Economics Bulletin

Volume 42, Issue 4

Stationary parameterization of GARCH processes

Tucker S McElroy
U.S. Census Bureau

Abstract

We propose using the multivariate logistic transform to re-parameterize the Autoregressive Conditionally Heteroscedastic model such that the necessary stationarity constraints are automatically imposed, thereby allowing for unconstrained optimization when computing quasi-maximum likelihood estimates. A few simulations and a standard R data set of daily closing prices (Germany DAX) provide illustrations of the re-parameterization. We offer some numerical comparisons to available R packages (fgarch and rugarch), and comment on the potential advantages of the new technique.

This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not those of the U.S. Census Bureau.

Citation: Tucker S McElroy, (2022) "Stationary parameterization of GARCH processes", *Economics Bulletin*, Volume 42, Issue 4, pages 1908-1930

Contact: Tucker S McElroy - tucker.s.mcelroy@census.gov.

Submitted: December 28, 2021. **Published:** December 30, 2022.

1 Introduction

The Autoregressive Conditionally Heteroscedastic (ARCH) process, and its generalization (GARCH), are useful for modeling time series data that exhibit stochastic volatility (Bollerslev, 1986). The basic conceit of the ARCH process is that its conditional variance, viewed as a stochastic process, is expressible as a linear combination of past squares of the data. Imposing that the coefficients be non-negative, and that their sum is less than one, is sufficient to guarantee positivity of the conditional variance and strict stationarity of the process. A more general condition for strict stationarity, involving the Lyapunov exponent, is discussed in Francq and Zakoian (2019), but the stricter summation condition of Bollerslev (1986) is necessary for the unconditional variance to be finite; thus, any pseudo-likelihood method for fitting a finite-variance ARCH or GARCH model must in principle enforce these positivity and summation conditions on the parameters.

State of the art software, such as the *rugarch* R package (Ghalanos, 2014), apparently approach the problem using solvers that allow for box and inequality constraints. However, as pointed out in Pinheiro and Bates (1996), nonlinear optimization with box constraints is a numerically inefficient method for fitting a stochastic model. It is preferable to re-parameterize the stochastic process, essentially by obtaining a bijection between the parameter manifold and some Euclidean space, and then proceed via unconstrained numerical optimization with respect to the new parameters. As an example, a variance σ^2 has the inequality constraint that $\sigma^2 > 0$; the new parameter $\psi \in \mathbb{R}$ defined by the logarithm bijection $\psi = \log(\sigma^2)$ is a commonly-used device.

A bijection (between the parameter manifold and Euclidean space) for ARCH and GARCH processes based on the multivariate logistic transform (Aitchison and Shen, 1980) was explored in Exercise 11.17 of McElroy and Politis (2020), and is here developed and implemented. This article describes the bijection in Section 2, makes some numerical comparisons (Section 3), and summarizes the implications (Section 4). R functions for simulation and fitting of ARCH and GARCH are available from the author's GitHub: `https://github.com/tuckermcelroy/GARCH-param`.

2 A Bijection for ARCH and GARCH

The ARCH process $\{X_t\}$ of order p is defined via

$$X_t = \sigma_t \, Z_t, \qquad \sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j \, X_{t-j}^2,$$

where $\{Z_t\}$ are i.i.d. $(0,1)$ random variables; letting $\mathcal{F}_{-\infty}^{t}$ be the information set of present and past value of $\{X_t\}$, we assume that Z_t is independent of $\mathcal{F}_{-\infty}^{t-1}$ for each t. The parameters of a stationary ARCH(p) process must satisfy $a_0 > 0$ and

$$a_j \in (0,1) \quad 1 \leq j \leq p, \quad \sum_{j=1}^{p} a_j \in (0,1). \tag{1}$$

The positivity condition in (1) can be relaxed, allowing zero values for a_j (for $0 \leq j \leq p$), but for purposes of re-parameterization we shall assume positivity of all the parameters. Let the space described by (1) be denoted $\mathcal{A}_p$. We propose a bijection Ω from $\mathcal{A}_p$ to $\mathbb{R}^p$, which is described in the case that $p > 1$ as follows. (Note that when $p = 1$, there is only the one constraint, and we set $a_1 = (1 + e^{-x_1})^{-1}$ for $x_1 \in \mathbb{R}$.)

Proposition 1 *For $p > 1$, the mapping $\Omega : \mathcal{A}_p \to \mathbb{R}^p$ is a bijection, where*

$$\Omega(\underline{a}) = \begin{bmatrix} -\log\left(1/\sum_{j=1}^{p} a_j - 1\right) \\ -\log(a_2/a_1) \\ \vdots \\ -\log(a_p/a_1) \end{bmatrix}, \quad \Omega^{-1}(\underline{x}) = \begin{bmatrix} \left(1 + \sum_{j=2}^{p} e^{-x_j}\right)^{-1} \left(1 + e^{-x_1}\right)^{-1} \\ e^{-x_2} \left(1 + \sum_{j=2}^{p} e^{-x_j}\right)^{-1} \left(1 + e^{-x_1}\right)^{-1} \\ \vdots \\ e^{-x_p} \left(1 + \sum_{j=2}^{p} e^{-x_j}\right)^{-1} \left(1 + e^{-x_1}\right)^{-1} \end{bmatrix}.$$

Proof of Proposition 1. That the inverse of Ω is the stated Ω^{-1} is easily checked. To see that both maps are surjective, we proceed by the following argument. First, take $\underline{a} \in \mathcal{A}_p$, and set $r = \sum_{j=1}^{p} a_j$, which will take values in $(0, 1)$. Let $\alpha_j = a_j/r$, and note that $\sum_{j=1}^{p} \alpha_j = 1$. Define $\underline{x}$ via

$$x_1 = -\log(r^{-1} - 1), \quad x_j = -\log(\alpha_j/\alpha_1) \quad 2 \leq j \leq p.$$

Clearly, as r ranges over $(0, 1)$ we see that x_1 ranges over $\mathbb{R}$, and we can obtain any value $\underline{x} \in \mathbb{R}^p$ by choice of a corresponding $\underline{a} \in \mathcal{A}_p$. Thus Ω is surjective. Inverting this mapping, we obtain for any given $\underline{x} \in \mathbb{R}^p$

$$\alpha_1 = \frac{1}{1 + \sum_{k=2}^{p} e^{-x_k}}, \quad \alpha_j = \frac{e^{-x_j}}{1 + \sum_{k=2}^{p} e^{-x_k}} \quad 2 \leq j \leq p,$$

along with $r = \left(1 + e^{-x_1}\right)^{-1}$, which is also in $(0, 1)$. It follows that each α_j belongs to $(0, 1)$ and $\sum_{j=1}^{p} \alpha_j = 1$ holds. Setting $a_j = r\,\alpha_j$ for $1 \leq j \leq p$, we verify that the resulting $\underline{a} \in \mathcal{A}_p$. This shows that Ω^{-1} is surjective. $\square$

We can write the ARCH pseudo-likelihood as a function of $\underline{x} \in \mathbb{R}^p$, where $\underline{a} = \Omega^{-1}(\underline{x})$ (and $a_0 = e^{x_0}$, for $x_0 \in \mathbb{R}$). This approach avoids two problems: (i) evaluating the pseudo-likelihood at values $\underline{a} \notin \mathcal{A}_p$ would result in an evaluation difficulty (or meaningless parameters), and (ii) the failure to explore the entire space $\mathcal{A}_p$.

Remark 1 Zero Parameters If some parameters are known to equal zero, then we apply the bijection to the remaining free parameters, noting that the summability condition is unchanged when some parameters are fixed to be zero. In practice, one can fit both an unconstrained ARCH model and a constrained model with some zero values, and then compare the models via information criteria – see Tsay (2005) for further discussion. Potential zero restrictions can first be identified by determining the asymptotic variances of QMLEs from the unconstrained model. Alternatively, if one wants to allow for zero estimation, i.e., taking each $a_j \in [0, 1)$, one could proceed by using the logistic transform on the set $(-\epsilon, 1)$ for some very small $\epsilon > 0$ (Nelson and Cao (1992) show that negative coefficients can still produce a stationary solution).

Remark 2 The GARCH Extension The GARCH(p,q) process (with $p > 0$ assumed) generalizes the ARCH(p), as GARCH(p,0) = ARCH(p), and the equation for the variance becomes

$$\sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j\, X_{t-j}^2 + \sum_{j=1}^{q} b_j\, \sigma_{t-j}^2.$$

In place of (1) we now have the constraints

$$a_j \in (0, 1) \quad 1 \leq j \leq p, \quad b_j \in (0, 1) \quad 1 \leq j \leq q, \quad \sum_{j=1}^{p} a_j + \sum_{j=1}^{q} b_j \in (0, 1). \tag{2}$$

Hence, the bijection Ω can also be used for the GARCH process, as the parameters a_j and b_j can be treated alike – we extend the re-parameterization to $\mathbb{R}^{p+q}$.

Viewing the pseudo-likelihood as a function of the new parameters $\underline{x}$, central limit theory can be established under standard conditions in the literature (Tsay, 2005). Applying the delta method with Ω^{-1} provides the central limit theorem for the original parameters. More precisely, suppose that $\widehat{\underline{x}}$ is the QMLE for a true parameter $\widetilde{\underline{x}}$, and correspondingly $\widehat{\underline{a}} = \Omega^{-1}(\widehat{\underline{x}})$ and $\widetilde{\underline{a}} = \Omega^{-1}(\widetilde{\underline{x}})$. From a sample of size T, suppose that $\sqrt{T}\,(\widehat{\underline{x}} - \widetilde{\underline{x}}) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V)$ as $T \to \infty$. Then with J the Jacobian matrix of Ω^{-1} (evaluated at $\widetilde{\underline{x}}$), by the delta method we obtain $\sqrt{T}\,(\widehat{\underline{a}} - \widetilde{\underline{a}}) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, J\,V\,J')$. Direct calculation yields

$$J = \begin{bmatrix} a_1\,(1 - \sum_{j=1}^p a_j) & a_1\,a_2/\sum_{j=1}^p a_j & a_1\,a_3/\sum_{j=1}^p a_j & \cdots & a_1\,a_p/\sum_{j=1}^p a_j \\ a_2\,(1 - \sum_{j=1}^p a_j) & a_2^2/\sum_{j=1}^p a_j - a_2 & a_2\,a_3/\sum_{j=1}^p a_j & \cdots & a_2\,a_p/\sum_{j=1}^p a_j \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_p\,(1 - \sum_{j=1}^p a_j) & a_p\,a_2/\sum_{j=1}^p a_j & a_p\,a_3/\sum_{j=1}^p a_j & \cdots & a_p^2/\sum_{j=1}^p a_j - a_p \end{bmatrix}.$$

In applications, V can be estimated as a numerical Hessian, a common by-product of finding the QMLEs with numerical optimization routines such as BFGS (Golub and Van Loan, 2012).

3 Numerical Illustrations

To illustrate the utility of the new method, we demonstrate that the new parameterization offers superior QMLEs to current methods. (Our code is not available in Rcpp, which would enable fair speed comparisons with available GARCH software.) The author's Github repo has a R Markdown Notebook detailing simulation and fitting of Gaussian and Student t ARCH and GARCH models.

First we simulate ARCH processes with Gaussian innovations, and fit correctly specified models using the stationary parameterization of this article, and compare to results from using the *tseries garch*, *fgarch*, and *rugarch* packages. The new method provides a comparable, or somewhat better, fit to these simulations. For illustration, with a particular simulation the value of the divergence (-2 times the log likelihood) for the Gaussian ARCH(10) evaluated at the true parameter is 7501.034, and the minimal value of the divergence (at the QMLE) is 7490.944. Fitting with *tseries garch* yields 7494.234 at the QMLE, and for *fgarch* 7490.977 is obtained; *rugarch* gives 7490.987. The values for *fgarch* and *rugarch* are close to our results, indicating that constrained optimization and stationary parameterization can yield similar results when the true parameters are in the stationary region.

Examining a Student t ARCH with 4 degrees of freedom, the divergence is 10439.47 at the true parameter, and is 10426.39 at the QMLE. The *tseries garch* is quite a bit higher at 10999.77, because it does not have capability to fit non-Gaussian distributions. However, *fgarch* and *rugarch* can be fitted with a scaled Student t distribution, and we obtain 10820.05 and 10866.61 respectively. Here we notice that estimates of α_0 are too high, but the degrees of freedom and other parameters are estimated well. In these comparisons, we allow *fgarch* and *rugarch* to estimate the mean as well, whereas in our code the mean is assumed to be zero.

The results for GARCH processes are similar. For Gaussian innovations, our simulation has divergence 2124.892 at the true parameter, and is 2124.695 at the QMLE. Results for the *tseries garch* package are slightly better, at 2123.873; 2121.842 is obtained by *fgarch* and 2124.861 by *rugarch*. These are in the same vicnity, and the discrepancies might be explicable through the fact that *fgarch* and *rugarch* estimate the mean. In any case, the new parameterization is quite accurate in its estimation of the GARCH parameters; in the other three packages the α_0 value is too low: 0.0529 for *tseries garch*, 0.0307 for *fgarch*, and 0.0003 for *rugarch*, whereas the true value is .1421, and our estimate is .1660.

In the Student t case (with 5 degrees of freedom) the divergence is 3019.559 at the true parameter and 3014.981 at the QMLE. The divergence for *tseries garch* is 3178.027 at the QMLE, which is substantially worse because of the non-Gaussian data. *fgarch* and *rugarch* perform similarly, with divergences of 3169.724 and 3198.343 respectively; although both are estimating the degrees of freedom well, some of the GARCH parameter estimates are somewhat far from the true values.

For an empirical illustration, we examine the EuStockMarkets data set of base R[1], consisting of daily closing prices of major European stock indices: Germany DAX (Ibis), Switzerland SMI, France CAC, and UK FTSE. The time period is 1991-1998, with weekends and holidays omitted. We fit a GARCH(1,1) using both the Gaussian *tseries garch* package, which yields a divergence of -11917.38 at the QMLE, with parameters estimated to be $\widehat{\alpha_0} = 0.000005$, $\widehat{\alpha_1} = 0.068329$, and $\widehat{\beta_1} = 0.889067$. Our own fit (using the above estimates as initial values in the optimization routine) has a divergence of -12110.84 at the QMLE, with parameters estimated to be $\widehat{\alpha_0} = 0.000001$, $\widehat{\alpha_1} = 0.050284$, $\widehat{\beta_1} = 0.908902$, and $\widehat{\nu} = 6.200205$, where ν is the degrees of freedom parameter for the Student t. For comparison, the results from *fgarch* are: divergence -12019.65 at the QMLE, with parameter estimates $\widehat{\alpha_0} = 0.000002$, $\widehat{\alpha_1} = 0.079022$, $\widehat{\beta_1} = 0.903585$, and $\widehat{\nu} = 6.038375$. For *rugarch* we have -12018.86 for the divergence, with parameter estimates $\widehat{\alpha_0} = 0.000002$, $\widehat{\alpha_1} = 0.079047$, $\widehat{\beta_1} = 0.903785$, and $\widehat{\nu} = 6.016333$. Again, our own encoding does not do a mean estimation, whereas the other packages provide this (omitted). We note that whereas there is good agreement about the degrees of freedom, ultimately the new parameterization provides a lower divergence.

4 Conclusion

This article presents a re-parameterization of the GARCH process using the multivariate logistic transform, such that the stationarity constraints of the finite-variance model are automatically enforced. We believe this is advantageous in numerical computation of QMLEs, because unconstrained optimization can then be utilized – and this is generally preferable to using constrained optimization. Essentially, the implicit constraints on the parameters are "built in" through the re-parameterization.

We remark that there are contexts where a larger parameter set might be entertained; the Lyapunov exponent condition of Nelson (1990) shows the summation constraint in (2) – in the case of $p = q = 1$ – can be violated, while still ensuring strict stationarity, but at the cost of the variance no longer being finite. Interestingly, results by Jensen and Rahbek (2004a, 2004b) and Francq and Zakoian (2012) show that larger compact super-sets of the true parameter spaces can be utilized in nonlinear optimization, in the sense that convergence of parameter estimates can still be guaranteed – albeit, the limiting value of the parameter estimates may correspond to an infinite variance process.

However, if finite-variance models are desired, then the parameter constraints employed in this article are appropriate. Our examples demonstrate the possible benefits: the new parameterization can yield improved fitting results, especially when working with non-Gaussian innovations. What has been proposed, of course, is not a competitor to established packages such as *fgarch* and *rugarch*, but rather this article's parameterization could be considered for incorporation into such software.

[1]https://www.rdocumentation.org/packages/datasets/versions/3.6.2/topics/EuStockMarkets

References

[1] Aitchison, J. and S. Shen (1980) "Logistic-normal distributions: Some properties and uses" *Biometrika* **67**(2): 261-272.

[2] Bollerslev, T. (1986) "Generalized autoregressive conditional heteroskedasticity" *Journal of Econometrics* **31**(3), 307-327.

[3] Francq, C. and J. Zakoian (2012) "Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models" *Econometrica* **80**(2), 821-861.

[4] Francq, C. and J. Zakoian (2019) *GARCH models: structure, statistical inference and financial applications*, John Wiley & Sons: New York.

[5] Ghalanos, A. (2014) "Univariate GARCH models" R package version 1.3-5.

[6] Golub, G. and C. Van Loan (2012) *Matrix computations* (Vol. 3), JHU Press: Baltimore.

[7] Jensen, S. and A. Rahbek (2004a) "Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case" *Econometrica* **72**(2), 641-646.

[8] Jensen, S. and A. Rahbek (2004b) "Asymptotic inference for nonstationary GARCH" *Econometric Theory* **20**(6), 1203-1226.

[9] McElroy, T. and D. Politis (2020) *Time Series: A First Course with Bootstrap Starter*, Chapman Hall: New York.

[10] Nelson, D. (1990) "Stationarity and persistence in the GARCH (1, 1) model" *Econometric Theory* **6**(3), 318-334.

[11] Nelson, D. and C. Cao (1992) "Inequality constraints in the univariate GARCH model" *Journal of Business & Economic Statistics* **10**(2), 229-235.

[12] Pinheiro, J. and D. Bates (1996) "Unconstrained parameterizations for variance-covariance matrices" *Statistics and computing* **6**(3), 289-296.

[13] Tsay, R. (2005) *Analysis of Financial Time Series*, John Wiley & Sons: New York.