AR(1) TIME SERIES PROCESS
Econometrics 7590

Zsuzsanna HORVÁTH and Ryan JOHNSTON

Abstract: We define the AR(1) process and its properties and applications. We demonstrate the applicability of our method to model time series data consisting of daily values of the interest rate on federal funds. We show that correctly identifying the distribution of the errors terms allows for correct modelling of the data. Furthermore, we show our conclusions concerning the hypothesis test for goodness-of-fit cannot be rejected.

Contents

1   Introduction

Any array of time and numbers that are associated can be considered a time series, however, we typically think of a time series as an ordered sequence of values (data points) of variables at equally spaced time intervals. Time series models are used in an attempt make sense of time series. They are used to obtain an understanding of the underlying factors and theory (where did the data come from? What are the statistical properties of the data? What trends are present?) that produce the observed data. The results are then used to fit these models for predictive forecasting and monitoring. Time series analysis is the study of these models and is used in many applications including budgetary analysis, census analysis, economic forecasting, inventory studies, process and quality control, stock market analysis, utility studies, workload projections, and yield projections.

There exist many models used for time series, however, there are three very broad classes that are used most often. These are the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models. These models are often intertwined to generate new models. For example, the autoregressive moving average model (ARMA) combines the (AR) model and the (MA) model. Another example of this is the autoregressive integrated moving average (ARIMA) model, which combine all three of the models previously mentioned. The most commonly used model for time series data is the autoregressive process. The autoregressive process is a difference equation determined by random variables. The distribution of such random variables is the key component in modeling time series. The time series considered in this paper is the first order autoregressive equation, written as AR(1). The AR(1) equation is a standard linear difference equation

$$X_k = \rho X_{k-1} + \varepsilon_k. \qquad k = 0, \pm 1, \pm 2, \ldots$$

where the ε_k are called the error terms or innovations and are what make up the variability in the time series.

For practical reasons, it is desirable to have a unique solution that is independent of time (stationary) and a function of the past error terms. A solution that is independent of time allows one to be able to avoid an initial condition, which may be difficult to find or at an inconvenient location in a time series. A solution as a function of the past error terms is necessary in models used to forecast.

It is important to note that the existence of a unique stationary solution is non-trivial. Assumptions about the error terms are made to guarantee a unique stationary solution. Much of the literature on AR models assume that the error terms are an uncorrelated sequence of random variables with a probability distribution that has zero for the mean and a finite variance. These assumptions limit our ability to model time series that exhibit

3

more volatile behavior such as the stock market or interest rates. Fortunately it has been shown that weaker assumptions can be made to allow the use of distributions that more closely model high volatility time series data without losing the guarantee that there exists a unique stationary solution.

2   AR(1) Time Series

The *pth* order autoregressive time series (often written as AR(p)) X_k is given by the following equation

$$\sum_{i=0}^{p} \rho_i X_{k-i} = \varepsilon_k, \qquad\qquad k = 0, \pm 1, \pm 2, \ldots$$

where $\rho_0 \neq 0, \rho_p \neq 0$ and the ε_t are typically assumed to be uncorrelated $(0, \sigma^2)$ random variables (i.e. $E[\varepsilon_0] = 0, E[\varepsilon_0^2] = \sigma^2$). Thus, the AR(1) process is a *first* order autoregressive time series and most commonly defined by the following equations

$$X_k = \rho X_{k-1} + \varepsilon_k \qquad\qquad k = 0, \pm 1, \pm 2, \ldots$$

which is simply a first order linear difference equation.

A *difference equation* is an expression relating a variable X_k to its previous values. For example, a difference equation is an equation directly relating the value X at time k to the value of X at a previous time period, plus another variable ε dependent on time k (ε_k). A difference equation written as

(2.1.1) $$X_k = \rho X_{k-1} + \varepsilon_k$$

is called a first-order difference equation because only the first backshift or lag of the variable appears in the equation. Note that X_k is represented as a linear function of X_{k-1} and ε_k. For the AR(1) process, the $\varepsilon_k's$ are regarded as random variables and are often referred to as the error terms or innovations. They generally make up the variability that is part of the system when it moves from one time period to the next. Nothing has yet been mentioned about the variable ρ, which is a constant. The value of ρ will be considered shortly.

Using the difference equation above, X_k is obtained from knowing the value of X_{k-1}, and in turn X_{k-1} is obtained from knowing X_{k-2} and so on. Observe that X_k can be obtained from an initial value X_0 that is k-time periods prior. Thus, a solution for X_k can be found through recursive substitution

$$
\begin{aligned}
X_k &= \rho X_{k-1} + \varepsilon_k \\
&= \rho(\rho X_{k-2} + \varepsilon_{k-1}) + \varepsilon_k \\
&= \rho^2 X_{k-2} + \rho \varepsilon_{k-1} + \varepsilon_k \\
&= \rho^3 X_{k-3} + \rho^2 \varepsilon_{k-2} + \rho \varepsilon_{k-1} + \varepsilon_k \\
&\;\;\vdots \\
&= \rho^k X_0 + \rho^{k-1} \varepsilon_1 + \cdots + \rho \varepsilon_{k-1} + \varepsilon_k,
\end{aligned}
$$

where X_0 is the initial value. This can be generalized to the case wherein the initial value is X_{k-N}. For example, consider two time periods, one at time k and the other counted N time periods back from k, denoted as $k - N$. The value of X at time $k - N$ is like the initial value, so X_{k-N} becomes the value of X known at time $k - N$. Thus, in the general case, the value of X at time k (X_k) can be obtained using recursive substitution

$$
\begin{aligned}
X_k &= \rho X_{k-1} + \varepsilon_k \\
&= \rho(\rho X_{k-2} + \varepsilon_{k-1}) + \varepsilon_k \\
&= \rho^2 X_{k-2} + \rho \varepsilon_{k-1} + \varepsilon_k \\
&= \rho^3 X_{k-3} + \rho^2 \varepsilon_{k-2} + \rho \varepsilon_{k-1} + \varepsilon_k \\
&\;\;\vdots \\
&= \rho^N X_{k-N} + \rho^{N-1} \varepsilon_{k-(N-1)} + \cdots + \rho \varepsilon_{k-1} + \varepsilon_k.
\end{aligned}
$$

Thus, for every N

(2.2.1)
$$
X_k = \rho^N X_{k-N} + \sum_{i=0}^{N-1} \rho^i \varepsilon_{k-i}.
$$

The limit as $N \to \infty$ of equation (2.2.1) and for $|\rho| < 1$ indicates $X_k = \sum_{i=0}^{\infty} \rho^i \varepsilon_{k-i}$ must be the solution of the difference equation, assuming that the infinite sum $\sum_{i=0}^{\infty} \rho^i \varepsilon_{k-i}$ exists. If ε_k is stationary, then $\sum_{i=0}^{\infty} \rho^i \varepsilon_{k-i}$ is also stationary. Recall that stationary implies that the time series is independent of time, in other words, the joint distribution of $(X_{k_1}, X_{k_2}, \ldots, X_{k_j})$ is the same as that of $(X_{k_1+h}, X_{k_2+h}, \ldots, X_{k_j+h})$. More specifically,

$$
F_{X_{k_1}, X_{k_2}, \ldots, X_{k_j}}(x_{k_1}, x_{k_2}, \ldots, x_{k_j}) = F_{X_{k_1+h}, X_{k_2+h}, \ldots, X_{k_j+h}}(x_{k_1+h}, x_{k_2+h}, \ldots, x_{k_j+h}).
$$

The joint distribution depends only on the difference h, not on the time $(k_1, \ldots, k_j)$.

It may also be observed that the effect of $\varepsilon_{k-(N-1)}$ on X_k is ρ^{N-1}. Thus for $|\rho| < 1$, ρ^{N-1} geometrically goes to zero. Whereas when $|\rho| > 1$, ρ^{N-1} grows exponentially over

time. Thus, for $|\rho| < 1$, the system is considered stable, or the further back in time a given change occurs, the less it will affect the present. The given change eventually dies out over time. For $|\rho| > 1$, the system blows up. A given change from the past increasingly affects the future as time goes on. For practical purposes, it is desirable to have a system that is less effected the further we go into the past. It is important to note that for $|\rho| < 1$, the solution is given as a function of the error terms from the past. For $|\rho| > 1$, it is possible to obtain a stable solution, but the solution is given as a function of the error terms from the future. For obvious reasons it is typically assumed the value of $|\rho|$ to be less than one since it would not serve to be practical to use a model that requires knowledge of observations from the future rather than the past.

We now take a look at a time series data set taken from the Economagic website (http://www.economagic.com). The data consists of roughly 18 months of daily values of the interest rate on federal funds. There are approximately 500 data points from October 29, 1992 to March 13, 1994. We desire to obtain the distribution of the solution to this time series. Thus, we need to obtain the distribution of the error terms since the AR(1) equation is determined simply by past observations of the random variable X and the error terms. So, if we knew the distribution of the error terms, we would have the distribution of the AR(1) equation. In addition, we would need to know the value of ρ, which we find an estimator for using the method of least squares. The chi-squared goodness-of-fit test is considered a good test in trying to determine the distribution of a random variable even when the paramaters are unknown, which is the case we are faced with. First, we observe a plot of the data to get an idea of what trends there might be. We hope for the data to be stationary, which in fact is the reason we chose the data set that we did. We also create a histogram of the observed data to begin to get an idea of what distribution we could use to test for goodness-of-fit. It may be necessary to "play around" with the parameters of the histogram by decreasing or increasing the interval length for the data to be binned. This is done in order to try to fit the shape of the histogram as closely to a distribution curve of a specific probability distribution. Now we calculute the estimators for ρ and our distribution function and use the chi-squared goodness-of-fit test to determing whether the observed data could be of the hypothesized distribution. The details of obtaining all of this information is explained in the following section.

3   Fitting the Data to the Model

The first step in fitting the collected data to the AR(1) model described in the previous section is to estimate the value of ρ. This will be accomplished using the least squares estimation. The error term is ϵ_k. We want to minimize the sum of the square of errors

for our observed values with respect to ρ. We take the derivative of the sum of squares to get

$$\frac{\partial}{\partial \rho} \sum_{k=2}^{n} (X_k - \rho X_{k-1})^2 = 2 \sum_{k=2}^{n} (X_k - \rho X_{k-1})(-X_{k-1}).$$

We set the derivative equal to 0 and obtain the following:

$$2 \sum_{k=2}^{n} (X_k - \rho X_{k-1})(-X_{k-1}) = 0$$

$$\sum_{k=2}^{n} \left(-X_k X_{k-1} + \rho X_{k-1}^2 \right) = 0$$

$$-\sum_{k=2}^{n} X_k X_{k+1} + \rho \sum_{k=2}^{n} X_{k-1}^2 = 0$$

$$\rho \sum_{k=2}^{n} X_{k-1}^2 = \sum_{k=2}^{n} X_k X_{k-1}$$

Hence, we obtain our least squares estimator for ρ:

(3.1)
$$\hat{\rho} = \frac{\sum_{k=2}^{n} X_k X_{k-1}}{\sum_{k=2}^{n} X_{k-1}^2}.$$

From our observed data, the least squares estimator yields that $\hat{\rho} = 0.6538525$.
Our next step in modelling the data is to determine the approximate distribution that the data may have come from. To find the distribution of our data, we need to first find the distribution of our error terms. Therefore, we solve for ϵ_k. Recall,

$$X_k = \rho X_{k-1} + \epsilon_k$$
$$\hat{\epsilon}_k = X_k - \hat{\rho} X_{k-1}$$

We now have all of the information necessary to generate the histogram. By estimating $\hat{\rho}$, we can see the correlation between each observation from time period to time period. The only thing that is different between the observations are the random error terms, ϵ_k. Hence, knowing the distribution of the $\epsilon_k's$ will give us the exact distribution of the observed data. To determine this, we create a histogram of $\hat{\epsilon}_1, \hat{\epsilon}_2 \ldots \hat{\epsilon_{500}}$. In order to obtain a detailed picture, we generate a histogram that has 40 equally spaced bins. The distribution is clearly Cauchy. The question now becomes estimating the parameters of the Cauchy to get a nice fit for the error terms. We find that a Cauchy with parameters,

location $= 1.02$, scale $= .0499$ provides a good fit. The histogram of the error terms, along with the density of a Cauchy can be found in Figure 4.1. To further strengthen the assumption that the $\epsilon'_k s$ came from a Cauchy distribution, we provide a χ^2 test.

4   Proof

H_0 : time series data has Cauchy Distribution with parameter: location=1.02, scale=.0499
H_a : H_0 is not true

We will divide the interval into K cells and calculate the number of observations that fall into each cell. We define this as $Y_1, Y_2, \ldots, Y_8$. We define the range of our 8 cells as

$$[-\infty, 0.96), [0.96, 0.97), [0.97, 0.98), [0.98, 1.02), [1.02, 1.5), [1.5, 1.8), [1.8, 2.8), [2.8, \infty)$$

Let $K = 8$ and $n = 500$.
Recall, the density of the Cauchy distribution:

$$f(x) = \frac{1}{\pi} \left(\frac{1}{1 + x^2} \right).$$

We define
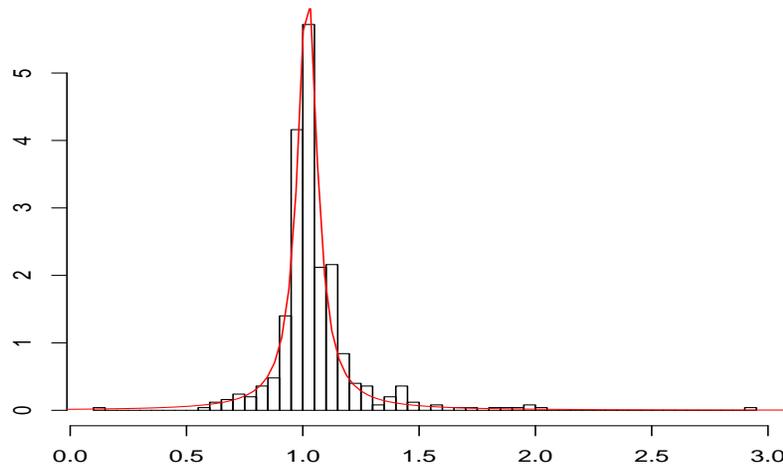
$$E_i = 500 \int_{t_{i-1}}^{t_i} f(x) dx.$$

Recall

$$\sum_{i=1}^{K} \frac{(Y_i - E_i)^2}{E_i} \sim \chi^2(K - 1)$$

We obtain that:

$$Y_1 = 90 \qquad\qquad\qquad Y_5 = 258$$
$$Y_2 = 10 \qquad\qquad\qquad Y_6 = 4$$
$$Y_3 = 20 \qquad\qquad\qquad Y_7 = 6$$
$$Y_4 = 111 \qquad\qquad\qquad Y_8 = 1$$

Using the definition of E_i gives:

$$E_1 = 110 \qquad\qquad\qquad E_5 = 234$$
$$E_2 = 14 \qquad\qquad\qquad E_6 = 6$$
$$E_3 = 18 \qquad\qquad\qquad E_7 = 6$$
$$E_4 = 107 \qquad\qquad\qquad E_8 = 4$$

FIGURE   4.1 Histogram of $\hat{\epsilon}_1, \hat{\epsilon}_2 \ldots \hat{\epsilon_{500}}$

Using the definition of χ^2 we get:

$$\frac{(90-110)^2}{110} + \frac{(10-14)^2}{14} + \frac{(20-18)^2}{18} + \frac{(111-107)^2}{107} +$$

$$\frac{(258-234)^2}{234} + \frac{(4-6)^2}{6} + \frac{(6-6)^2}{6} + \frac{(1-4)^2}{4} \sim \chi^2(7)$$

More succinctly, we have just computed

$$\sum_{i=1}^{K} \frac{(Y_i - E_i)^2}{E_i} = 10.51.$$

Referencing a standard χ^2 table shows that $\chi^2(7)$ can be as large as 14.07 with 95% probability. Hence, the null hypothesis that these values are Cauchy cannot be rejected (at a significance level of 5%). Furthermore, the corresponding p-value is .161466, which strengthens our conclusion that the null hypothesis cannot be rejected at any reasonable levels of significance.

5   Time Series

We generate $X_1, X_2, \ldots, X_n$ satisfying the equation,
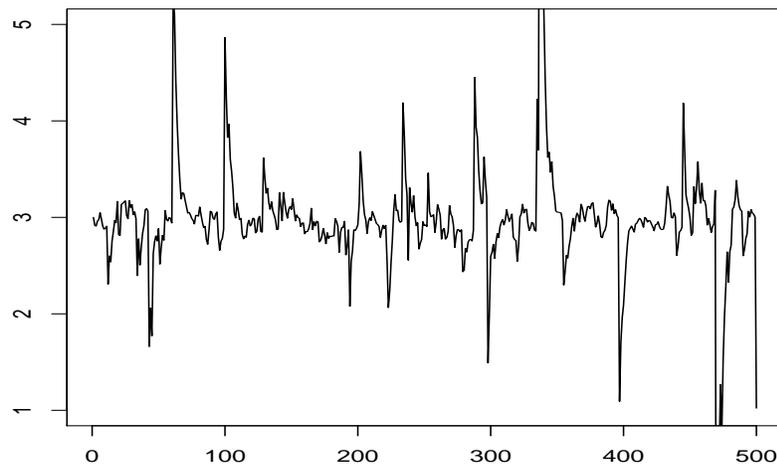
$$X_n = \rho X_{n-1} + \epsilon_n.$$

FIGURE 5.1 500 simulated values of $X_1, X_2, \ldots, X_n$

As previously discussed, the random error term, ϵ_n, was generated using a Cauchy distribution with parameters: location = 1.02, scale = .0499.

The plot of the 500 simulated values of $X_1, X_2, \ldots, X_n$ can be found in Figure 5.1.Essentially, Figure 5.1 is a simulation of the AR(1) time series for the value of $\rho = 0.6538525$. In each case the error terms are Cauchy random variables. The original 500 observed values from the data set can be found in Figure 5.2. To further compare Figures 5.1 and 5.2 we superimpose the two graphs. This result is shown on Figure 5.3.

6 Conclusion: Modeling High Volatility

We have capitalized upon the properties of the Cauchy distribution, namely that it does not have a defined mean or variance. These characteristics are essential for the modelling of volatile data. Having such a distribution has not only become convenient, but necessary in forecasting and monitoring data sets in economics, finance and other various fields. The properties of the Cauchy distribution satisfy the requirements from the literature, which guarantee a stationary solution to the time series. We have shown that correctly identifying the distribution of the error terms, allows for the correct modelling of our data. Due to our extemely high p-value, our conclusions concerning the hypothesis test for goodness-of-fit cannot be rejected at any reasonable level of significance.
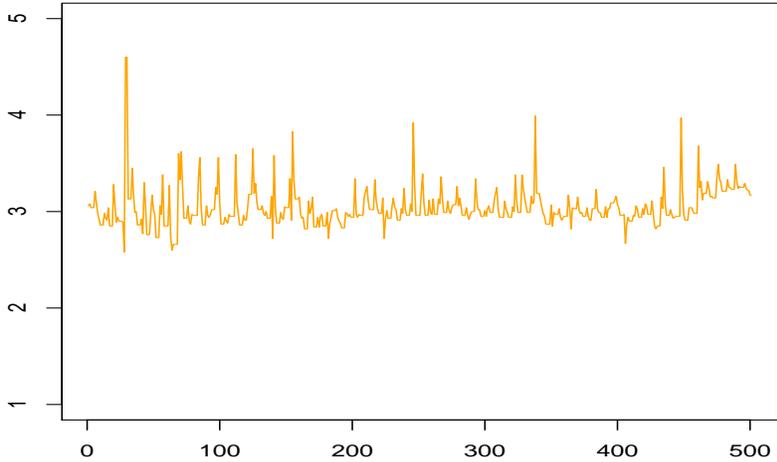
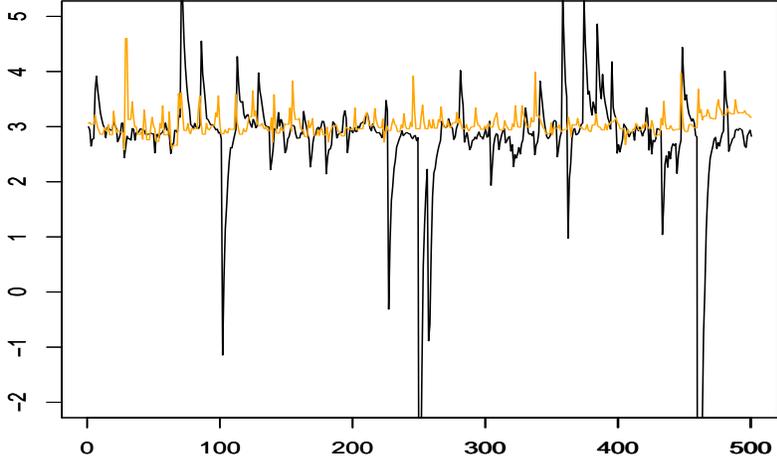FIGURE 5.2 500 observed values from original dataset

FIGURE 5.3 Simultaneous plot of observed and simulated values

7 Bibliography

- Brockwell, Peter J. and Richard A. Davis. *Time Series: Theory and Methods*. New York: Springer–Verlag, 1987.

- Chatfield, Christopher. *The Analysis of Time Series: An Introduction*. New York, NY: Chapman and Hall, 1984.

- Chung, Kai Lai. *A Course in Probability Theory*. New York: Academic Press, 1974.

- Fuller, Wayne A. *Introduction To Statistical Time Series*. New York: John Wiley & Sons Inc., 1976.

- Gunnip, Jon. *Analyzing Aggregated AR(1) Processes*. University of Utah: 2006.

- Hamilton, James D. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.

- Harris, Richard and Robert Sollis. *Applied Time Series Modelling and Forecasting*. Hoboken, NJ: John Wiley & Sons Inc., 2003.

- Horvath, Lajos and Remigijus Leipus. *Effect of Aggregation On Estimators in AR(1) Sequence*. Preprint, 2005.

- Williams, David. *Probability with Martingales*. New York, NY: Cambridge University Press, 1991.