

TS Module 15 Forecasting basics

(The attached PDF file has better formatting.)

Time series practice problems: sum of squared errors

The optimal ARIMA model has the lowest mean squared error for its forecasts. The variance of the error terms is not known exactly, since the residuals depend on fitted values from the ARIMA process.

Illustration: For an AR(1) process, if we select μ and ϕ , we know the expected values for periods 2 and subsequent. For an MA(1) process, even if we select μ and θ , we don't know the residual in Period 1, so we don't know the expected value in Period 2, the residual in Period 2, and so forth. In practice, this is not a material problem. The uncertainty in the Period 1 residual does not have a material effect on the expected values several periods later.

We estimate the variance from the observed values and the assumed ARIMA parameters. The exercise below calculates the sum of squared errors. Focus on the following items:

- If the time series is stationary, no differences are needed.
- If the time series is a non-stationary ARIMA process, convert it to an ARMA process by taking first differences. In some cases, one might need second differences.

STARTING THE COMPUTATIONS

The exam problem will give values for all periods and μ , ϕ , and θ parameters.

- The estimate for the first period depends on the previous values (for an autoregressive process) or the previous residuals (for a moving average process).
- We can't estimate the residuals for the first period.
 - The problem may give this residual.
 - If the problem does not give this residual, assume it is zero.
- The exam problem will say to compute the sum of squared errors for Periods 2–N.

We can compute forecasts and residuals for Periods 2–N for an autoregressive process.

The Period 2 forecast for a moving average process depends on the residual in Period 1. We don't know the Period 1 residual, so we can not compute the forecast or residual for Period 2. Similarly, we can not compute the *exact* residuals for any future period.

After several periods, the forecast depends only slightly on the residual in Period 1. We assume that all values before Period 1 were the mean and all residuals before Period 1 were zero. This is the simplest way to start, though not the most accurate.

Using these assumptions, the residual for Period 1 is too large.

- We don't use the residual computed for Period 1 in the sum of squared errors, since it would over-state the result.
- The over-statement decreases for later residuals (Periods 2+), unless θ_1 is large.
 - If $\theta_1 = 1$, the error repeats in each period.
 - Practical moving average processes have low θ_1 .
- Know that the result is slightly over-stated. Advanced techniques exist for a better estimate of the early residuals, but they are not covered in the on-line course.

Know the logic of the estimation for an MA(1) process.

- We don't have an estimate for Period 1, so we don't have a residual.
- Without the residual for the first term, we can not estimate the second term.
- To simplify the estimate of σ^2 , we assume the first residual is zero.
- This assumption slightly over-states the residual for the second term.
- If the first residual is zero, the second residual does not depend on θ_1 .

In practice, we don't know θ_1 *a priori*. We select θ_1 to minimize the sum of squared errors.

- The exam problems compute the sum of squared errors for a given θ_1 .
- Your student project may use statistical software to minimize the squared errors.

Expect an exam problem giving you an ARMA or ARIMA process and asking for the sum of squared errors. The process may be autoregressive, moving average, or both.

Exercise 15.1: Sum of Squared Errors, ARIMA(0,1,1) Model

We use an ARIMA(0,1,1) model $y_t - y_{t-1} = \mu + \epsilon_t - \theta_1 \epsilon_{t-1}$.

t	y_t
0	7.50
1	8.00
2	12.00
3	10.30
4	2.00
5	7.50

- An ARMA process is $y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1}$.
- An ARIMA process is $y_t - y_{t-1} = \mu + \epsilon_t - \theta_1 \epsilon_{t-1}$.
- The first differences of the ARIMA model form an MA(1) process.
- The ARIMA time series has six observations: Periods 0 through 5.
- The ARMA model of first differences has five values.

t	y_t	$y_t - y_{t-1}$
0	7.50	
1	8.00	0.50
2	12.00	4.00
3	10.30	-1.70
4	2.00	-8.30
5	7.50	5.50
Total		0.00

- The MA(1) process underlying the ARIMA(0,1,1) time series has a mean $\mu = \delta$.
- The mean of *first differences* is the expected *change in the values of the time series*.

- What is the estimated mean of the MA(1) model of the first differences?
- What is the forecasted first difference for period 2?
- What is the forecasted value for Period 2?
- What is the residual in Period 2?
- What is the forecasted first difference for period 3?
- What is the forecasted value for period 3?
- What is the error term for period 3?
- What are the forecasts and residual for periods 4 and 5?
- What is the sum of squared errors for periods 2 through 5?

{The following paragraphs explain why we use the sum of squared errors starting with Period 2 instead of Period 1. The exam does not test optimal fitting of a moving average process.}

Terms: The sum of squared errors, error sum of squares, and residual sum of squares are synonyms. We use the acronym ESS, or error sum of squares, since it is not confusing. The acronym RSS stands for regression sum of squares (most authors) and regression sum of squares (Fox textbook).

We did not predict the Period 0 value with the ARIMA model, so we do not include it in the sum of squared errors. To predict Period 0, we need values for Period -1 . We may assume a residual of zero for this period to predict future periods, but we don't include this assumed residual of zero to measure the goodness-of-fit.

The first term for the underlying ARMA process is Period 1. To predict this value, we need the ARMA residual for Period 0, which we do not have. We assume the expected value for Period 1 is the observed value, so its residual is zero. This assumption slightly distorts the sum of squared errors for the subsequent terms, but the error is not material.

We minimize the sum of squared errors to select the optimal θ_1 . But the residual in Period 1 *does not depend on* θ_1 , since the previous residual is assumed to be zero. Some statisticians do not include Period 1 in the residual sum of squares, since it does not help determine θ_1 . Excluding Period 1 reduces the degrees of freedom.

The sum of squared errors starting at the second period, $S(\theta_1) = \sum_{t=2}^5 \hat{\varepsilon}_t^2$, depends on θ_1 .

To fit the ARIMA model, we choose θ_1 to minimize the sum of squared errors. We write the residual sum of squares as a function of θ_1 and find the minimum of the function.

Solving for θ_1 requires non-linear regression. Linear regression solves for the response variable (the forecasts) based on known explanatory variables (residuals). For a moving average process:

- The residuals depend on the estimates (fitted values) for each period.
- The fitted values depend on the residuals and θ_1 .
- We estimate θ_1 by minimizing the sum of squared errors.

Contrast this with fitting an AR(1) autoregressive process:

- The fitted values depend on the past values and ϕ_1 .
- The fitted values do not depend on past residuals.

This differences between autoregressive and moving average processes is less relevant for modern statistical software. We now use numerical techniques to optimize θ_1 .

- We evaluate the sum of squared errors at various values of θ_1 .
- We choose the value of θ_1 that minimizes the sum of squared errors.

Exam problems do not solve for θ_1 by numerical methods.

- They specify a value for θ_1 and solve for the sum of squared errors.
- Know how to compute the sum of squared errors and the degrees of freedom.

Part A: The time series is ARIMA(0,1,1), or IMA(1), so the *first differences* are an MA(1) model, for which $\mu = \delta$. The last term of the time series equals the first term. The sum of the first differences is zero, so the mean of the first differences, $\hat{\delta} = \mu = 0$.

Note: Cryer and Chan use the symbol θ_0 in place of δ .

Take heed: This practice problem has a drift of zero for the ARIMA process. In practice, we use ARIMA processes when the drift is not zero.

Part B: By assumption, the error term in Period 0 is zero. The *expected* error term in any period is also zero. The forecasted first difference for Period 1 is $0 - 0.8 \times 0 = 0$. In general, the forecasted first difference for Period 1 is the mean of the MA(1) model.

Part C: The forecasted value for Period 1 is the actual value in Period 0 plus the mean of the MA(1) model of first differences. The forecasted value for Period 1 is $7.5 + 0 = 7.5$.

Part D: The residual is the actual value minus the forecasted value. The residual in Period 1 is $8.0 - 7.5 = 0.5$.

Part E: This forecast depends on θ_1 . We evaluate the residual sum of squares at $\theta_1 = 0.8$. The first difference for period 1 has an error term of 0.5, so the forecasted first difference for period 2 is $-0.8 \times 0.5 = -0.4$.

Part F: The value for period 1 is 8. The forecasted first difference is -0.4 , so the forecasted value for period 2 is $8 - 0.4 = 7.6$.

Part G: The observed value in period 2 is 12, so the residual is $12 - 7.6 = 4.4$. If the value is 1 higher, the first difference is also 1 higher, so the residual for the first differences in period 2 is 4.4.

Part H: We compute the forecasts and residuals for each period in the same fashion, shown in the table below.

Part I: The right-most column of the table shows the squares of the residuals. The sum of squares is the 69.513 in the last row.

The table below shows all the values.

t	y_t	ϵ_t	\hat{y}_t	ϵ_t^2
0	7.5	–		
1	8.0	0.5		
2	12.0	4.4	7.600	19.360
3	10.3	1.82	8.480	3.312
4	2.0	-6.844	8.844	46.840
5	7.5	0.0248	7.475	0.001
Total		-0.5992		69.513

For *regression analysis*, the error sum of squares depends on the regression coefficients. We choose coefficients to minimize the error sum of squares. The explanatory variables for each data point are known, so the error sum of squares is clearly defined.

For *time series analysis*, the residual sum of squares depends on unknown prior values.

- For an $AR(p)$ model, we don't know the p values before the time series begins, so we don't know the residuals for the first p values.
- For an $MA(q)$ model, we don't know the q expected values before the time series begins, so we don't know the residuals for the first q values.

One might say that there are no actual values or expected values before the time series begins. This is not correct: the time series is a sample of values.

Illustration: We have a time series of average daily temperature for 1/1/20X7 – 12/31/20X7. If we use an $ARMA(1,1)$ process, the daily temperature on day t is a function of the daily temperatures on day $t-1$ and the expected daily temperature on day $t-1$ (which gives the residual for 12/31/20X7).

- Without the actual daily temperature on 12/31/20X6 and the expected daily temperature on 12/31/20X6, we can not derive the expected daily temperature on 1/1/20X7.
- Without the expected daily temperature on 1/1/20X7, we don't know the residual on 1/1/20X7, so we can not derive the expected daily temperature on 1/2/20X7.