

TS Module 16 ARIMA Forecasting

(The attached PDF file has better formatting.)

Time Series Practice Problems Random Walk

Many actuarial and financial series are random walks. Know how to distinguish a random walk from a risk-neutral probability, how to forecast with a random walk, and how to compute the standard error of a forecast from a random walk.

*Question 16.1: Random Walk

We observe 12 values of a random walk process as

100, 104, 101, 102, 105, 107, 109, 105, 108, 108, 113, 111

What is the estimated *drift* of this random walk process?

- A. -1
- B. $-\frac{1}{2}$
- C. 0
- D. $\frac{1}{2}$
- E. 1

Answer 16.1: E

The random walk is $y_t = \text{drift} + y_{t-1} + \epsilon$. To estimate the drift, we use linear regression. This scenario is simple enough for a rough estimate. We have 12 values, which give 11 changes (first differences). The average first difference is +1. The twelfth entry is 11 more than the first entry, which also gives an average first difference of +1.

The table below shows the values of the random walk and the first differences.

1	2	3	4	5	6	7	8	9	10	11	12	<i>Avg</i>
100	104	101	102	105	107	109	105	108	108	113	111	
	4	-3	1	3	2	2	-4	3	0	5	-2	1

*Question 16.2: Random Walk vs White Noise

We are deciding whether a time series of N values is a random walk with a drift of zero or white noise. We know it is one of these two, but we don't know which one. Let $N = 10,000$.

We compute the mean, residuals, and variance of the time series. The residuals are the observed values minus the mean of the values.

We form a correlogram, and we compute the sample autocorrelations for the first ten lags.

We take the first differences of the time series and we examine if it is stationary.

Which of the following is true? (*Close to means not significantly different from.*)

- A. If the mean of the process is close to zero, it is probably white noise; if the mean is significantly different from zero, it is probably a random walk.
- B. If the first differences are stationary, the time series is probably white noise; if the first differences are not stationary, the time series is probably a random walk.
- C. If the mean of the residuals is close to zero, it is probably white noise; if the mean of the residuals is significantly different from zero, it is probably a random walk.
- D. If the variance of the entries is close to zero, it is probably white noise; if the variance of the entries is close to one, it is probably a random walk.
- E. If the sample autocorrelations are close to zero, it is probably white noise; if the sample autocorrelations are high and do not decline rapidly to zero, it is probably a random walk.

Answer 16.2: E

Statement A: A white noise process may have any mean, not just zero; a random walk has no mean.

Jacob: If the current value of a random walk is Z and the drift is zero, the expected value at all future dates is Z . Doesn't this mean that the mean is Z ?

Rachel: We normally define the mean as the expected value. For a time series, we speak of a mean only if the expected value of an infinite time forecast does not depend on the present value or past values. Statisticians are not consistent in these definitions. The course textbook says that a time series is stationary if the mean is the same at all points.

Jacob: So you agree that the expected value at all future dates is Z ?

Rachel: The expected value at any future date T is Z . But as $T \rightarrow \infty$, the distribution of the future values becomes infinitely diffuse; that is, the variance becomes infinite, and the mean no longer exists.

Jacob: If the mean is Z for all T , no matter how large, the mean must also be Z when T is ∞ .

Rachel: Let us rephrase this concept in discrete terms, which are easier to grasp. What is the mean of the integers from $-T$ through T ?

Jacob: This distribution is symmetric about zero, so the mean is zero.

Rachel: As $T \rightarrow \infty$, this set becomes all the integers. What is the mean of all integers?

Jacob: This set is also symmetric about zero, so the mean must also be zero.

Rachel: If this set had a mean of zero, then if we add 1 to each element of the set, the new mean should be 1. But when we add one, we get the same set of all integers, whose mean we just said is zero. The mean can't be both zero and one; rather, it has no mean.

Jacob: But isn't the set of all integers symmetric about zero?

Rachel: It is also symmetric about 1, about -1 , and about every other integer. It is symmetric also about every half integer, such as $3\frac{1}{2}$.

Statement B: For a white noise process, the mean of the first differences is zero, but their absolute values may be large.

Jacob: For a white noise process, the standard deviation of the autocorrelations goes to zero as the series becomes large. If the mean is zero and the standard deviation is zero, how can the absolute value be large?

Rachel: The standard deviation of the sample autocorrelation goes to zero. The autocorrelation itself is a scalar and doesn't have a standard deviation.

We give a discrete example. Suppose the white noise process is a random draw from two values: +1,000 and -1,000. We evaluate the mean of the first differences, their average absolute value, and the sample autocorrelation of lag 1.

Suppose we have $N+1$ values, so we have N first differences. Only four scenarios exist:

- We start at -1,000 and end at -1,000; the mean first difference is zero.
- We start at +1,000 and end at +1,000; the mean first difference is zero.
- We start at -1,000 and end at +1,000; the mean first difference is $2,000 / N$, which goes to zero as N becomes large.
- We start at +1,000 and end at -1,000; the mean first difference is $-2,000 / N$, which goes to zero as N becomes large.

For the mean absolute value, we reason that

- If the current value is -1,000, the next value is either -1,000 or +1,000, so the mean absolute value of the first difference is $\frac{1}{2} \times (0 + 2,000) = 1,000$.
- If the current value is +1,000, the next value is either -1,000 or +1,000, so the mean absolute value of the first difference is $\frac{1}{2} \times (2,000 + 0) = 1,000$.

Jacob: Does that mean that the average absolute value is always 1,000?

Rachel: Suppose we have a series of 101 values.

- If the values are $\{-1,000, +1,000, -1,000, +1,000, \dots, -1,000, +1,000, -1,000\}$, the average absolute value of the first differences is 2,000.
- If the values are $\{-1,000, -1,000, -1,000, -1,000, \dots, -1,000, -1,000, -1,000\}$, the average absolute value of the first differences is zero.

The mean absolute value over all scenarios is 1,000, but the average absolute value in any one scenario may range from zero to 2,000.

Jacob: What is the sample autocorrelation of lag 1? Is it zero?

Rachel: The sample autocorrelation ranges from -1 to +1. Consider the two scenarios above:

- If the values are $\{-1,000, +1,000, -1,000, +1,000, \dots, -1,000, +1,000, -1,000\}$, the sample autocorrelation of lag 1 is -1.
- If the values are $\{-1,000, -1,000, -1,000, -1,000, \dots, -1,000, -1,000, -1,000\}$, the sample autocorrelation of lag 1 is not defined, since the variance of this time series is zero. But if the last value is +1,000 instead of -1,000, the sample autocorrelation is close to +1.

Jacob: The sample autocorrelation ranges from -1 to (almost) $+1$ for all values of T . Why do we say that the standard deviation of the sample autocorrelation goes to zero as $T \rightarrow \infty$?

Rachel: Suppose $T = 4$, so we have 16 possible scenarios for the time series. In two of these, or 25%, the sample autocorrelation is -1 . In four of these, the sample autocorrelation is positive and relatively large. The standard deviation of the sample autocorrelation is large.

Suppose $T = 1,000$, so we have $2^{1,000}$ possible scenarios. The sample autocorrelation is -1 in two of these, or about 0%. In almost all of the scenarios, the sample autocorrelation is close to zero, so the standard deviation is about zero.

Statement C: The mean of the residuals of any correctly specified time series should be zero.

Statement D: The sample autocorrelation is near zero for a white noise process and it is high for a random walk.

Jacob: We just said that the sample autocorrelation for a white noise process can range from -1 to $+1$; now you say it is zero.

Rachel: It is possible for white noise to have a sample autocorrelation from -1 to $+1$. As T becomes large, the probability of a sample autocorrelation different from zero becomes very small. If T is large and the sample autocorrelation is not close to zero, we can not say with certainty that the time series is not white noise, but the probability of its being white noise is close to zero. The reverse is true for a random walk. Even if T is large, it is possible for the sample autocorrelation of lag 1 to be very small, but this becomes unlikely as T becomes large.

For stochastic processes, we can make statistical statements. We can never know with certainty that the time series is a random walk or white noise.

Statement E: A white noise process has a constant variance, not zero variance.

Jacob: Didn't we just say that the white noise process has a standard deviation that approaches zero as T become large?

Rachel: The sample autocorrelation has a variance that approaches zero as T becomes large. The white noise process itself has a constant variance.

*Question 16.3: Random Walk

Which of the following is a random walk with a drift of -0.05 ?

A. $y_t = +0.05 - 0.05y_{t-1} + \varepsilon_t$

B. $y_t = -0.05 + 1.00y_{t-1} + \varepsilon_t$

C. $y_t = -0.05 - 1.00y_{t-1} + \varepsilon_t$

D. $y_t = -0.05 - 0.05y_{t-1} + \varepsilon_t$

E. $y_t = +1.00 - 0.05y_{t-1} + \varepsilon_t$

Answer 16.3: B

A random walk has first differences that are white noise plus a drift. This implies that the autoregressive coefficient is one and the constant term is the drift.

Jacob: I though a random walk is not an autoregressive process?

Rachel: A random walk is not a *stationary* autoregressive process. It is an ARIMA(1,1,0) process.

*Question 16.4: ARIMA Process

Suppose a time series is an autoregressive process of order 1:

$$Y_t = \alpha + \phi_1 Y_{t-1} + \epsilon_t.$$

with $\alpha = 1$ and $\phi_1 = 1$. This time series is

- A. A white noise process with a drift of zero
- B. A white noise process with a drift of one
- C. A random walk with a drift of zero
- D. A random walk with a drift of one
- E. A stationary process

Answer 16.4: D

Jacob: If the drift of the random walk is zero, does that mean that its mean is constant?

Rachel: On the contrary: if the drift is zero, the forecast for all future periods is the most recent value, which changes all the time. A random walk is not stationary and does not have a mean.

*Question 16.5: Random Walk with No Drift

Fertility rates follow a random walk with no drift: $y_t = y_{t-1} + \epsilon_t$, where ϵ_t has a normal distribution with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2 = 0.00005$.

- Fertility rates were 1.8% in 20X6 and 1.7% in 20X7.
- What is the forecasted fertility rate for 20X9 (two periods ahead, or \hat{y}_{T+2})?

- A. 1.80%
- B. 1.75%
- C. 1.70%
- D. 1.65%
- E. 1.60%

Answer 16.5: C

With a drift of zero, the most recent value is the best estimate for all future values.

*Question 16.6: Random Walk with No Drift

Fertility rates follow a random walk with no drift: $y_t = y_{t-1} + \epsilon_t$, where ϵ_t has a normal distribution with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2 = 0.00005$.

- Fertility rates were 1.8% in 20X6 and 1.7% in 20X7.
- What is the *standard error* of the forecast for 20X9 (two periods ahead, or \hat{y}_{T+2})?

(Note: The model and the σ^2 are **not** uncertain. The standard error is process risk only.)

- A. 0.0100
- B. 0.0141
- C. 0.100
- D. 0.141
- E. 0.200

Answer 16.6: A

- The variance for the one period ahead forecast is σ^2 .
- The variance for the two periods ahead forecast is $\sigma^2 + \sigma^2 = 2 \sigma^2$.

The standard error for the two periods ahead forecast is $\sqrt{2} \sigma = \sqrt{2} \times \sqrt{0.00005} = 0.010$

*Question 16.7: Random Walk with Drift

Fertility rates follow a random walk with drift: $y_t = y_{t-1} + d + \epsilon_t$, where $d = -0.0005$ (-0.05%) and ϵ_t follows a normal distribution with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2 = 0.00005$.

- Fertility rates were 1.8% in 20X6 and 1.7% in 20X7.
- What is the forecasted fertility rate for 20X9 (two periods ahead, or \hat{y}_{T+2})?

- A. 1.80%
- B. 1.750%
- C. 1.70%
- D. 1.65%
- E. 1.60%

Answer 16.7: E

We add $d \times t$ to the most recent value for the forecast in Period t .

*Question 16.8: Random Walk with Drift

Fertility rates follow a random walk with drift: $y_t = y_{t-1} + d + \epsilon_t$, where $d = -0.0005$ (-0.05%) and ϵ_t follows a normal distribution with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2 = 0.00005$.

- Fertility rates were 1.8% in 20X6 and 1.7% in 20X7.
- What is the *standard error* of the forecast for 20X9 (two periods ahead, or \hat{y}_{T+2})?

(Note: The model and the σ^2 are **not** uncertain. The standard error is process risk only.)

- A. 0.0100
- B. 0.0141
- C. 0.100
- D. 0.141
- E. 0.200

Answer 16.8: A

- The variance for the one period ahead forecast is σ^2 .
- The variance for the two periods ahead forecast is $\sigma^2 + \sigma^2 = 2 \sigma^2$.

The standard error for the two periods ahead forecast is $\sqrt{2} \sigma = \sqrt{2} \times \sqrt{0.00005} = 0.010$

The deterministic drift does not affect the standard error.