**DEFINITION 1.** (See Medeiros and Veiga, 2004.) A time series $\{y_t\}$ follows a first-order flexible coefficient GARCH model with $m = H + 1$ limiting regimes, $\text{FCGARCH}(m, 1, 1)$, if

$$y_t = h_t^{1/2} \varepsilon_t,$$

$$h_t = G(w_t; \psi) = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \sum_{i=1}^{H} [\alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2] f(s_t; \gamma_i, c_i), \quad t = 1, \ldots, T,$$

where $\{\varepsilon_t\}$ is a sequence of identically and independently distributed zero mean and unit variance random variables, $\varepsilon_t \sim \text{IID}(0, 1)$, $G(w_t; \psi)$ is a nonlinear function of the vector of variables $w_t = [y_{t-1}, h_{t-1}, s_t]'$, and is indexed by the vector of parameters

$$\psi = [\alpha_0, \beta_0, \lambda_0, \alpha_1, \ldots, \alpha_H, \beta_1, \ldots, \beta_H, \lambda_1, \ldots, \lambda_H, \gamma_1, \ldots, \gamma_H, c_1, \ldots, c_H]',$$

and $f(s_t; \gamma_i, c_i), i = 1, \ldots, H$, is the logistic function defined as

$$f(s_t; \gamma_i, c_i) = \frac{1}{1 + e^{-\gamma_i (s_t - c_i)}}.$$  

It is clear that $f(s_t; \gamma_i, c_i)$ is a monotonically increasing function, such that $f(s_t; \gamma_i, c_i) \to 1$ as $s_t \to \infty$ and $f(s_t; \gamma_i, c_i) \to 0$ as $s_t \to -\infty$. The parameter $\gamma_i, i = 1, \ldots, H$, is called the slope parameter and determines the speed
of the transition between two limiting regimes. The number of limiting regimes is defined by the hyper-parameter $H$. When $\gamma_i \rightarrow \infty$, the logistic function becomes a step function, and the FCGARCH model becomes a threshold-type specification. The variable $s_t$ is known as the transition variable.

There are many possible choices for $s_t$. For example, if $s_t = y_{t-1}$, then we model the differences in the dynamics of the conditional variance according to the sign and size of shocks in past returns, which represent previous “news.” More specifically, suppose that we set $H = 2$ in (6), $c_1$ is highly negative, and $c_2$ is large and positive. Then the resulting FCGARCH model will have 3 limiting regimes that can be interpreted as follows. The first regime may be related to extremely low negative shocks (“very bad news”) and the dynamics of the volatility are driven by $h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2$ as $f(y_{t-1}; \gamma_1, c_1) \approx 0$, $i = 1, 2$. In the middle regime, which represents low absolute returns (“tranquil periods”), $h_t = \alpha_0 + \alpha_1 + (\beta_0 + \beta_1) h_{t-1} + (\lambda_0 + \lambda_1) y_{t-1}^2$ as $f(y_{t-1}; \gamma_1, c_1) \approx 1$ and $f(y_{t-1}; \gamma_2, c_2) \approx 0$. Finally, the third regime is related to high positive shocks (“very good news”) and $h_t = \alpha_0 + \alpha_2 + (\beta_0 + \beta_1 + \beta_2) h_{t-1} + (\lambda_0 + \lambda_1 + \lambda_2) y_{t-1}^2$, as $f(y_{t-1}; \gamma_1, c_1) \approx 1, i = 1, 2$. The speed of the transitions between different limiting GARCH models is determined by the parameter $\gamma_i$, $i = 1, 2$. The more abrupt the transitions ($\gamma_i \gg 0$), the more straightforward is the interpretation of the FCGARCH specification as a multiple regime model.

Other choices of the transition variable are possible. For example, Audrino and Trojani (2003) and Chen et al. (2003) consider past returns on the S&P 500 as possible determinants of asymmetrical behavior in the conditional volatility of financial time series. Long-term returns are also a possibility. Past cumulated returns are another possible choice for transition variables, which may represent long-run “trends” in the markets. However, in this chapter we will focus on the past returns as transition variables only. The reason is twofold: First, in order to capture both size and sign asymmetries as discussed in the literature, the past daily return is the natural choice of transition variable; second, all the theoretical results in Medeiros and Veiga (2004), especially the stationarity condition, are derived under this particular choice of transition variable. The stationarity condition is very important as it allows the FCGARCH model to have unstable regimes. In the empirical examples we claim and provide evidence that this is one of the causes why the estimated persistence in GARCH models is close to one. Deriving new stationarity conditions for other choices of the transition variable is an interesting topic for future research, but it is beyond the scope of this chapter.

To ensure strictly positive variances and the identifiability of the FCGARCH model, Medeiros and Veiga (2004) impose the following set of restrictions:

**Restriction 1 (Identifiability).** The parameters $c_i$ and $\gamma_i$, $i = 1, \ldots, H$, satisfy the conditions:

1. $-\infty < c_1 < \cdots < c_H < \infty$;
2. $\gamma_i > 0$. 

Eric Hillebrand and Marcelo C. Medeiros
RESTRICTION 2 (Strictly positive variances). The parameters $\gamma_i$ and $c_i$, $i = 1, \ldots, H$, and $\alpha_j$, $\beta_j$, and $\lambda_j$, $j = 0, \ldots, H$, are such that

$$f(s_t; \gamma_1, c_1) \geq f(s_t; \gamma_2, c_2) \geq \cdots \geq f(s_t; \gamma_H, c_H), \forall t \in [0, T];$$

(R.3) $\sum_{j=0}^{K} \alpha_j > 0, \forall K = 0, \ldots, H;$

(R.4) $\sum_{j=0}^{K} \beta_j > 0, \forall K = 0, \ldots, H.$

The FCGARCH model has interesting dynamic properties. For the particular case when $s_t = y_{t-1}$, Medeiros and Veiga (2004) show that

$$E \left\{ \log \left[ \left( \beta_0 + \sum_{i=1}^{H} \beta_i f_{i,t-1} \right) + \left( \lambda_0 + \sum_{i=1}^{H} \lambda_i f_{i,t-1} \right) \varepsilon_{t-1}^2 \right] \right\} < 0, \forall t, \quad (8)$$

where $f_{i,t-1} \equiv f(y_{t-1}; \gamma_i, c_i), i = 1, \ldots, H$, is the necessary and sufficient log-moment condition for strict stationarity and ergodicity of the FCGARCH$(m, 1, 1)$ model. The log-moment condition is important, as condition (8) can be satisfied even in the absence of finite second-moments of $y_t$; see McAleer (2005) for a comprehensive discussion of log-moment conditions for volatility models. A sufficient condition for strict stationarity and ergodicity of the FCGARCH$(m, 1, 1)$ model in terms of the parameters is

$$\frac{1}{2} (\beta_0 + \lambda_0) + \frac{1}{2} \sum_{i=0}^{H} (\beta_i + \lambda_i) \leq 1.$$

In addition, a sufficient condition for the existence of the second-order moment of $y_t$ is

$$\frac{1}{2} (\beta_0 + \lambda_0) + \frac{1}{2} \sum_{i=0}^{H} (\beta_i + \lambda_i) < 1. \quad (9)$$

Furthermore, define $\beta_U \equiv \sum_{i=1}^{H} \beta_i$ and $\lambda_U \equiv \sum_{i=1}^{H} \lambda_i$. As shown in Medeiros and Veiga (2004), the fourth-order moment of $y_t$ exists if $E[\varepsilon_t^4] = \mu_4 < \infty$, (9) holds, and

$$\beta_0^2 + \beta_0 \beta_U + \beta_U^2 + \mu_4 \left[ \lambda_0 + \lambda_0 \lambda_U + \frac{\lambda_U^2}{2} \right] + 2\lambda_0 \beta_0 + \beta_0 \lambda_U + \lambda_0 \beta_U + \lambda_U \beta_U < 1. \quad (10)$$

It is important to notice that even with explosive regimes the FCGARCH$(m, 1, 1)$ process may still be strictly stationary, ergodic, and have a finite fourth moment. Furthermore, for some regimes the sum of the autoregressive GARCH parameters may exceed one. This flexibility generates models with higher kurtosis than the standard GARCH$(1, 1)$, even with Gaussian errors. The IGARCH model with Gaussian errors is also capable of generating data with
Fig. 3. Simulated path, conditional variance, and autocorrelation function of a FCGARCH(1, 1) process with 3 regimes, Gaussian errors, and parameters $\alpha_0 = 0.0002, \alpha_1 = -0.00012, \alpha_2 = 0.0001, \lambda_0 = 0.15, \lambda_1 = -0.07, \lambda_2 = 0.01, \beta_0 = 0.95, \beta_1 = -0.25, \beta_2 = 0.1, \gamma_1 = 5000, \gamma_2 = 5000, c_1 = -0.018,$ and $c_2 = 0.015.$

high kurtosis. However, contrary to the FCGARCH model, it does not have a finite fourth moment.

EXAMPLE 1. Consider a FCGARCH(1, 1) process with 3 regimes, Gaussian errors, and parameters $\alpha_0 = 0.0002, \alpha_1 = -0.00012, \alpha_2 = 0.0001, \lambda_0 = 0.15, \lambda_1 = -0.07, \lambda_2 = 0.01, \beta_0 = 0.95, \beta_1 = -0.25, \beta_2 = 0.1, \gamma_1 = 5000, \gamma_2 = 5000, c_1 = -0.018,$ and $c_2 = 0.015.$ The condition for geometric ergodicity and strict stationarity holds given that

$$\frac{1}{2}(\beta_0 + \lambda_0) + \frac{1}{2} \sum_{i=0}^{H} (\beta_i + \lambda_i) = 0.9950,$$

despite the explosiveness of the first regime. Figure 3 shows a simulated path of such a process, its conditional variance, and the corresponding autocorrelation function up to the 30th lag. It is apparent that the simulated model has long-range dependence.

3.2. Parameter estimation

As the distribution of $\varepsilon_t$ is unknown, the parameters of the FCGARCH model are estimated by quasi-maximum likelihood (QML). The quasi-log-likelihood
function of the FCGARCH model is given by

$$L_T(\psi) = \frac{1}{T} \sum_{t=1}^{T} l_t(\psi),$$

where $l_t(\psi) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(h_t) - \frac{y_t^2}{2h_t}$. Under mild regularity conditions, Medeiros and Veiga (2004) prove that

$$T^{1/2}(\hat{\psi}_T - \psi_0) \overset{D}{\to} N(0, A(\psi_0)^{-1}B(\psi_0)A(\psi_0)^{-1}),$$

where

$$A(\psi_0) = \mathbb{E} \left[ \frac{\partial^2 l_{u,t}(\psi)}{\partial \psi \partial \psi'} \bigg| \psi_0 \right]$$

and

$$B(\psi_0) = \mathbb{E} \left[ \frac{T \partial L_{u,T}(\psi)}{\partial \psi} \bigg| \psi_0 \right] \frac{\partial L_{u,T}(\psi)}{\partial \psi'} \bigg| \psi_0 \right].$$

$$= \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \left( \frac{\partial l_{u,t}(\psi)}{\partial \psi} \bigg| \psi_0 \right) \frac{\partial l_{u,t}(\psi)}{\partial \psi'} \bigg| \psi_0 \right).$$

$A(\psi_0)$ and $B(\psi_0)$ are consistently estimated by

$$A_T(\psi) = \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{1}{2h_t^2} \frac{\partial t}{\partial \psi} \frac{\partial h_t}{\partial \psi'} \left( \frac{y_t^2}{h_t} \right) - \left( \frac{y_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \psi'} \left( \frac{1}{2h_t} \frac{\partial h_t}{\partial \psi} \right) \right\}$$

and

$$B_T(\psi) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial l_t(\psi)}{\partial \psi} \frac{\partial l_t(\psi)}{\partial \psi'} = \frac{1}{4T} \sum_{t=1}^{T} \frac{1}{h_t^2} \left( \frac{y_t^2}{h_t} - 1 \right)^2 \frac{\partial h_t}{\partial \psi} \frac{\partial h_t}{\partial \psi}.$$ (13)

### 3.3. Determining the number of regimes

The number of regimes in the FCGARCH model, as represented by the number of transition functions in (6), is not known in advance and should be determined from the data. One possibility is to begin with a small model (such as GARCH(1,1) or white noise) and add regimes sequentially. The decision to add another regime may be based on the use of model selection criteria (MSC) or cross-validation. However, this has the following drawback. Suppose the data have been generated by an FCGARCH model with $m$ regimes ($m-1$ transition functions). Applying MSC to decide whether or not another regime should be added to the model requires estimation of a model with $m+1$ regimes. A typical MSC comparison of the two models is then equivalent to a likelihood ratio test of $m$ against $m+1$ regimes; see Terasvirta and Mellin (1986) for...
a discussion. The choice of MSC determines the (asymptotic) significance level of the test. When the larger model is not identified under the null hypothesis, the likelihood ratio statistic does not have an asymptotic \( \chi^2 \) distribution under the null.

Medeiros and Veiga (2004) tackle the problem of determining the number of regimes of the FCGARCH model with a “specific-to-general” modeling strategy, circumventing the problem of identification in a way that enables us to control the significance level of the tests in the sequence, and compute an upper bound to the overall significance level.

Consider an FCGARCH with \( H \) limiting regimes, defined as

\[
y_t = h_t^{1/2} \epsilon_t,
\]

\[
h_t = \alpha_0 + \beta_0 h_{t-1} + \lambda_0 y_{t-1}^2 + \sum_{i=1}^{H-1} \left[ \alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2 \right] f(y_{t-1}; \gamma_i, c_i).
\]

An additional regime is represented by an extra term in (14) of the form,

\[
\left[ \alpha_H + \beta_H h_{t-1} + \lambda_H y_{t-1}^2 \right] f(y_{t-1}; \gamma_H, c_H).
\]

A convenient null hypothesis is

\[
H_0: \gamma_H = 0,
\]

against the alternative \( H_1: \gamma_H > 0 \), since the logistic function becomes a constant independent of the state variable if \( \gamma_H = 0 \). Note that the model is not identified under the null hypothesis. In order to remedy this problem, we follow Luukkonen et al. (1988) and expand the logistic function \( f(y_{t-1}; \gamma_H, c_H) \) into a first-order Taylor expansion around the null hypothesis \( \gamma_H = 0 \). After merging terms, the resulting model for \( h_t \) is

\[
h_t = \tilde{\alpha}_0 + \tilde{\beta}_0 h_{t-1} + \tilde{\lambda}_0 y_{t-1}^2 + \sum_{i=1}^{H-1} \left[ \alpha_i + \beta_i h_{t-1} + \lambda_i y_{t-1}^2 \right] f(y_{t-1}; \gamma_i, c_i)
\]

\[
+ \pi y_{t-1} + \delta h_{t-1} y_{t-1} + \rho y_{t-1}^3 + R,
\]

where \( R \) is the remainder, \( \tilde{\alpha}_0 = \alpha_0 - \frac{\alpha_H \gamma_H c_H y_0^2}{4}, \) \( \tilde{\beta}_0 = \beta_0 - \frac{\beta_H \gamma_H c_H y_0}{4}, \) \( \tilde{\lambda}_0 = \lambda_0 - \frac{\lambda_H \gamma_H c_H}{4} \), \( \pi = \gamma_H y_0, \) \( \delta = \frac{\beta_H \gamma_H y_0}{4}, \) and \( \rho = \frac{\lambda_H \gamma_H y_0}{4} \).

Under \( H_0 \), \( R = 0 \) and, if \( E[|y_0|] < \infty \), a Lagrange multiplier (LM) type of test can be derived. Although the test statistic is constructed under the assumption of normality, a robust version of the LM test against nonnormal errors is available and can be carried out in stages as follows:

1. Estimate model (6) under the null, call the estimated variance \( \hat{h}_{0,t}, \) and compute \( \text{SSR}_0 = \sum_{t=1}^{T} (y_t^2 / \hat{h}_{0,t} - 1)^2 \).
2. Set

\[
\hat{z}_t = \frac{1}{\hat{h}_{0,t}} \left\{ \hat{s}_t + \sum_{k=1}^{t-1} \left[ \prod_{j=k+1}^{t} \left( \hat{\beta}_0 + \sum_{i=1}^{H-1} \hat{\beta}_i f_{i,j-1} \right) \right] \hat{x}_k \right\},
\]
\[ \hat{\mu}_t = \frac{1}{h_{0,t}} \left\{ \hat{\nu}_t + \sum_{k=1}^{t-1} \left[ \prod_{j=k+1}^{t} \left( \hat{\beta}_0 + \sum_{i=1}^{H-1} \hat{\beta}_i f_{i,j-1} \right) \right] \hat{\nu}_k \right\}, \]

where \( \hat{\nu}_t = [y_{t-1}, \hat{h}_{0,t-1}, y_{t-1}, y_{t-1}^3]' \) and

\[
\hat{x}_t = [1, \hat{h}_{0,t-1}, y_{t-1}^2, f_{1,t-1}, \ldots, f_{H-1,t-1}, f_{1,t-1} \hat{h}_{0,t-1}, f_{1,t-1} y_{t-1}, \ldots, f_{H-1,t-1} y_{t-1}, \hat{\gamma}_1 y_{t-1}^2, \ldots, \]

\[
(\hat{\alpha}_1 + \hat{\beta}_1 h_{0,t-1} + \hat{\lambda}_1 y_{t-1}^2) \frac{\partial f_{1,t-1}}{\partial \gamma_1}, \ldots,
\]

\[
(\hat{\alpha}_H-1 + \hat{\beta}_H-1 h_{0,t-1} + \hat{\lambda}_H-1 y_{t-1}^2) \frac{\partial f_{H-1,t-1}}{\partial \gamma_{H-1}},
\]

\[
(\hat{\alpha}_1 + \hat{\beta}_1 h_{0,t-1} + \hat{\lambda}_1 y_{t-1}^2) \frac{\partial f_{1,t-1}}{\partial c_1}, \ldots,
\]

\[
(\hat{\alpha}_H-1 + \hat{\beta}_H-1 h_{0,t-1} + \hat{\lambda}_H-1 y_{t-1}^2) \frac{\partial f_{H-1,t-1}}{\partial c_{H-1}} \].

Regress \((y_t^2/\hat{h}_{0,t} - 1)\) on \(\hat{z}_t\) and compute the sum of the squared residuals, SSR₁.

3. Regress \(\hat{u}_t\) on \(\hat{z}_t\) and compute the residual vectors, \(\hat{r}_t, t = 1, \ldots, T\).

4. Regress 1 on \((y_t^2/\hat{h}_{0,t} - 1)\hat{r}_t\), and compute the residual sum of squares, SSR. The test statistic given by

\[ LM_R = T - SSR \]  

(16)

has an asymptotic \(\chi^2\) distribution with 3 degrees of freedom under the null hypothesis.

A remark concerning the computational implementation of the test is important. If \(\hat{\gamma}_i, i \in [1, H]\), is very large, the gradient matrix becomes near-singular and the test statistic numerically unstable, which distorts the size of the test. The reason is that the vectors corresponding to the partial derivatives with respect to \(\gamma_i\) and \(c_i\), respectively, tend to be almost perfectly linearly correlated. This is due to the fact that the time series of those elements of the gradient resemble dummy variables that are constant most of the time and nonconstant during the same spells. In those cases, one solution is to omit the terms that depend on the derivatives of the logistic function; see Eitrheim and Terasvirta (1996) for a complete discussion. This can be done without significantly affecting the value of the test statistic.

We are now ready to combine the above statistical ingredients into a practical modeling strategy. We begin by testing linearity against an ARCH(\(q\)) model at significance level \(\delta\). The model under the null hypothesis is a homoskedastic model. If the null hypothesis is not rejected, the homoskedastic model is considered as the data generating process. In case of rejection, a GARCH(1, 1) model...
is estimated and tested against an $\text{FCGARCH}(1, 1, 1)$ model with two or more regimes at the significance level $\delta \varrho$, $0 < \varrho < 1$. Another rejection leads to estimating a model with two regimes and testing it against a model with three or more regimes at the significance level $\delta \varrho^2$. The sequence is terminated at the first non-rejection of the corresponding null hypothesis. The significance level is reduced at each step of the sequence and converges to zero, thereby avoiding excessively large models and controlling the overall significance level. An upper bound for the overall significance level may be obtained using the Bonferroni bound (Gourieroux and Monfort, 1995, p. 203). The selection of the parameter $\varrho$ is ad hoc. In order to avoid selecting small models (few regimes), it is good practice to carry out the modeling cycle with different values of $\varrho$.

REMARK 1. It is important to stress that all the parameters of FCGARCH model are re-estimated when a new regime is added. This avoids possible biases and overestimation of the number of regimes. Although we did not show theoretically that the specific-to-general modeling strategy is consistent, simulation results in Medeiros and Veiga (2004) show that the methodology advocated here works well in typical sample sizes found in financial applications.

4. Forecast comparison of locally stationary $\text{GARCH}(1, 1)$ vs. $\text{FCGARCH}(1, 1)$ models

In this section, we will set up an forecast exercise to compare local $\text{GARCH}(1, 1)$ models for financial volatility data with $\text{FCGARCH}(1, 1)$ models. All the models are estimated in MATLAB. The estimation of the FCGARCH model is computational demanding, especially the selection of start values. For this paper, we use a method based on a genetic algorithm to estimate the initial values for the parameters. With this careful selection of initial values, convergence is usually achieved.

4.1. Data

Our data set consists of six daily financial time series, four of which cover stock markets and two cover foreign exchange markets. The stock market indices are the Dow Jones Industrial Average (DJIA), the Japanese Nikkei 300 (NIKKEI), the German stock index DAX and the French stock index CAC 40; the exchange rate series are the British Pound (GBP) and the Swiss Franc (CHF), both against the US Dollar. The sample periods are 2 January 1991 through 19 January 2006 for the Dow Jones and Nikkei, 3 January 1991 through 22 January 2006 for the DAX and CAC40, and 2 January 1991 through 2 February 2006 for the British Pound and Swiss Franc. We use the beginning of the series through 31 December 2003 as the estimation sample and 1 January 2004 through the end of the available sample as the holding sample to evaluate our forecasts. The data were provided by Bloomberg. We scale all return series by $\sqrt{250}$ to obtain annualized volatilities.
4.2. Local GARCH(1, 1) models

The forecast of a locally stationary GARCH(1, 1) model is performed by estimating the segments of constant volatility parameters in a first step and then estimating a simple GARCH(1, 1) model with constant mean return on the last of the resulting segments. Thus, only the last regime is used in generating the forecast. We use the Kokoszka and Leipus (1999, ?) change-point detector to identify the regime changes. The detector statistic is given by

$$R_t = \frac{t(T - t)}{T^2} \left( \frac{1}{t} \sum_{\tau=1}^{t} r_{\tau}^2 - \frac{1}{T - t} \sum_{\tau=t+1}^{T} r_{\tau}^2 \right), \quad t = 1, \ldots, T,$$

and the estimated change point is chosen at $\hat{k} = \min\{k: |R_k| = \max_{1 \leq t \leq T} |R_t|\}$. The statistic converges in distribution to a Brownian Bridge, thus enabling closed form asymptotic inference. Table 1 reports the estimated change points at the 99% confidence level and the local GARCH(1, 1) estimates. Figure 4 shows the return series and three standard deviation confidence bands for the local volatility regimes. Some of the regimes exhibit relatively low persistence, with the sum $\hat{\alpha} + \hat{\beta}$ of the estimated autoregressive parameters around 0.90. The first segment of the Dow Jones has an estimated $\alpha$ of zero, so the GARCH conditional volatility equation is not defined and volatility seems to be constant at 14 percent.

Since we restrict the change point detection to the estimation sample through December 2003, the detector does not pick up the transition of the stock markets to a tranquil regime after 2003. If we run the detector on our entire sample through 2006, we find a significant change point in October 2003 for the DJIA and in June 2004 for the DAX and CAC 40. There is no such change found in the Nikkei series.

The forecast is obtained using an expanding window scheme. For the observations 1 January 2004 through the end of the holding sample, we estimate GARCH(1, 1) on the span from the last detected change point through the day before the forecast. Using the estimated parameters, we obtain the one-day-ahead forecast. This gives a time series of one-day-ahead forecasts for the holding sample sizes.

4.3. FCGARCH(1, 1) models

We specify the FCGARCH models using the GARCH(1, 1) specification as a basis model. Applying the robust version of the LM test as discussed in Section 3, the null hypothesis is rejected for all series. At each step of the testing sequence, we multiply the significance level of the test by one half ($\varrho = 1/2$). For the stock indices, we start the sequence of LM tests at a 99% confidence level. When the foreign exchange markets are considered, the confidence level is lowered to 90%. We use a relatively low confidence level for
Table 1. Change points and local GARCH models

<table>
<thead>
<tr>
<th>Change point</th>
<th>Stat</th>
<th>Prob</th>
<th>( \hat{\omega} )</th>
<th>( \hat{\alpha} )</th>
<th>( \hat{\beta} )</th>
<th>( \hat{\sigma} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DJIA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 February 1992</td>
<td>2.12</td>
<td>0.000</td>
<td>7e-4 (5e-5)</td>
<td>0.0423 (0.0180)</td>
<td>0.8740 (0.0678)</td>
<td>0.09</td>
</tr>
<tr>
<td>15 December 1995</td>
<td>2.10</td>
<td>0.000</td>
<td>9e-4 (1.2e-3)</td>
<td>0.0404 (0.0314)</td>
<td>0.8961 (0.1064)</td>
<td>0.12</td>
</tr>
<tr>
<td>26 March 1997</td>
<td>3.62</td>
<td>0.000</td>
<td>0.0011 (4e-4)</td>
<td>0.0802 (0.0154)</td>
<td>0.8924 (0.0219)</td>
<td>0.20</td>
</tr>
<tr>
<td>NIKKEI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 January 1994</td>
<td>1.75</td>
<td>0.004</td>
<td>6e-4 (5e-4)</td>
<td>0.0625 (0.0224)</td>
<td>0.9121 (0.0362)</td>
<td>0.15</td>
</tr>
<tr>
<td>21 October 1997</td>
<td>1.87</td>
<td>0.002</td>
<td>0.0021 (8e-4)</td>
<td>0.0627 (0.0125)</td>
<td>0.8947 (0.0244)</td>
<td>0.22</td>
</tr>
<tr>
<td>DAX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 August 1991</td>
<td>1.65</td>
<td>0.009</td>
<td>2e-4 (1e-4)</td>
<td>0.0314 (0.0089)</td>
<td>0.9619 (0.0122)</td>
<td>0.16</td>
</tr>
<tr>
<td>15 October 1997</td>
<td>2.91</td>
<td>0.000</td>
<td>0.0018 (6e-4)</td>
<td>0.0971 (0.0149)</td>
<td>0.8803 (0.0181)</td>
<td>0.29</td>
</tr>
<tr>
<td>CAC40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 April 1998</td>
<td>2.35</td>
<td>0.000</td>
<td>0.0012 (4e-4)</td>
<td>0.0753 (0.0140)</td>
<td>0.9063 (0.0170)</td>
<td>0.26</td>
</tr>
<tr>
<td>GBP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 October 1993</td>
<td>3.82</td>
<td>0.000</td>
<td>9e-4 (7e-4)</td>
<td>0.0548 (0.0266)</td>
<td>0.8938 (0.0614)</td>
<td>0.13</td>
</tr>
<tr>
<td>CHF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 October 1995</td>
<td>3.07</td>
<td>0.000</td>
<td>0.0018 (0.0017)</td>
<td>0.0214 (0.0077)</td>
<td>0.9630 (0.0169)</td>
<td>0.08</td>
</tr>
<tr>
<td>26 August 1998</td>
<td>2.02</td>
<td>0.001</td>
<td>5e-4 (5e-4)</td>
<td>0.0157 (0.0115)</td>
<td>0.9457 (0.0514)</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Notes: Change points estimated at the 99% confidence level and local GARCH(1, 1) models. “Stat” is the test statistic proposed by Kokoszka and Leipus (1999, ?); “Prob” is the probability for the statistic that asymptotically follows a standard Brownian Bridge under the null hypothesis; numbers in parentheses are standard errors according to Bollerslev and Wooldridge (1992).
GARCH Models in the Presence of Structural Breaks and Regime Switches

Fig. 4. Change points in the return series identified by the Kokoszka and Leipus (1999, ?) detector and plus/minus three standard deviation bands for the local volatility.

Table 2. FCGARCH models: number of estimated regimes

<table>
<thead>
<tr>
<th></th>
<th>DJIA</th>
<th>NIKKEI</th>
<th>DAX</th>
<th>CAC40</th>
<th>GBP</th>
<th>CHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regimes</td>
<td>3∗∗∗</td>
<td>2∗∗∗</td>
<td>3∗∗∗</td>
<td>2∗∗∗</td>
<td>2∗</td>
<td>2∗</td>
</tr>
</tbody>
</table>

*Significance at the 90% confidence level.

**Significance at the 99% confidence level.

the exchange rate series since it is well established that these data exhibit little if any non-linearities (Andersen et al., 2001; Kilian and Taylor, 2003; Engel and West, 2005; Hansen and Lunde, 2005). There is no evidence of multiple-regime structures at higher confidence levels. Table 2 shows the estimated number of regimes for each series. The sequence of robust LM tests shows evidence of two limiting regimes for four series: Nikkei, CAC 40, GBP, and CHF. For the DJIA and DAX, three limiting regimes are found. Figure 5 shows, for each series, the sum of the transition function versus the transition variable.
Table 3 shows the “local persistence” associated with each limiting regime in the estimated FCGARCH models. In the FCGARCH specification, the sum $\beta_0 + \lambda_0$ is the local persistence in the first extreme regime that can be associated with “bad” or “very bad” news depending if the estimated model has two or three limiting regimes. The sum $\beta_0 + \beta_1 + \lambda_0 + \lambda_1$ is the local persistence either in the “tranquil period” or in “very good news regime.” The fourth column in the table shows the local persistence of the last limiting regime in the FCGARCH model and is associated with “good” or “very good” news, depending on whether the estimated model has two or three regimes. We use the term local persistence because the transition variable is $y_{t-1}$, possibly causing very frequent regime-switching. A given regime may thus not be persistent at all.

Some interesting facts emerge from the table. First, when stock indexes are considered, the regime associated with negative returns is highly persistent in the FCGARCH model. Furthermore, the GARCH effect seems to dissipate when returns become more positive, especially when there are three regimes and not only two. On the other hand, in the case of exchange rates, the local persistence does not change much between regimes: it is very high in both of them. Finally, even with highly persistent regimes, all models are stationary, as restriction (9) is met for all cases, as reported in the last column of the table.
Table 3. FCGARCH models: local persistence in each regime

<table>
<thead>
<tr>
<th>Series</th>
<th>$\beta_0 + \lambda_0$</th>
<th>$\beta_0 + \beta_1 + \lambda_0 + \lambda_1$</th>
<th>$\beta_0 + \beta_1 + \beta_2 + \lambda_0 + \lambda_1 + \lambda_2$</th>
<th>Stationarity condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DJIA</td>
<td>1.1535</td>
<td>0.9458</td>
<td>0.7966</td>
<td>0.9750</td>
</tr>
<tr>
<td>CAC 40</td>
<td>1.0604</td>
<td>0.8856</td>
<td>–</td>
<td>0.9730</td>
</tr>
<tr>
<td>DAX</td>
<td>1.1504</td>
<td>0.9290</td>
<td>0.8192</td>
<td>0.9848</td>
</tr>
<tr>
<td>Nikkei</td>
<td>1.0749</td>
<td>0.8390</td>
<td>–</td>
<td>0.9569</td>
</tr>
<tr>
<td>GBP</td>
<td>0.9935</td>
<td>0.9924</td>
<td>–</td>
<td>0.9930</td>
</tr>
<tr>
<td>CHF</td>
<td>0.9865</td>
<td>0.9846</td>
<td>–</td>
<td>0.9856</td>
</tr>
</tbody>
</table>

*Notes:* The table shows the persistence associated with each limiting regime in the FCGARCH model. The sum $\beta_0 + \lambda_0$ is the local persistence in the first extreme regime that can be associated with “bad” or “very bad” news depending if the estimated model has two or three limiting regimes. The sum $\beta_0 + \beta_1 + \lambda_0 + \lambda_1$ is the local persistence either in the “tranquil period” or in the “very good news regime.” The fourth column in the table shows the local persistence of the last limiting regime in the FCGARCH model and is associated with “good” or “very good” news depending if the estimated model has two or three regimes. The last column reports the stationarity condition.

4.4. Forecast comparison

The forecasts from the FCGARCH model are the one-day-ahead forecasts using the actual lagged squared excess returns of the holding sample. The FCGARCH(1, 1) model is estimated on the estimation sample only, not on an expanding window scheme. The FCGARCH structure specifies the regime changes in dependence of the state variable, thus inherently capturing the transition dynamics. We use the daily squared returns as a proxy for the true volatility.

Figure 6 plots the differences in the square root of the squared forecast errors between the local GARCH(1, 1) expanding window scheme (positive sign) and FCGARCH(1, 1) (negative sign). Therefore, a positive difference means that the FCGARCH model performed better; a negative difference indicates better performance by the local GARCH(1, 1) expanding window model.

Noticeably, the FCGARCH forecasts perform better for the Dow Jones and for the Nikkei series than for all other series. Here, the gains in the forecast accuracy range between zero and one percent of volatility. For the DAX and the British Pound, the gains are of the magnitude of $1/10$ of a percentage point; for the CAC 40 and the Swiss Franc, they are only of the magnitude of $1/100$ of a percentage point. For three series, the FCGARCH forecast performs better on the entire holding sample. For the DAX, the FCGARCH forecasts are predominantly better; for the CAC 40, they are better roughly on the first year of the holding sample and worse on the second year. For the British pound, the FCGARCH forecasts are better at a short horizon (10 days), then the local GARCH approach becomes better. In interpreting the small forecast gains in particular for the CAC 40 and the Swiss Franc, we emphasize that the FCGARCH model is estimated on the estimation sample only and is here on par with an expanding window GARCH model. With this in mind, the forecast gains for the Dow Jones
and Nikkei seem substantial. Figure 7 compares the forecasts for the Dow Jones and for the Nikkei. In the case of the Dow Jones, the FCGARCH model fairly consistently forecasts volatility lower than the local GARCH approach. For the Nikkei, on the other hand, the FCGARCH model predicts lower volatility than the local GARCH model in the tranquil regime while it predicts higher volatility than the local GARCH model in the volatile regime.

We also estimate a “stable” GARCH(1, 1) model using the full estimation sample (without breaks). Table 4 reports the root mean squared error (RMSE) and the mean absolute error (MAE) for the forecasts from the “stable” and local GARCH(1, 1) models as well as the FCGARCH(1, 1) model. From the table, it is clear that the FCGARCH model produces more accurate forecasts for the DJIA, DAX, Nikkei, and CHF series. For the GBP series the results are mixed. The RMSE indicates that the FCGARCH model is often the best model, while the MAE often points to the local GARCH model as the best.

In order to examine whether the forecast differences are significant, we apply the modified Diebold–Mariano test (Diebold and Mariano, 1995) of Harvey et al. (1997) to these series of forecasts. The null hypothesis is no difference in expected absolute or squared errors. The results are reported in Table 5.

Fig. 6. Comparison of the forecast errors from the FCGARCH(1, 1) and local GARCH(1, 1) models. The figure shows the difference in the forecast of the annualized standard deviation.
Fig. 7. Comparison of forecasts from the FCGARCH(1, 1) and local GARCH(1, 1) models for the Dow Jones and the Nikkei.

Table 4. Forecast comparison: RMSE and MAE

<table>
<thead>
<tr>
<th>Series</th>
<th>RMSE</th>
<th>MAE</th>
<th></th>
<th>RMSE</th>
<th>MAE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FCGARCH</td>
<td>Local GARCH</td>
<td>GARCH</td>
<td>FCGARCH</td>
<td>Local GARCH</td>
<td>GARCH</td>
</tr>
<tr>
<td>DJIA</td>
<td>0.0155</td>
<td>0.0165</td>
<td>0.0185</td>
<td>0.1106</td>
<td>0.1173</td>
<td>0.1273</td>
</tr>
<tr>
<td>CAC 40</td>
<td>0.0311</td>
<td>0.0307</td>
<td>0.0316</td>
<td>0.1518</td>
<td>0.1478</td>
<td>0.1539</td>
</tr>
<tr>
<td>DAX</td>
<td>0.0392</td>
<td>0.0393</td>
<td>0.0397</td>
<td>0.1646</td>
<td>0.1670</td>
<td>0.1691</td>
</tr>
<tr>
<td>Nikkei</td>
<td>0.0457</td>
<td>0.0471</td>
<td>0.0469</td>
<td>0.1561</td>
<td>0.1648</td>
<td>0.1629</td>
</tr>
<tr>
<td>GBP</td>
<td>0.0120</td>
<td>0.0121</td>
<td>0.0121</td>
<td>0.0902</td>
<td>0.0893</td>
<td>0.0904</td>
</tr>
<tr>
<td>CHF</td>
<td>0.0180</td>
<td>0.0181</td>
<td>0.0180</td>
<td>0.1118</td>
<td>0.1130</td>
<td>0.1120</td>
</tr>
</tbody>
</table>

Notes: The table reports the root mean squared errors (RMSE) and the mean absolute errors (MAE) for the FCGARCH, Local GARCH, and GARCH models.

When the quadratic loss function is used, the test shows that the FCGARCH model is superior to the local GARCH model for the DJIA and Nikkei series at the 5% significance level and for the GBP and CHF series at the 10% level. The test also indicates that the FCGARCH model produces better forecasts than the “stable” GARCH for the DJIA, CAC 40, DAX, Nikkei, and GBP series.
Table 5.  Forecast comparison: modified Diebold–Mariano test

<table>
<thead>
<tr>
<th>Series</th>
<th>Squared errors Local GARCH</th>
<th>Squared errors GARCH</th>
<th>Absolute errors Local GARCH</th>
<th>Absolute errors GARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DJIA</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>CAC 40</td>
<td>0.9983</td>
<td>0.0000</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>DAX</td>
<td>0.4184</td>
<td>0.0033</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Nikkei</td>
<td>0.0042</td>
<td>0.0081</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>GBP</td>
<td>0.0877</td>
<td>0.0633</td>
<td>0.9984</td>
<td>0.0000</td>
</tr>
<tr>
<td>CHF</td>
<td>0.0701</td>
<td>0.4787</td>
<td>0.0000</td>
<td>0.0868</td>
</tr>
</tbody>
</table>

Notes: The table reports the $p$-value of the modified Diebold–Mariano test. The null hypothesis is that the FCGARCH model is not superior to the column model (local GARCH or “stable” GARCH) in terms of expected absolute or squared loss. 0.0000 indicates less than 0.00005.

When absolute errors are considered, the FCGARCH model is superior to the local GARCH model for the DJIA, DAX, Nikkei, and CHF series. When the FCGARCH model is compared to the “stable” GARCH model, the former is superior for all series with the exception of GBP.

5. Conclusion

In estimating and forecasting with GARCH models, one has to be careful not to ignore possible parameter regime changes in the data. The GARCH model is very sensitive to these changes and exhibits a distinct estimation bias in this case: The sum of the estimated autoregressive parameters is close to one regardless of the data-generating parameters. Therefore, it is desirable to have a variation of the GARCH model that allows for time dependent parameter values that can capture different volatility regimes. The FCGARCH model of Medeiros and Veiga (2004) is such a model. It features different states of volatility (bad news, tranquil, good news) depending on a transition variable that governs the changes. We compare the forecast performance of the FCGARCH model against a local GARCH model, where the segments of constant GARCH parameters are identified by a change point detector, and a simple GARCH(1, 1) model. On a data set of four daily stock index time series and two exchange rate series between 1991 and 2006, we find that the FCGARCH model outperforms the local GARCH approach as well as the standard GARCH model on most indices. For the exchange rates, the evidence is mixed.

References


