Neglecting Parameter Changes in GARCH Models

Eric Hillebrand

Louisiana State University*

March 29, 2004

Abstract

If a GARCH model is estimated on a time series that contains parameter changes in the conditional volatility process and these parameter changes are not accounted for, a distinct error in the estimation occurs: The sum of the estimated autoregressive parameters of the conditional variance converges to one. In finite samples, the sum of the estimated autoregressive parameters is heavily biased towards one. This paper shows that this convergence holds for all common estimators of GARCH. Simulations of the GARCH model show that the effect occurs for realistic parameter changes and sample sizes for financial volatility data.

JEL-codes: C22, C51.

Key words: GARCH, volatility persistence, spurious estimation, spurious unit root, spurious long memory

]

*Department of Economics, 2126 CEBA Bldg., Baton Rouge, LA 70803, phone: (225) 578-3795, fax (225) 578-3807, e-mail erhil@lsu.edu

1 Introduction: Time Scales and Persistence in Financial Volatility Data

Estimations of GARCH models on long time series usually indicate high persistence. Many studies report the sum of the estimated autoregressive parameters of the conditional variance equation to be close to unity.[1] This means that after a shock the volatility reverts only very slowly to its long term mean. This result is not unexpected because volatility shows strong serial correlation and clustering. These empirical findings suggest that volatility either has a unit root or is fractionally integrated, so that it has either indefinite or long memory. The Integrated GARCH model (IGARCH, Engle and Bollerslev 1986) and the Fractionally Integrated GARCH model (FIGARCH, Baillie et al. 1996) were proposed to reflect this.

The concern that structural changes may obfuscate persistence estimation in GARCH models was raised early. In a comment to the original IGARCH paper by Engle and Bollerslev (1986), Francis Diebold suggested that the failure to accommodate shifts in monetary policy regimes, reflected by changes in the constant term in the conditional variance equation, might result in a spurious estimate of integrated interest rate volatility (Diebold 1986). Lamoureux and Lastrapes (1990) explored Diebold's conjecture by allowing for changing states of the constant term of the conditional variance of a GARCH(1,1) model. Hamilton and Susmel (1994) used the Markov-switching model and improved volatility forecasts of ARCH models by

[1] Engle and Bollerslev (1986): weekly returns on exchange rates over 12 years; Baillie and DeGennaro (1990): daily returns on stock index over 18 years; Bollerslev and Engle (1993): daily returns on exchange rates over 5 years; Baillie et al. (1996): daily returns on exchange rates over 13 years; Ding and Granger (1996): daily returns on stock index over 63 years; Andersen and Bollerslev (1997): 5 minute returns on exchange rates over one year and on stock index future prices over 4 years; Engle and Patton (2001): daily returns on stock index over 12 years.

incorporating regime changes. Gray (1996) extended the regime switching approach to GARCH. These locally stationary models that segment financial data obtained significantly lower estimates of persistence.

Francq et al. (2001) examined ARCH processes which are subject to Markov-switching parameters in simulations. These simulations indicate that as a result of the stochastic nature of the Markov-switching process, the ARCH parameters will be estimated in the neighborhood of integration when the changes are not accounted for.

A number of studies find strong support for parameter changes in financial data, for example Bos, Franses, and Ooms (1998) and Andreou and Ghysels (2002), who implement the ARCH-specific change-point detection methods proposed by Kokoszka and Leipus (1999, 2000) and the least-squares detection method of Lavielle and Moulines (2000).

One possible interpretation of these findings is that in financial markets a low-persistence volatility process that is disturbed by occasional parameter regime shifts generates volatility data that appear to be highly persistent when the parameter shifts are not accounted for in global estimations.

An alternative interpretation is that several processes of different time scales act continuously and simultaneously. These may again be interrupted by parameter changes. Two different model classes are commonly proposed to capture the presence of several scales: multi-scale stochastic volatility models and fractionally integrated, or long memory models. An alternative is Engle and Lee (1999). They construct a multiple-process GARCH model that amounts to a GARCH(2,2) specification and find strong evidence in its favor.

In the context of stochastic volatility models, Fouque et al. (2003) and Chernov et al. (2003) suggest multiple-scales models. LeBaron (2001) demonstrates in a model with three short time scales that the aggregation of the three factors implies long memory properties. Gallant and Tauchen (2001) estimate a two-scale volatility model and find a long and a short correlation structure in daily stock returns.

Fractionally integrated processes capture the multi-scale nature of financial volatility data well (e.g., Andersen et al. 2001, Baillie et al. 1996, Ding et al. 1993, Ding and Granger 1996). Granger (1980) shows that aggregation of processes with different autoregressive parameters, and thereby different persistence structures, can induce long memory properties. Therefore, long memory models are natural candidates for capturing multiple scales. The connection between structural breaks and the parameter of fractional integration has been discussed recently (Bos et al. 1999, Choi and Zivot 2002, Diebold and Inoue 2001, Granger and Hyung 1999, Granger and Teräsvirta 2001, Lobato and Savin 1998, Ohannissian et al. 2003, Sakoulis and Zivot 2000). The complex multiple scale structure of financial volatility has also been investigated from the viewpoint of scaling according to power laws (Müller et al. 1997, Ghashgaie et al. 1996).

In summary, financial volatility comprises multiple time scales. Disputed is the nature and length of these different scales. Using high-frequency data of 5 minute returns estimated at different frequencies, Andersen and Bollerslev (1997) give a concise overview of the irregular picture of scale length estimation. They obtain estimates for the mean reversion time ranging from two hours to 7 days for the half-life of the S&P500 volatility between 1986 and 1989 using GARCH models. Using spectral methods, Fouque et al. (2003) estimate an average mean reversion length of about 1.5 days for high-frequency S&P500 data between 1994 and 1998.

In this paper, I will show that parameter regime changes in GARCH models that are not accounted for in global estimations cause the sum of the estimated autoregressive parameters to converge to one. This is usually taken as evidence of high persistence, but in the case of neglected parameter changes, it is spurious and a possible confounding factor in volatility analysis. Simulations will show that this effect occurs for jump sizes and sample sizes that are realistic for financial data. Regardless of the lengths of the data-generating persistence time scales, in the presence of neglected parameter changes, GARCH will indicate high persistence. Therefore, in this case, GARCH is not a suitable model to measure persistence. Prior to any GARCH estimation, a change-point detection study using, for example, the method of Kokoszka and Leipus (1999, 2000) is necessary.

This phenomenon of spurious high persistence measurement is a consequence of the geometry of the estimation problem. It is independent of the estimation method. It is also independent of the statistical properties of the parameter changes. In this paper, I label this effect of spurious high persistence due to neglected parameter changes in GARCH estimations "spurious almost-integration", for lack of a better expression.

The closest related study is Mikosch and Starica (2000). They demonstrate that the Whittle estimate of the autoregressive parameter of the ARMA(1,1) representation of GARCH(1,1) will be close to unity when there are change-points that are not accounted for. Here, I will provide a proof that the "spurious almost-integration" effect is general enough to hold for all estimators of GARCH whose variance vanishes with the sample size, thereby including the commonly used maximum likelihood estimator. Also, the result generalizes to GARCH(p,q), showing that higher-order GARCH models like the one of Engle and Lee (1999) are not a remedy for the problem.

The rest of the paper is organized as follows. Section 2 briefly describes persistence estimation in GARCH models. Section 3 presents the proof that parameter changes that are unaccounted for push the sum of the estimated autoregressive parameters of GARCH models to unity. Section 4 explores the phenomenon in simulations with realistic sample and jump sizes. Section 5 concludes.

2 Persistence Estimation with GARCH Models

Engle (1982) and Bollerslev (1986) suggested the following estimation approach. The return r_t from a stock with price S_t at time t is modeled as

$$r_t := \log(S_{t+1}) - \log(S_t) = \mathbb{E}(r_t|\mathcal{F}_{t-1}) + \varepsilon_t = \mu(b) + \varepsilon_t \; t = 1, \ldots, N. \tag{1}$$

Here, $\mathcal{F}_t$ denotes the filtration modeling the information set and μ is the conditional mean function with argument b, for example a regression $\mu(b) = x_t'b$, where x_t denotes a set of independent variables. Another example of the conditional mean is simply a constant $\mu(b) \equiv \mu$. In this paper, we are interested in the parameters of the conditional variance equation only, so we can neglect this issue.

The disturbance ε_t is assumed to be normally distributed, conditional on the information available at time $t-1$:

$$\varepsilon_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t), \tag{2}$$

i.e. $\varepsilon_t = \eta_t\sqrt{h_t}$, $\eta_t \sim \mathcal{N}(0,1)$, where h_t denotes the conditional variance. The latter is determined by the difference equation

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \tag{3}$$

where $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i \leq 1$. This is the GARCH(p,q) model for the conditional variance. In the case where equality holds, we have the integrated GARCH or IGARCH model.

Consider the expected value of (3) in the GARCH(1,1) case conditional on the information set $\mathcal{F}_{t-2}$,

$$\mathbb{E}_{t-2}h_t := \mathbb{E}(h_t|\mathcal{F}_{t-2}) = \omega + \alpha\mathbb{E}_{t-2}\varepsilon_{t-1}^2 + \beta\mathbb{E}_{t-2}h_{t-1}.$$

It follows from the distribution assumption in equation (2) that $\mathbb{E}_{t-2}\varepsilon_{t-1}^2 = h_{t-1}$. From the definition of h_t, $\mathbb{E}_{t-2}h_{t-1} = h_{t-1}$, because only lagged information enters the conditional variance equation. Therefore,

$$\mathbb{E}_{t-2}h_t = \omega + (\alpha + \beta)\mathbb{E}_{t-2}h_{t-1}.$$

The conditional expected value of the variance process h_t is an AR(1) process with first-order autoregressive coefficient $\alpha + \beta$. The usual stationarity conditions of autoregressive models apply. That is, in the GARCH(1,1) case the sum $\alpha + \beta$ must be less than one. In the GARCH(p,q) case, the sum $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j$ must be smaller than one. The closer it comes to one, the higher is the persistence of the process.

To obtain the unconditional expected variance of ε_t, assume that the process h_t is covariance-stationary. Then, $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ holds and

$$\mathbb{E}h_t = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}. \tag{4}$$

For ease of exposition I will consider the GARCH(1,1) case only. In this case,

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1} \tag{5}$$

and

$$\mathbb{E}h_t = \omega/(1 - \lambda), \tag{6}$$

where $\lambda = \alpha+\beta$. This simplification does not come at the cost of a loss of generality. All results presented in this paper generalize to the GARCH(p,q) case.

Collect all model parameters in the vector $\theta = (b,\, \omega,\, \alpha_1,\, \ldots,\, \alpha_q,\ \beta_1,\, \ldots,\, \beta_p)$. The log-likelihood function is given by

$$L_N(\theta) := -\frac{1}{2N}\left[N \log(2\pi) + \sum_{t=1}^{N}\left(\log h_t(\theta) + \frac{\varepsilon_t^2(\theta)}{h_t(\theta)}\right)\right]. \tag{7}$$

The conditional variance process h_t is not observable and therefore has to be estimated at the initial parameter estimate and then updated at every iteration in the maximization. The GARCH model is not restricted to the conditionally normal case. Bollerslev (1987) suggests using the t-distribution and treating the number of degrees of freedom as additional parameter.

Consistency and asymptotic normality of the maximum likelihood estimator can only be proven for the GARCH(1,1) case so far. The main results can be found in Weiss (1986), Bollerslev and Wooldridge (1992), and Lumsdaine (1996). There are no closed analytical expressions for the estimators.

In practice, the likelihood is maximized by numerical optimization methods. Most software packages implement a quasi-Newton method using line search and Hessian updating algorithms. For purposes of this paper, I maximize (7) using the 'dfpmin' routine from the "Numerical Recipes" (Press et al. 2002) which implements a quasi-Newton method. The gradients are computed using analytical expressions; the Hessians are approximated by finite differencing.

Simulation studies show that the log-likelihood function does not always yield a unique maximum (Doornik and Ooms 2000, Zumbach 2000). In Section 4, I find a strong small sample bias in the estimates of the autoregressive parameters α and β. This is in accordance with the findings of Lumsdaine (1995). As the sample size increases, the parameter estimates become very accurate, though. Therefore, I use the standard estimation approach in this paper.

As an example, I estimate GARCH on daily closing data of the Dow Jones Industrial Average between Dec 7, 1987, and October 31, 2003. The data come from Datastream and the sample size is 4000 observations. Within the class of model orders up to GARCH(3,3), the Bayes Information Criterion favors GARCH(1,1). The estimation result for the conditional variance equation is

$$h_t = \underset{(3.388\text{e-}7)}{1.17\text{e-}6} + \underset{(0.011)}{0.054}\,\varepsilon_{t-1}^2 + \underset{(0.010)}{0.936}\,h_{t-1}.$$

The numbers in parentheses are heteroskedasticity-consistent standard errors according to Bollerslev and Wooldridge (1992).

This is a result commonly reported by the GARCH studies discussed in the introduction. If time series that cover several years are considered, α is usually estimated in the region between 5 and 10 percent and β is usually estimated in the region above 90 percent. The estimated sum $\hat{\lambda} = \hat{\alpha} + \hat{\beta}$ of the autoregressive parameters, here equal to 0.99, is typically almost one. The t-statistic of β is mostly as large as we find it here. This common result is obtained not only from various stock price data but also from exchange rate or interest rate data.

3 Neglecting Parameter Changes in GARCH Estimations

In this section, I provide a proof that if a GARCH model is estimated on data that contain switches in the data-generating parameters of the conditional variance equation and these switches are not accounted for, the sum of the estimated autoregressive parameters converges to one. This phenomenon is independent of the estimation method because it stems from the geometry of the estimation problem. Also, the parameter switches need not have a specific stochastic structure; a single deterministic change-point is sufficient for the effect to occur.

Lemma 1. *Denote by $\mathbb{E}_0 h_t$ the expected value of a stationary GARCH(1,1) model conditional on the start value $h_0 \in \mathbb{R}$. Then, the relation*

$$\mathbb{E}_0 h_t = \mathbb{E} h_t + o(1)_N \tag{8}$$

holds for $t \in \{1, \dots, N\}$, where $\mathbb{E} h_t = \omega/(1-\lambda)$ and $\lambda = \alpha + \beta < 1$.

Proof. The expected value conditional on the initial value h_0 is given by

$$\mathbb{E}_0 h_t = \omega + \mathbb{E}_0(\alpha \eta_{t-1}^2 + \beta)\mathbb{E}_0 h_{t-1} = \omega + \lambda \mathbb{E}_0 h_{t-1} = \omega \frac{1-\lambda^t}{1-\lambda} + \lambda^t h_0,$$

as $\mathbb{E}_0 \eta_t^2 = 1$ for all t and the η_t are independent. Thus, substituting from equation

(6),

$$|\mathbb{E}_0 h_t - \mathbb{E}h| = \left|\omega\frac{1-\lambda^t}{1-\lambda} + \lambda^t h_0 - \frac{\omega}{1-\lambda}\right| = \lambda^t\left|h_0 - \frac{\omega}{1-\lambda}\right| = o(1)_N. \qquad \square$$

Assumption 1. *The processes $\{h_t\}$ and $\{\varepsilon_t\}$ are observable without measurement error, or at least with a measurement error that is independent of the parameter estimator $\hat{\theta}$ and that vanishes with increasing sample size.*

This assumption is, of course, unrealistic. The process h_t is not observable and in real estimation problems h_t is estimated by $\hat{h}_t(\hat{\theta})$ and ε_t by $\hat{\varepsilon}_t(\hat{\theta})$. I make the conjecture that if "spurious almost-integration" occurs when h_t and ε_t are observable, it will also occur when there is less perfect information about the volatility condition of the market. Section 4 provides simulation evidence to support this conjecture. For the sake of notational brevity, I assume that the measurement is error-free. The case of an error that is independent of the parameter estimates does not alter the argument.

Consider the case where there are $k-1$ points in time where the variance parameters in the data-generating θ change:

$$h_t(\theta_i) = \omega_i + \alpha_i\varepsilon_{t-1}(\theta_i)^2 + \beta_i h_{t-1}(\theta_i),\ t = N_{i-1}+1,\ \ldots,\ N_i, \tag{9}$$

where $i = 1, \ldots, k$, and let $N_0 = 0$, and $N_k = N$. Denote the parameter vector within segments θ_i. It contains the parameters b of the conditional mean equation, which do not change from segment to segment, and the segment-dependent parameters of the conditional variance ω_i, α_i, β_i.

This regime-changing structure of the data-generating process is not accounted for in the estimation. The estimated variance model is

$$h_t = \hat{\omega} + \hat{\alpha}\varepsilon_{t-1}^2 + \hat{\beta}h_{t-1}. \tag{10}$$

Let $\mathbb{E}_{(i)}h_t$ denote the expected values with respect to the initial value in segment

$i = 1, \ldots, k$, so that

$$\mathbb{E}_{(i)} h_t := \mathbb{E}(h_t | \mathcal{F}_{N_{i-1}}).$$

From Lemma 1 it follows that within each segment, the h_t cluster around

$$\mathbb{E} h_t(\theta_i) = \omega_i / (1 - \lambda_i), \tag{11}$$

where $\lambda_i = \alpha_i + \beta_i < 1$, because $\mathbb{E}_{(i)} h_t = \mathbb{E} h_t(\theta_i) + o(1)_{N_i - N_{i-1}}$.

Lemma 2. *Let $\mathbb{E} h_t(\theta_i)$ denote the unconditional expected value of a variance series generated by the parameter vector θ_i. Let h_t be generated according to* (9). *Then,*

$$\begin{aligned} \bar{h} &= \frac{1}{N} \sum_{i=1}^{k} (N_i - N_{i-1}) \mathbb{E} h_t(\theta_i) + o(1)_N, \\ \overline{\varepsilon^2} &= \frac{1}{N} \sum_{i=1}^{k} (N_i - N_{i-1}) \mathbb{E} h_t(\theta_i) + o(1)_N, \end{aligned} \tag{12}$$

where it is assumed that as $N \to \infty$, the segment lengths $N_i - N_{i-1} \to \infty$.

Proof. For $t \in \{N_{i-1} + 1, \ldots, N_i\}$, write $h_t = \mathbb{E}_{(i)} h_t + x_t$, x_t being the deviation from the expected value conditional on the initial values within segments. Then, by the law of large numbers,

$$\frac{1}{N_i - N_{i-1}} \sum_{t=N_{i-1}+1}^{N_i} x_t = o(1)_{N_i - N_{i-1}}.$$

The sample mean of h over the entire sample can be written as

$$\begin{aligned} \bar{h} &= \frac{1}{N} \sum_{t=1}^{N} h_t, \\ &= \frac{1}{N} \sum_{t=1}^{N_1} \mathbb{E}_{(1)} h_t + \ldots + \frac{1}{N} \sum_{t=N_{k-1}+1}^{N} \mathbb{E}_{(k)} h_t \\ &\quad + \frac{1}{N} \sum_{t=1}^{N_1} x_t + \ldots + \frac{1}{N} \sum_{t=N_{k-1}+1}^{N} x_t, \\ &= \frac{1}{N} \sum_{t=1}^{N_1} \mathbb{E}_{(1)} h_t + \ldots + \frac{1}{N} \sum_{t=N_{k-1}+1}^{N} \mathbb{E}_{(k)} h_t + \sum_{i=1}^{k} o(1)_{N_i - N_{i-1}}. \end{aligned}$$

I assume that $N_i - N_{i-1} \to \infty$ as $N \to \infty$, so I can write the last term as $o(1)_N$.

Applying Lemma 1, I obtain

$$\begin{aligned}\bar{h} &= \frac{1}{N}\sum_{t=1}^{N_1}(\mathbb{E}h_t(\theta_1)+o(1)_{N_1})+\ldots+\frac{1}{N}\sum_{t=N_{k-1}+1}^{N}\left(\mathbb{E}h_t(\theta_k)+o(1)_{N-N_{k-1}}\right)+o(1)_N\\ &= \frac{1}{N}\sum_{i=1}^{k}(N_i-N_{i-1})\mathbb{E}h_t(\theta_i)+\frac{1}{N}\sum_{i=1}^{k}\sum_{t=N_{i-1}+1}^{N_i}o(1)_{N_i-N_{i-1}}+o(1)_N.\end{aligned}$$

The second term on the right-hand side is of the order $O(1/N)$. Therefore, the last two terms can be written as $o(1)_N$ and the statement for the sample mean of the h is proven.

In the same manner, write $\varepsilon_t^2 = \mathbb{E}_{(i)}\varepsilon_t^2 + y_t = \mathbb{E}_{(i)}h_t + y_t$ by the distribution assumption in equation (2). Then,

$$\overline{\varepsilon^2} = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_t^2 = \frac{1}{N}\sum_{i=1}^{k}(N_i-N_{i-1})\mathbb{E}h_t(\theta_i)+o(1)_N. \qquad \square$$

Assumption 2. *The influence of a single realization of the processes ε_t^2 and h_t on the estimator $\hat{\theta}$ vanishes with growing sample size:*

$$cov(\hat{\theta},\varepsilon_t^2) = o(1)_N \;\forall\, t$$

$$cov(\hat{\theta},h_t) = o(1)_N \;\forall\, t.$$

"Spurious almost-integration" is essentially a geometric phenomenon and occurs regardless of the estimation method. The only assumption I make about the estimator $\hat{\theta}$ is that its covariance with a single observation in the ε_t^2 series or in the h_t series vanishes with growing sample size. This assumption is not restrictive, it holds for a very general class of estimators. In essence, it assumes that some central limit theorem holds. For example, if we apply the Cauchy-Schwarz inequality to the covariance of $\hat{\theta}$ and ε_t^2,

$$\text{cov}(\hat{\theta},\varepsilon_t^2) = \mathbb{E}\left[(\hat{\theta}-\mathbb{E}\hat{\theta})(\varepsilon_t^2-\mathbb{E}\varepsilon_t^2)\right] \le \sqrt{\text{Var}(\hat{\theta})\,\text{Var}(\varepsilon_t^2)},$$

we see that the assumption is tantamount to a vanishing variance of the estimator as the sample size increases, given that the fourth moment of the ε_t series is finite (Bollerslev 1986).

For instance, the asymptotic distribution of the maximum likelihood estimator $\hat{\theta}$ of the GARCH(1,1) parameters $\theta = (\mu, \alpha_0, \alpha, \beta)^T$ is given by

$$\sqrt{N}(\hat{\theta} - \theta) \sim_{N\to\infty} \mathcal{N}(0, \Sigma)$$

where Σ is a symmetric positive definite matrix that arises from the outer product of the likelihood score (Weiss 1986, Bollerslev and Wooldridge 1992, Lumsdaine 1996). Hence, the variance of $\hat{\theta}$ vanishes with N and satisfies the assumption. For higher order GARCH models there is as yet no asymptotic distribution theory. However, it is commonly assumed that the estimators have an asymptotically normal distribution.

There is a twofold dependency in the estimation of a GARCH model, one being the dependency of the estimated volatility process $\hat{h}_t$ on the estimated parameters, the other being the dependency of the estimators on the data h_t and ε_t^2. The former dependency is addressed by the assumption of perfect measurement, the latter by the assumption of the vanishing influence of a single realization on the estimator.

Proposition 1. *If there are k switches in the data-generating parameters of the conditional variance equation of a GARCH(1,1) model, as specified in equation* (9), *and the model is estimated on the entire series without accounting for the parameter changes, then, under Assumptions 1 and 2, the condition*

$$\mathbb{E}_{(i)}\hat{\lambda} = \mathbb{E}_{(i)}(\hat{\alpha} + \hat{\beta}) = 1$$

must hold in every single segment i, up to terms that vanish with growing segment sizes $N_i - N_{i-1}$, $i = 1, \ldots, k$.

Proof. Subtract the sample mean from (10) and apply Assumption 1:

$$h_t - \bar{h} = \hat{\alpha}(\varepsilon_{t-1}^2 - \overline{\varepsilon^2}) + \hat{\beta}(h_{t-1} - \bar{h}). \tag{13}$$

Take expectations of (13) conditional on the initial value within segments, we have

$$\mathbb{E}_{(i)}h_t - \mathbb{E}_{(i)}\bar{h} = \mathbb{E}_{(i)}(\hat{\alpha}\varepsilon_{t-1}^2) - \mathbb{E}_{(i)}(\hat{\alpha}\overline{\varepsilon^2}) + \mathbb{E}_{(i)}(\hat{\beta}h_{t-1}) - \mathbb{E}_{(i)}(\hat{\beta}\bar{h}).$$

Use Assumption 2 and the distribution assumption in equation (2):

$$\mathbb{E}_{(i)}h_t - \mathbb{E}_{(i)}\bar{h} = \mathbb{E}_{(i)}\hat{\alpha}\mathbb{E}_{(i)}h_{t-1} - \mathbb{E}_{(i)}(\hat{\alpha}\overline{\varepsilon^2}) + \mathbb{E}_{(i)}\hat{\beta}\mathbb{E}_{(i)}h_{t-1} - \mathbb{E}_{(i)}(\hat{\beta}\bar{h}) + o(1)_N. \quad (14)$$

Plug in $\bar{h}$ and $\overline{\varepsilon^2}$ from Lemma 2, and use Lemma 1 to obtain:

$$\begin{aligned}
&\mathbb{E}h_t(\theta_i) - \frac{1}{N}\sum_{j=1}^{k}(N_j - N_{j-1})\mathbb{E}h_t(\theta_j) + o(1)_N \\
&\quad = \mathbb{E}_{(i)}\hat{\alpha}\left(\mathbb{E}h_t(\theta_i) + \lambda^{t-N_{i-1}-1}c_i\right) \\
&\quad\quad - \mathbb{E}_{(i)}\left[\hat{\alpha}\left(\frac{1}{N}\sum_{j=1}^{k}(N_j - N_{j-1})\mathbb{E}h_t(\theta_j) + \frac{1}{N}\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j}\lambda^{t-N_{j-1}-1}c_j \right.\right. \\
&\quad\quad\quad \left.\left. + \frac{1}{N}\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j} y_t\right)\right] \\
&\quad\quad + \mathbb{E}_{(i)}\hat{\beta}\left(\mathbb{E}h_t(\theta_i) + \lambda^{t-N_{i-1}-1}c_i\right) \\
&\quad\quad - \mathbb{E}_{(i)}\left[\hat{\beta}\left(\frac{1}{N}\sum_{j=1}^{k}(N_j - N_{j-1})\mathbb{E}h_t(\theta_j) + \frac{1}{N}\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j}\lambda^{t-N_{j-1}-1}c_j \right.\right. \\
&\quad\quad\quad \left.\left. + \frac{1}{N}\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j} x_t\right)\right] + o(1)_N, \qquad (15)
\end{aligned}$$

where the last $o(1)_N$ term on the left-hand side stems from the applications of Lemma 1 to $\mathbb{E}_{(i)}h_t$ and Lemma 2 to $\mathbb{E}_{(i)}\bar{h}$. The last $o(1)_N$ term on the right-hand side comes from the application of Assumption 2 in equation (14) above. The decompositions $h_t = \mathbb{E}_{(i)}h_t + x_t$ and $\varepsilon_t^2 = \mathbb{E}_{(i)}h_t + y_t$ were introduced in the proof of Lemma 2. The coefficient

$$c_i = h_{N_{i-1}+1} - \frac{\omega_i}{1-\lambda_i} = h_{N_{i-1}+1} - \mathbb{E}h_t(\theta_i)$$

is the distance of the initial value of segment i from the unconditional mean implied by θ_i.

Except for the sums of the x_t and y_t, all the terms in parentheses in (15) are deterministic.

Assumption 2 implies that

$$\operatorname{cov}(\hat{\alpha}, \varepsilon_t^2) = \operatorname{cov}(\hat{\alpha}, \mathbb{E}_{(i)}h_t) + \operatorname{cov}(\hat{\alpha}, y_t) = o(1)_N,$$

and therefore

$$\frac{1}{N}\mathbb{E}_{(i)}(\hat{\alpha}\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j} y_t) = \frac{1}{N}\mathbb{E}_{(i)}\hat{\alpha}\,\mathbb{E}_{(i)}(\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j} y_t) + \frac{1}{N}\,\mathrm{cov}(\hat{\alpha}, \sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j} y_t)$$
$$= 2\,\frac{1}{N}\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j} o(1)_N = o(1)_N,$$

and analogously,

$$\frac{1}{N}\mathbb{E}_{(i)}(\hat{\beta}\sum_{j=1}^{k}\sum_{t=N_{j-1}+1}^{N_j} x_t) = o(1)_N.$$

Plugging into (15) and arranging terms, I find

$$\mathbb{E}h_t(\theta_i) - \frac{1}{N}\sum_{j=1}^{k}(N_j - N_{j-1})\mathbb{E}h_t(\theta_j)$$
$$= \mathbb{E}_{(i)}(\hat{\alpha}+\hat{\beta})\left[\mathbb{E}h_t(\theta_i) - \frac{1}{N}\sum_{j=1}^{k}(N_j - N_{j-1})\mathbb{E}h_t(\theta_j)\right] + o(1)_N \qquad (16)$$

That is, as $N \to \infty$ and $N_i - N_{i-1} \to \infty$, in every segment i the condition

$$\mathbb{E}_{(i)}\hat{\lambda} = \mathbb{E}_{(i)}(\hat{\alpha}+\hat{\beta}) = 1$$

must hold up to terms of order $o(1)_N$. Note that as long as one single change-point occurs,

$$\mathbb{E}h_t(\theta_i) - \frac{1}{N}\sum_{j=1}^{k}(N_j - N_{j-1})\mathbb{E}h_t(\theta_j) \neq 0. \qquad \square$$

This proof generalizes to the GARCH(p,q) case. This generalization can be obtained from the author upon request.

In the remainder of this section, I will provide some intuition for Proposition 1. Consider the case of a single change-point in the conditional variance parameters of a GARCH(1,1) model. In each of the two segments, the realizations of the conditional volatility process are centered approximately around the unconditional, stationary mean corresponding to the parameters of that segment.[2] If a GARCH(1,1) model is estimated globally without accounting for the segmentation, the resulting estimation

[2] "Approximately" means up to terms that vanish with growing segment length. Strictly, the stationary measure is not defined in the case of segmented data. For this reason, the largest part of the formal argument is concerned with making this approximation precise.

hyperplane (parameterized by $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$) must go through both segment means of h_t. As the mean of the $\{h_t\}$ and the mean of the $\{h_{t-1}\}$ is the same for sufficiently long segments, a line connecting two different means in the (h_t, h_{t-1})-subspace is close to the identity. Therefore, the estimator of β will pick up the slope of the identity and be close to one. The remaining autoregressive parameter $\hat{\alpha}$ is chosen residually such that $\hat{\alpha}+\hat{\beta}\approx 1$. The sum will always stay slightly below one to keep the estimated process $\hat{h}_t$ from exploding.

Write down the estimated GARCH(1,1) equation with the correctly measured h_t and ε_t:

$$h_t = \hat{\omega} + \hat{\alpha}\varepsilon_{t-1}^2 + \hat{\beta}h_{t-1}, \tag{17}$$

and subtract the sample mean

$$h_t - \bar{h} = \hat{\alpha}(\varepsilon_{t-1}^2 - \overline{\varepsilon^2}) + \hat{\beta}(h_{t-1} - \bar{h}). \tag{18}$$

One might argue that with an exact measurement of h_t and ε_t, it makes sense to plug in these values into (17) and just back out the parameters α and β, thereby possibly finding parameter switches. However, we are primarily interested in the case where h_t is unobservable, so the trivial back-out approach does not work. Therefore, I proceed with the estimation of (17). The reader may think of a measurement error in h_t and ε_t that is independent of the estimators and vanishes with N. In that case, the parameters cannot be backed out as well.

Applying the Assumptions 1 and 2 and taking expectations of (18), in the proof of Proposition 1 I obtain an expression similar to the following:

$$\mathbb{E}(h_t - \bar{h}) \approx \mathbb{E}\hat{\alpha}\,\mathbb{E}(\varepsilon_{t-1}^2 - \overline{\varepsilon^2}) + \mathbb{E}\hat{\beta}\,\mathbb{E}(h_{t-1} - \bar{h}). \tag{19}$$

According to the distribution assumption in equation (2),

$$\mathbb{E}\varepsilon_{t-1}^2 = \mathbb{E}(\mathbb{E}_{t-2}\varepsilon_{t-1}^2) = \mathbb{E}h_{t-1} \tag{20}$$

for $t \geq 2$. For essentially the same reason, $\overline{\varepsilon^2} \approx \bar{h}$, as it is shown in Lemma 2. From this, $\mathbb{E}h_t \approx \mathbb{E}h_{t-1}$, and (20) we can rewrite (19) as follows

$$\mathbb{E}(h_t - \bar{h}) \approx \mathbb{E}(\hat{\alpha} + \hat{\beta})\mathbb{E}(h_t - \bar{h}). \tag{21}$$

If there are no parameter switches in $\theta = (\mu, \omega, \alpha, \beta)'$ and (17) is the correct specification, then $\mathbb{E}h_t = \mathbb{E}\bar{h}$. The condition (21) is trivial. However, in the case where there are switches in θ, equation (17) is misspecified. Then, the expected value of $\bar{h}$ is a weighted average of the means of h_t in the different segments, as it is shown in Lemma 2. These means are the $\mathbb{E}h_t(\theta_i) = \omega_i/(1 - \alpha_i - \beta_i)$ plus terms that vanish with growing segment length, where θ_i denotes the parameter vector in segment i.

That is, within each segment the difference between the expected value of h_t and the expected value of the global sample mean $\bar{h}$ of h_t is non-zero. Hence, condition (21) is not trivial and thus

$$\mathbb{E}(\hat{\alpha} + \hat{\beta} \,|\, \text{initial value of segment } i) \approx 1,$$

up to terms that take care of the initial values of each segment and that vanish with growing segment sizes. This is the "spurious almost-integration" effect.

4 Simulations

In this section, I explore the implications of Proposition 1 in simulations. The simulations illustrate that the convergence stated in the proposition does occur for sample sizes and parameter jumps that are realistic for financial data.

The simulation object is a GARCH(1,1) model that is estimated on time series of different lengths that have a single parameter change-point occurring in the middle of the sample. All three parameters of the conditional variance equation of the GARCH(1,1) model, ω, α, and β, are changed separately.

To study parameter changes in ω, I consider the model

$$\begin{aligned} r_t &= \varepsilon_t, \\ \varepsilon_t|_{\mathcal{F}_{t-1}} &\sim \mathcal{N}(0,\, h_t) \\ h_t &= \omega_i + 0.10\,\varepsilon_{t-1}^2 + 0.70\,h_{t-1}. \end{aligned} \tag{22}$$

The model will be estimated on four different sample sizes, $N = 200$, 800, 4000, and 10000. The single change-point always occurs in the middle of the sample.

I consider ten parameter switches of different sizes. I try to cover a range of volatility that is realistic for financial data. Therefore, the annualized volatility σ_1 in the first segment of the simulated data, before the change-point, is always equal to 10 percent. For σ_2 I then take an equidistant grid of ten points ranging from 10 to 25 percent annualized volatility. In the GARCH(1,1) model, the annualized volatility σ is given by $\sqrt{250\omega/(1-\alpha-\beta)}$. Inserting the grid for σ_2 and backing out the parameter under study give the grid for the data-generating parameter value in the second segment. For each jump size and sample size, I simulate ten thousand time series and estimate GARCH(1,1) globally on each series.

There is an intuitive reason to expect that the "spurious almost-integration" effect will depend on the size of the jump: In the case of a very small size of the parameter change, the two clusters of (h_{t-1}, h_t)-points around $\mathbb{E}h_t(\theta_1)$ and $\mathbb{E}h_t(\theta_2)$ will be very close to each other and the support of the point clusters on the h_{t-1}-axis will be larger than on the h_t-axis. The estimation hyperplane will be influenced more by the data-generating slope in the $(h_t,\, h_{t-1})$-subspace than by the slope of the identity. Therefore, the smaller the jump the closer the estimation of β will be to the data-generating parameter value.

Table 1 shows the result for the change in ω. The first row shows results of the series without a parameter change. It is apparent that there is substantial small sample bias in the estimates, consistent with the bias reported in Lumsdaine (1995).

For sample sizes 200 and 800, β is largely underestimated and ω is accordingly overestimated. As we have

$$\mathbb{E}\varepsilon_t^2 = \mathbb{E}h_t = \frac{\omega}{1-\alpha-\beta}$$

and the variance of the ε_t is fixed by the sample average $\sum \hat{\varepsilon}_t^2/N$ from the data, an underestimation of β or α must always be compensated by an overestimation of ω and vice versa. For larger samples, the estimates of all three parameters are quite accurate.

In the case of parameter changes, the estimated sum $\hat{\lambda}$ of the autoregressive parameters converges to one as the sample size increases, as stated in Proposition 1. For jumps of sufficient size, this causes "spurious almost-integration" even in samples of rather small sizes. For example, for a jump from 10 to 18 percent volatility and sample size $N = 800$ representing 3 years of daily data, the sum of the estimated autoregressive parameters is 0.99. For the sample size $N = 4000$ representing 16 years of daily data, a single jump from 10 to 15 percent volatility pushes the sum of the autoregressive parameters to 0.98. Many of the long term GARCH studies cited in the introduction cover time spans like these. It is likely that there will be at least one change-point in such a long time span, so GARCH is very likely to pick up "spurious almost-integration".

The sum λ of the autoregressive parameters is never estimated at exactly one. Where the tables state "1", the estimate is larger than 0.995. The convergence is always from the left and never attains unity. The reason lies in the estimation algorithm. The conditional volatility process has to be updated for every new parameter estimate obtained in an iteration of the maximization, so any estimate that implies a $\hat{\lambda} \geq 1$ would blow up the estimated conditional volatility process and the negative log-likelihood with it. So the algorithm avoids $\hat{\lambda} \geq 1$.

To study parameter changes in α, I consider model (22) with the conditional variance equation

$$h_t = 1.6\text{e-}5 + \alpha_i\, \varepsilon_{t-1}^2 + 0.50\, h_{t-1}. \tag{23}$$

In the first segment, α_1 is always equal to 0.10, yielding 10 percent annualized volatility. The jump size grid for α_2 is obtained as before. Table 2 shows the simulation results.

As before, the estimated sum $\hat{\lambda}$ of the autoregressive parameters moves towards unity with growing sample size and with growing jump size. It is estimated above the mean of the in-segment data-generating λ's and grows monotonically with sample and jump size. However, the speed of the convergence is much slower than in the case of jumps in the constant. For the sample and jump sizes considered here, $\hat{\lambda}$ does not come close to unity.

From the distribution assumption in equation (2) we have that

$$h_t = \omega + \alpha\, \eta_{t-1}^2\, h_{t-1} + \beta h_{t-1},$$

where η_t is a standard normal random variable. From Section 3 we know that the position of the point clusters in the $(h_t,\, h_{t-1})$-subspace is the cause of "spurious almost-integration". From the conditional variance equation above we see that the relation between h_t and h_{t-1} is much more direct for the term with the β coefficient than for the term with the α coefficient. The α coefficient is cushioned by the random process, which may be one reason why changes in α have a less dramatic impact on $\hat{\lambda}$.

To study parameter changes in β, I consider model (22) with the conditional variance equation

$$h_t = 1.6\text{e-}5 + 0.10\, \varepsilon_{t-1}^2 + \beta_i\, h_{t-1}. \tag{24}$$

As before, the range of considered volatilities in the second segment is 10 to 25 percent and the β_2 values for the second segment are chosen accordingly. Table 3 shows that changes in β cause "spurious almost-integration" very quickly. For example, for the sample size $N = 4000$ and a switch from 10 to 13 percent volatility the mean of the data-generating λ's within segments is 0.69. However, the switch moves $\hat{\lambda}$ to 0.93, whereas a switch to 15 percent causes "spurious almost-integration".

I ran the same simulations for changes in the constant ω in the data generating process $h_t = \omega_i + 0.30\,\varepsilon_{t-1}^2 + 0.50\,h_{t-1}$. For changes in α and β, I started at the process $h_t = 1.6\text{e-}5 + 0.30\,\varepsilon_{t-1}^2 + 0.30\,h_{t-1}$, which implies 10 percent volatility, and increased the autoregressive parameters in ten steps until the implied volatility reached 20 percent. The results were very similar to the ones reported here.

Also, I simulated changes in all parameters of two GARCH(2,2) models, following the same jump size grid for the annualized volatility. The "spurious almost-integration" results were very similar to the GARCH(1,1) case. These simulations are available from the author upon request. This demonstrates that in the presence of parameter changes that are not accounted for, higher-order GARCH models are not suited to capturing multiple short time scales.

In summary, "spurious almost-integration" occurs for sample sizes and jump sizes of realistic magnitudes. Even when the estimated sum of the autoregressive parameters is not pushed close to one, it is severely overestimated when there are neglected change-points in the data. In the cases of parameter changes in the constant ω and in the second autoregressive parameter β of the conditional variance equation, the speed of the convergence of the estimated sum of the autoregressive parameters to one is faster than in the case of changes in the first autoregressive parameter α.

5 Conclusion

Neglecting changes in the data-generating parameters of the conditional variance of GARCH models causes a substantial overestimation of the autoregressive parameters of the conditional variance. In this paper, I show that this effect stems from the geometry of the estimation problem. Different parameter regimes imply different means of volatility. In the space spanned by the time series h_t and its own lags, the different means of the segments are aligned along the identity. Therefore, when GARCH is estimated with data that contain different means, the estimated autoregressive parameters pick up the slope of the identity. This effect causes the estimated sum of the autoregressive parameters of the conditional variance in GARCH to grow close to unity as both the sample size and the parameter jump size increases. The effect does not depend on the estimation method. It is not necessary to have a certain underlying stochastic structure that drives the parameter changes; a single deterministic change-point is sufficient.

In simulations of the GARCH model, I show that this effect occurs for sample sizes and jump sizes that are realistic for financial market volatility. Therefore, whenever GARCH is estimated on long-range data sets that contain change-points, it will show an estimated sum of the autoregressive parameters close to one. It will indicate high persistence even though the average persistence within segments of constant parameters may be low.

These results do not imply that the long memory found in financial volatility data is in fact spurious. It may well be that a genuine long-memory process and occasional change-points co-exist in the data. The findings show, however, that in the presence of neglected parameter change-points, GARCH is an inappropriate model for measuring volatility persistence. Before carrying out GARCH estimations on financial volatility data, a thorough change-point study of the data is necessary in order to avoid this “spurious almost-integration” effect.

Acknowledgements

The main research for this paper was done while the author was visiting the Department of Mathematics at Stanford University. I am deeply indebted to George Papanicolaou for his hospitality and many intense discussions. The paper benefited from discussions with Caio Almeida, Arnold Kim, Mordecai Kurz, Jonathan Mattingly, and Knut Sølna, the participants of the ESCR conference on long memory, structural breaks and stock market volatility, and from the comments of the anonymous referees. Thanks to David Brasington for a native speaker's eye on the text. Remaining errors of any kind are mine.

References

Andersen, T. G. and T. Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, Journal of Empirical Finance 4, 115–158.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and H. Ebens, 2001, The distribution of stock return volatility, Journal of Financial Economics 61, 43–76.

Andreou, E. and E. Ghysels, 2002, Detecting multiple breaks in financial market volatility dynamics, Journal of Applied Econometrics 17, 579–600.

Baillie, R. T. and R. P. DeGennaro, 1990, Stock returns and volatility, Journal of Financial and Quantitative Analysis 25, 203–214.

Baillie, R. T., Bollerslev, T. and H. O. Mikkelsen, 1996, Fractionally integrated generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 74, 3–30.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 31, 307–327.

Bollerslev, T., 1987 , A conditionally heteroskedastic time series model for speculative prices and rates of return, Review of Economics and Statistics 69, 542–547.

Bollerslev, T. and J. M. Wooldridge, 1992, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, Econometric Reviews 11, 143–172.

Bollerslev, T. and R. F. Engle, 1993, Common persistence in conditional variances, Econometrica 61, 167–186.

Bos, C. S., Franses, P. H. and M. Ooms, 1999, Long memory and level shifts: Re-analyzing inflation rates, Empirical Economics 24, 427–449.

Chernov, M., Gallant, A.R., Ghysels, E. and G. Tauchen, 2003, Alternative models of stock-price dynamics, Journal of Econometrics 116, 225-257.

Choi, K. and E. Zivot, 2002, Long memory and structural changes in the forward discount: An empirical investigation, mimeo. `http://faculty.washington.edu/ezivot/ezresearch.htm`.

Diebold, F. X., 1986, Modeling the persistence of conditional variances: A comment, Econometric Reviews 5, 51–56.

Diebold, F. X. and A. Inoue, 2001, Long memory and regime switching, Journal of Econometrics 105, 131–159.

Ding, Z., Granger, C. W. J. and R. F. Engle, 1993, A long memory property of stock market returns and a new model, Journal of Empirical Finance 1, 83–106.

Ding, Z. and C. W. J. Granger, 1996, Modeling volatility persistence of speculative returns: A new approach, Journal of Econometrics 73, 185–215.

Doornik, J. A. and M. Ooms, 2000, Multimodality and the GARCH likelihood, Working paper. `http://www.feweb.vu.nl/econometriclinks/ooms/papers/`

Engle, R. F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica 50, 987–1007.

Engle, R. F. and T. Bollerslev, 1986, Modelling the persistence of conditional variances, Econometric Reviews 5, 1–50.

Engle, R. F. and G. G. J. Lee, 1999, A long-run and short-run component model of stock return volatility, In: Engle, R. F. and H. White, eds., Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger (Oxford University Press, Oxford).

Engle, R. F. and A. J. Patton, 2001, What good is a volatility model? Quantitative Finance 1, 237–245.

Fouque, J. P., Papanicolaou, G., Sircar, K. R. and K. Sølna, 2003, Short time-scale in S&P 500 volatility, Journal of Computational Finance 6, 1–23.

Francq, C., Roussignol, M. and J. M. Zakoïan, 2001, Conditional heteroskedasticity driven by hidden Markov chains, Journal of Time Series Analysis 22, 197–220.

Gallant, A.R. and G. Tauchen, 2001, Efficient method of moments, mimeo. `http://www.unc.edu/~arg`.

Ghashgaie, S., Breymann, W., Peinke, J., Talkner, P. and Y. Dodge, 1996, Turbulent cascades in foreign exchange markets, Nature 381, 767–770.

Granger, C. W. J., 1980, Long memory relationships and the aggregation of dynamic models, Journal of Econometrics 14, 227–238.

Granger, C. W. J. and N. Hyung, 1999, Occasional structural breaks and long memory, UC San Diego Discussion Paper 99-14.

Granger, C. W. J. and T. Teräsvirta, 2001, A simple nonlinear time series model with misleading linear properties, Economics Letters 62, 161–165.

Gray, S. F., 1996, Modeling the conditional distribution of interest rates as a regime-switching process, Journal of Financial Economics 42, 27–62.

Hamilton, J. D. and R. Susmel, 1994, Autoregressive conditional heteroskedasticity and changes in regime, Journal of Econometrics 64, 307–333.

Kokoszka, P. and R. Leipus, 1999, Testing for parameter changes in ARCH models, Lithuanian Mathematical Journal 39, 182–195.

Kokoszka, P. and R. Leipus, 2000, Changepoint estimation in ARCH models, Bernoulli 6, 513–539.

Lamoureux, C. G. and W. D. Lastrapes, 1990, Persistence in variance, structural change, and the GARCH model, Journal of Business and Economic Statistics 8, 225–234.

Lavielle, M. and E. Moulines, 2000, Least-squares estimation of an unknown number of shifts in time series, Journal of Time Series Analysis 20, 33–60.

LeBaron, B., 2001, Stochastic volatility as a simple generator of apparent financial power laws and long memory, Quantitative Finance 1, 621–631.

Lobato, I. N. and N. E. Savin, 1998, Real and spurious long-memory properties of stock-market data, Journal of Business and Economic Statistics 16, 261–268.

Lumsdaine, R., 1995, Finite-sample properties of the maximum-likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: A Monte Carlo investigation, Journal of Business and Economic Statistics 13, 1–10.

Lumsdaine, R., 1996, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, Econometrica 64, 575–596.

Mikosch, T. and C. Starica, 2000, Change of structure in financial time series, long range dependence and the GARCH model, Forthcoming in The Review of Economics and Statistics. `http://www.math.ku.dk/ mikosch/preprint.html`

Mueller, U. A., Dacorogna, M. M., Dave, R. D., Olsen, R. B., Pictet, O. V., and J. E. von Weizsaecker, 1997, Volatilities of different time resolutions - Analyzing the dynamics of market components, Journal of Empirical Finance 4, 213–239.

Ohannissian, A., Russell, J. and R. Tsay, 2003, Using temporal aggregation to distinguish between true and spurious long memory, Working paper. `http://gsb-www.uchicago.edu/fac/jeffrey.russell/research/wp.htm`.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and B. P. Flannery, 2002, Numerical Recipes in C++ (Cambridge University Press, Cambridge).

Sakoulis, G. and E. Zivot, 2000, Time-variation and structural change in the forward discount: Implications for the forward rate unbiasedness hypothesis, mimeo.`http://faculty.washington.edu/ezivot/ezresearch.htm`.

Weiss, A. A., 1986, Asymptotic theory for ARCH models: Estimation and testing, Econometric Theory 2, 107–131.

Zumbach, G., 2000, The pitfalls in fitting GARCH(1,1) processes, In: Dunis, C., ed., Advances in Quantitative Asset Management (Kluwer, Dordrecht).

Table 1: Effects of a single change-point in the constant ω of a GARCH(1,1) model with $\mu = 0$ and $h_t = \omega_i + 0.10\varepsilon_{t-1}^2 + 0.70h_{t-1}$ on the estimators $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$ when the change-point is not accounted for and a GARCH(1,1) model is estimated on the entire series. The change-point always occurs in the middle of the sample. The annualized standard deviation σ_1 implied by the first segment is always 0.10. The entries are the means of 10,000 simulations.

$N(N_1)$			200(100)				800(400)				4000(2000)				10000(5000)			
σ_2	ω_1	ω_2	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
.100	8e-6	8.0e-6	2e-5	.11	.47	.58	1e-5	.10	.64	.74	8e-6	.10	.69	.79	8e-6	.10	.70	.80
.117	8e-6	1.1e-5	2e-5	.11	.50	.62	1e-5	.10	.70	.80	8e-6	.10	.74	.84	7e-6	.10	.74	.84
.133	8e-6	1.4e-5	2e-5	.12	.59	.71	6e-6	.10	.80	.89	4e-6	.09	.85	.93	4e-6	.09	.85	.93
.150	8e-6	1.8e-5	1e-5	.13	.68	.80	3e-6	.08	.87	.95	1e-6	.06	.92	.98	1e-6	.05	.93	.98
.167	8e-6	2.2e-5	1e-5	.13	.74	.87	2e-6	.08	.90	.98	6e-7	.05	.94	.99	4e-7	.04	.95	.99
.183	8e-6	2.7e-5	7e-6	.13	.78	.92	1e-6	.08	.91	.99	4e-7	.05	.95	.99	3e-7	.04	.96	1
.200	8e-6	3.2e-5	5e-6	.14	.81	.95	1e-6	.08	.91	.99	4e-7	.05	.95	1	2e-7	.04	.96	1
.217	8e-6	3.8e-5	4e-6	.14	.82	.96	1e-6	.08	.91	.99	3e-7	.05	.95	1	2e-7	.04	.96	1
.233	8e-6	4.4e-5	4e-6	.15	.82	.97	1e-6	.09	.91	.99	3e-7	.05	.95	1	2e-7	.04	.96	1
.250	8e-6	5.0e-5	4e-6	.15	.82	.97	1e-6	.09	.90	.99	3e-7	.05	.94	1	2e-7	.04	.96	1

Table 2: Effects of a single change-point in the α parameter of a GARCH(1,1) model with $\mu = 0$ and $h_t = 1.6\text{e-}5 + \alpha_i \varepsilon_{t-1}^2 + 0.50 h_{t-1}$ on the estimators $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$ when the change-point is not accounted for and a GARCH(1,1) model is estimated on the entire series. The change-point always occurs in the middle of the sample. The annualized standard deviation σ_1 implied by the first segment is always 0.10. In the first segment, the parameter is always $\alpha_1 = 0.10$ so that the sum λ_1 of the autoregressive parameters in the first segment is always 0.60. The column titled $\bar{\lambda}$ reports the average $.5\lambda_1 + .5\lambda_2$. The entries are the means of 10,000 simulations.

$N(N_1)$				200(100)				800(400)				4000(2000)				10000(5000)			
σ_2	α_1	α_2	$\bar{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
.100	.10	.10	.60	2e-5	.11	.36	.46	2e-5	.10	.43	.54	2e-5	.10	.49	.59	2e-5	.10	.49	.59
.117	.10	.21	.66	2e-5	.16	.41	.57	2e-5	.16	.50	.66	1e-5	.16	.53	.69	1e-5	.16	.53	.70
.133	.10	.28	.69	2e-5	.21	.45	.66	1e-5	.21	.54	.75	1e-5	.21	.56	.77	1e-5	.21	.56	.77
.150	.10	.32	.71	2e-5	.24	.48	.72	1e-5	.24	.56	.80	1e-5	.25	.57	.82	1e-5	.25	.57	.82
.167	.10	.36	.73	2e-5	.26	.51	.76	1e-5	.27	.56	.83	1e-5	.28	.57	.85	1e-5	.28	.58	.85
.183	.10	.38	.74	1e-5	.28	.51	.79	1e-5	.29	.57	.86	1e-5	.30	.58	.88	1e-5	.30	.58	.88
.200	.10	.40	.75	1e-5	.29	.52	.81	1e-5	.30	.57	.87	1e-5	.31	.58	.89	1e-5	.32	.58	.90
.217	.10	.42	.76	1e-5	.30	.52	.83	1e-5	.31	.57	.89	1e-5	.32	.58	.91	1e-5	.33	.58	.91
.233	.10	.43	.77	1e-5	.31	.53	.84	1e-5	.32	.57	.90	1e-5	.33	.58	.92	1e-5	.34	.58	.92
.250	.10	.44	.77	1e-5	.32	.53	.85	1e-5	.33	.57	.90	1e-5	.34	.58	.92	1e-5	.34	.58	.93

Table 3: Effects of a single change-point in the β parameter of a GARCH(1,1) model with $\mu = 0$ and $h_t = 1.6\text{e-}5 + 0.10\varepsilon_{t-1}^2 + \beta_i h_{t-1}$ on the estimators $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$ when the change-point is not accounted for and a GARCH(1,1) model is estimated on the entire sample. The change-point always occurs in the middle of the series. The annualized standard deviation σ_1 implied by the first segment is always 0.10. In the first segment, the parameter is always $\beta_1 = 0.50$ so that the sum λ_1 of the autoregressive parameters in the first segment is always 0.60. The column titled $\bar{\lambda}$ reports the average $.5\lambda_1 + .5\lambda_2$. The entries are the means of 10,000 simulations.

$N(N_1)$				200(100)				800(400)				4000(2000)				10000(5000)			
σ_2	β_1	β_2	$\bar{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
.100	.50	.50	.60	2e-5	.11	.36	.46	2e-5	.10	.43	.54	2e-5	.10	.49	.59	2e-5	.10	.49	.59
.117	.50	.61	.66	2e-5	.11	.43	.54	1e-5	.10	.58	.68	1e-5	.10	.64	.74	1e-5	.10	.64	.74
.133	.50	.68	.69	2e-5	.12	.55	.67	8e-6	.09	.76	.86	4e-6	.07	.85	.93	3e-7	.07	.87	.94
.150	.50	.72	.71	1e-5	.12	.65	.78	3e-6	.08	.87	.95	1e-6	.05	.94	.98	6e-7	.04	.96	.99
.167	.50	.76	.73	1e-5	.13	.73	.86	2e-6	.08	.90	.98	6e-7	.05	.95	.99	4e-7	.04	.96	.99
.183	.50	.78	.74	7e-6	.13	.78	.91	1e-6	.08	.91	.98	5e-7	.05	.94	.99	4e-7	.04	.95	1
.200	.50	.80	.75	6e-6	.14	.80	.94	1e-6	.08	.91	.99	5e-7	.06	.94	.99	4e-7	.05	.95	1
.217	.50	.81	.76	5e-6	.14	.81	.95	1e-6	.09	.90	.99	5e-7	.06	.94	1	4e-7	.05	.94	1
.233	.50	.83	.77	4e-6	.14	.82	.97	1e-6	.09	.90	.99	5e-7	.06	.93	1	4e-7	.06	.94	1
.250	.50	.84	.77	4e-6	.15	.82	.97	1e-6	.10	.90	.99	5e-7	.07	.93	1	4e-7	.06	.94	1