



Misspecification and domain issues in fitting Garch(1,1) models: a Monte Carlo investigation

Leonardo Bottolo

► To cite this version:

Leonardo Bottolo. Misspecification and domain issues in fitting Garch(1,1) models: a Monte Carlo investigation. Communications in Statistics - Simulation and Computation, 2008, 38 (01), pp.31-45. 10.1080/03610910802395653 . hal-00514333

HAL Id: hal-00514333

<https://hal.science/hal-00514333>

Submitted on 2 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Misspecification and domain issues in fitting Garch(1,1) models: a Monte Carlo investigation

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2007-0055.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	01-Aug-2008
Complete List of Authors:	Bottolo, Leonardo; Imperial College, Institute for Mathematical Sciences
Keywords:	Innovation distribution, IGARCH effect, Monte Carlo simulations, Weak and strong stationarity, volatility forecasting
Abstract:	In this work we investigate the impact of misspecification of the innovations in fitting Garch\$(1,1)\$ models. We show that an incorrect specification of the innovations together with the reduction of the parameter space to the weak stationarity region, can give rise to a spurious IGARCH effect. We address this point through an extensive Monte Carlo simulation study. We also analyse the impact of misspecification on forecasted volatilities, showing that innovations with light tails can lead to a remarkable overestimate of volatilities.
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
Bellini-Bottolo-miss-revision.zip	



For Peer Review Only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Misspecification and domain issues in fitting Garch(1, 1) models: a Monte Carlo investigation

Fabio Bellini

University of Milano-Bicocca, Italy

fabio.bellini@unimib.it

Leonardo Bottolo*

Imperial College London, UK

l.bottolo@imperial.ac.uk

Abstract

In this work we investigate the impact of misspecification of the innovations in fitting Garch(1, 1) models. We show that an incorrect specification of the innovations together with the reduction of the parameter space to the weak stationarity region, can give rise to a spurious IGARCH effect. We address this point through an extensive Monte Carlo simulation study. We also analyse the impact of misspecification on forecasted volatilities, showing that innovations with light tails can lead to a remarkable overestimate of volatilities.

Keywords: Innovation distribution; IGARCH effect; Monte Carlo simulations; Weak and strong stationarity; volatility forecasting.

1 Introduction

It is usually considered a general fact that volatility, at least to some degree, can be forecasted. To this aim, the most popular class of models both in literature and amongst practitioners are the GARCH models. Despite their simplicity, they are generally believed to capture some of the basic properties of financial time series. Moreover, their exact structure is nontrivial and begins to be fully understood only in some recent probabilistic works; see Berkes, Horváth and Kokoszka, 2003 and 2004.

The aim of this paper is to investigate some aspects of GARCH modelling and fitting that are often omitted but could lead to a significative impact on real data applications.

The first step in fitting GARCH models is the choice of the distribution of the innovations. Limiting the analysis to the univariate case, many different densities have been used: standard normal

*Address for correspondence: Institute for Mathematical Sciences, Imperial College, 53 Prince's Gate, London SW7 2PG, UK.

(Bollerslev, 1986), Student- t (Bollerslev, 1987), generalized error (Engle and González-Rivera, 1991 and Nelson, 1991), gamma (Engle and González-Rivera, 1991), α -stable (Mittnik, Paoella and Rachev, 2002), max-entropic (Rockinger and Jondeau, 2002) and many others. Recently a more flexible family of innovations based on mixture models has been introduced both in the classical (Haas, Mittnik and Paoella, 2004) and bayesian analysis (Ausín and Galeano, 2007).

However the general view is that the correct distribution of the innovations is not so important: this is based on asymptotic consistency results that, under mild assumptions, ensure that with Gaussian innovations the estimates are asymptotically unbiased and normally distributed (Bollerslev and Wooldridge, 1992). The most general version of these results, reviewed in Section 2, is given in Berkes, Horváth and Kokoszka (2003).

From a practical perspective, it is considered not realistic to fit a GARCH model on time periods longer than one year (roughly) otherwise non-stationarity could play a significant role (Mikosch and Stărică, 2004a and 2004b, Stărică, Herzel and Nord, 2006). Since the general asymptotic results do not hold in this case, an incorrect specification of the innovations could impact the parameters estimates: this “misspecification issue” has been investigated for example in Engle and González-Rivera (1991).

After selecting a suitable model for the innovations, a second step is the choice of the domain where the Quasi-Maximum Likelihood estimate is performed. It is well known, see for example Nelson (1990), Bougerol and Picard (1992), Mittnik, Paoella and Rachev (2002) and Bellini and Bottolo (2007) that the domain of strong stationarity of a Garch(1,1) process is strictly bigger than the domain of covariance stationarity: this unusual behaviour (with respect to other time series) corresponds to the case where, for some sets of parameters, there exist stationary solutions with infinite unconditional variance. Although this may seem unrealistic, we notice that the observed “IGARCH effect” would also imply an infinite unconditional variance.

Apart from being bigger than the weak stationarity domain, the strong stationarity domain depends crucially on the (unknown) density of the innovations. Analytic expressions for both domains are known only in the Gaussian and in the Cauchy case, but numerical simulations show that the heavier the tails of the innovations, the bigger the domains, Mikosch (2004). In practical applications this issue is usually not considered and the estimates are performed simply on the weak stationarity domain. However this is a rigid assumption since the strong stationarity domain can be much bigger, causing differences in the parameters estimates compared to the weak stationarity constraint.

Quite often when a Garch(1,1) model is estimated on real data, the parameters lie exactly on the boundary of the weak stationarity domain. This situation corresponds to the aforementioned “IGARCH effect” frequently found in the empirical literature, and it has been interpreted as an extreme persistence of shocks in the volatility; see Bollerslev, Chou and Kroner (1992) and references therein. However its interpretation is not unique: it could represent a sort of “long memory of the

volatility shocks”, see for example Ding, Engle and Granger (1993) or the “IGARCH effect” and more generally long memory processes, could be an artifact due to non-stationarity in the model’s parameters caused by excessively long time series, see for example Mikosch and Stărică (2004a).

In this paper we suggest that the IGARCH effect could be differently explained by a combined effect of misspecification of the innovations that increases the variability of the estimates and the weak stationarity constraint that “ties” all the estimates to the boundary. In order to test this hypothesis we performed an extensive Monte Carlo study of Garch(1, 1) time series of different lengths simulated under different innovations, i.e. generalised error, normal and t distribution respectively. Each time series is then estimated under a Garch(1, 1) model with different fitting innovations, adding/removing the weak stationary constraint.

The paper is organised as follows. In Section 2 we review the general results about GARCH models and in particular the domain of weak and strong stationarity, WS and SS henceforth. Then we compare the exact domain of SS with the simulated one for normal innovations and present the simulated SS domain for the generalised error distribution, GED henceforth, and the t distribution. In Section 3 we discuss in detail the results of the simulation study: in particular we report the marginal and joint empirical distribution of the parameters estimates under misspecification of the innovations as well as the impact of misspecification on the estimated volatilities. Finally Section 4 contains some concluding remarks and guidance for practitioners.

2 General properties of Garch(1, 1) process

The Garch(1, 1) process is described by the equations

$$X_t = \sigma_t Z_t \quad (1)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (2)$$

with $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and where Z_t , $t \in \mathbb{Z}$, is an i.i.d. sequence such that $E(Z_0) = 0$ and $E(Z_0^2) = 1$ with Z_0 the common distribution of Z_t . It has been proved in Nelson (1990) that (1) and (2) have a unique strong stationary and ergodic solution if and only if

$$E[\log(\alpha_1 Z_0^2 + \beta_1)] < 0. \quad (3)$$

The above result is a special case of the general necessary and sufficient condition for the SS of a Garch(p, q) process provided in Bougerol and Picard (1992). A trivial necessary condition for (3) is $\beta_1 < 1$, while a sufficient condition is

$$\alpha_1 + \beta_1 < 1. \quad (4)$$

Condition (4) is also necessary and sufficient for the existence of a WS solution of (1) and (2) and in particular it is a necessary and sufficient condition for the variance of X_t to be finite. We define the (α_1, β_1) region specified by (3) as the SS region, while we refer to the region defined by (4) as the WS region. For GARCH models this corresponds to the unusual situation (with respect to other time series) where the SS region is strictly bigger than the WS region, i.e. there are strong stationary solutions that are not weakly stationary since the unconditional variance becomes infinite in the SS region. The exact form of the SS domain (3) depends on the distribution of the (unknown) innovations Z_0 and has been computed explicitly only in the Gaussian and in the Cauchy case (Nelson, 1990).

It is a well known fact that the tails of the unconditional distribution can be Paretian even if the distribution of the innovations has light tails; in Nelson (1990) an exact characterisation of the *tail index* of the unconditional variance σ^2 is given: if

$$E[(\alpha_1 Z_0^2 + \beta_1)^{\tau/2}] = 1 \quad (5)$$

and $E[(\alpha_1 Z_0^2 + \beta_1)^p] < 1$ for each $0 < p < \tau/2$, then

$$E(\sigma^2)^p = \begin{cases} < +\infty & \text{if } p < \tau/2 \\ = \infty & \text{if } p \geq \tau/2. \end{cases} \quad (6)$$

Under the additional hypothesis that $E(|Z_0|^{\tau+\varepsilon}) < +\infty$, $\varepsilon > 0$, it is possible to show that X_t has tail index τ : this suggests that knowing the parameters α_1 and β_1 and the distribution of the innovations Z_0 , it is possible to calculate theoretically the value of the tail index τ and check if the GARCH model can reproduce the empirical tails, as suggested in Mikosch and Stărică (2004a). In the particular case $\alpha_1 + \beta_1 = 1$, it is easy to see from (5) and (6) that $\tau = 2$, hence leading to an infinite unconditional variance for the IGARCH model.

2.1 Strong stationarity domains: normal case

In Nelson (1990) the integral in (3) is explicitly calculated when Z_0 has a normal distribution. We recall here Nelson's results and compare the exact SS region with the corresponding region derived numerically. If $Z_0 \sim N(0, 1)$, $\beta_1 > 0$, then

$$E[\log(\alpha_1 Z_0^2 + \beta_1)] = \log(2\alpha_1) + \varphi\left(\frac{1}{2}\right) + \left(\frac{2\pi\beta_1}{\alpha_1}\right)^{1/2} \Phi\left(\frac{1}{2}, \frac{3}{2}; \frac{\beta_1}{2\alpha_1}\right) - \left(\frac{\beta_1}{\alpha_1}\right) {}_2F_2\left(1, 1; 2, \frac{3}{2}; \frac{\beta_1}{2\alpha_1}\right) \quad (7)$$

and

$$E(\alpha_1 Z_0^2 + \beta_1)^p = (2\alpha_1)^{-\frac{1}{2}} \beta_1^{p+1/2} \Psi\left(\frac{1}{2}, p + \frac{3}{2}; \frac{\beta_1}{2\alpha_1}\right),$$

where $\Phi(a, b; z) = \sum_{k=0}^{\infty} \frac{(a)_k}{(b)_k} \frac{z^k}{k!}$, with increasing factorial $(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}$, ${}_2F_2(a, b; c, d; z)$ is a Gaussian hypergeometric function defined as ${}_2F_2(a, b; c, d; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k (d)_k} \frac{z^k}{k!}$ (Abramowitz and Stegun, 1970, Section 15), $\Psi(a, b; z)$ is confluent hypergeometric function of type 2 and φ is Euler psi function.

Since we are interested in the general shape of SS region under different innovations, for the normal case we perform a Monte Carlo integration of (3) as $n^{-1} \sum_{i=1}^n \log(\alpha_1 z_n^2 + \beta_1) \xrightarrow{SLLN} E[\log(\alpha Z_0^2 + \beta)]$. To verify the accuracy of the approximation, in Figure 1 (a) we plot the exact and simulated SS region for different samples size, $n = 1,000, 10,000$ and $100,000$ respectively: as expected with a relative small sample size the simulated SS region is a coarse approximation of the analytic one (7). This is not very important in the normal case as an exact expression for the SS domain exists, but this could represent an obstacle for other distributions where an analytic expression does not.

Figure 1 about here

2.2 Strong stationarity domains: non-normal case

Here we focus on two classical families of innovations:

- i) standardised t distribution with δ degrees of freedom, DoF henceforth,

$$f(x) = \frac{\Gamma(\frac{\delta+1}{2})}{\Gamma(\frac{\delta}{2})} \sqrt{\frac{\delta-2}{\pi\delta^2}} \left\{ 1 + \frac{(\delta-2)x^2}{\delta^2} \right\}^{-(\delta+1)/2};$$

- ii) standardised GED with shape parameter ν

$$f(x) = \frac{\nu \exp\{-\frac{1}{2}|x|^\nu\}}{\beta \Gamma(\frac{1}{\nu}) 2^{(1+\frac{1}{\nu})}}$$

with

$$\beta = \left(\frac{2^{-2/\nu} \Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})} \right)^{1/2}.$$

The t family is chosen as a prototypical “supergaussian” family (with tails heavier than the normal distribution, including the normal for $\delta \rightarrow \infty$). The GED family incorporates the normal distribution as a special case for $\nu = 2$, while is “supergaussian” for $\nu < 2$ and “subgaussian” if $\nu > 2$.

For each distribution and for different values of the shape parameters δ and ν , we compute numerically the SS domain approximating the integral in (3) drawing 100,000 random numbers from Z_0 in the same fashion we did for the normal case. The results, depicted in Figure 1 (b), show that, at least for the set of distributions considered, the heavier the tails, the bigger the SS domain. This could be simply explained taking into account the approximation $\log(1+x) \sim x - x^2/2$ of (3)

$$E[\log(\alpha_1 Z_0^2 + \beta_1)] \approx (\alpha_1 + \beta_1 - 1) - \frac{1}{2}(\alpha_1 + \beta_1 - 1)^2 - \frac{1}{2}\alpha_1^2 [E(Z_0^4) - 1] \quad (8)$$

showing that as $E(Z_0^4)$ increases, the stationarity domain increases. As a consequence if the “true” or “underlying” innovations have heavy tails, than the SS region can be bigger than the WS region.

2.3 Properties of QML estimate

The estimate of the model (1) and (2) is often performed *via* Maximum Likelihood with Gaussian innovations. This means that even if the distributional assumption of the innovations are not correct, ML with Gaussian innovations provides reasonable parameter estimates. We refer to this procedure with Quasi-Maximum Likelihood, QML henceforth.

More precisely the estimates $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\beta}_1$ are obtained maximising the Gaussian log-likelihood

$$\ell_T = -\frac{1}{2} \sum_{t=1}^T (\log \sigma_t^2 + \sigma_t^{-2} x_t^2) \tag{9}$$

in the WS domain. This automatically guarantees that the unconditional variance σ^2 is finite and that the maximisation is performed on a compact set independent from the distribution of the underlying innovations. The solution of (9) has attractive asymptotic properties: for example, it has been proved in Berkes, Horváth and Kokoszka (2003) that, under the very mild conditions $E(|Z_0|^{2+\varepsilon}) < +\infty$, $\varepsilon > 0$ and $\lim_{t \rightarrow 0} t^{-\mu} \Pr \{Z_0 \leq t\} = 0$ for some $\mu > 0$, provided that $E(Z_0^2) = 1$, the QML estimates with Gaussian innovations are asymptotically unbiased. This result refines previous results of Lumsdaine (1996) and Lee and Hansen (1994). Under the more restrictive hypothesis $E(Z_0^4) < +\infty$, it is possible to prove the asymptotic normality of the QML estimators.

In practical applications the residuals display some departure from normality, see for example Bollerslev, Chou and Kroner (1992). However the theoretical asymptotic properties of the QML estimators for non-normal innovations and under misspecification are much less clear. One crucial result can be found in Newey and Steigerwald (1997) that once again guarantees asymptotic unbiasedness and normality if both underlying and fitting innovations are unimodal and symmetric around 0. This covers the normal, t and GED distributions that we used in the simulation study, but does not prevent a systematic bias for example in the case of skewed innovations.

3 Simulation study

In this Section we describe the simulation study we performed in order to investigate the effect of the misspecification in the parameters estimates and forecasted volatilities. We considered four different Garch(1,1) processes with the following parameters:

	A	B	C	D
α_0	0.1	0.02	0.1	0.1
α_1	0.1	0.2	0.15	1.6
β_1	0.5	0.75	0.85	0.1

(10)

Models A and B lie inside the WS region, while C is on the WS boundary and D is outside the WS region, but inside the SS domain for every considered innovation distribution Z_0 , see Figure 1 (b). Models A and C have been considered in Lumsdaine (1995), while Nakatsuma and Tsurumi (1996) analysed model D .

For each model 5,000 time series of length $T = 250$ and $T = 1,000$ have been simulated each of which with three possible innovations: standard normal, N henceforth, standardised t with $\delta = 3$, t_3 henceforth, and standardised GED with shape parameter $\nu = 6$, GED_6 henceforth. Then for each time series the MLE of a Garch(1,1) has been calculated under three different fitting innovations, N , standardised t with δ DoF, t_δ henceforth, and standardised GED with shape parameter ν , GED_ν henceforth. We checked *a posteriori* if for each simulated case the estimated parameters laid inside the SS region: in principle one should impose the SS constraint in the maximisation algorithm, but this is not easy since, as noticed before, we have an explicit (although quite involved) description of the SS domain only for normal innovations, see Figure 1.

All simulations and estimates has been performed in MATLAB: in particular we used the UCSD GARCH Toolbox (with minor corrections) by Sheppard (2002). In the UCSD GARCH Toolbox the initial values of $(\alpha_0, \alpha_1, \beta_1)$ are generated automatically and are inside the WS region. Moreover in order to prevent transient effects on the estimated volatilities, the algorithm oversamples 500 observations and deletes the same number from the beginning of the simulated series of the volatilities.

Apart from the fitted parameter values, we are also interested in the error on the forecasted volatilities, that could give rise to errors on the forecasted Value-at-Risk. In order to measure these errors we use two indices:

- i) Mean Relative Error on volatilities, MRE henceforth,

$$\text{MRE} = \frac{1}{T} \sum_{t=1}^T \frac{|\sigma_t - \hat{\sigma}_t|}{\sigma_t}, \quad (11)$$

where σ_t and $\hat{\sigma}_t$ are the sequences of simulated and estimated volatilities respectively. Since (11) does not depend on the absolute order of magnitude of the volatilities, it is an effective tool to compare the impact of misspecification on $\hat{\sigma}_t$ under different models;

- ii) Frequency of Underestimated Volatilities, FUV henceforth,

$$\text{FUV} = \frac{1}{T} \sum_{t=1}^T I_{\hat{\sigma}_t < \sigma_t} \quad (12)$$

that measures the fraction of times in which the estimated volatilities are smaller than the simulated ones.

In the following we focus the analysis on: i) the marginal empirical distribution of the parameter estimates $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$ and $\hat{\delta}$ or $\hat{\nu}$, see Table 1; ii) the joint empirical distribution of $(\hat{\alpha}_1, \hat{\beta}_1)$, see Figure 2; iii) the empirical distribution of MRE and FUV, see Table 2.

3.1 Marginal empirical distribution

The results on the marginal distribution of the estimated parameters are presented in Table 1 for model *A* and it is constructed as follows: the left part (columns “ GED_ν ” “ N ” and “ t_δ ”) refers to the estimates with the WS constraint, while the right part (columns “ GED_ν^u ” “ N^u ” and “ t_δ^u ”, where u stands for unconstrained) refers to the estimates performed without it. The upper part of the table indicates simulations with $T = 250$, while the lower part with $T = 1,000$. Besides, each upper and lower part is divided according to the innovation distribution used in the simulation step (rows “ GED_6 ” “ N ” and “ t_3 ”). Therefore each table can be thought made of four subtables formed by nine “squares”. Each square refers to a particular pair of simulated *vs.* fitting distributions, the fitting distributions in the columns and the simulated distributions in the rows. We kept this structure for comparison purpose across the various models in all the tables and figures. For simplicity of exposition we omit tables related to models *B*, *C* and *D* although they are available from the authors on request.

Let us consider a single subtable made of nine “squares”. Where do we have misspecification? On the principal diagonal (squares GED_6-GED_ν , $N-N$, t_3-t_δ and analogously $GED_6-GED_\nu^u$, $N-N^u$, $t_3-t_\delta^u$) we have no misspecification. For the other squares there are three cases:

- i) the fitting distribution belongs to a family that incorporates the underlying distribution of the innovations (squares $N-GED_\nu$, $N-t_\delta$ and analogously $N-GED_\nu^u$, $N-t_\delta^u$): we call this case “type-1 misspecification”;
- ii) the fitting distribution has heavier tails than the underlying one (squares GED_6-N , GED_6-t_δ and analogously GED_6-N^u , $GED_6-t_\delta^u$): we call this case “type-2 misspecification”;
- iii) the fitting distribution has lighter tails than the underlying one (squares t_3-GED_ν and t_3-N and analogously $t_3-GED_\nu^u$ and t_3-N^u): we call this case “type-3 misspecification”.

In Table 1 for each square, we show the median and in brackets the median absolute deviation from the median, MAD henceforth, of the 5,000 fitted parameters. There is an extra parameter $\hat{\nu}$ when the model is fitted with the GED family (columns GED_ν and GED_ν^u) as well as $\hat{\delta}$ for the t case (columns t_δ and t_δ^u).

Let us consider for simplicity model *A*, Table 1, whose underlying parameters α_1 and β_1 are well inside the WS region, see Figure 1. Even when there is no misspecification (squares GED_6-GED_ν , $N-N$, t_3-t_δ , $N-GED_\nu$, $N-t_\delta$), there is a clear tendency to *underestimate* parameters α_1 and β_1 and

overestimate α_0 even if this behaviour is less marked for parameter α_1 . Moreover it seems to be much more pronounced when $T = 250$, but is still clearly present for $T = 1,000$. This is not surprising since this phenomenon has been already noticed when the QML method is performed on relative small samples, see Bollerslev and Wooldrige (1992). In order to overcome this problem alternative methods have been proposed in the literature: see for example Mikosch and Straumann (2002) or, for a more promising approach, Baille and Chung (2001). The shape parameters $\hat{\nu}$ and $\hat{\delta}$ are instead estimated reasonable well.

As expected the main difference between the case $T = 250$ and $T = 1,000$ refers to the estimated standard errors that tend to reduce roughly with a factor 0.5. In both cases the presence of misspecification both of type-2 and type-3 does not seem to impact deeply on the median: the negative bias on $\hat{\alpha}_1$ and $\hat{\beta}_1$ seems more a small sample size effect then a problem of correct specification. Instead the misspecification increases the standard error of the estimates: this is apparent for example comparing the cases GED_6-N and GED_6-t_δ and the cases t_3-GED_ν and t_3-N with $T = 1,000$. Things are less clear when the sample size is $T = 250$: consider for example the case GED_6 for which the minimum MAD, across the fitting innovations, is attained at least for one parameter when the distribution is normal. However a clearer picture of the effects of misspecification on the estimates is provided in the next Subsection where the empirical joint distribution of the parameter is presented.

Surprisingly, the relaxation of the WS constraint does not make a big difference, at least for the median of the estimated parameters: the maximisation process works pretty well, converging soon to the correct values even without the boundary not only for model A , which simulated parameter are well inside the boundary of SS, but also for the other models (data not shown).

3.2 Joint distribution under misspecification

The joint empirical distribution of parameters $\hat{\alpha}_1$ and $\hat{\beta}_1$ are plotted for model A in Figure 2. In order to have a clearer picture, we smoothed the joint distribution for all models through a Gaussian kernel density, see for example Silverman (1986) for a general introduction to kernel estimates.

Let us first consider the estimates *without* the WS constraint. Even when there is no misspecification, in the case $T = 250$ with simulated t_3 innovations, we observe a non negligible density outside the WS boundary (see square $t_3-t_\delta^u$). Although this is expected in model D , where the parameters are simulated outside the WS boundary, for model A (but the same happens for models B and C , data not shown), it can give rise to a spurious “IGARCH effect”: adding the WS boundary the estimates are “pushed” toward the boundary (see square t_3-t_δ). Moving to the case $T = 1,000$, we notice that this effect is considerably but not totally reduced, particularly in model A where the simulated parameters are well inside the WS domain. Also models B and C show a similar behaviour.

This situation becomes even more evident in type-3 misspecification: for example compare squares t_3-N^u with t_3-N in model A with $T = 1,000$, Figure 2. We claim that this behaviour is due to a *misspecification effect*: if the fitting innovations have lighter tails than the underlying ones, the estimated parameters are tied on the WS boundary in the constrained maximisation, whereas they spread around in the unconstrained one. Therefore it seems that the so called “IGARCH effect”, often found in the empirical literature, can be simply an artifact generated by two main causes:

- i) small sample size, for example $T = 250$, and fat tails innovations, for example t_3 distribution, even under a correct specification of the innovations;
- ii) misspecification of type-3, namely underlying innovations with fatter tails than the fitting innovations.

Moreover it is interesting to note that misspecification of the type-2 does not have the same impact on the empirical distribution of $\hat{\alpha}_1$ and $\hat{\beta}_1$: for example consider squares GED_6-N^u and GED_6-N for model A in Figure 2. Same remarks can be extended to models B, C , data not shown.

3.3 Errors on forecasted volatilities

In the two previous Subsections we analysed the impact of misspecification on the parameters estimates: we roughly found that marginally the misspecification does not impact deeply on the median of the estimated parameters, according to the general theoretical results that assure asymptotic unbiasedness for different families of fitting innovations (Newey and Steigerwald, 1997). Then we observed that results are less clear for the joint distribution of $\hat{\alpha}_1$ and $\hat{\beta}_1$, where misspecification of type-3 combined with the WS boundary can give rise to a spurious “IGARCH effect”.

In this Subsection we analyse how misspecification impacts on the forecasted volatilities. To this aim we computed MRE (11) and FUV (12) on volatilities: the idea behind is that MRE measures the average relative error on volatilities, while FUV measures the fraction of times that the “true” (underlying) volatility is beyond the estimated one. Therefore the index FUV does not carry any information about the magnitude of the error on volatility, but only about the sign of the difference. The median and MAD estimators of the two indexes for the 9 cases of model A are shown in Table 2.

Let us start with the no misspecification case (GED_6-GED_ν , $N-N$, t_3-t_δ) in model A , Table 2. The median of MRE, first row of each square, increases moving from GED_6-GED_ν to $N-N$ and finally to t_3-t_δ : as the tails of the underlying innovations become fatter, the median increases and it is roughly double in the case $T = 250$ than in the case $T = 1,000$. The median of FUV is always around 0.5, with the exception of the t_3 simulated case with $T = 250$. This means that in general under no misspecification there is no systematic underestimating or overestimating of volatilities, with the

exception of the Student t case with $T = 250$. Again, we see that if the underlying innovations have fat tails, then even with no misspecification we can have a systematic bias on volatilities. Moving to $T = 1000$, we observe that the median of FUV is around 0.5 even in the t_3-t_δ case. Analysing the forecasted volatilities in the absence of misspecification, we meet the same pattern that we had with the estimated parameters.

Misspecification of type-2 does not seem to have a great impact on the median values of both MRE and FUV, for example compare squares GED_6-N , GED_6-t_δ . Misspecification of type-3 can be seen analysing squares t_3-GED_ν and t_3-N . We notice two main effects: i) the median of MRE is higher in the misspecified cases t_3-GED_ν and t_3-N than t_3-t_δ , particularly when $T = 1,000$; when $T = 250$ it is very high in all cases; ii) the median of the FUV is very low in the t_3-GED_ν case, low in the t_3-N case and around 0.5 in the t_3-t_δ case. Therefore with this type of misspecification we have a *systematic overestimate of volatilities*, that is remarkable in the t_3-GED_ν case. Similar comments apply also for models B , C and D , data not shown. Summarising also for the forecasted volatilities we discovered the same asymmetric pattern highlighted in the previous Subsections:

- i) if the tails of the fitting distribution are heavier than the underlying ones, we don't see big effects on the median of MRE and FUV;
- ii) if the tails of the fitting distribution are too light then the median of MRE increases, but more strikingly the median of FUV becomes very low, indicating that the estimated volatilities are typically higher than the underlying ones. This behaviour could be simply described as follows: if the tails of the fitting distribution are too light then to match the time series of the unconditional distribution (1), the volatility need to be systematically overestimated.

4 Concluding remarks

In this work we investigated two general issues in fitting GARCH models: the impact of misspecification and the relevance of the WS boundary " $\alpha_1 + \beta_1 = 1$ ". For what concerns the first issue, the situation is quite complex. A first general remark is that both ML and QML estimates show some instability even if the case of no misspecification with samples size $T = 250$, especially when the tails of the underlying innovations are Paretian (t case). Therefore it seems necessary to develop alternative parameter estimate methods in order to overcome these difficulties: Baille and Chung (2001) and Mikosch and Straumann (2002) are promising new approaches.

The impact of the misspecification generates an even greater variability on the estimated parameters. Firstly we saw that its effect is "asymmetric": if the tails of the underlying innovations are lighter than the fitting ones then there isn't any noticeable effect, while if the opposite holds the

impact is severe. Moreover we noticed that this greater variability often gives rise to a spurious “IGARCH effect” when the estimate is performed under the WS constraint. This should be stressed since many empirical works and many numerical implementations of GARCH models add the WS boundary without even mentioning it. In our opinion it would be worth to investigate on real data sets all the cases for which the estimated parameters satisfy exactly the WS boundary.

In our opinion this leads directly to an open methodological problem: how to perform ML estimates on the SS domain that depends on the unknown shapes of the innovations? If we agree that the WS boundary is too restrictive, then a possible crude solution could be a post-validation of the estimated parameters by calculating numerically the SS region with the estimated innovations, and verifying *ex-post* that they lie inside it. The proposed solution, here applied, has the disadvantage that the boundary is not implemented inside the maximisation process.

The second interesting issue is the impact of misspecification on the volatilities. Our main result is that if the tails of the underlying innovations are heavier than the fitting innovations, there is a systematic overestimate of volatilities. The intuitive idea behind this empirical evidence is the following: in order to compensate the light tailedness of the fitting innovations one should have systematically larger volatilities. This phenomenon can be easily identified both through a high value of MRE and a small frequency in FUV.

Finally extensions of this analysis to Garch(p, q) models and for a broader class of innovations that preserve asymptotic unbiasedness and normality are of course possible.

References

- Abramowitz, M. and Stegun, I. (1970). *Handbook of Mathematical Functions*. New York: Dover Publications, Inc.
- Ausín, M.C. and Galenao, P. (2007). Bayesian estimation of the Gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, **51**, 2636-2652.
- Baille, T.R. and Chung, H. (2001). Estimation of GARCH from the autocorrelations of squares of a process. *Journal of Times Series Analysis*, **22**, 631-650.
- Bellini, F. and Bottolo, L. (2007). Stationarity domains for δ -power GARCH process with heavy tails. *Statistics and Probability Letters*, **77**, 1418-1427.
- Berkes, I., Horváth, L. and Kokoszka, P.S. (2003). GARCH processes: structure and estimation. *Bernoulli*, **9**, 201-227.
- Berkes, I., Horváth, L. and Kokoszka, P.S. (2004). Probabilistic and statistical properties of GARCH processes. In *Asymptotic Methods in Stochastics: Festschrift for Miklós Csörgő* (Eds. L. Horváth and B. Szyszkowicz), Fields Institute Communications, **19**, 409-429.

- 1
- 2
- 3
- 4 Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econo-*
- 5 *metrics*, **31**, 307-327.
- 6
- 7 Bollerslev, T. (1987). A conditionally heteroscedastic time series model for speculative prices and
- 8 rates of return. *Review of Economics and Statistics*, **69**, 542-547.
- 9
- 10 Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modelling in finance. *Journal of Econo-*
- 11 *metrics*, **52**, 5-55.
- 12
- 13 Bollerslev, T. and Wooldrige, J.M. (1992). Quasi maximum likelihood estimation and inference in
- 14 dynamic models with time varying covariance. *Econometrics Review*, **11**, 143-172.
- 15
- 16 Bougerol, P. and Picard, N. (1992). Strict stationarity of Generalized Autoregressive Processes. *An-*
- 17 *nals of Probability*, **20**, 1714-1730.
- 18
- 19 Ding, Z., Engle, R.F. and Granger, C.W.J. (1993). A long memory property of stock markets returns
- 20 and a new model. *Journal of Empirical Finance*, **1**, 83-106.
- 21
- 22 Engle, R.F. and González-Rivera, G. (1991). Semiparametric ARCH models. *Journal of Business*
- 23 *and Economic Statistics*, **9**, 345-359.
- 24
- 25 Haas, M., Mittnik, S. and Paolella, M.S. (2004). Mixed normal conditional heteroscedasticity. *Journal*
- 26 *of Financial Econometrics*, **2**, 211-250.
- 27
- 28 Lee, S.W. and Hansen, B.E. (1994). Asymptotic theory for the Garch(1, 1) quasi-maximum likelihood
- 29 estimator. *Econometric Theory*, **10**, 29-52.
- 30
- 31 Lumsdaine, R.L. (1995). Finite-sample properties of the maximum likelihood estimator in Garch(1, 1)
- 32 models: a Monte Carlo investigation. *Journal of Business and Economics Statistics*, **13**, 1-10.
- 33
- 34 Lumsdaine, R.L. (1996). Consistency and asymptotic normality of the quasi-likelihood estimator in
- 35 IGarch(1, 1) and covariance stationary Garch(1, 1) models. *Econometrica*, **64**, 575-596.
- 36
- 37 Mikosch, T. (2004). Modelling dependence and tails of financial time series. In *Extreme Values*
- 38 *in Finance, Telecommunications and the Environment*. (Eds. B. Finkenstädt and H. Rootzén),
- 39 Chapman and Hall CRC, London, 185-286.
- 40
- 41 Mikosch, T. and Stărică, C. (2004a). Changes of structure in financial time series and the GARCH
- 42 model. *RevStat Statistical Journal*, **2**, 41-73.
- 43
- 44 Mikosch, T. and Stărică, C. (2004b). Non-stationarities in financial time series, the long-range depen-
- 45 dence and the IGARCH effect. *Review of Economics and Statistics*, **86**, 378-390.
- 46
- 47 Mikosch, T. and Straumann, D. (2002). Whittle estimation in a heavy-tailed Garch(1, 1) model.
- 48 *Stochastic Processes and their Applications*, **100**, 187-222.
- 49
- 50 Mittnik, S., Paolella, M.S. and Rachev, S.T. (2002). Stationarity of stable power-Garch processes.
- 51 *Journal of Econometrics*, **106**, 97-107.
- 52
- 53 Nakatsuma, T. and Tsurumi, H. (1996). ARMA-GARCH models: Bayes estimation versus MLE and
- 54 Bayes non-stationarity test. Available at the Web site:
- 55
- 56
- 57
- 58
- 59
- 60

<http://snade.rutgers.edu/Rutgers/wp/rutgers-wplist.html>

Nelson, D.B. (1990). Stationarity and persistence in the Garch(1,1) model. *Econometric Theory*, **6**, 318-334.

Nelson, D.B. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica*, **59**, 347-370.

Newey, W.K. and Steigerwald, D.G. (1997). Asymptotic bias for quasi-maximum likelihood estimators in conditional heteroscedasticity models. *Econometrica*, **65**, 587-599.

Rockinger, M. and Jondeau, E. (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, **106**, 119-142.

Sheppard, K. (2002). *UCSD GARCH Toolbox*. Version 2.0.9. Available at the Web site: http://www.kevinshppard.com/wiki/UCSD_GARCH

Stărică, C., Herzel, S. and Nord, T. (2006). The impact of the IGARCH effect on long-horizon volatility forecasting. Available at the Web site: <http://www.unipg.it/herzel>

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, NY.

T	GED_ν	N	t_δ	GED_ν^u	N^u	t_δ^u
250	GED_6	0.1501 (0.0910) 0.1028 (0.0551) 0.2732 (0.4024) 6.4410 (1.5140)	0.1223 (0.0739) 0.1016 (0.0713) 0.3980 (0.3036) 4737.5 (6811.0)	0.1489 (0.0905) 0.1030 (0.0550) 0.2800 (0.4084) 6.4475 (1.5167)	0.1913 (0.0877) 0.0884 (0.1103) 0.0396 (0.0587) 3831.1 (5550.8)	0.1229 (0.0743) 0.1034 (0.0711) 0.3942 (0.3026) 3831.1 (5550.8)
	N	0.1268 (0.0910) 0.0983 (0.0775) 0.3632 (0.4350) 2.0690 (0.3088)	0.1406 (0.0933) 0.1008 (0.0814) 0.2999 (0.4302) 62.506 (61.142)	0.1269 (0.0911) 0.0983 (0.0774) 0.3631 (0.4369) 2.0703 (0.3085)	0.1689 (0.0975) 0.0959 (0.0958) 0.1576 (0.2336) 60.435 (57.026)	0.1419 (0.0940) 0.1015 (0.0827) 0.2938 (0.4228) 60.435 (57.026)
	t_3	0.1335 (0.0654) 0.0579 (0.0856) 0.0618 (0.0917) 1.1000 - ^a	0.1488 (0.0831) 0.0992 (0.1170) 0.1879 (0.2786) 3.1221 (0.7133)	0.1341 (0.0650) 0.0566 (0.0838) 0.0602 (0.0892) 1.1000 - ^a	0.1388 (0.0874) 0.0648 (0.0961) 0.0461 (0.0683) 3.1238 (0.7170)	0.1489 (0.0855) 0.0988 (0.1186) 0.1921 (0.2848) 3.1238 (0.7170)
1000	GED_6	0.1073 (0.0391) 0.1004 (0.0273) 0.4665 (0.1684) 6.1261 (0.6800)	0.1166 (0.0569) 0.0985 (0.0355) 0.4214 (0.2442) 7120.4 (8792.1)	0.1074 (0.0393) 0.1004 (0.0274) 0.4661 (0.1689) 6.1244 (0.6836)	0.1215 (0.0687) 0.0965 (0.0436) 0.3922 (0.3022) 7708.8 (9220.0)	0.1044 (0.0428) 0.1010 (0.0344) 0.4786 (0.1855) 7708.8 (9220.0)
	N	0.1073 (0.0514) 0.1002 (0.0393) 0.4630 (0.2238) 2.0176 (0.1462)	0.1123 (0.0570) 0.1004 (0.0397) 0.4339 (0.2477) 198.70 (252.16)	0.1074 (0.0517) 0.1003 (0.0394) 0.4621 (0.2245) 2.0172 (0.1463)	0.1098 (0.0540) 0.1016 (0.0393) 0.4467 (0.2332) 192.52 (244.90)	0.1062 (0.0494) 0.1016 (0.0395) 0.4661 (0.2168) 192.52 (244.90)
	t_3	0.0988 (0.0543) 0.0749 (0.0485) 0.3789 (0.3075) 1.1000 - ^a	0.1201 (0.0790) 0.0852 (0.0753) 0.3598 (0.4054) 3.0117 (0.3242)	0.0985 (0.0546) 0.0745 (0.0490) 0.3794 (0.3074) 1.1000 - ^a	0.1183 (0.0784) 0.0853 (0.0756) 0.3644 (0.4037) 3.0089 (0.3233)	0.1151 (0.0565) 0.1035 (0.0549) 0.4391 (0.2392) 3.0089 (0.3233)

Table 1: Model A, $\alpha_0 = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.5$: median and MAD in brackets of 5,000 MLEs of $(\alpha_0, \alpha_1, \beta_1)$ for various choices of the fitting innovation distribution with/without the WS constraint.

^aUSCD Garch Toolbox automatically sets the lower bound of the parameter ν equal to 1.1. When the simulated innovation distribution has heavy tails as in the t_3 case, the maximisation routine always reaches the lower bound showing no variability in the estimates.

T		GED_ν	N	t_δ	GED_ν^u	N^u	t_δ^u
250	GED_6	0.0717 (0.0302)	0.0863 (0.0291)	0.0783 (0.0341)	0.0717 (0.0303)	0.0871 (0.0291)	0.0781 (0.0342)
		0.4960 (0.3677)	0.4920 (0.2965)	0.4800 (0.3558)	0.4960 (0.3736)	0.4920 (0.2965)	0.4800 (0.3499)
	N	0.1032 (0.0466)	0.1055 (0.0437)	0.1035 (0.0459)	0.1032 (0.0466)	0.1061 (0.0444)	0.1041 (0.0455)
		0.4640 (0.3855)	0.4840 (0.3677)	0.4800 (0.3795)	0.4640 (0.3855)	0.4760 (0.3617)	0.4720 (0.3795)
	t_3	0.2564 (0.1113)	0.2582 (0.1396)	0.2490 (0.1642)	0.2562 (0.1093)	0.2588 (0.1409)	0.2474 (0.1640)
		0.0120 (0.0178)	0.1200 (0.1779)	0.2960 (0.4388)	0.0120 (0.0178)	0.1240 (0.1838)	0.2960 (0.4388)
1000	GED_6	0.0364 (0.0164)	0.0437 (0.0207)	0.0408 (0.0184)	0.0365 (0.0164)	0.0479 (0.0267)	0.0405 (0.0185)
		0.5040 (0.3529)	0.5080 (0.2758)	0.5030 (0.3202)	0.5040 (0.3514)	0.5260 (0.2520)	0.5030 (0.3217)
	N	0.0565 (0.0244)	0.0579 (0.0253)	0.0561 (0.0246)	0.0566 (0.0244)	0.0568 (0.0248)	0.0562 (0.0246)
		0.4935 (0.3529)	0.5000 (0.3410)	0.5000 (0.3529)	0.4940 (0.3514)	0.4940 (0.3499)	0.4955 (0.3573)
	t_3	0.2229 (0.0584)	0.1704 (0.0931)	0.1299 (0.0839)	0.2228 (0.0586)	0.1703 (0.0928)	0.1298 (0.0849)
		0.0030 (0.0044)	0.1670 (0.2461)	0.4670 (0.6716)	0.0030 (0.0044)	0.1680 (0.2476)	0.4825 (0.6842)

Table 2: Model A , $\sigma_0 = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.5$: median and MAD in brackets for the 5, 000 estimates of MRE and FUV. Estimation performed with different fitting innovations with/without the WS boundary.

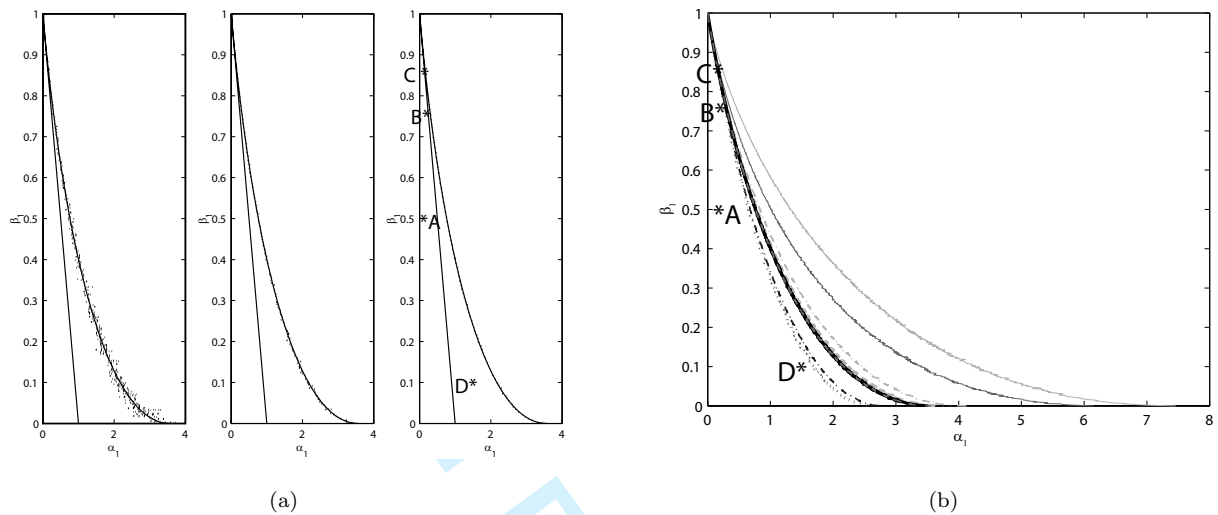


Figure 1: (a) SS domain for normal innovations derived through a Monte Carlo integration, dotted lines, with different samples size: left panel $n = 1,000$, central panel $n = 10,000$ and right panel $n = 100,000$. Superimposed the WS domain, solid straight line, and the SS domain, solid line, using (4) and (7) respectively. Capital letters on the right panel show the coordinates in the stationarity domain of the simulated models. (b) SS domain for different innovations derived through a Monte Carlo integration with $n = 100,000$: normal distribution, black bold solid line; standardised t distribution, light grey lines and from right to left with $\delta = 60$, dotted line, $\delta = 30$, dashed line, $\delta = 9$, dashed-dotted line and $\delta = 3$ solid line, outermost line; standardised GED, dark grey lines and from left to right with $\nu = 1$, solid line, $\nu = 2$, dashed line, $\nu = 4$ dashed-dotted line and $\nu = 6$ dotted line, innermost line. Capital letters show the coordinates in the stationarity domain of the simulated models.

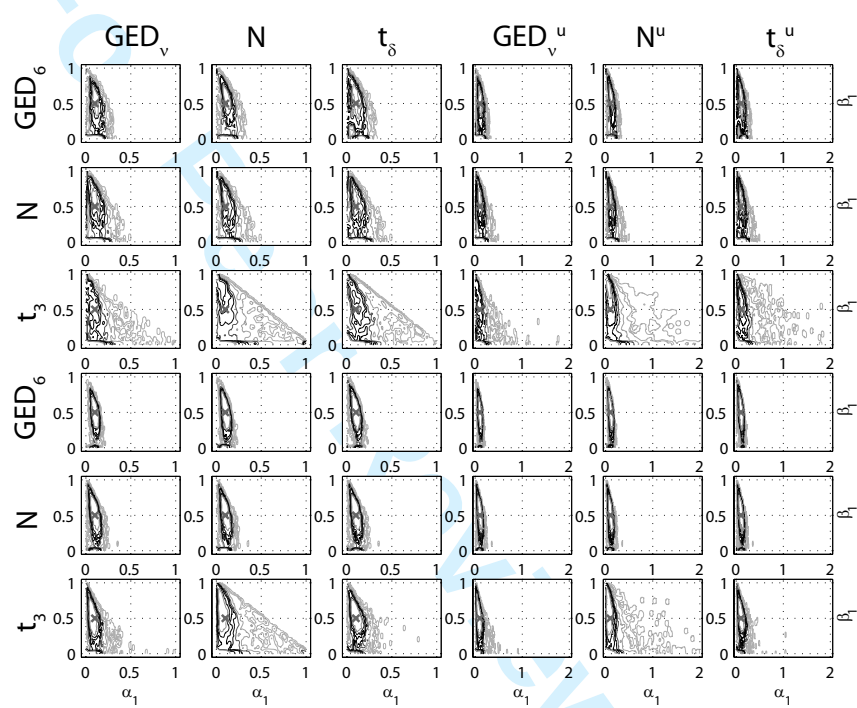


Figure 2: Model A, $\alpha_0 = 0.1$, $\alpha_1 = 0.1$ and $\beta_1 = 0.5$. Normal kernel contour plot of the joint distribution $(\hat{\alpha}_1, \hat{\beta}_1)$ with dark grey cross for the simulated parameters.