

Mixtures of Distributions and Volatility

Juan Carlos Abril¹, María de las Mercedes Abril² and Carlos Ismael Martínez³

Universidad Nacional de Tucumán, Facultad de Ciencias Económicas and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). Tucumán. Argentina.

¹jabril@herrera.unt.edu.ar, mabrilblanco@hotmail.com² and cimartinez@herrera.unt.edu.ar³

Using Monte Carlo methods we generate time series with the following features: a) series with distributions that are the mix of two normal distributions with different variances, b) series that satisfy volatility models, c) series that satisfy an AR(1) model but with contaminated errors that follow the same distribution as the mixes given in a) and d) series that follow the same distribution as the mixes given in a) but with conditional heterocedasticity. From the analysis we see that it is difficult to identify in practical situations the real generating process of the series. In fact, the processes that come from distribution mixes have many similar characteristics to the ones that satisfy the volatility scheme. We use the corresponding theoretical considerations and also the usual tools in the identifying process of any time series; that is, series graphs, histograms, the corresponding sampling distributions, correlograms and partial correlograms.

Keywords: Autoregression, Contaminated errors, Distribution mixes, AR(1) models, ARCH models, Volatility.

Introduction

Combination of Normal Distributions

Let's assume that we have the process $\{u_t\}$ whose distribution is a mixture of normal densities, that is $N(\mu, \sigma^2)$ and $N(\mu, k\sigma^2)$, where $0 < \sigma^2 < \infty, k > 0$ and without loss of generality we can suppose that $\mu = 0$. Let's assume also that the random variables u_t are independent between them. Following Lindsay (1995), this mixture can be written as $p N(0, k\sigma^2) + (1-p) N(0, \sigma^2)$, $0 < p < 1$. So the density of u_t is

$$g(u_t) = \frac{p}{\sqrt{2\pi}\sqrt{k}\sigma} \exp\left(-\frac{u_t^2}{2k\sigma^2}\right) + \frac{1-p}{\sqrt{2\pi}\sigma} \exp\left(-\frac{u_t^2}{2\sigma^2}\right),$$

where the parameters are now σ^2 , p , and k , $0 < p < 1$. We can consider p as the proportion of "contamination" among the variables with $N(0, \sigma^2)$ density.

From above we can see that the process $\{u_t\}$ is stationary (in the wide sense). According to Durbin and Koompan (2001, y 2012), this is a process whose distribution has "heavy tails" and can be used to explain some events that can take extreme values with a probability larger than the normal.

We consider in our case that $\sigma^2 = 1$.

Volatility

Volatility can be define as the variance of a random variable, normally a return in economic applications, conditional to all past information. As volatility is not measured directly, it can manifest in many ways when we study any financial series.

If x_t is the series under study, let's set

$$\mu_t = E(x_t | F_{t-1}) = E_{t-1}(x_t), \quad (1)$$

$$\begin{aligned} h_t &= \text{var}(x_t | F_{t-1}) = E\{(x_t - \mu_t)^2 | F_{t-1}\} \\ &= E_{t-1}(y_t - \mu_t)^2 = \text{var}_{t-1}(x_t), \end{aligned} \quad (2)$$

as the conditional mean and the conditional variance of x_t given the information until the instant $t-1$ contained in F_{t-1} .

A typical volatility model has the following form

$$x_t = \mu_t + \sqrt{h_t} \varepsilon_t \quad (3)$$

where $E_{t-1}(\varepsilon_t) = 0$, $\text{var}_{t-1}(\varepsilon_t) = 1$ and the ε_t are independent and identically distributed (*IID*) with an F distribution. The mean and the unconditional variance of x_t will be denoted as $\mu_x = E(x_t)$ and $\sigma_x^2 = \text{var}(x_t)$ respectively, and G is the distribution of x_t . It is clear that (1) and (2) and F determine μ_x , σ_x^2 and G but not the other way.

Method

Data Generation of the Mixes

(1) We generate pseudo random independent numbers $N(0,1)$ denoted by $\varepsilon_1, \dots, \varepsilon_T$. The sequence starts with a value automatically provided by the software. This value can also be selected by the user of the software.

(2) We generate $\hat{p}_t, t=1, \dots, T$, from the uniform distribution in the interval $(0,1)$, that is, from a $U(0,1)$.

(3) We transform ε_t to produce the mixture

$$u_t = \begin{cases} \sqrt{k} \varepsilon_t, & \text{if } \hat{p}_t \leq p \\ \varepsilon_t, & \text{if } \hat{p}_t > p \end{cases} \quad (4)$$

for $t=1, \dots, T$, where the values of k and p were given in the process of generating the mixture. In our case we took $k=9, 16, 25$ and 100 ; $p=0.10, 0.15, 0.20$ and 0.30 having work with combinations of given (k, p) values. In practical situations we may need to estimate the values of k and p , since they have the unknown parameters characteristics. The formula (4) may be written as

$$u_t = \sqrt{kb(\hat{p}_t - p)} \varepsilon_t, \quad (5)$$

where

$$b(a_t) = \begin{cases} 1, & \text{if } a_t \leq 0 \\ k^{-1}, & \text{if } a_t > 0. \end{cases} \quad (6)$$

Generation of the volatility data

We generate the series $\{x_t\}$ that satisfies the model

$$x_t = \sqrt{h_t} \varepsilon_t \quad (7)$$

$$h_t = \alpha_0 + \alpha_1 x_{t-1}^2, t = 1, \dots, T. \quad (8)$$

where ε_t is generated according to what was stated above and α_0 and α_1 were given in the process of generating the series. In our case we consider $\alpha_0 = 1$ and $\alpha_1 = 0.5$. It is clear that (7) together with (8) define a series that follows an ARCH(1) model.

We use $\{x_t\}$ to estimate the values of α_0 and α_1 .

Generation of the data for the AR(1) model

For a given value of the parameter ϕ we generate the series $\{y_t\}$ that follows the model

$$y_t = \phi y_{t-1} + u_t \quad t = 1, \dots, T, \quad (9)$$

where $\{u_t\}$ was defined in (5) and (6). In our case we suppose that $\phi = 0.6$.

We use y_t to obtain the estimates $\hat{\phi}$ of ϕ . These estimates are obtained using maximum likelihood under the assumption of normality.

We define the residuals e_t of the series $\{y_t\}$ that satisfies (9) as

$$e_t = y_t - \hat{y}_t, \text{ where } \hat{y}_t = \hat{\phi} y_{t-1}. \quad (10)$$

Generation of the data for the mixes with conditional heterocedasticity

From the $\varepsilon_1, \dots, \varepsilon_T$ generated in a) and the $\hat{p}_t, t = 0, 1, \dots, T-1$ generated in b) we create the series

$$w_t = \sqrt{k c(\hat{p}_{t-1} - p)} \varepsilon_t \quad (11)$$

where

$$c(a_{t-1}) = \begin{cases} 1, & \text{if } a_{t-1} \leq 0 \\ k^{-1}, & \text{if } a_{t-1} > 0. \end{cases} \quad (12)$$

and the values of k and p were given in c).

Sample

When we work with a single series, without replications, we consider a sample of size $T = 2000$. When we make n replications, we take a sample of size $T = 200$ and the number of replications is $n = 1000$. From this, we generate empirical time series of size T using the steps specified on the previous sections.

Discussion

Series analysis and final remarks

The structure of the mixes defined in (5) and (6) is similar to a typical volatility model like the one in (3);

but when we analyse the correlograms and the partial correlograms of $\{u_t\}$ and $\{u_t^2\}$ we see that in both cases we can accept the hypothesis of lack of serial correlation, which matches the theory and the generating process of the series. The only thing that captures our attention is that for $\{u_t^2\}$ when k and p take high values, the autocorrelation and the partial autocorrelation of a certain order (9 in the case of $k=100$ and $p=0.30$ as we see in Figure 4), or close to it, lead us to reject the hypothesis that their parameters are zero, and this does not have a theoretical explanation. This kind of situation could lead us identify for $\{u_t\}$ an ARCH model of higher order with some of its coefficients of lower order equal to zero and its errors proportionally distributed as a χ_1^2 . When we estimate the ARCH model, the obtained residuals satisfy the hypothesis that they come from a process of random variables independent and identically distributed with a constant mean and variance. We can also see that when p and k increase, the respective distributions of $\{u_t\}$ have a higher probability for extreme values, that is, they become variables with “heavy tails” distributions, which in some cases can be indistinguishable with the corresponding volatility models. All of this is seen in Figure 1, Figure 2, Figure 3 and Figure 4.

The series in (7) and (8) were generated following an ARCH(1) model. When we analyze the correlogram and the partial correlogram of $\{x_t\}$ as it is seen in Figure 5 we can accept the hypothesis of lack of serial correlation. On the other side, when we consider $\{x_t^2\}$ as it is seen in Figure 6 we may identify an AR(2) model, when the generating process corresponds to an AR(1) model, with errors proportionally distributed as χ_1^2 . This feature matches the theory. As for the estimation of α_0 and α_1 , despite we consider normality and it is done by maximum likelihood, we find that the empirical distributions of those estimators obtained by Monte Carlo methods are centered around the true value of the parameters. A further analysis of the distributions for the estimators of α_0 and α_1 is left for the future.

The series in (11) and (12) satisfy the definition of volatility and they fulfill everything in (1) and (2). Apart from that, the series $\{w_t\}$ is, from a practical point of view, indistinguishable from $\{u_t\}$ defined in (5) and (6) which corroborates the fact that mixture of distributions with different variances and volatility are closely related that in some cases can be indistinguishable one from the other.

The series $\{y_t\}$ defined in (9) corresponds to an AR(1) model with errors whose distribution is a mixture of normals with different variances. Those errors $\{u_t\}$ were generated from (4). The estimations of the parameter φ despite we consider normality and they are done using maximum likelihood, show that the empirical distribution of the estimator done by Monte Carlo methods is centered in the true value of the parameter. In Figure 7 we see $\{y_t\}$, its sample density function, its sample autocorrelation function and its partial sample autocorrelation function for $p=0.10$ and $k=9$, and in Figure 8 we see the same for $p=0.30$ and $k=100$. It is clear that despite the generating model for $\{y_t\}$ is an AR(1), this is not the one we identify from looking at the figures. Even so, we can see that when p and k from the mixes increase, it becomes harder to identify the AR(1) model, and we can even accept that the generating process is white noise. Analyzing the residuals $\{e_t\}$ of the series $\{y_t\}$ presented and described before, we see that probably we can identify them as an AR(2) model. For the series $\{e_t^2\}$ the autocorrelation and the partial autocorrelation of a certain order (3 in our case as we see in Figure 11 and Figure 12), or close to it, can lead us to reject the hypothesis that the respective parameters are zero. This last thing may also lead us to identify in practical cases

for the errors $\{u_t\}$ of the AR(1) process in (9), an ARCH model of higher order with some of its coefficients of lower order equal to zero and with its errors distributed as a χ_1^2 . This will make us think again as before that we can confuse models with “heavy tails” with volatily models.

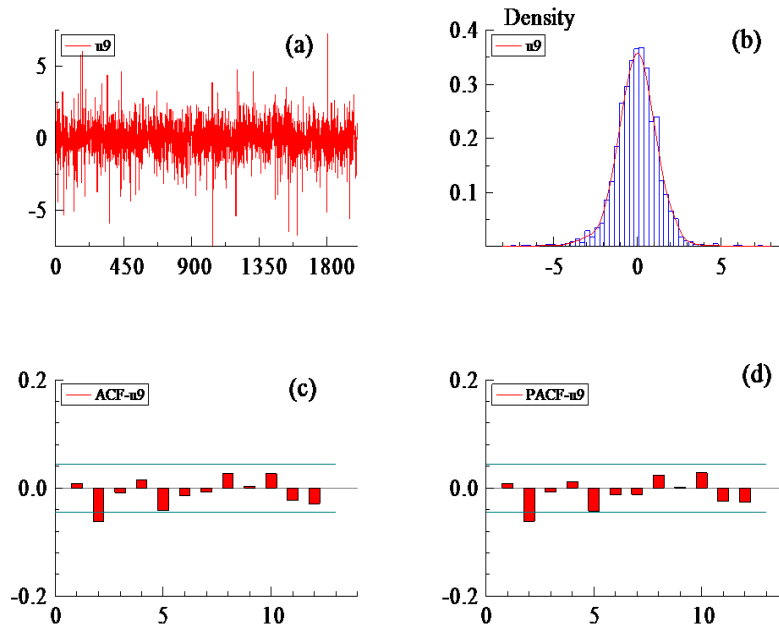


Figure 1. Series u_t generated according to formula (4) as a mixture of normal random variables, with $p = 0.10$ and $k = 9$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

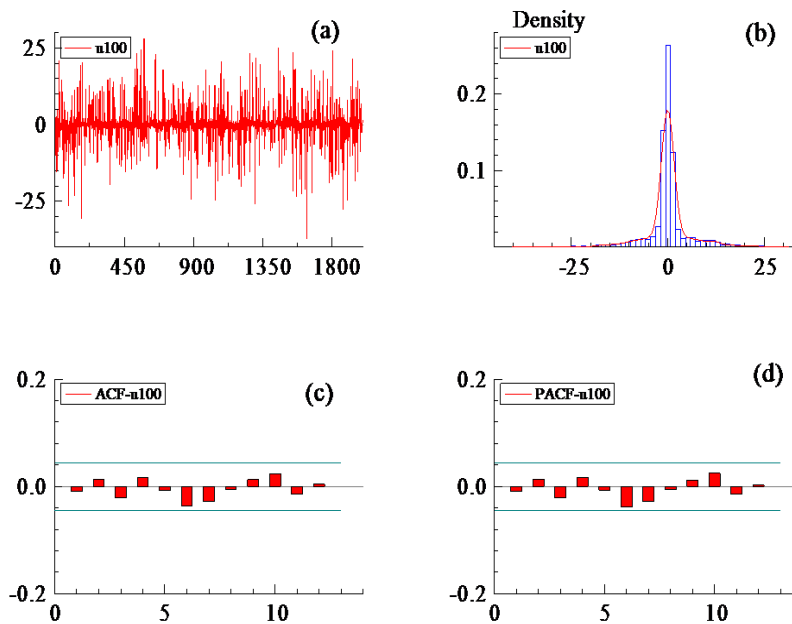


Figure 2. Series u_t generated according to formula (4) as a mixture of normal random variables, with $p = 0.30$ and $k = 100$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

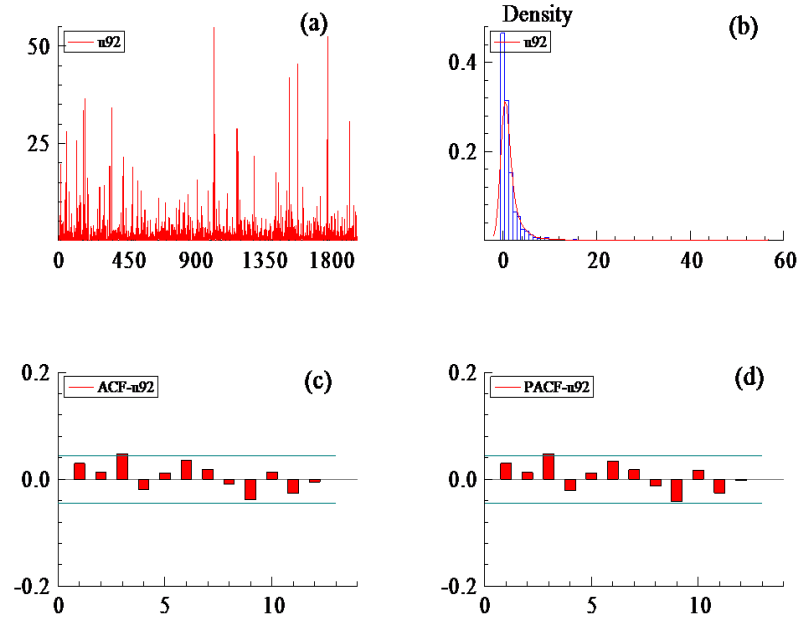


Figure 3. Series u_t^2 where u_t was generated according to formula (4) as a mixture of random normal variables with $p = 0.10$ and $k = 9$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

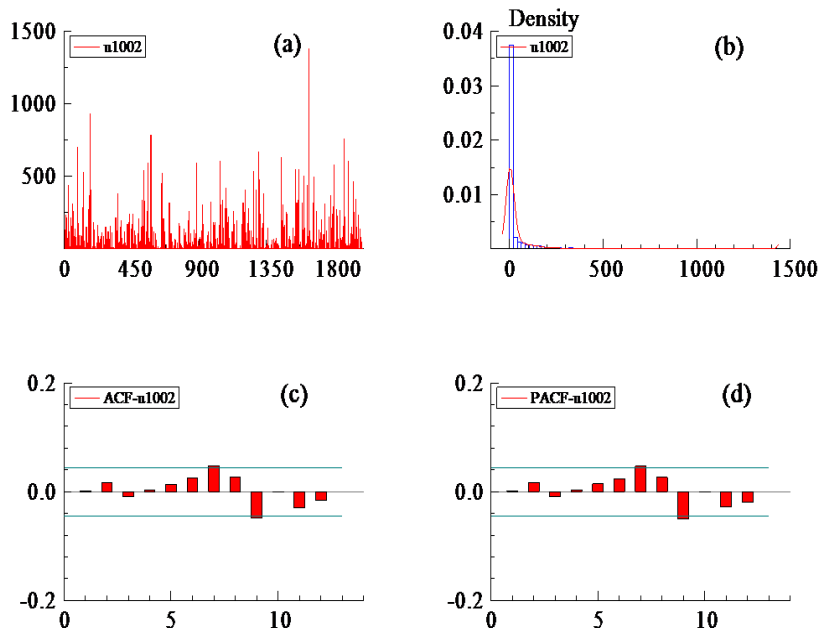


Figure 4. Series u_t^2 where u_t was generated according to formula (4) as a mixture of random normal variables with $p = 0.30$ and $k = 100$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

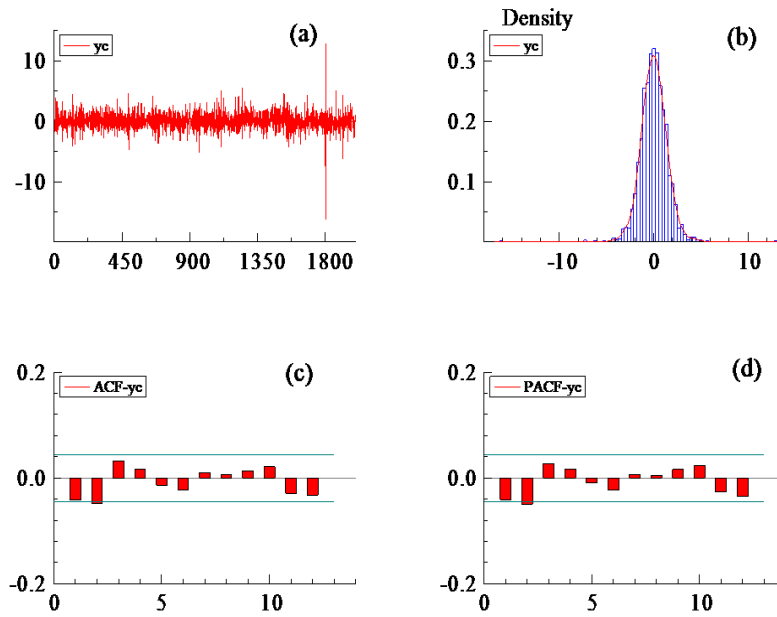


Figure 5. Series x_t generated according to formulas (7) and (8) as a series that satisfies an ARCH(1) model, with $\alpha_0 = 1$ and $\alpha_1 = 0.5$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

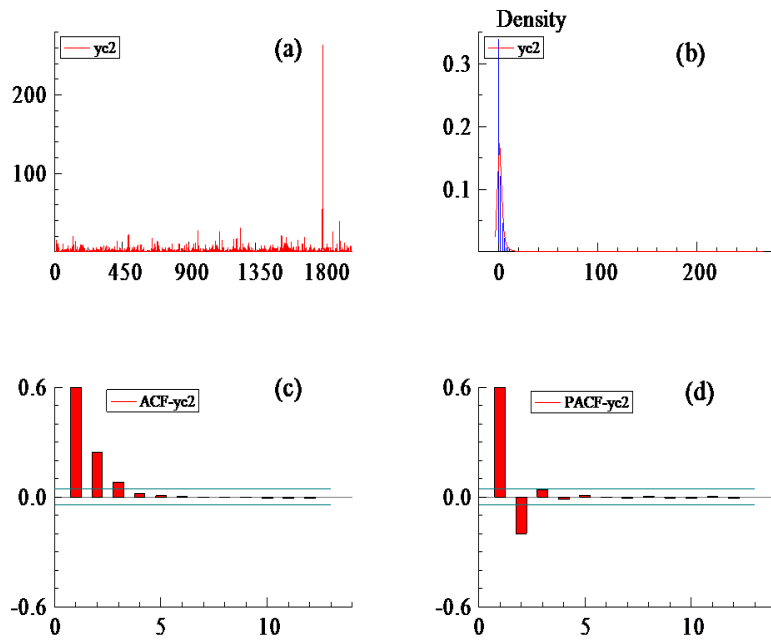


Figure 6. Series x_t^2 generated according to formulas (7) and (8) as a series that satisfies an ARCH(1) model, with $\alpha_0 = 1$ and $\alpha_1 = 0.5$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

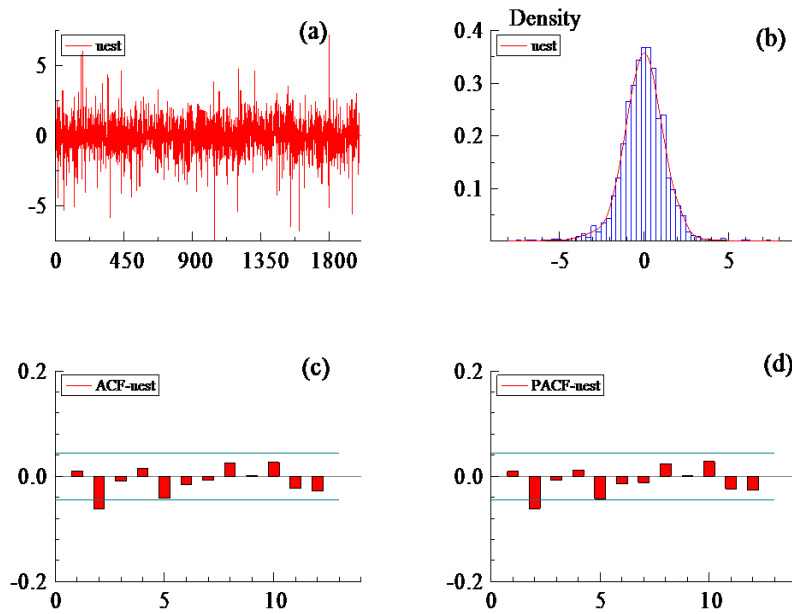


Figure 7. Series y_t generated according formula (9) with errors u_t generated according to formula (4) as a mixture of random normal variables with $p = 0,10$ and $k = 9$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

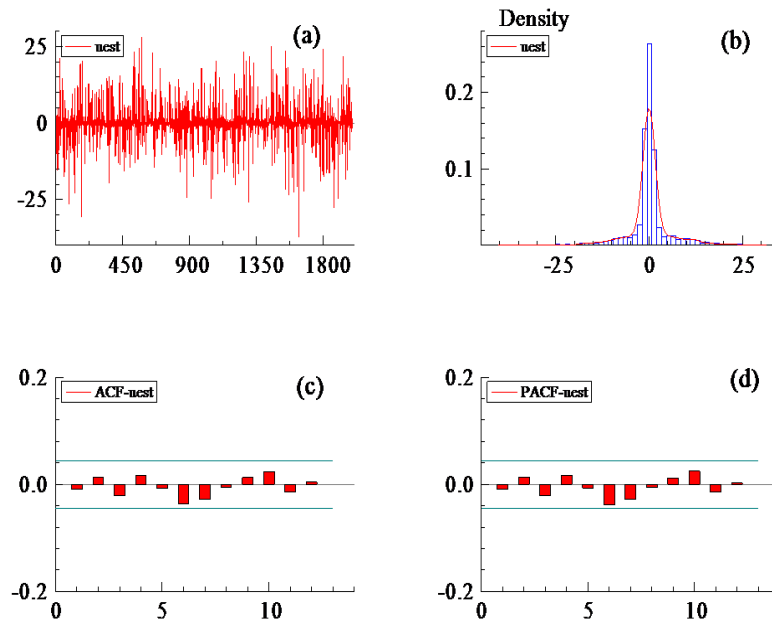


Figure 8. Series y_t generated following formula (9) with errors u_t generated according to formula (4) as a mixture of normal random variables with $p = 0.30$ and $k = 100$. Original series (a), its sample density function (b), its sample autocorrelation function (c) and its partial sample autocorrelation function (d).

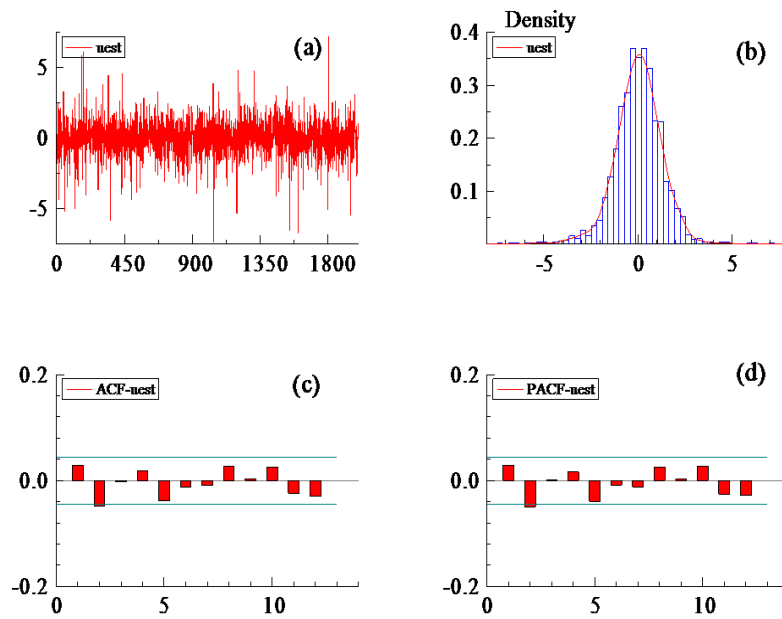


Figure 9. Residuals e_t satisfying (10), of the series y_t generated according to formula (9) with error u_t generated according to the formula (4) as a mixture of normal random variables with $p = 0.10$ and $k = 9$. Original series of residuals (a), its simple density function (b), its sample autocorrelation function (c) and its simple partial autocorrelation function (d).

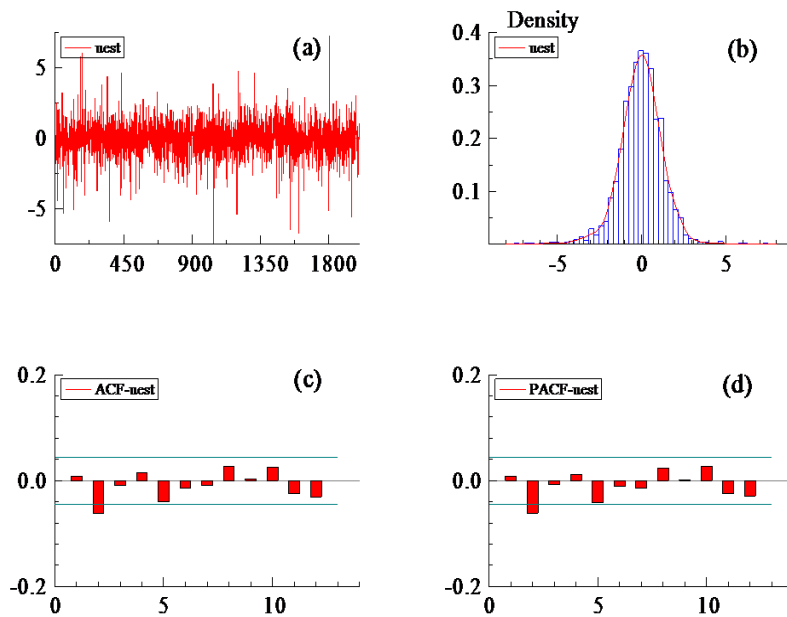


Figure 10. Residuals e_t that satisfy (10), from the series y_t generated according to formula (9) whose errors u_t were generated following formula (4) as a mixture of normal random variables, with $p = 0.30$ and $k = 100$. Original series of residuals (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

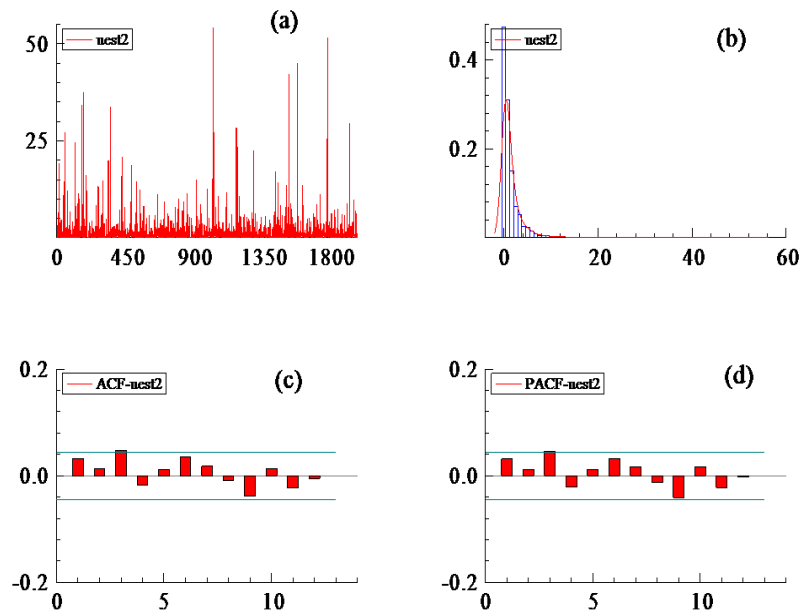


Figure 11. Series e_t^2 of residuals e_t that satisfy formula (10), of the series y_t generated according to formula (9) with errors u_t generated following formula (4) as a mixture of normal random variables with $p = 0.10$ and $k = 9$. Original series of squared residuals (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d).

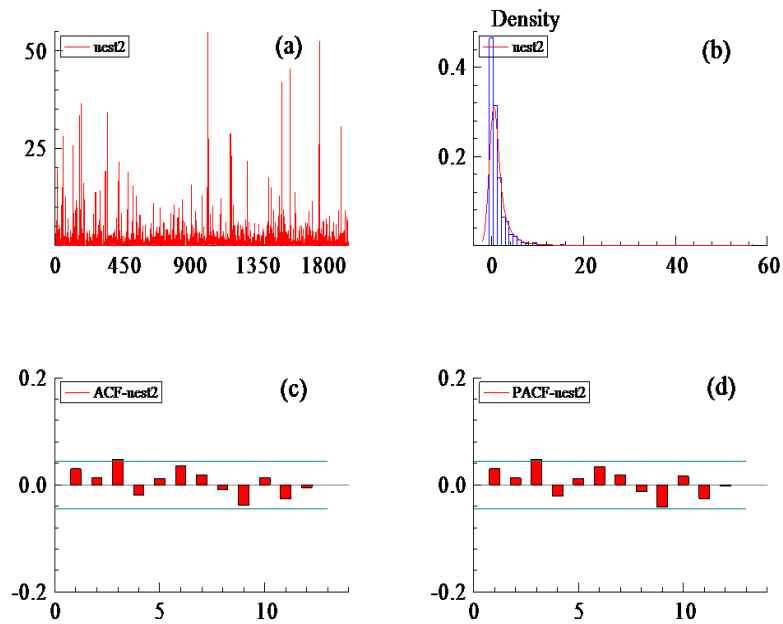


Figure 12. Series e_t^2 of residuals e_t that satisfy (10), from the series y_t generated following formula (9) with errors u_t generated according to formula (4) as a mixture of normal random variables with $p = 0.30$ and $k = 100$. Original series of squared residuals (a), its sample density function (b), its sample autocorrelation function (c) and its sample partial autocorrelation function (d)

References

- Abril, Juan Carlos. (1999), *Análisis de Series de Tiempo Basado en Modelos de Espacio de Estado*, EUDEBA: Buenos Aires.
- Abril, Juan Carlos. (2004). *Modelos para el Análisis de las Series de Tiempo*. Ediciones Cooperativas, Buenos Aires.
- Abril, María de las Mercedes. (2014). *El Enfoque de Espacio de Estado de las Series de Tiempo para el Estudio de los Problemas de Volatilidad*. Phd. Thesis. Universidad Nacional de Tucumán. Argentina.
- Andersen, T.G., T. Bollerslev, F. X. Diebold, and P. Labys. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 579-625.
- Andersen, T.G., R.A. Davis, J.P. Kreiss, and T. Mikosch, T. (2009). *Handbook of Financial Time Series*. Springer: Berlin.
- Bandi, F.M. and J. R. Russell. (2006). *Volatility*. Working paper, GSB, University of Chicago.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307-27.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative process and rates of return. *Review of Economics and Statistics*, 69, 542—547.
- Bollerslev, T. (1988). On the correlation structure for the generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis*, 9, 121—132.
- Bollerslev, T., R. Y. Chou, and K. F. Kroner. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52, 5—59.
- Bollerslev, T., R. F. Engle, and D. B. Nelson. (1994). Arch Models. In *Handbook of Econometrics, Vol. IV* (eds. R.F. Engle and D.L. McFadden), 2959—3038. North Holland: New York.
- Durbin, J. and S. J. Koopman (2001), *Time Series Analysis by State Space Methods*, Oxford University Press: Oxford.
- Durbin, J. and S. J. Koopman. (2012). *Time Series Analysis by State Space Methods (2nd Edition)*. Oxford University Press: Oxford.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, 50, 987-1007.
- Ghysels, E., A.C. Harvey and E. Renault. (1996). Stochastic volatility. In C. R. Rao and G. S. Maddala (eds.), *Statistical Methods in Finance*, pp. 119-91. North-Holland: Amsterdam.
- Lindsay, B. G. (1995). *Mixture models: theory, geometry and applications*. Regional conference series in probability and statistics, Institute of Mathematical Statistics.
- Poon, S.H. and C.W.J. Granger. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41, 478-539.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen (eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1-67. Chapman and Hall: London.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press: Oxford.
- Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley: Chichester.
- Venables, W. N. and B. D. Ripley. (1997). *Modern Applied Statistics with S-Plus (2nd Edition)*. Springer-Verlag: New York.
- Verbeek, M. (2008). *Modern Econometrics (3rd Edition)*. Wiley: New York.
- Zivot, E. and J. Wang. (2006). *Modelling Financial Time Series With SPLUS*. Second Edition. Springer: New York.