
The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

CHARACTERIZING CYCLES

Topic 26

EXAM FOCUS

This topic focuses on ways to characterize a cycle in forecasting models. Along with seasonal and trend components, cycles constitute an essential third component in a forecasting model. Cyclical dynamics captures the dynamics of a data series *outside* of trend or seasonal data. Thus, the complexity of cyclical dynamics demands a more robust forecasting model. For the exam, understand the concept of covariance stationary and the requirements for a time series to exhibit covariance stationarity. Also, be able to define a white noise process and know how a lag operator works. The concepts introduced in this topic serve as a foundation for the material in the next topic on modeling cycles.

COVARIANCE STATIONARY

LO 26.1: Define covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function, and autoregression.

To forecast a time series, one needs to understand and characterize its structure. The following terminology relates to modeling data interrelationships and stability over time.

- **Autoregression** refers to the process of regressing a variable on lagged or past values of itself. As you will see in the next topic, when the dependent variable for a time series is regressed against one or more lagged values of itself, the resultant model is called as an **autoregressive (AR) model**. For example, the sales for a firm could be regressed against the sales for the firm in the previous month. Thus, in an autoregressive time series, past values of a variable are used to predict the current (and hence future) value of the variable.
- A time series is **covariance stationary** if its mean, variance, and covariances with lagged and leading values do not change over time. Covariance stationarity is a requirement for using AR models.
- **Autocovariance function** refers to the tool used to quantify stability of the covariance structure. Its importance lies in its ability to summarize cyclical dynamics in a series that is covariance stationary.
- **Autocorrelation function** refers to the degree of correlation and interdependency between data points in a time series. It recognizes the fact that correlations lend themselves to clearer interpretation than covariances. Recall that the degree of correlation is measured on a continuum from -1 to 1 , whereas degrees of covariance employ a much wider range, which can be unwieldy in determining levels of association.
- **Partial autocorrelation function** refers to the partial correlation and interdependency between data in a time series that measures the association between data in a series after controlling for the effects of lagged observations.

LO 26.2: Describe the requirements for a series to be covariance stationary.

A time series is covariance stationary if it satisfies the following three conditions:

1. *Constant and finite expected value.* The expected value of the time series is constant over time.
 2. *Constant and finite variance.* The time series volatility around its mean (i.e., the distribution of the individual observations around the mean) does not change over time.
 3. *Constant and finite covariance between values at any given lag.* The covariance of the time series with leading or lagged values of itself is constant.
-

LO 26.3: Explain the implications of working with models that are not covariance stationary.

Requirements for covariance stationarity of a time series, though strict in appearance, make allowances for many series that are not covariance stationary. This is achieved by working with models that provide special treatment to trend and seasonality components that are stationary, which allows the remaining, or residual, cyclical component to be covariance stationary.

Note that forecasting models whose “probabilistic nature” changes (i.e., lacks covariance stationarity) would not lend themselves well to predicting the future. Such a trait would make the process of characterizing a cycle difficult, if not impossible. However, a nonstationary series can be transformed to appear covariance stationary by using transformed data, such as growth rates.

WHITE NOISE

LO 26.4: Define white noise, and describe independent white noise and normal (Gaussian) white noise.

LO 26.5: Explain the characteristics of the dynamic structure of white noise.

A time series process with a zero mean, constant variance, and no serial correlation is referred to as a **white noise process** (or *zero-mean white noise*). This is the simplest type of time series process and it is used as a fundamental building block for more complex time series processes. Even though a white noise process is serially uncorrelated, it may not be serially independent or normally distributed.

Variants of a white noise process include independent white noise and normal white noise. A time series process that exhibits both serial independence and a lack of serial correlation is referred to as **independent white noise** (or *strong white noise*). A time series process that exhibits serial independence, is serially uncorrelated, and is normally distributed is referred to as **normal white noise** (or *Gaussian white noise*).

The dynamic structure of a white noise process includes the following characteristics:

- The unconditional mean and variance must be constant for any covariance stationary process.
- The lack of any correlation in white noise means that all autocovariances and autocorrelations are zero beyond displacement zero (displacement refers to the distance of a moving body from a central point). This same result holds for the partial autocorrelation function of white noise.
- Both conditional and unconditional means and variances are the same for an independent white noise process (i.e., they lack any forecastable dynamics).
- Events in a white noise process exhibit no correlation between the past and present.

LAG OPERATORS

LO 26.6: Explain how a lag operator works.

A **lag operator** quantifies how a time series evolves by lagging a data series. It enables a model to express how past data links to the present and how present data links to the future. For example, a lag operator, L , operates on series, y_t , by lagging it as follows:

$$Ly_t = y_{t-1}$$

Another example of a common lag operator is a *first-difference operator* (Δ), which applies a polynomial in the lag operator as follows:

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}$$

A key component of an operator is the **distributed lag**, which is a weighted sum of present and past values in a data series, achieved by lagging present values upon past values.

WOLD'S REPRESENTATION THEOREM

LO 26.7: Describe Wold's theorem.

LO 26.8: Define a general linear process.

LO 26.9: Relate rational distributed lags to Wold's theorem.

Wold's representation theorem is a model for the covariance stationary residual (i.e., a model that is constructed after making provisions for trends and seasonal components). Thus, the theorem enables the selection of the correct model to evaluate the evolution of covariance stationarity. Wold's representation utilizes an infinite number of distributed lags, where the one-step-ahead forecasted error terms are known as *innovations*.

The **general linear process** is a component in the creation of forecasting models in a covariance stationary time series. It uses Wold's representation to express innovations that

capture an evolving information set. These evolving information sets move the conditional mean over time (recall that a requirement of stationarity is a constant unconditional mean). Thus, it can model the dynamics of a times series process that is outside of covariance stationarity (i.e., unstable).

As mentioned, applying Wold's representation requires an infinite number of distributed lags. However, it is not practical to model an infinite number of parameters. Therefore, we need to restate this lag model as infinite polynomials in the lag operator because infinite polynomials do not necessarily contain an infinite number of parameters. Infinite polynomials that are a ratio of finite-order polynomials are known as **rational polynomials**. The distributed lags constructed from these rational polynomials are known as **rational distributed lags**. With these lags, we can approximate Wold's representation. In the next topic, we'll examine the properties of an autoregressive moving average (ARMA) process, which is a practical approximation for Wold's representation.

ESTIMATING THE MEAN AND AUTOCORRELATION FUNCTIONS

LO 26.10: Calculate the sample mean and sample autocorrelation, and describe the Box-Pierce Q-statistic and the Ljung-Box Q-statistic.

LO 26.11: Describe sample partial autocorrelation.

Sample data for a time series forms the basis for estimating the sample mean and sample autocorrelation of a covariance stationary series. With these estimated parameters, an analyst can study the dynamics that underpin the dataset and find a model that best fits the data. Sample data can be used to estimate the sample mean and the sample autocorrelation.

The **sample mean** is an approximation of the mean of the population and can be used to estimate the autocorrelation function. The sample mean, given a sample size of T , is computed as follows:

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

The **sample autocorrelation** estimates the degree to which white noise characterizes a series of data. Recall that for a time series to be classified as a white noise process, all autocorrelations must be zero in the population dataset. The sample autocorrelation, as a function of displacement τ , is computed as follows:

$$\hat{\rho}(\tau) = \frac{\sum_{t=\tau+1}^T [(y_t - \bar{y})(y_{t-\tau} - \bar{y})]}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Similar to sample autocorrelation, the **sample partial autocorrelation** can also be used to determine whether a time series exhibits white noise. It differs from sample autocorrelation in that it performs linear regression on a finite or feasible data series. However, the outcome of sample partial autocorrelation is typically identical to that achieved through sample

Topic 26**Cross Reference to GARP Assigned Reading – Diebold, Chapter 7**

autocorrelation. Sample partial autocorrelations usually plot within two-standard-error bands (i.e., 95% confidence interval) when the time series is white noise.

A **Q-statistic** can be used to measure the degree to which autocorrelations vary from zero and whether white noise is present in a dataset. This can be done by evaluating the overall statistical significance of the autocorrelations. This statistical measure is approximately chi-squared distributed with m degrees of freedom in large samples under the null hypothesis of no autocorrelations.

The **Box-Pierce Q-statistic** reflects the absolute magnitudes of the correlations, because it sums the squared autocorrelations. Thus, the signs do not cancel each other out, and large positive or negative autocorrelation coefficients will result in large Q-statistics. The **Ljung-Box Q-statistic** is similar to the Box-Pierce Q-statistic except that it replaces the sum of squared autocorrelations with a weighted sum of squared autocorrelations. For large sample sizes, weights for both statistics are roughly equal.

KEY CONCEPTS

LO 26.1

The terms covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function, and autoregression relate to the degree of data interrelationships and their stability. A time series is covariance stationary if its mean, variance, and covariances with lagged and leading values do not change over time.

LO 26.2

A time series is covariance stationary if it satisfies the following three conditions:
(1) constant and finite expected value, (2) constant and finite variance, and (3) constant and finite covariance between values at any given lag.

LO 26.3

Models that lack covariance stationarity are unstable and do not lend themselves to meaningful forecasting.

LO 26.4

A time series process with a zero mean, constant variance, and no serial correlation is referred to as white noise. This is the simplest type of time series process and is used as a building block for more complex time series processes.

LO 26.5

The lack of any correlation in a white noise process means that all autocovariances and autocorrelations are zero beyond displacement zero. The past is not correlated with the present which, in turn, is not correlated with the future.

LO 26.6

A lag operator enables a forecasting model to express how past data links to the present and how present data links to the future.

LO 26.7

Wold's representation theorem evaluates covariance stationarity as a prerequisite for time series modeling. It utilizes an infinite number of distributed lags.

LO 26.8

The general linear process is intended to capture an information set that evolves.

LO 26.9

The distributed lags constructed from rational polynomials are known as rational distributed lags. With these lags, Wold's representation can be approximated.

LO 26.10

Understanding the degree of data correlation and dynamics that underpin the dataset is critical to the characterization of a cycle. If white noise is present, then there should be no forecastable events. Q-statistics further refine the measurement of the degree to which autocorrelations vary from zero and whether white noise is present in the dataset.

LO 26.11

Sample partial autocorrelation is a somewhat simplified version of sample autocorrelation in that it uses a finite data series.

CONCEPT CHECKERS

1. All of the following traits characterize the covariance stationarity of a time series process, except:
 - A. stability of the mean.
 - B. stability of the covariance structure.
 - C. a nonconstant variance in the time series.
 - D. stability of the autocorrelation.

2. Which of the following features correctly characterizes a white noise process?
 - A. Conditional mean in the dataset.
 - B. Minimal variance.
 - C. No correlation between data points.
 - D. Partial autocorrelations are greater than zero.

3. Which of the following statements is most likely correct regarding lag operators? Lag operators:
 - A. consider only infinite-order polynomials.
 - B. quantify how a time series evolves by lagging a data series.
 - C. are of limited use in modeling a time series.
 - D. only use lagged future values.

4. Regarding Q-statistics, the Box-Pierce and Ljung-Box Q-statistics:
 - A. produce different results.
 - B. are more accurate for smaller datasets.
 - C. essentially yield the same result.
 - D. both use an unweighted sum of squared autocorrelations.

5. Regarding sample partial autocorrelations, which of the following statements is true?
A sample partial autocorrelation:
 - A. is identical to sample autocorrelation.
 - B. differs from sample autocorrelation in the size of the dataset to which it applies.
 - C. utilizes non-linear regressions.
 - D. typically falls within a one-standard-error band.

CONCEPT CHECKER ANSWERS

1. C The time series volatility around its mean (i.e., the distribution of the individual observations around the mean) does not change over time.
2. C The lack of any correlation in white noise means that all autocovariances and autocorrelations are zero.
3. B Lag operators may use finite-order polynomials and are an essential tool to model a time series. They quantify how a time series evolves by typically lagging present values upon past values.
4. C Both Q-statistics typically arrive at the same result. The Ljung-Box statistic works better with smaller samples of data and replaces the sum of squared autocorrelations in the Box-Pierce statistic with a weighted sum of squared autocorrelations.
5. B The linear regression that is part of the sample partial autocorrelation process takes place on a feasible data sample, which differs from the infinite data sample for partial autocorrelations. Sample partial autocorrelations should fall within two-standard-error (standard deviation) bands.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

MODELING CYCLES: MA, AR, AND ARMA MODELS

Topic 27

EXAM FOCUS

Moving average (MA) processes can be used to capture the relationship between a time series variable and its current and lagged random shocks. This is useful for researchers if an event is mostly described by random shocks. However, it becomes even more useful when it is transformed into an autoregressive representation. An autoregressive (AR) process attempts to capture how a time series variable's lagged observations of itself combine with random shocks to forecast a variable. Sometimes forecasters need a combination of these two concepts to improve the usefulness of a forecasting model, which results in an autoregressive moving average model (ARMA). For the exam, understand the properties of an MA(1) process and an AR(1) process and how they can be broadened to incorporate additional lag operators. Also, be able to describe an ARMA process and understand its applications.

FIRST-ORDER MOVING AVERAGE PROCESS

LO 27.1: Describe the properties of the first-order moving average (MA(1)) process, and distinguish between autoregressive representation and moving average representation.

Conceptually, a moving average process is a linear regression of the current values of a time series against both the current and previous unobserved white noise error terms, which are random shocks. The first-order moving average [MA(1)] process has a mean of zero and a constant variance and can be defined as:

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}$$

where:

y_t = the time series variable being estimated

ε_t = current random white noise shock

ε_{t-1} = one-period lagged random white noise shock

θ = coefficient for the lagged random shock

The MA(1) process is considered to be first-order because it only has one lagged error term (ε_{t-1}). This yields a very short-term memory because it only incorporates what happened one period ago. If we ignore the lagged error term for a moment and assume that $\varepsilon_t > 0$, then $y_t > 0$. This is equivalent to saying that a positive error term will yield a positive dependent variable (y_t). When adding back the lagged error term, we are now saying that the dependent variable is impacted by not only the current error term, but also the previous

Topic 27**Cross Reference to GARP Assigned Reading – Diebold, Chapter 8**

period's unobserved error term, which is amplified by a coefficient (θ). Consider an example using daily demand for ice cream (y_t) to better understand how this works:

$$y_t = \varepsilon_t + 0.3\varepsilon_{t-1}$$

In this equation, the error term (ε_t) is the daily change in temperature. Using only the current period's error term (ε_t), if the daily change in temperature is positive, then we would estimate that daily demand for ice cream would also be positive. But, if the daily change yesterday (ε_{t-1}) was also positive, then we would expect an amplified impact on our daily demand for ice cream by a factor of 0.3.

One key feature of moving average processes is called the *autocorrelation (ρ) cutoff*. We would compute the autocorrelation using the following formula:

$$\rho_1 = \frac{\theta_1}{1 + \theta_1^2}; \text{ where } \rho_\tau = 0 \text{ for } \tau > 1$$

Using the previous example of estimating daily demand for ice cream with $\theta = 0.3$, we would compute the autocorrelation to be 0.2752 as follows:

$$0.2752 = \frac{0.3}{1 + 0.3^2}$$

For any value beyond the first lagged error term, the autocorrelation will be zero in an MA(1) process. This is important because it is one condition of being covariance stationary (i.e., mean = 0, variance = σ^2), which is a condition of this process being a useful estimator.

It is also important to note that this **moving average representation** has both a current random shock (ε_t) and a lagged unobservable shock (ε_{t-1}) on the independent side of this equation. This presents a problem for forecasting in the real world because it does not incorporate observable shocks. The solution for this problem is known as an **autoregressive representation** where the MA(1) process formula is inverted so we have a lagged shock and a lagged value of the time series itself. The condition for inverting an MA(1) process is $|\theta| < 1$. The autoregressive representation, which is an algebraic rearrangement of the MA(1) process formula, is expressed in the following formula:

$$\varepsilon_t = y_t - \theta\varepsilon_{t-1}$$

This process of inversion enables the forecaster to express current observables in terms of past observables.

MA(q) PROCESS

LO 27.2: Describe the properties of a general finite-order process of order q (MA(q)) process.

The MA(1) process is a subset of a much larger picture. Forecasters can broaden their horizon to a finite-order moving average process of order q , which essentially adds lag operators out to the q^{th} observation and potentially improves on the MA(1) process. The MA(q) process is expressed in the following formula:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

where:

- y_t = the time series variable being estimated
- ε_t = current random white noise shock
- ε_{t-1} = one-period lagged random white noise shock
- ε_{t-q} = q^{th} -period lagged random white noise shock
- θ = coefficients for the lagged random shocks

The MA(q) process theoretically captures complex patterns in greater detail, which can potentially provide for more robust forecasting. This also lengthens the memory from one period to the q^{th} period. Returning to the previous example, using the demand for ice cream, a forecaster could use not only the current and previous day's changes in temperature to predict ice cream demand, but also the entire previous week's demand to enhance the informational value of the estimation.

Just as the MA(1) process exhibits autocorrelation cutoff after the first lagged error term, the MA(q) process experiences autocorrelation cutoff after the q^{th} lagged error term. Again, this is important because covariance stationarity is essential to the predictive ability of the model.

FIRST-ORDER AUTOREGRESSIVE PROCESS

LO 27.3: Describe the properties of the first-order autoregressive (AR(1)) process, and define and explain the Yule-Walker equation.

We have seen that when a moving average process is inverted it becomes an autoregressive representation, and is, therefore, more useful because it expresses the current observables in terms of past observables. An autoregressive process does not need to be inverted because it is already in the more favorable rearrangement, and is, therefore, capable of capturing a more robust relationship compared to the unadjusted moving average process. The first-order autoregressive [AR(1)] process must also have a mean of zero and a constant variance.

Topic 27**Cross Reference to GARP Assigned Reading – Diebold, Chapter 8**

It is specified in the form of a variable regressed against itself in a lagged form. This relationship can be shown in the following formula:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

where:

y_t = the time series variable being estimated

y_{t-1} = one-period lagged observation of the variable being estimated

ε_t = current random white noise shock

ϕ = coefficient for the lagged observation of the variable being estimated

Just like the moving average process, the predictive ability of this model hinges on it being covariance stationary. In order for an AR(1) process to be covariance stationary, the absolute value of the coefficient on the lagged operator must be less than one (i.e., $|\phi| < 1$).

Using our previous example of daily demand for ice cream, we would forecast our current period daily demand (y_t) as a function of a coefficient (ϕ) multiplied by our lagged daily demand for ice cream (y_{t-1}) and then add a random error shock (ε_t). This process enables us to use a past observed variable to predict a current observed variable.

In order to estimate the autoregressive parameters, such as the coefficient (ϕ), forecasters need to accurately estimate the autocovariance of the data series. The **Yule-Walker equation** is used for this purpose. When using the Yule-Walker concept to solve for the autocorrelations of an AR(1) process, we use the following relationship:

$$\rho_t = \phi^t \text{ for } t = 0, 1, 2, \dots$$

The Yule-Walker equation is used to reinforce a very important distinction between autoregressive processes and moving average processes. Recall that moving average processes exhibit autocorrelation cutoff, which means the autocorrelations are essentially zero beyond the order of the process [an MA(1) process shows autocorrelation cutoff after time 1]. The significance of the Yule-Walker equation is that for autoregressive processes, the autocorrelation decays very gradually. Consider an AR(1) process that is specified using the following formula:

$$y_t = 0.65y_{t-1} + \varepsilon_t$$

The coefficient (ϕ) is equal to 0.65, and using the concept derived from the Yule-Walker equation, the first-period autocorrelation is 0.65 (i.e., 0.65^1), the second-period autocorrelation is 0.4225 (i.e., 0.65^2), and so on for the remaining autocorrelations.



Professor's Note: While autocorrelation cutoff is a hallmark of moving average processes, a gradual decay in autocorrelations is a sure sign that a forecaster is dealing with an autoregressive process.

It should also be noted that if the coefficient (ϕ) were to be a negative number, perhaps -0.65 , then the decay would still occur but the graph would oscillate between negative and positive numbers. This is true because $-0.65^3 = -0.2746$, $-0.65^4 = 0.1785$, and $-0.65^5 = -0.1160$. You would still notice the absolute value decaying, but the actual autocorrelations would alternate between positive and negative numbers over time.

AR(p) PROCESS

LO 27.4: Describe the properties of a general p^{th} order autoregressive (AR(p)) process.

Just as the MA(1) process was described as a subset of the much broader MA(q) process, so is the relationship between the AR(1) process and the AR(p) process. The AR(p) process expands the AR(1) process out to the p^{th} observation as seen in the following formula:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

where:

y_t = the time series variable being estimated

y_{t-1} = one-period lagged observation of the variable being estimated

y_{t-p} = p^{th} -period lagged observation of the variable being estimated

ε_t = current random white noise shock

ϕ = coefficients for the lagged observations of the variable being estimated

The AR(p) process is also covariance stationary if $|\phi| < 1$ and it exhibits the same decay in autocorrelations that was found in the AR(1) process. However, while an AR(1) process only evidences oscillation in its autocorrelations (switching from positive to negative) when the coefficient is negative, an AR(p) process will naturally oscillate as it has multiple coefficients interacting with each other.

AUTOREGRESSIVE MOVING AVERAGE PROCESS

LO 27.5: Define and describe the properties of the autoregressive moving average (ARMA) process.

So far, we have examined moving average processes and autoregressive processes assuming they interact independently of each other. While this may be the case, it is possible for a time series to show signs of both processes and theoretically capture a still richer relationship. For example, stock prices might show evidence of being influenced by both unobserved shocks (the moving average component) and their own lagged behavior (the

Topic 27**Cross Reference to GARP Assigned Reading – Diebold, Chapter 8**

autoregressive component). This more complex relationship is called an **autoregressive moving average (ARMA) process** and is expressed by the following formula:

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

where:

y_t = the time series variable being estimated
 ϕ = coefficient for the lagged observations of the variable being estimated

y_{t-1} = one-period lagged observation of the variable being estimated

ε_t = current random white noise shock

θ = coefficient for the lagged random shocks

ε_{t-1} = one-period lagged random white noise shock

You can see that the ARMA formula merges the concepts of an AR process and an MA process. In order for the ARMA process to be covariance stationary, which is important for forecasting, we must still observe $|\theta| < 1$. Just as with the AR process, the autocorrelations in an ARMA process will also decay gradually for essentially the same reasons.

Consider an example regarding sales of an item (y_t) and a random shock of advertising (ε_t). We could attempt to forecast sales for this item as a function of the previous period's sales (y_{t-1}), the current level of advertising (ε_t), and the one-period lagged level of advertising (ε_{t-1}). It makes intuitive sense that sales in the current period could be affected by both past sales and by random shocks, such as advertising. Another possible random shock for sales could be the seasonal effects of weather conditions.



Professor's Note: Just as moving average models can be extrapolated to the q^{th} observation and autoregressive models can be taken out to the p^{th} observation, ARMA models can be used in the format of an ARMA(p,q) model. For example, an ARMA(3,1) model means 3 lagged operators in the AR portion of the formula and 1 lagged operator on the MA portion. This flexibility provides the highest possible set of combinations for time series forecasting of the three models discussed in this topic.

APPLICATION OF AR AND ARMA PROCESSES

LO 27.6: Describe the application of AR and ARMA processes.

A forecaster might begin by plotting the autocorrelations for a data series and find that the autocorrelations decay gradually rather than cut off abruptly. In this case, the forecaster should rule out using a moving average process. If the autocorrelations instead decay gradually, he should consider specifying either an autoregressive (AR) process or an autoregressive moving average (ARMA) process. The forecaster should especially consider these alternatives if he notices periodic spikes in the autocorrelations as they are gradually decaying. For example, if every 12th autocorrelation jumps upward, this observation indicates a possible seasonality effect in the data and would heavily point toward using either an AR or ARMA model.

Another way of looking at model applications is to test various models using regression results. It is easiest to see the differences using data that follows some pattern of seasonality, such as employment data. In the real world, a moving average process would not specify a very robust model, and autocorrelations would decay gradually, so forecasters would be wise to consider both AR models and ARMA models for employment data.

We could begin with a base AR(2) model that adds in a constant value (μ) if all other values are zero. This is shown in the following generic formula:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Applying actual coefficients, our real AR(2) model might look something like:

$$y_t = 101.2413 + 1.4388y_{t-1} - 0.4765y_{t-2} + \varepsilon_t$$

We could also try to forecast our seasonally impacted employment data with an ARMA(3,1) model, which might look like the following formula:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \theta \varepsilon_{t-1} + \varepsilon_t$$

Applying actual coefficients our real ARMA(3,1) model might look something like:

$$y_t = 101.1378 + 0.5004y_{t-1} + 0.8722y_{t-2} - 0.4434y_{t-3} + 0.9709\varepsilon_{t-1} + \varepsilon_t$$

In practice, researchers would attempt to determine whether the AR(2) model or the ARMA(3,1) model provides a better prediction for the seasonally impacted data series.

KEY CONCEPTS

LO 27.1

The first-order moving average process enables forecasters to consider the likely current effect on a dependent variable of current and lagged white noise error terms. While this is a useful process, it is most useful when inverted as an autoregressive representation so that current observables can be explained in terms of past observables.

LO 27.2

While the first-order moving average process does provide useful information for forecasting, the q th-order moving average process allows for a richer analysis because it incorporates significantly more lagged error terms all the way out to the order of q .

LO 27.3

The first-order autoregressive process incorporates the benefits of an inverted MA(1) process. Specifically, the AR(1) process seeks to explain the dependent variable in terms of a lagged observation of itself and an error term. This is a better forecasting tool if the autocorrelations decay gradually rather than cut off immediately after the first observation with a first-order process.

LO 27.4

The p th-order autoregressive process adds additional lagged observations of the dependent variable and enhances the informational value relative to an AR(1) process in much the same way that an MA(q) process adds a richer explanation to the MA(1) process.

LO 27.5

The autoregressive moving average (ARMA) process has the potential to capture more robust relationships. The ARMA process incorporates the lagged error elements of the moving average process and the lagged observations of the dependent variable from the autoregressive process.

LO 27.6

Both autoregressive (AR) and autoregressive moving average (ARMA) processes can be applied to time series data that show signs of seasonality. Seasonality is most apparent when the autocorrelations for a data series do not abruptly cut off, but rather decay gradually with periodic spikes.

CONCEPT CHECKERS

1. In practice, the moving average representation of a first-order moving average [MA(1)] process presents a problem. Which of the following statements represents that problem and how can it be resolved? The problem is that a moving average representation of an MA(1) process:
 - A. does not incorporate observable shocks, so the solution is to use a moving average representation.
 - B. incorporates only observable shocks, so the solution is to use a moving average representation.
 - C. does not incorporate observable shocks, so the solution is to use an autoregressive representation.
 - D. incorporates only observable shocks, so the solution is to use an autoregressive representation.

2. Which of the following statements is a key differentiator between a moving average (MA) representation and an autoregressive (AR) process?
 - A. A moving average representation shows evidence of autocorrelation cutoff.
 - B. An autoregressive process shows evidence of autocorrelation cutoff.
 - C. An unadjusted moving average process shows evidence of gradual autocorrelation decay.
 - D. An autoregressive process is never covariance stationary.

3. The purpose of a q^{th} -order moving average process is to:
 - A. add exactly two additional lagged variables to the original specification.
 - B. add a second error term to an MA(1) process.
 - C. invert the moving average process to make the formula more useful.
 - D. add as many additional lagged variables as needed to more robustly estimate the data series.

4. Which of the following statements about an autoregressive moving average (ARMA) process is correct?
 - I. It involves autocorrelations that decay gradually.
 - II. It combines the lagged unobservable random shock of the MA process with the observed lagged time series of the AR process.
 - A. I only.
 - B. II only.
 - C. Both I and II.
 - D. Neither I nor II.

5. Which of the following statements is correct regarding the usefulness of an autoregressive (AR) process and an autoregressive moving average (ARMA) process when modeling seasonal data?
 - I. They both include lagged terms and, therefore, can better capture a relationship in motion.
 - II. They both specialize in capturing only the random movements in time series data.
 - A. I only
 - B. II only.
 - C. Both I and II.
 - D. Neither I nor II.

CONCEPT CHECKER ANSWERS

1. C The problem with a moving average representation of an MA(1) process is that it attempts to estimate a variable in terms of unobservable white noise random shocks. If the formula is inverted into an autoregressive representation, then it becomes more useful for estimation because an observable item is now being used.
2. A A key difference between a moving average (MA) representation and an autoregressive (AR) process is that the MA process shows autocorrelation cutoff while an AR process shows a gradual decay in autocorrelations.
3. D The whole point of using more independent variables in a q^{th} -order moving average process is to capture a better estimation of the dependent variable. More lagged operators often provide a more robust estimation.
4. C The autoregressive moving average (ARMA) process is important because its autocorrelations decay gradually and because it captures a more robust picture of a variable being estimated by including both lagged random shocks and lagged observations of the variable being estimated. The ARMA model merges the lagged random shocks from the MA process and the lagged time series variables from the AR process.
5. A Both autoregressive (AR) models and autoregressive moving average (ARMA) models are good at forecasting with seasonal patterns because they both involve lagged observable variables, which are best for capturing a relationship in motion. It is the moving average representation that is best at capturing only random movements.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

VOLATILITY

Topic 28

EXAM FOCUS

Traditionally, volatility has been synonymous with risk. Thus, the accurate estimation of volatility is crucial to understanding potential risk exposure. This topic pertains to methods that employ historical data when generating estimates of volatility. Simplistic models tend to generate estimates assuming volatility remains constant over short time periods. Conversely, complex models account for variations over time. For the exam, be able to estimate volatility using both the exponentially weighted moving average (EWMA) and the generalized autoregressive conditional heteroskedasticity [GARCH(1,1)] models.

VOLATILITY, VARIANCE, AND IMPLIED VOLATILITY

LO 28.1: Define and distinguish between volatility, variance rate, and implied volatility.

The volatility of a variable, σ , is represented as the standard deviation of that variable's continuously compounded return. With option pricing, volatility is typically expressed as the standard deviation of return over a one-year period. This differs from risk management, where volatility is typically expressed as the standard deviation of return over a one-day period.

The traditional measure of volatility first requires a measure of change in asset value from period to period. The calculation of a continuously compounded return over successive days is as follows:

$$u_i = \ln\left(\frac{S_i}{S_{i-1}}\right)$$

where:

S_i = asset price at time i

This is similar to the proportional change in an asset, which is calculated as follows:

$$u_i = \frac{S_i - S_{i-1}}{S_{i-1}}$$

From a risk management perspective, the daily volatility of an asset usually refers to the standard deviation of the daily proportional change in asset value.

By assuming daily returns are independent with the same level of variation, daily volatility can be extended over a number of days, T , by multiplying the standard deviation of the return by the square root of T . This is known as the *square root of time rule*. For example, if the daily volatility is 1.5%, the standard deviation of the return (compounded continuously) over a 10-day period would be computed as $1.5\% \times \sqrt{10} = 4.74\%$. Note that when converting daily volatility to annual volatility, the usual practice is to use the square root of 252 days, which is the number of business days in a year, as opposed to the number of calendar days in a year.

Risk managers may also compute a variable's **variance rate**, which is simply the square of volatility (i.e., standard deviation squared: σ^2). In contrast to volatility, which increases with the square root of time, the variance of an asset's return will increase in a linear fashion over time. For example, if the daily volatility is 1.5%, the variance rate is $1.5\%^2 = 0.0225\%$. Thus, over a 10-day period, the variance will be 0.225% (i.e., $0.0225\% \times 10$).

In addition to variance and standard deviation, which are computed using historical data, risk managers may also derive implied volatilities. The **implied volatility** of an option is computed from an option pricing model, such as the Black-Scholes-Merton (BSM) model. The volatility of an asset is not directly observed in the BSM model, so we compute implied volatility as the volatility level that will result when equating an option's market price to its model price.



Professor's Note: Computing option prices using the BSM model will be demonstrated in Book 4.

The most widely used index for publishing implied volatility is the Chicago Board Options Exchange (CBOE) Volatility Index (ticker symbol: VIX). The VIX demonstrates implied volatility on a wide variety of 30-day calls and puts on the S&P 500 Index. Note that trading in futures and options on the VIX is a bet on volatility only. Since its inception, the VIX has mainly traded between 10 and 20 (which corresponds to volatility of 10%–20% on the S&P 500 Index options), but it reached a peak of close to 80 in October 2008, after the collapse of Lehman Brothers. The VIX is often referred to as the fear index by market participants because it reflects current market uncertainties.

THE POWER LAW

LO 28.2: Describe the power law.

It is typically assumed that the change in asset prices is normally distributed. This makes it convenient to apply standard deviation when determining confidence intervals for an asset's price. For example, by assuming an asset price of \$50 and a volatility of 4.47%, we can compute a one-standard-deviation move as $50 \times 0.0447 = 2.24$. With this information, we can define the 95% confidence interval as $50 \pm 1.96 \times 2.24$.

In practice, however, the distribution of asset price changes is more likely to exhibit fatter tails than the normal distribution. Thus, heavy-tailed distributions can be used to better capture the possibility of extreme price movements (e.g., a five-standard-deviation move). An alternative approach to assuming a normal distribution is to apply the power law.

The power law states that when X is large, the value of a variable V has the following property:

$$P(V > X) = K \times X^{-\alpha}$$

where:

V = the variable

X = large value of V

K and α = constants

Example: The power law

Assume that data on asset price changes determines the constants in the power law equation to be the following: $K = 10$ and $\alpha = 5$. Calculate the probability that this variable will be greater than a value of 3 and a value of 5.

Answer:

$$P(V > 3) = 10 \times 3^{-5} = 0.0412 \text{ or } 4.12\%$$

$$P(V > 5) = 10 \times 5^{-5} = 0.0032 \text{ or } 0.32\%$$

By taking the logarithm of both sides in the power law equation, we can perform regression analysis to determine the power law constants, K and α :

$$\ln[P(V > X)] = \ln(K) - \alpha \ln(X)$$

In this case, the dependent variable, $\ln[P(V > X)]$, can be plotted against the independent variable, $\ln(X)$. Furthermore, if we assume that X represents the number of standard deviations that a given variable will change, we can determine the probability that V will exceed a certain number of standard deviations. For example, if regression analysis indicates that $K = 8$ and $\alpha = 5$, the probability that the variable will exceed four standard deviations will be equal to $8 \times 4^{-5} = 0.0078$ or 0.78%. The power law suggests that extreme movements have a very low probability of occurring, but this probability is still higher than what is indicated by the normal distribution.

ESTIMATING VOLATILITY

LO 28.3: Explain how various weighting schemes can be used in estimating volatility.

By collecting continuously compounded return data, u_i , over a number of days (as shown in LO 28.1), we can compute the mean return of the individual returns as follows:

$$\bar{u} = \frac{1}{m} \sum_{i=1}^m u_{n-i}$$

where:

m = number of observations leading up to the present period

If we assume that the mean return is zero, which would be true when the mean is small compared to the variability, we obtain the maximum likelihood estimator of variance:

$$\sigma_n^2 = \frac{1}{m} \sum_{i=1}^m u_{n-i}^2$$

In simplest terms, historical data is used to generate returns in an asset-pricing series. This historical return information is then used to generate a volatility parameter, which can be used to infer expected realizations of risk. However, the straightforward approaches just presented weight each observation equally in that more distant past returns have the same influence on estimated volatility as observations that are more recent. If the goal is to estimate the current level of volatility, we may want to weight recent data more heavily. There are various weighting schemes, which can all essentially be represented as:

$$\sigma_n^2 = \sum_{i=1}^m \alpha_i u_{n-i}^2$$

where:

α_i = weight on the return i days ago

The weights (the α 's) must sum to one, and if the objective is to generate a greater influence on recent observations, then the α 's will decline in value for older observations.

One extension to this weighting scheme is to assume a long-run variance level in addition to the weighted squared return observations. The most frequently used model is an **autoregressive conditional heteroskedasticity model**, ARCH(m), which can be represented by:

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^m \alpha_i u_{n-i}^2 \text{ with } \gamma + \sum \alpha_i = 1 \text{ so that}$$

$$\sigma_n^2 = \omega + \sum_{i=1}^m \alpha_i u_{n-i}^2$$

where:

$\omega = \gamma V_L$ (long-run variance weighted by the parameter γ)

Therefore, the volatility estimate is a function of a long-run variance level and a series of squared return observations, whose influence declines the older the observation is in the time series of the data.

THE EXPONENTIALLY WEIGHTED MOVING AVERAGE MODEL

LO 28.4: Apply the exponentially weighted moving average (EWMA) model to estimate volatility.

LO 28.8: Explain the weights in the EWMA and GARCH(1,1) models.

The exponentially weighted moving average (EWMA) model is a specific case of the general weighting model presented in the previous section. The main difference is that the weights are assumed to decline exponentially back through time. This assumption results in a specific relationship for variance in the model:

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2$$

where:

λ = weight on previous volatility estimate (λ between zero and one)

The simplest interpretation of the EWMA model is that the day- n volatility estimate is calculated as a function of the volatility calculated as of day $n - 1$ and the most recent squared return. Depending on the weighting term λ , which ranges between zero and one, the previous volatility and most recent squared returns will have differential impacts. High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

Example: EWMA model

The decay factor in an exponentially weighted moving average model is estimated to be 0.94 for daily data. Daily volatility is estimated to be 1%, and today's stock market return is 2%. What is the new estimate of volatility using the EWMA model?

Answer:

$$\sigma_n^2 = 0.94 \times 0.01^2 + (1 - 0.94) \times 0.02^2 = 0.000118$$

$$\sigma_n = \sqrt{0.000118} = 1.086\%$$

One benefit of the EWMA is that it requires few data points. Specifically, all we need to calculate the variance is the current estimate of the variance and the most recent squared return. The current estimate of variance will then feed into the next period's estimate, as will this period's squared return. Technically, the only "new" piece of information for the volatility calculation will be that attributed to the squared return.

THE GARCH(1,1) MODEL

LO 28.5: Describe the generalized autoregressive conditional heteroskedasticity (GARCH (p,q)) model for estimating volatility and its properties.

LO 28.6: Calculate volatility using the GARCH(1,1) model.

One of the most popular methods of estimating volatility is the **generalized autoregressive conditional heteroskedastic** (GARCH)(1,1) model. A GARCH(1,1) model not only incorporates the most recent estimates of variance and squared return, but also a variable that accounts for a long-run average level of variance.



Professor's Note: In the GARCH(p,q) notation, the p stands for the number of lagged terms on historical returns squared, and the q stands for the number of lagged terms on historical volatility.

The best way to describe a GARCH(1,1) model is to take a look at the formula representing its determination of variance, which can be shown as:

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

where:

α = weighting on the previous period's return

β = weighting on the previous volatility estimate

ω = weighted long-run variance = γV_L

V_L = long-run average variance = $\frac{\omega}{1 - \alpha - \beta}$

$\alpha + \beta + \gamma = 1$

$\alpha + \beta < 1$ for stability so that γ is not negative

The EWMA is nothing other than a special case of a GARCH(1,1) volatility process, with $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$. Similar to the EWMA model, β represents the exponential decay rate of information. The GARCH(1,1) model adds to the information generated by the EWMA model in that it also assigns a weighting to the average long-run variance estimate. An additional characteristic of a GARCH(1,1) estimate is the implicit assumption that variance tends to revert to a long-term average level. Recognition of a mean-reverting characteristic in volatility is an important feature when pricing derivative securities such as options.

Example: GARCH(1,1) model

The parameters of a generalized autoregressive conditional heteroskedastic (GARCH)(1,1) model are $\omega = 0.000003$, $\alpha = 0.04$, and $\beta = 0.92$. If daily volatility is estimated to be 1%, and today's stock market return is 2%, what is the new estimate of volatility using the GARCH(1,1) model, and what is the implied long-run volatility level?

Answer:

$$\sigma_n^2 = 0.000003 + 0.04 \times 0.02^2 + 0.92 \times 0.01^2 = 0.000111$$

$$\sigma_n = \sqrt{0.000111} = 1.054\%$$

$$\text{long-run average variance} = \frac{\omega}{(1 - \alpha - \beta)} = \frac{0.000003}{(1 - 0.04 - 0.92)} = 0.000075$$

$$\bar{\sigma} = \sqrt{0.000075} = 0.866\% = \text{long-run volatility}$$

Mean Reversion**LO 28.7: Explain mean reversion and how it is captured in the GARCH(1,1) model.**

Empirical data indicates that volatility exhibits a mean-reverting characteristic. Given that stylized fact, a GARCH model tends to display a better theoretical justification than the EWMA model. The method for estimating the GARCH parameters (or weights), however, often generates outcomes that are not consistent with the model's assumptions. Specifically, the sum of the weights of α and β are sometimes greater than one, which causes instability in the volatility estimation. In this case, the analyst must resort to using an EWMA model.

The sum of $\alpha + \beta$ is called the **persistence**, and if the model is to be stationary over time (with reversion to the mean), the sum must be less than one. The persistence describes the rate at which the volatility will revert to its long-term value following a large movement. The higher the persistence (given that it is less than one), the longer it will take to revert to the mean following a shock or large movement. A persistence of one means that there is no reversion, and with each change in volatility, a new level is attained.

ESTIMATION AND PERFORMANCE OF GARCH MODELS

As was previously mentioned, one way to estimate volatility (e.g., variance) is to use a **maximum likelihood estimator**. Maximum likelihood estimators select values of model parameters that maximize the likelihood that the observed data will occur in a sample. Any variable of interest can be estimated via the maximum likelihood method, which requires formulating an expression or function for the underlying probability distribution of the data and then searching for the parameters that maximize the value generated by the expression.

One important consideration relates to which distribution is chosen when calculating probability. The most popular is the normal distribution, but normally distributed data are not often found in financial markets.

GARCH models are estimated using maximum likelihood techniques. The estimation process begins with a guess of the model's parameters. Then a calculation of the likelihood function based on those parameter estimates is made. The parameters are then slightly adjusted until the likelihood function fails to increase, at which time the estimation process assumes it has maximized the function and stops. The values of the parameters at the point of maximum value in the likelihood function are then used to estimate GARCH model volatility.

LO 28.9: Explain how GARCH models perform in volatility forecasting.**LO 28.10: Describe the volatility term structure and the impact of volatility changes.**

One of the useful features of GARCH models is that they do a very good job at modeling volatility clustering when periods of high volatility tend to be followed by other periods of high volatility and periods of low volatility tend to be followed by subsequent periods of low volatility. Thus, there is autocorrelation in u_i^2 . If GARCH models do a good job of explaining volatility changes, there should be very little autocorrelation in u_i^2 / σ_i^2 . GARCH models appear to do a very good job of explaining volatility.

The question then arises, if GARCH models do a good job at explaining past volatility, how well do they forecast future volatility? The simple answer to this question is that GARCH models do a fine job at forecasting volatility from a volatility term structure perspective (e.g., estimates of volatility given time to expiration for options). Even though the actual volatility term structure figures are somewhat different from those forecasted by GARCH models, GARCH-generated volatility data does an excellent job in predicting how the volatility term structure responds to changes in volatility. This modeling tool is quite frequently used by financial institutions when estimating exposure to various option positions.

KEY CONCEPTS

LO 28.1

The volatility of a variable is the standard deviation of that variable's continuously compounded return. The variance rate of a variable is the square of its standard deviation. Variance and standard deviation are computed using historical data. Risk managers may also compute implied volatility, which is the volatility that forces a model price (i.e., option pricing model) to equal the market price.

LO 28.2

The power law is an alternative approach to using probabilities from a normal distribution. It states that when X is large, the value of a variable V has the following property, where K and α are constants:

$$P(V > X) = K \times X^{-\alpha}$$

LO 28.3

Historical price data is used to generate return estimates, which are then used to estimate volatility. Traditional volatility estimation methods weight past information equally across time. Weighting schemes can be used to weight recent information more heavily than distant data.

LO 28.4

The EWMA model generates volatility estimates based on weightings of the last estimate of volatility and the latest current price change information. The objective is to account for previous volatility estimates, as well as to account for the latest return information.

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2$$

where:

λ = weight on previous volatility estimate (λ between zero and one)

LO 28.5

One of the most popular methods of estimating volatility is the generalized autoregressive conditional heteroskedastic (GARCH)(p,q) model. In a GARCH(p,q) model, the p stands for the number of lagged terms on historical returns squared, and the q stands for the number of lagged terms on historical volatility.

LO 28.6

GARCH(1,1) models not only incorporate the most recent estimates of volatility and return, but also incorporate a long-run average level of variance.

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2$$

where:

α = weighting on the previous period's return

β = weighting on the previous volatility estimate

ω = weighted long-run variance = γV_L

$$V_L = \text{long-run average variance} = \frac{\omega}{1 - \alpha - \beta}$$

$$\alpha + \beta + \gamma = 1$$

$\alpha + \beta < 1$ for stability so that γ is not negative

GARCH(1,1) estimates of volatility have a better theoretical justification than the EWMA model. In the event that model parameter estimates indicate instability, however, EWMA volatility estimates may be used.

LO 28.7

In a GARCH(1,1) model, the sum of $\alpha + \beta$ is called the persistence. The persistence describes the rate at which the volatility will revert to its long-term value. A persistence equal to one means there is no mean reversion.

LO 28.8

The EWMA is nothing other than a special case of a GARCH(1,1) volatility process, with $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$. Similar to the EWMA model, β in the GARCH(1,1) equation represents the exponential decay rate of information. The GARCH(1,1) model adds to the information generated by the EWMA model in that it also assigns a weighting to the average long-run variance estimate.

LO 28.9

GARCH models do a very good job at modeling volatility clustering when periods of high volatility tend to be followed by other periods of high volatility and periods of low volatility tend to be followed by subsequent periods of low volatility.

LO 28.10

When forecasting future volatility, GARCH-generated volatility data does an excellent job in predicting the volatility term structure (i.e., differing volatilities for options given differing maturities). This modeling tool is quite frequently used by financial institutions when estimating exposure to various option positions.

CONCEPT CHECKERS

1. An analyst is attempting to compute a confidence interval for asset Z prices. Assume a daily volatility of 1% and a current asset price of 100. What is the 95% confidence interval for asset price at the end of five days, assuming price changes are normally distributed?
 - A. 100 ± 1.96 .
 - B. 100 ± 2.24 .
 - C. 100 ± 4.39 .
 - D. 100 ± 9.80 .
2. The parameters of a generalized autoregressive conditional heteroskedastic (GARCH)(1,1) model are $\omega = 0.00003$, $\alpha = 0.04$, and $\beta = 0.92$. If daily volatility is estimated to be 1.5%, and today's stock market return is 0.8%, what is the new estimate of the standard deviation?
 - A. 1.68%.
 - B. 1.55%.
 - C. 1.45%.
 - D. 2.74%.
3. The λ of an exponentially weighted moving average (EWMA) model is estimated to be 0.9. Daily standard deviation is estimated to be 1.5%, and today's stock market return is 0.8%. What is the new estimate of the standard deviation?
 - A. 1.68%.
 - B. 1.55%.
 - C. 1.45%.
 - D. 2.74%.
4. The parameters of a GARCH(1,1) model are $\omega = 0.00003$, $\alpha = 0.04$, and $\beta = 0.92$. These figures imply a long-run daily standard deviation of:
 - A. 1.68%.
 - B. 1.55%.
 - C. 1.45%.
 - D. 2.74%.
5. GARCH(1,1) models can only be used to estimate volatility in the case where:
 - A. $\alpha + \beta > 0$.
 - B. $\alpha + \beta < 1$.
 - C. $\alpha > \beta$.
 - D. $\alpha < \beta$.

CONCEPT CHECKER ANSWERS

1. C First, convert daily volatility to weekly volatility using the square root to time: $1\% \times \sqrt{5} = 2.24\%$. Next, compute the one-standard-deviation move: $100 \times 0.0224 = 2.24$. Finally, derive the confidence interval: $100 \pm 1.96 \times 2.24 = 100 \pm 4.39$.

2. B $\sigma_n^2 = 0.00003 + (0.008)^2 \times 0.04 + (0.015)^2 \times 0.92 = 0.00023956$

$$\sigma_n = \sqrt{0.00023956} = 0.0155 = 1.55\%$$

3. C $\sigma_n^2 = 0.9 \times (0.015)^2 + (1 - 0.9) \times (0.008)^2 = 0.0002089$

$$\sigma_n = \sqrt{0.0002089} = 0.0145 = 1.45\%$$

4. D The long-run variance rate can be estimated by dividing the ω of a GARCH(1,1) model by $1 - \alpha - \beta$. This yields $0.00003 / (1 - 0.04 - 0.92) = 0.00075$; long-run standard deviation = $\sqrt{0.00075} = 0.0274 = 2.74\%$.

5. B Stable GARCH(1,1) models require $\alpha + \beta < 1$; otherwise the model is unstable.