Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597248

A note on adaptation in garch models

Gloria González-Rivera[a]
[a] Department of Economics, University of California, Riverside, CA

To cite this Article González-Rivera, Gloria(1997) 'A note on adaptation in garch models', Econometric Reviews, 16: 1, 55 — 68
To link to this Article: DOI: 10.1080/07474939708800372
URL: http://dx.doi.org/10.1080/07474939708800372

ECONOMETRIC REVIEWS, 16(1), 55-68 (1997)

A NOTE ON ADAPTATION IN GARCH MODELS

Gloria González-Rivera

Department of Economics, University of California, Riverside, CA 92521

Key Words and Phrases: Adaptation, Generalized Autoregressive Conditional Heteroscedasticity (GARCH), maximum likelihood, semiparametric estimator.

JEL: C22, C14.

ABSTRACT

In the framework of the Engle-type (G)ARCH models, I demonstrate that there is a family of symmetric and asymmetric density functions for which the asymptotic efficiency of the semiparametric estimator is equal to the asymptotic efficiency of the maximum likelihood estimator. This family of densities is bimodal (except for the normal). I also characterize the solution to the problem of minimizing the mean squared distance between the parametric score and the semiparametric score in order to search for unimodal densities for which the semiparametric estimator is likely to perform well. The LaPlace density function emerges as one of these cases.

1 Introduction

The most preferred estimation method used in models with Generalized Autoregressive Conditional Heteroscedasticity, (G)ARCH, (Engle 1982), (Bollerslev 1986) is maximum likelihood (MLE). It produces consistent, efficient and asymptotically normal estimators if the mean and the variance equations as well as the probability density function of the error term are correctly specified. In many empirical studies, the most common assumption about the distribution of the error is conditional normality. But this assumption is rejected very often. Consequently, researchers are using other density functions that are closer to the data under study. Among others, Baillie and Bollerslev (1989) and Hsieh (1989) used a Student-t to approximate the conditional distribution of exchange rates, and Nelson (1991) used the Generalized Error distribution to model the conditional probability density function of stock returns. Since we do not have guidelines on how to choose the density function, Engle and González-Rivera (1991) proposed a semiparametric estimator, which consists of a parsimonious parametric specification of the mean and variance equations and, even though the conditional probability density function of the innovations is not known, it is assumed that is sufficiently smooth to be approximated by a non-parametric density estimator.

Consider a random variable y_t that is conditionally heteroscedastic with conditional variance h_t, such that

$$y_t = \sqrt{h_t} u_t$$

where u_t are independent and identically distributed random variables, known as standardized innovations. This paper focuses in the Engle-type specifications of the conditional variance h_t. The most popular specifications in the financial literature are:

i. Bollerslev (1986), GARCH(p,q),

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

ii. Nelson (1991), EGARCH,

$$ln\ h_t = \omega + (1 + \sum_{i=1}^{q} \alpha_i L^i)(1 - \sum_{j=1}^{p} \beta_j L^j)^{-1}(\theta u_{t-1} + \gamma(|u_{t-1}| - E|u_{t-1}|))$$

iii. Ding, Granger and Engle (1993), asymmetric power ARCH

$$h_t^d = \omega + \sum_{i=1}^{q} \gamma_i(|y_{t-i}| - \alpha_i y_{t-i})^d + \sum_{j=1}^{p} \beta_j h_{t-j}^d$$

These models are summarized as $h_t = h(y_{t-1}, y_{t-2}, ...; \theta)$, the conditional variance is a function of the past information y_{t-i} and the parameter vector θ. The objective is to estimate the vector θ. The semiparametric estimator is a two-step estimator. In the first step, consistent estimates of the parameters of interest are obtained through quasi-maximum likelihood estimation, where the likelihood function is written under the assumption of conditional normality of the innovations, even though this may be a wrong assumption. With the initial consistent estimates of θ, a non-parametric density of the standardized innovations is constructed. The second step consists of using this non-parametric density to build a non-parametric likelihood that is maximized with respect to θ (as is explained in more detail in the next section). The goal of the semiparametric estimator is to recapture the asymptotic efficiency losses due to quasi-maximum likelihood estimation, which can be large when the true probability density of the standardized innovations is far away from normality.

An issue very closely related to the semiparametric estimation of GARCH models is the property of such an estimator to be *adaptive* An estimator, which is constructed with a data-based procedure, is said to be *adaptive* if knowledge of the error probability density function does not help to improve the asymptotic efficiency of the estimator. If the error probability density function is fully known, the maximum likelihood estimator will have an asymptotic variance which achieves a lower bound, the Cramer-Rao bound, for all regular estimators. If the semiparametric estimator is adaptive (the density is unknown), it means that the semiparametric efficiency bound equals the asymptotic variance of the MLE.

For the Engle-type (G)ARCH models mentioned above, the variance parameters θ are not generally adaptively estimable (Engle and González-Rivera 1991), (Steigerwald 1994), (Drost and Klaassen 1993), (Linton 1993), and further research has been directed towards the search for different specifications of the variance equation that bring back the property of adaptation. In particular, Linton (1993) provides a different specification of the variance equation in which the scale effect is separated from the rest of the parameters. Under the assumption of symmetric densities, he shows that the relative scale parameters are adaptively estimable.

In this paper I prove that there is a family of probability density functions for which the asymptotic variance of the MLE estimator equals the semiparametric efficiency bound for the Engle-type GARCH models, and consequently adaptive estimators exist within this family. This result holds for symmetric as well as asymmetric densities. In the next section, I explain the background and notation used in the paper. Section 3 contains the main results and provides some guidelines about the density functions in which the semiparametric estimator is more likely to succeed and section 4 concludes.

2 Background and notation

Conditioning on the information set ψ_{t-1}, that contains information up to time $t-1$, the random variable y_t is distributed as $f(0, h_t)$, where h_t is the conditional variance which is a function of past information and of a set of parameters θ, $h_t = h(y_{t-1}, y_{t-2}, ...; \theta)$. The probability density function f is not specified. In order to estimate this model, consider the standardized variable $u_t(\theta) \equiv y_t/\sqrt(h_t)$. The random variables u_t are assumed to be i.i.d. with continuous density function g with mean zero and variance one. The shape of g will be obtained independently from location and scale. The parameters that define the shape, say η, are considered nuisance parameters, but they may have relevant information for the estimation of the parameters of interest θ. For a sample of length T, the log-likelihood function can be written as

$$\mathcal{L}_T(\theta) = -\frac{1}{2}\sum_{t=1}^{T} log\ h_t + \sum_{t=1}^{T} log\ g(u_t(\theta)) \tag{1}$$

To abbreviate notation I write u_t for $u_t(\theta)$. The sample score function is

$$S_T(\theta) = \frac{\partial \mathcal{L}_T}{\partial \theta} = \sum_{t=1}^{T} -\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}(1 + u_t\frac{g'}{g}(u_t)) \tag{2}$$

g' is the derivative of g respect to u_t. Note that the first factor in the score function, $\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}$, depends solely on past information; it relies on the specification of the variance equation. The second factor, $(1 + u_t\frac{g'}{g})$, is a function of u_t; it depends on the shape of the density. The factors are independent of each other. The expectation of the score is zero for any density function, since integration by parts results in $E(u_t\frac{g'}{g}) = -1$. There is a complete description of the estimation of this semiparametric model in Engle and González-Rivera (1991).

Let $S(\eta)$ be the population score vector for the nuisance parameters and $S(\theta)$ the population score vector for the parameters of interest. The set of parameters η is unknown and consequently the semiparametric estimator of θ cannot exploit the information contained in η. If η contains any information about θ, the efficient score for θ is found by calculating the residual vector, $R(\theta)$, from the projection of $S(\theta)$ on the closure of the set of all linear combinations of $S(\eta)$, or tangent set $\mathcal{T}$. The tangent set consists of linear combinations of $S(\eta)$ and because the u_t's are random variables with mean zero and variance one, the elements of the tangent set are orthogonal to the function vector $(u_t, u_t^2 - 1)'$. Through the projection, all the variation of $S(\theta)$ due to $S(\eta)$ is removed (Newey 1990). The residual vector, $R(\theta)$, is the difference between $S(\theta)$ and the projection and, by construction, is orthogonal to the projection. Hence, $R(\theta)$ is the efficient score for θ and the semiparametric bound is

$$V = (E(R(\theta)R(\theta)'))^{-1} \tag{3}$$

For a sample of length T, the sample vector residual is

$$R_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} R_t(\theta)$$

and

$$R_t(\theta) = -\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}(1 + u_t\frac{g'}{g}(u_t)) + E\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\right)\Big\{(1 + u_t\frac{g'}{g}(u_t))$$

$$-E\{(1 + u_t\frac{g'}{g}(u_t))F(u_t)'\}\{E(F(u_t)F(u_t)')\}^{-1}F(u_t)\Big\} \tag{4}$$

where $F(u_t) = (u_t, u_t^2 - 1)'$. A proof of this result can be found in Bickel et al. (1993) and Steigerwald (1994).

3 Probability Density Functions and Variance Parameters

In equation (4), $R_t(\theta)$ is composed of two terms. The first one is easily recognized as the score vector $S_t(\theta)$. The second term is a weighted residual of the projection of a function of the shape of the density, i.e. $(1+u_t\frac{g'}{g})$, on the set $F(u_t)$. The smaller the residual the closer $R_t(\theta)$ will be to the fully parametric score, $S_t(\theta)$. The limiting case happens when the residual of the projection of $(1+u_t\frac{g'}{g})$ on $F(u_t)$ is zero for all t. If this is so, then adaptation is possible for the variance parameters, within a family of probability density functions. I distinguish between the symmetric and the asymmetric class.

3.1 Symmetric Densities

If the residual of the projection of $(1+u_t\frac{g'}{g})$ on $F(u_t)$ is zero for all t, then

$$1+u_t\frac{g'}{g}(u_t)=E\{(1+u_t\frac{g'}{g}(u_t))F(u_t)'\}\{E(F(u_t)F(u_t)')\}^{-1}F(u_t) \qquad (5)$$

Let us call κ, the coefficient of kurtosis under the density $g(.)$ Considering that u_t has mean equal to zero and variance equal to one and that integration by parts gives $E(u_t^2\frac{g'}{g})=0$ equation (5) is written as

$$1+u_t\frac{g'}{g}(u_t)=(E(u_t^3\frac{g'}{g}(u_t))+1)\frac{u_t^2-1}{\kappa-1} \qquad (6)$$

Rearranging terms and integrating by parts gives $E(u_t^3\frac{g'}{g})=-3$ and the following differential equation is obtained

$$\frac{g'}{g}(u_t)=-\lambda u_t+(\lambda-1)\frac{1}{u_t} \qquad (7)$$

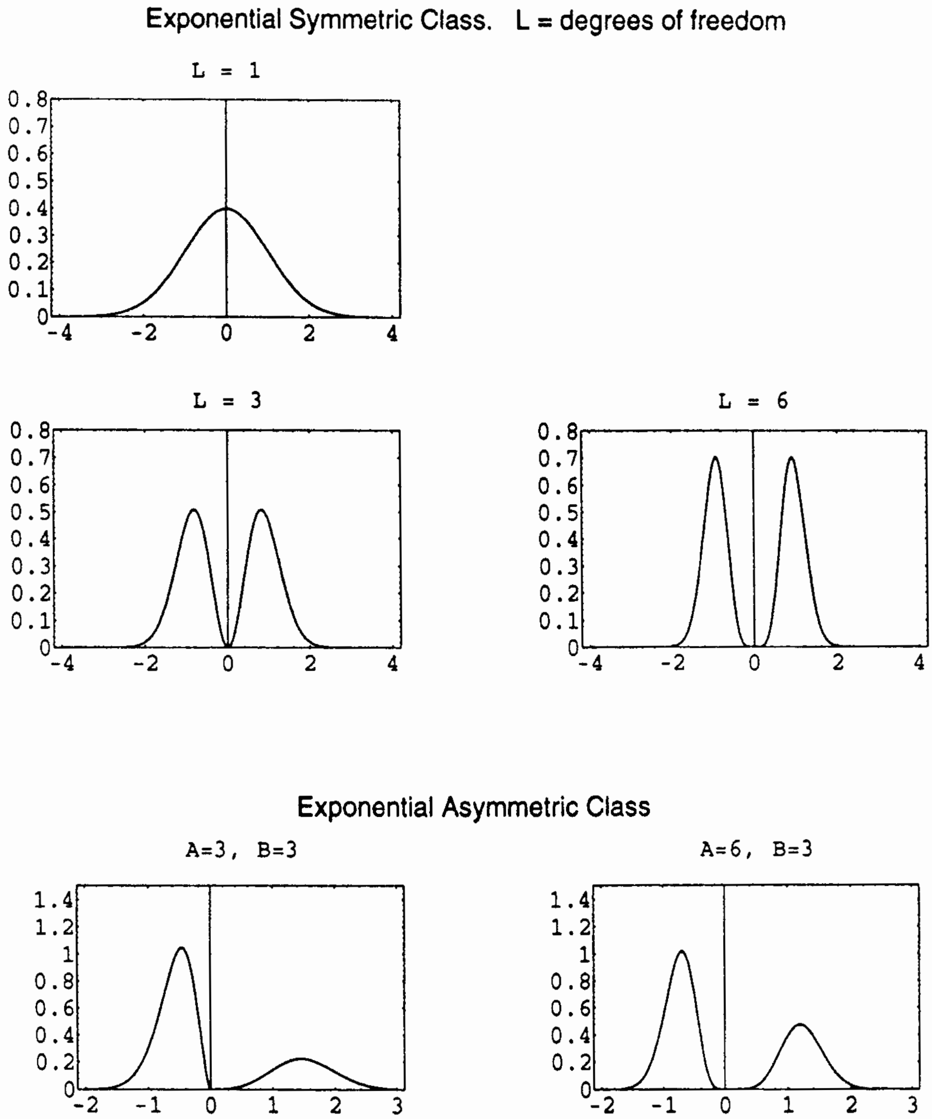
Exponential Symmetric Class. L = degrees of freedom
L = 1
L = 3
L = 6
Exponential Asymmetric Class
A=3, B=3
A=6, B=3

FIG. 1

where $\lambda \equiv \frac{2}{\kappa-1}$. The solution to equation (7) is

$$g(u_t) = K\,|u_t|^{(\lambda-1)}\,e^{-\frac{\lambda}{2}u_t^2} \tag{8}$$

Since $g(u_t)$ has to be a density function, the constant K is found by making the integral of $g(u_t)$ equal to one. Furthermore, to solve this integral, it is required that $\lambda > 0$. If we add continuity and differentiability of the density function for $-\infty < u_t < +\infty$, in particular, $g(.)$ should be differentiable at $u_t = 0$, then $\lambda > 2$ or $\lambda = 1$. With these restrictions in mind the final solution is

$$g(u_t) = \frac{\lambda^{\lambda/2}}{2^{\lambda/2}\Gamma(\frac{\lambda}{2})}\,|u_t|^{(\lambda-1)}\,e^{-\frac{\lambda}{2}u_t^2} \tag{9}$$

It is easy to check that the random variable u_t with density function (9) has, in fact, mean zero and variance equal to 1. The density $g(u_t)$ belongs to the exponential family. In particular, g is the density of a random variable u_t that is the ($\pm$) square root of a χ^2_λ divided by the degrees of freedom λ. The restrictions on λ impose restrictions on the degree of kurtosis of u_t. In fact, κ has to be bounded between 1 and 2, for $\lambda > 2$, that is, u_t has to be platykurtic. The limiting case happens when $\lambda = 1$ ($\kappa = 3$), that is, $g(u_t)$ is normal. It can be concluded that independently of the specification of the variance equation, the variance parameters can be estimated adaptively in the family of functions described by equation (9). Figure 1 contains plots of the density (9) for different values of λ.

3.2 Asymmetric Densities

For asymmetric densities, the same line of argument can be followed. In this class $E(u_t^3) = \varsigma$, where ς is the coefficient of skewness.

The differential equation that follows from (5) looks like

$$\frac{g\prime}{g} = -Au_t + \frac{1}{u_t}(A-1) + B \tag{10}$$

where

$$A = \frac{2}{\kappa - 1 - \varsigma^2} \qquad B = \frac{2\varsigma}{\kappa - 1 - \varsigma^2}$$

Among the possible solutions to (10), the solution implied by $\int g(u)du = 1$ is

$$g(u) = \begin{cases} K_1(-u)^{A-1}e^{-\frac{1}{2}Au^2+Bu}, & u \leq 0 \\ K_2 u^{A-1}e^{-\frac{1}{2}Au^2+Bu}, & u > 0 \end{cases} \tag{11}$$

where K_1 and K_2 are the normalizing constants. It should be noted that these two constants must be different in order to ensure that the u_t's have mean zero. This function will be asymmetric for $B \neq 0$. It is bimodal and because continuity and differentiability of the density are required, $A > 2$. As in the symmetric case, adaptation will be possible in two-parameter family of densities if and only if $\kappa - \varsigma^2 < 2$. Figure 1 contains plots of the density (11) for different values of A and B.

3.3 Unimodal Densities

The previous subsections have dealt with instances in which the semiparametric score equals the fully parametric score, i.e. $R_t(\theta) = S_t(\theta)$ and as a result a class of densities emerge for which adaptive estimators do exist. These densities share the property of being bimodal, with the exception of the normal density that is a limiting case. In this subsection, I focus on the search of unimodal densities for which the semiparametric estimator is likely to perform well.

The performance of the semiparametric estimator depends on the "closeness" of the semiparametric score $R(\theta)$ to the parametric score $S(\theta)$. Using equation (2) and substituting in (4), I write

$$R_t(\theta) - S_t(\theta) = E\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\right)\left\{(1 + u_t\frac{g'}{g}(u_t))\right.$$

$$\left. -E\{(1 + u_t\frac{g'}{g}(u_t))F(u_t)'\}\{E(F(u_t)F(u_t)')\}^{-1}F(u_t)\right\} \qquad (12)$$

The goal is to minimize $R_t(\theta) - S_t(\theta)$ according to a mean squared criterium,

$$min_g\, Q = E(R_t(\theta) - S_t(\theta))^2 \qquad (13)$$

respect to a density $g(.)$.

To characterize the solution to this problem, consider that the quantity $E\left(\frac{1}{2h_t}\frac{\partial h_t}{\partial \theta}\right)$ does not depend on the density $g(.)$ and the minimum of Q will be given by the minimum of

$$H \equiv E\left\{(1 + u_t\frac{g'}{g}(u_t)) - E\{(1 + u_t\frac{g'}{g}(u_t))F(u_t)'\}\{E(F(u_t)F(u_t)')\}^{-1}F(u_t)\right\}^2$$

Working out the previous expression,

$$0 \leq H = E(1 + u_t\frac{g'}{g}(u_t))^2 - \frac{4}{\kappa - 1 - \varsigma^2}$$

and, since H is non-negative, the solution to the problem (13) is characterized by the conditions

$$E(1 + u_t\frac{g'}{g}(u_t))^2 = \frac{4}{\kappa - 1} \qquad (C1)$$

for symmetric densities; and

$$E(1 + u_t\frac{g'}{g}(u_t))^2 = \frac{4}{\kappa - 1 - \varsigma^2} \qquad (C2)$$

for asymmetric densities.

The probability density functions given in (8) and (11) satisfy (C1) and (C2) respectively. If I add the restriction that $g(.)$ should be unimodal, a solution to the problem (13) is the normal density. For other unimodal

TABLE I

Symmetric Unimodal Densities

Density	d	κ	$d - \frac{4}{\kappa-1}$
Student-t			
$\nu = 5$	1.25	9.0	0.75
$\nu = 8$	1.45	4.5	0.31
$\nu = 12$	1.60	3.75	0.14
LaPlace (0,1)	1.0	6.0	0.20
Logistic (0,1)	1.43	4.2	0.18

Asymmetric Unimodal Densities

Density	d	κ	ς	$d - \frac{4}{\kappa-1-\varsigma^2}$
Lognormal (0,1)	35.72	116.9	6.18	35.65
Chi-Square				
$\nu = 5$	10.0	5.4	1.26	8.57
$\nu = 6$	6.0	5.0	1.15	4.50
$\nu = 8$	4.0	4.5	1.0	3.40
$\nu = 12$	3.0	4.0	0.81	1.30

ν is degrees of freedom.
All densities have been standardized, the random variables u_t's have mean zero and variance equal to one

density functions, the approach I follow is to pick up the function and evaluate the conditions (C1) and (C2). To abbreviate notation, define $d \equiv E(1 + u_t \frac{g'}{g}(u_t))^2$. The closer $d - \frac{4}{\kappa-1-\varsigma^2}$ is to zero, the better the performance of the semiparametric estimator. Table I summarizes the findings.

Among the symmetric densities considered, the semiparametric estimator is likely to perform well within the family of densities as the Student-t with more than 12 degrees of freedom, LaPlace distributions and Logistic distributions. Among the asymmetric densities, the Chi-Square distributions with more than 12 degrees of freedom are also good candidates. These results are not surprising because the larger the degrees of freedom, the closer to the normal density. However, the LaPlace result is interesting because of the high kurtosis and the relative small distance $d - \frac{4}{\kappa-1}$.

4 Conclusions

In the framework of the Engle-type (G)ARCH models, I have shown that there is a family of density functions for which the efficiency of the semiparametric estimator is equal to the efficiency of the maximum likelihood estimator, either with symmetric or asymmetric densities. This family of densities share the property of being bimodal (except for the normal) and that explains part of the Monte-Carlo results found in Steigerwald (1994), where the performance of the semiparametric estimator under bimodality is consistently superior to the one under more standard distributions as the Student-t and Lognormal.

I have characterized the solution to the problem of minimizing the mean squared distance between the semiparametric score and the fully parametric score, $R(\theta) - S(\theta)$. Conditions (C1) and (C2) should be satisfied by those densities for which the asymptotic efficiency of the semiparametric estimator equals to the efficiency of the maximum likelihood estimator. Among unimodal densities, I have searched for instances in which, even though (C1) and (C2) are not satisfied, the semiparametric estimator is likely to perform well. I found that the LaPlace density function emerges as one of those cases.

ACKNOWLEDGEMENTS

I would like to thank Douglas Steigerwald and Feico Drost for useful conversations and an anonymous referee who made a valuable comment. I benefit from the suggestions of the participants at the European Econometric Society Meeting. I gratefully acknowledge the Intramural Research Grant of the Academic Senate of the University of California, Riverside. Any remaining errors are mine.

References

Baillie, R., and T. Bollerslev, (1989). The message in daily exchange rates: A conditional variance tale. *Journal of Business and Economic Statistics* 7:297–305.

Bickel, P., (1982). On adaptive estimation. *Annals of Statistics* 10:647–671.

Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner, (1993). *Efficient and Adaptive Statistical Inference for Semiparametric Models.* John Hopkins University Press.

Bollerslev, T., (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31:307–327.

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

Drost, F., and C. Klaassen, (1993). Adaptivity in semiparametric GARCH models. Tilburg University.

Engle, R., and G. González-Rivera, (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9:345–360.

Engle, R. F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50:987–1008.

Hsieh, D., (1989). Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business and Economic Statistics* 7:307–317.

Linton, O., (1993). Adaptive estimation in ARCH models. *Econometric Theory* 9:539–569.

Nelson, D., (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59:307–346.

Newey, W., (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5:99–135.

Steigerwald, D., (1994). Efficient estimation of models with conditional heteroscedasticity. Working Paper. Department of Economics. University of California, Santa Barbara.