Analytical derivates of the APARCH model

Sébastien Laurent[1]

Forthcoming in *Computational Economics*

October 24, 2003

Abstract

This paper derives analytical expressions for the score of the APARCH model of Ding, Granger, and Engle (1993). Interestingly, doing so we derive the analytical score of a broad range of GARCH model since the APARCH model nests at least seven specifications. The use of the APARCH model is now widespread in the literature. However, all the existing applications rely on numerical techniques to calculate the gradients. The paper shows that analytical gradients highly speed-up maximum-likelihood estimation.

Keywords: APARCH, analytical score.

1 Introduction

Traditional regression tools have shown their limitation in the modelling of high-frequency (weekly, daily or intra-daily) data (say y_t). Assuming that only the mean response could be changing with covariates while the variance remains constant over time often revealed to be an unrealistic assumption in practice. This fact is particularly obvious in series of financial data where clusters of volatility can be detected visually. Indeed, it is now widely accepted that high frequency financial returns are heteroskedastic.

Since the seminal paper of Engle (1982), autoregressive moving average (ARMA) models have been extended to essentially equivalent models for the variance. Autoregressive Conditional Heteroscedasticity (ARCH) models have been extensively used in the literature. [1]

[1]CeReFim (Université de Namur) and CORE (Université catholique de Louvain). E-mail: Sebastien.Laurent@fundp.ac.be. Correspondence to CeReFim, 8 rempart de la vièrge, B5000 Namur, Belgium. Phone: +32 (0) 81 724869. Fax: +32 (0) 81 724840.

[1]For a survey on ARCH-type models, see Bollerslev, Engle, and Nelson (1994) and Palm (1996) among others.

The Asymmetric Power ARCH (APARCH) model of Ding, Granger, and Engle (1993) is certainly one of the most promising ARCH-type model. Indeed, this model nests at least seven ARCH-type model (see below) and was found to be particularly relevant in many recent applications (see Giot and Laurent, 2001 and Mittnik and Paolella, 2000 among others).

The common point of all the applications dealing with the APARCH model (and most of the ARCH-type models) is that they are estimated by maximum likelihood methods and use numerical techniques to approximate the derivatives of the likelihood function with respect to the parameter vector (the score or gradient vector). However, as shown by Fiorentini, Calzolari, and Panattoni (1996), Gable, Van Norden, and Vigfusson (1997) (for Markov Switching Models) and McCullough and Vinod (1999), using analytical scores in the estimations procedure should improve the numerical accuracy of the resulting estimates and speed-up maximum-likelihood estimation.

For this reason, we propose to derive analytical expressions for the score of the APARCH model. Interestingly, doing so we derive the analytical score of a broad range of GARCH model (the seven models nested by the APARCH).

The rest of the paper is organized in the following way. Section 2 presents the APARCH models and the associated gradients. Section 3 provides an empirical application and Section 4 concludes.

2 APARCH Specification

The APARCH(p,q) model of Ding, Granger, and Engle (1993) can be defined as follows:

$$y_t = x'_{1,t}\mu + \varepsilon_t \tag{1}$$

$$\varepsilon_t = \sigma_t z_t \tag{2}$$

$$\sigma_t^\delta = x'_{2,t}\omega + \sum_{i=1}^{q}\alpha_i k(\varepsilon_{t-i})^\delta + \sum_{j=1}^{p}\beta_j \sigma_{t-j}^\delta \tag{3}$$

$$k(\varepsilon_{t-i}) = |\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}, \tag{4}$$

where $x_{1,t}$ and $x_{2,t}$ are two vectors of respectively n_1 and n_2 weakly exogenous variables (including the intercept), μ, ω, α_i's,γ_i's,β_j's and δ are parameters (or vectors of parameters) to be estimated. δ ($\delta > 0$) plays the role of a Box-Cox transformation of the conditional standard deviation σ_t, while the γ_i's reflect the so-called leverage effect. A positive (resp. negative) value of the γ_i's means that past negative (resp. positive) shocks have a deeper impact on current conditional volatility than past positive shocks (see Black, 1976).

The properties of the APARCH model have been studied recently by He and Teräsvirta (1999a, 1999b). Note that it is convenient to start the recursion of Eq. (3) by setting unobserved components to their sample average, i.e. setting $k(\varepsilon_{t-i})^\delta = \frac{1}{T}\sum_{s=1}^{T}(|\varepsilon_s| - \gamma_i\varepsilon_s)^\delta$ for $t \leq i$ and $\sigma_t^\delta = \left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_s^2\right)^{\frac{\delta}{2}}$ for $t \leq 0$.

This model couples the flexibility of a varying exponent with the asymmetry coefficient (to take the "leverage effect" into account). The APARCH includes seven other ARCH extensions as special cases:[2]

- The ARCH of Engle (1982) when $\delta = 2$, $\gamma_i = 0$ $(i = 1, \ldots, p)$ and $\beta_j = 0$ $(j = 1, \ldots, p)$.
- The GARCH of Bollerslev (1986) when $\delta = 2$ and $\gamma_i = 0$ $(i = 1, \ldots, p)$.
- Taylor (1986)/Schwert (1990)'s GARCH when $\delta = 1$, and $\gamma_i = 0$ $(i = 1, \ldots, p)$.
- The GJR of Glosten, Jagannathan, and Runkle (1993) when $\delta = 2$.
- The TARCH of Zakoian (1994) when $\delta = 1$.
- The NARCH of Higgins and Bera (1992) when $\gamma_i = 0$ $(i = 1, \ldots, p)$ and $\beta_j = 0$ $(j = 1, \ldots, p)$.
- The Log-ARCH of Geweke (1986) and Pentula (1986), when $\delta \rightarrow 0$.

To the best of our knowledge, up to now, the analytical gradients of the APARCH model have not been provided in the literature. This is probably due to the high degree of nonlinearity of this specification which makes their computation less trivial than in the ARCH case.

To achieve this goal, let us define $\gamma = (\gamma_1, \ldots, \gamma_q,)$ the vector of q "leverage effect" parameters, $d_t = \left(x_{2,t}, k(\varepsilon_{t-1})^\delta, \ldots, k(\varepsilon_{t-q})^\delta, \sigma_{t-1}^\delta, \ldots, \sigma_{t-p}^\delta\right)'$ and $\eta = (\mu', \vartheta', \gamma', \delta)'$ the vector of $(n_1 + n_2 + 2q + p + 1)$ unknown parameters of the conditional mean and the conditional dispersion equations, where $\vartheta = (\omega', \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)'$.

To estimate the APARCH by maximum likelihood, one has to make an additional assumption on the innovation process by choosing a density function, denoted $g(z_t; \tau)$ where τ is an extra parameter vector to be estimated. The problem to solve is thus to maximize the sample log-likelihood function $L_T(\eta, \tau)$ for the T observations $(t = 1, \ldots, T)$, with respect to the vector of parameters (η, τ), where $L_T(\eta, \tau) = \sum_{t=1}^{T}\log f(y_t|\eta, \tau, I_{t-1})$, $f(y_t|\eta, \tau, I_{t-1}) = \sigma_t^{-1}g(z_t|\tau)$, I_t is the information set at time t and $z_t = \frac{y_t - x_{1,t}'\mu}{\sigma_t}$.

It is usual to assume that z_t is normally distributed since the Gaussian Quasi Maximum Likelihood (QML) method can provide consistent estimates in the general framework of a dynamic model under correct specification of both the conditional mean and the conditional variance (see

[2]Complete developments leading to these conclusions are available in Ding, Granger, and Engle (1993).

Weiss, 1986 and Bollerslev and Wooldridge, 1992 among others). In this case, the log-likelihood function is defined as:

$$L_T(\eta,\tau) = L_T(\eta) = -\frac{1}{2}\left[Tln(2\pi) + \sum_{t=1}^{T} ln(\sigma_t^2) + \sum_{t=1}^{T}\frac{\varepsilon_t^2}{\sigma_t^2}\right]. \tag{5}$$

Differentiating with respect to the full set of parameters η yields:

$$\frac{\partial L_T(\eta)}{\partial \eta} = -\frac{\varepsilon_t}{\sigma_t^2}\frac{\partial \varepsilon_t}{\partial \eta} - \frac{1}{2}\frac{1}{\sigma_t^4}(\sigma_t^2 - \varepsilon_t^2)\frac{\partial \sigma_t^2}{\partial \eta}. \tag{6}$$

Differentiating the log-likelihood function with respect to μ requires an analytical expression for $\frac{\partial \varepsilon_t}{\partial \mu}$. In our case, the solution is trivial and is: $\frac{\partial \varepsilon_t}{\partial \mu} = -x'_{1,t}$. Differentiating the log-likelihood function with respect to η also requires the computation of $\frac{\partial \sigma_t^2}{\partial \mu}$ and $\frac{\partial \sigma_t^2}{\partial \eta}$ while in the APARCH specification, a power transform of the conditional variance is modelled (σ_t^δ). One can solve this problem by re-writing σ_t^2 as $\left(\sigma_t^\delta\right)^{\frac{2}{\delta}}$ which leads to:

$$\frac{\partial \sigma_t^2}{\partial(\mu',\vartheta',\gamma')} = \frac{2\sigma_t^2}{\delta\sigma_t^\delta}\frac{\partial \sigma_t^\delta}{\partial(\mu',\vartheta',\gamma')} \tag{7}$$

and

$$\frac{\partial \sigma_t^2}{\partial \delta} = \frac{2\sigma_t^2}{\delta\sigma_t^\delta}\left(\frac{\partial \sigma_t^\delta}{\partial \delta} - \frac{\sigma_t^\delta \ln(\sigma_t^\delta)}{\delta}\right). \tag{8}$$

Our goal is thus to find a tractable solution of $\frac{\partial \sigma_t^\delta}{\partial \eta}$ which can be done in four steps.

- First step. Given the choice we made for the initial values of the pre-sample terms $k(\varepsilon_{t-i})^\delta$ and σ_t^δ, differentiating with respect to the conditional mean parameters (μ) gives:

$$\begin{aligned}\frac{\partial \sigma_t^\delta}{\partial \mu} &= \delta\sum_{i=1}^{q}\alpha_i\left[k(\varepsilon_{t-i})^{\delta-1}(I_{t-i}^* + \gamma_i)x_{1,t-i}\right]^{\Im(t-i)} \\ &\times \left[\frac{1}{T}\sum_{s=1}^{T}(|\varepsilon_s - \gamma_i\varepsilon_s|)^{\delta-1}(I_s^* + \gamma_i)x_{1,s}\right]^{1-\Im(t-i)} \\ &+ \sum_{j=1}^{p}\beta_j\left(\frac{\partial \sigma_{t-j}^\delta}{\partial \mu}\right)^{\Im(t-j)}\left[-\frac{\delta}{T}\left(\frac{1}{T}\sum_{s=1}^{T}\varepsilon_s^2\right)^{\frac{\delta-2}{2}}\sum_{s=1}^{T}\varepsilon_s x_{1,s}\right]^{1-\Im(t-j)},\end{aligned} \tag{9}$$

where

$$I_t^* = \begin{cases} -1 \text{ if } \varepsilon_t > 0 \\ 1 \text{ if } \varepsilon_t < 0 \end{cases} \text{ and } \Im_t = \begin{cases} 1 \text{ if } t > 0 \\ 0 \text{ if } t \leq 0 \end{cases}.$$

Note that $I_t^* = \frac{\partial|\varepsilon_t|}{\partial \mu}$ and is not defined for $\varepsilon_t = 0$. However, even is situation is possible, it is almost unlikely in practice.

- Second step.

$$\frac{\partial \sigma_t^\delta}{\partial \vartheta} = d_t + \sum_{j=1}^{p} \beta_j \frac{\partial \sigma_{t-j}^\delta}{\partial \vartheta}, \tag{10}$$

where $\frac{\partial \sigma_t^\delta}{\partial \vartheta} = 0$ for $t \leq 0$.

- Third step. In a similar way, one can show that:

$$\frac{\partial \sigma_t^\delta}{\partial \gamma} = d_t^* + \sum_{j=1}^{p} \beta_j \frac{\partial \sigma_{t-j}^\delta}{\partial \gamma}, \tag{11}$$

where d_t^* is a $(1 \times q)$ vector whose i^{th} element is $\alpha_i \frac{\partial k(\varepsilon_{t-i})^\delta}{\partial \gamma_i}$ with

$$\frac{\partial k(\varepsilon_{t-i})^\delta}{\partial \gamma_i} = \begin{cases} -\delta k(\varepsilon_{t-i})^{\delta-1} \varepsilon_{t-i} \text{ if } t > 0 \\ -\frac{\delta}{T} \sum_{s=1}^{T} \left(|\varepsilon_s - \gamma_i \varepsilon_s|\right)^{\delta-1} \varepsilon_s \text{ if } t \leq 0 \end{cases} \tag{12}$$

and $\frac{\partial \sigma_t^\delta}{\partial \gamma} = 0$ for $t \leq 0$.

- Last step. Finally, differentiating with respect to δ gives:

$$\begin{aligned} \frac{\partial \sigma_t^\delta}{\partial \delta} &= \sum_{i=1}^{q} \alpha_i \left[k(\varepsilon_{t-i})^\delta \ln k(\varepsilon_{t-i})\right]^{\Im_{(t-i)}} \left[\frac{1}{T} \sum_{s=1}^{T} (|\varepsilon_s| - \gamma_i \varepsilon_s)^\delta \ln(|\varepsilon_s| - \gamma_i \varepsilon_s)\right]^{1-\Im_{(t-i)}} \\ &+ \sum_{j=1}^{p} \beta_j \left(\frac{\partial \sigma_{t-j}^\delta}{\partial \delta}\right)^{\Im_{(t-j)}} \left[0.5 \left(\frac{1}{T} \sum_{s=1}^{T} \varepsilon_s^2\right)^{\frac{\delta}{2}} \ln\left(\frac{1}{T} \sum_{s=1}^{T} \varepsilon_s^2\right)\right]^{1-\Im_{(t-j)}}. \end{aligned} \tag{13}$$

3 Empirical application

In this empirical application we consider daily data for a stock market indexes, i.e. the NIKKEI stock index for the period 4/1/1984 - 21/12/2000 (4246 observations, source: Datastream). Daily returns (in %) are defined as $y_t = 100\,[ln(p_t) - ln(p_{t-1})]$, where p_t is the price at day t.

We consider an APARCH$(1, 1)$ specification:

$$\begin{aligned} y_t &= \mu + \varepsilon_t \\ \sigma_t^\delta &= \omega + \alpha_1 \left(|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1}\right)^\delta + \beta_1 \sigma_{t-1}^\delta \\ z_t &\sim i.i.d.\ N(0, 1). \end{aligned}$$

Estimation has been done first by using numerical gradients. In a second step (using the same starting values as in the first case), estimation has been carried out using the analytical gradients presented in the previous sections.

Table 1: Gaussian QML estimation results of the APARCH(1,1)

	starting values	Numerical Score	Analytical Score
μ	0.1	0.04016 (0.01409)	0.04016 (0.01408)
ω	0.1	0.04028 (0.00558)	0.04028 (0.00558)
α_1	0.85	0.84713 (0.01095)	0.84713 (0.01096)
γ	0.04	0.15189 (0.01188)	0.15189 (0.01188)
β_1	0.2	0.46892 (0.04967)	0.46892 (0.04969)
δ	1.5	1.33403 (0.13810)	1.33403 (0.13814)
Time (in sec.)		154.51	60.53

Robust standard errors are reported in parentheses.

Table 1 presents Gaussian QML estimation results of the APARCH(1,1).

First, we see that the extra flexibility of the APARCH specification is required. Both the asymmetry coefficient (γ) and the power (δ) estimates suggest that a usual GARCH model is not appropriate to model the NIKKEI. This is also confirmed by Likelihood Ratio (LR) tests for the null hypothesis $H_0 : \delta = 1$ and $\gamma = 0$ (not reported here to save space). Comparing columns 3 and 4, one can see that numerical scores give very similar results to the analytical ones.[3] However, it is clear that using the analytical scores highly speeds up the estimation procedure (it is about 2.5 times faster).

4 Conclusion

In this paper, we derive analytical expressions for the score of the APARCH model. Doing so we derive the analytical score of a broad range of ARCH-type models since the APARCH nests at least seven models. Using a real dataset, we show that analytical scores highly speed up the maximum-likelihood estimation.

[3]Estimation has been done using Gauss 3.5 and the maximization package maxlik. Table 1 reports the results based on the BHHH algorithm of Berndt, Hall, and Hausman (1974). The same conclusion holds if we use the Newton-Raphson algorithm.

References

Berndt, E., H. R. Hall, B.H., and J. Hausman (1974): "Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement*.

Black, F. (1976): "Studies of Stock Market Volatility Changes," *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pp. 177–181.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T., R. Engle, and m. Nelson, D.B. (1994): *ARCH Models*chap. 4, Handbook of Econometrics. North Holland Press, R.F. Engle, D., McFadden (eds.), Amsterdam.

Bollerslev, T., and J. Wooldridge (1992): "Quasi-maximum Likelihood Estimation and Inference in Dynamic Models with Time-varying Covariances," *Econometric Reviews*, 11, 143–172.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Engle, R. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Fiorentini, G., G. Calzolari, and L. Panattoni (1996): "Analytic Derivatives and the Computation of GARCH Estimates," *Journal of Applied Econometrics*, 11, 399–417.

Gable, J., S. Van Norden, and R. Vigfusson (1997): "Analytical Derivatives for Markov Switching Models," *Computational Economics*, 10, 187–194.

Geweke, J. (1986): "Modeling the Persistence of Conditional Variances: A Comment," *Econometric Review*, 5, 57–61.

Giot, P., and S. Laurent (2001): "Value-at-Risk for long and short positions," Forthcoming in *Journal of Applied Econometrics*.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the Relation Between Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

He, C., and T. Teräsvirta (1999a): "Higher-order Dependence in the General Power ARCH Process and a Special Case," Stockholm School of Economics, Working Paper Series in Economics and Finance, No. 315.

——— (1999b): *Statistical Properties of the Asymmetric Power ARCH Process,* chap. 19, pp. 462–474, Cointegration, causality, and forecasting. Festschrift in honour of Clive W.J. Granger. in Engle, Robert F. and Halbert White, oxford university press edn.

Higgins, M., and A. Bera (1992): "A Class of Nonlinear ARCH Models," *International Economic Review*, 33, 137–158.

McCullough, B., and H. Vinod (1999): "The Numerical Reliability of Econometric Software," *Journal of Economic Literature*, 37, 633–665.

Mittnik, S., and M. Paolella (2000): "Conditional Density and Value-at-Risk Prediction of Asian Currency Exchange Rates," *Journal of Forecasting*, 19, 313–333.

Palm, F. (1996): "GARCH Models of Volatility," *in Maddala, G.S., Rao, C.R., Handbook of Statistics*, pp. 209–240.

Pentula, S. (1986): "Modeling the Persistence of Conditional Variances: A Comment," *Econometric Review*, 5, 71–74.

Schwert, W. (1990): "Stock Volatility and the Crash of '87," *Review of Financial Studies*, 3, 77–102.

Taylor, S. (1986): *Modelling Financial Time Series.* Wiley, New York.

Weiss, A. (1986): "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107–131.

Zakoian, J.-M. (1994): "Threshold Heteroskedasticity Models," *Journal of Economic Dynamics and Control*, 15, 931–955.