DIEBOLD, CHAPTER 7

Francis X. Diebold, *Elements of Forecasting*, 4th Edition (Mason, Ohio: Cengage Learning, 2006). Chapter 7. Characterizing Cycles

After completing this reading you should be able to:

- Define covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function, and autoregression.
- Describe the requirements for a series to be covariance stationary.
- Explain the implications of working with models that are not covariance stationary.
- Define white noise, describe independent white noise and normal (Gaussian) white noise.
- Explain the characteristics of the dynamic structure of white noise.
- Explain how a lag operator works.
- Describe Wold's theorem.
- Define a general linear process.
- Relate rational distributed lags to Wold's theorem.
- Calculate the sample mean and sample autocorrelation, and describe the Box-Pierce Q-statistic and the Ljung-Box Q-statistic.
- Describe sample partial autocorrelation.

Learning objective: Define covariance stationary, autocovariance function, autocorrelation function, partial autocorrelation function, and autoregression.

Learning objective: Describe the requirements for a series to be covariance stationary.

Learning objective: Explain the implications of working with models that are not covariance stationary.

COVARIANCE STATIONARY FUNCTION

When an independent variable in the regression equation is a lagged value of the dependent variable (as is the case in autoregressive time series models) statistical inferences based on OLS regression are not always valid. In order to conduct statistical inference based on these models, we must assume that the time series is covariance stationary or weakly stationary. There are three basic requirements for a time series to be covariance stationary:

1. The expected value or mean of the time series must be constant and finite in all periods.
2. The variance of the time series must be constant and finite in all periods.
3. The covariance of the time series with itself for a fixed number of periods in the past or future must be constant and finite in all periods.

It is a good time to note what type of time series is not mean stationary: if the time series has a trend line such as GDP growth or employment, by definition that time series is not mean stationary. Same thing with variance—if the time series has a trending or changing variance, it is not variance stationary.

For notation in this reading we are going to use

$$E(y_t) = \mu$$

to say the expectation of our sample path *y* is the mean of *y*. In other words, rule number 1.

AUTOCOVARIANCE FUNCTION

To determine if our series has a stable covariance structure over different periods of time we have to use an autocovariance function. Series analysis is usually limited to time series in finance but it can apply to any type of data. We refer to a "displacement" as the next step in a series be it three-dimensional data or linear time series data.

We need to include displacement and time in our autocovariance function like:

$$\gamma(t, \tau) = \gamma(\tau)$$

This just says our function has two variables, time and displacement, and the covariance between those two equals the expectation of each of them minus the mean, which from step one we assume to be zero. In order to be covariance stationary, the series cannot be dependent upon time, which reduces our autocorrelation equation to reduce to be dependent just on displacement:

$$\gamma(t, \tau) = \text{cov}(y_t, y_{t-\tau}) = E(y_t - \mu)(y_{t-\tau} - \mu)$$

AUTOCORRELATION FUNCTION

Now, just like covariance in other areas of analysis doesn't tell us much—there are no units, it can vary widely, etc.—we use correlation to normalize covariance. Recall that we calculate correlation by dividing covariance by standard deviation, and just in the same way we create the correlation function, we divide our autocovariance function by standard deviation.

When the autocorrelations are plotted by time step—we can graphically see how the dependence pattern changes or alters by lag step.

Learning objective: Define white noise, describe independent white noise and normal (Gaussian) white noise.

White noise is a special characterization of a type of error terms within a time series. Recall we use "y" to denote an observed time series and we further want to say the error terms have a mean 0 and some known, constant variance.

Formulaically,

$$y_t = \epsilon_t$$
$$\epsilon_t \sim (0, \sigma^2)$$

Furthermore, if the error terms are uncorrelated over time, that process is called white noise. If we can further show that the series *y* is independent and identically distributed, the white noise becomes independent white noise. Lastly, if we can show i.i.d as well as normally distributed, we say the white noise is Gaussian (normally distributed).

Learning objective: Explain the characteristics of the dynamic structure of white noise.

It is important to understand that white noise is uncorrelated over time, has zero autocorrelation, and is basically a random stochastic process. The take-away from this reading is that forecast errors should be white noise—and this is counterintuitive—because if the errors aren't white noise then the errors are serially correlated, which also means the errors are forecastable, which also means the forecast itself is flawed or unreliable.

Learning objective: Explain how a lag operator works.

Since we are manipulating time series data to explain how the past evolves into the future, we have to manipulate the forecast model and a lag operator will turn the current one into the previous observation like this:

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

A lag of two steps would have *L* raised to the second power and so forth. For now, just know the notation. We will get into the reason, use, and meaning later.

$$L^2 y_t = LLy_t = Ly_{t-1} = y_{t-2}$$

so the first lag of our autocovariance equation would look like this:

$$\gamma(1) = \text{cov}(Y_t, Y_{t-1})$$
$$= E((Y_t - \mu)(Y_{t-1} - \mu))$$

You can visualize a lag by taking any graph and "shifting" the entire chart by one or more time steps—however large the lag may be.

Now, we can also lag a time series by a polynomial where we set the observation at *t* plus some fraction of prior observations like this:

$$(1 + .14L + .06^2)\, y_t = y_t + .14_{t-1} + 0.6_{t-2}$$

Learning objective: Describe Wold's theorem.

Time series analysis is all about picking the right model to fit a series of data. When we have a set of data that is covariance stationary there are a lot of model choices that could fit the data with different degrees of effectiveness. This alone doesn't tell us anything about whether or not this is the right model. Think of this idea as analogous to correlation doesn't equal causation. This breaks down the time series into two pieces—one deterministic and one stochastic (a.k.a. random or white noise) so Wold's theorem is the model of the covariance stationary residual.

Time series models are constructed as linear functions of fundamental forecasting errors, also called innovations or shocks.

These basic building blocks of models are the same: mean 0 with some known variance, serially uncorrelated, a.k.a. white noise. In this sense, all errors are white noise and unforecastable.

Now, in this reading the error terms are often referred to as "innovations," which gets very confusing. The reason this came about is because if we have an actual error, not white noise, in a time series, then that actually introduces new information to the system that can be modeled, can be predicted, and is in some way correlated to the rest of the time series.

Furthermore, a "distributed lag" is a weighted sum of previous values that factor in some way to our estimation of the current value of the time series, which is exactly what the equation from the last learning objective is—a distributed lag, meaning the current value weight is distributed over several previous values of the time series.

Recall the exponentially weighted model where we set lambda at some value to drive the "decay" of the informational value of the historical data. If we want to grab all historical value, it's called infinitely distributable.

The formula looks like this:

$$y_t = B(L) \in_t = \sum_{i=0}^{\infty} b_i \in_{t-1}$$

$$\in_t \sim WN(O, \sigma^2)$$

Learning objective: Define a general linear process.

This reading is quite complicated and it is easy to get bogged down in formulas when all they are asking for is a definition. In this case, a general linear process describes Wold's theorem because the previous formula is a series of linear functions and its "innovations." This is probably a low probability for the exam.

Learning objective: Relate rational distributed lags to Wold's theorem.

There are two types of lags in time series analysis: infinite and finite. In an infinite series, we assume all data in the past has an impact on the dependent variable we are modeling so it is "infinitely distributed." Same for a finite distribution, we just have to define how many lagged periods impact the time series. In these types of lagged models, we assume there is some weight applied to each lag but there can also be polynomials in the lag factor. The problem arises because models with an infinite number of parameters can't be estimated from a finite sample of data. However, an infinite polynomial in the lag operator won't necessarily have infinite free parameters. We can have an infinite series of polynomials that only depend on, say, two parameters. A rational distributed lag is the ratio of the parameters in the infinitely distributed lag so that we can approximate an infinitely lagged series from a finite sample of data, which is how we recover Wold's theorem from rational distributed lags.

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

Recall that when we are dealing with any type of "sample" we are dealing with estimators, not parameters, and we extend the mean and autocorrelation to accommodate that we know there is some degree of error in the estimator. This is called replacing expectations with sample averages. If we have a sample size of T, the sample mean is calculated as:

$$\bar{y} = \frac{1}{T} \sum_{t=2}^{T} y_t$$

This is not interesting in itself but we can use it to calculate the sample autocorrelation function, which we calculate as:

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

Learning objective: Describe sample partial autocorrelation.

This is a low priority for the exam and is only a "describe learning statement." Instead of obtaining partial autocorrelations in a "thought experiment" of infinite regressions on an infinite data set, we now perform the same on a thought experiment on a more manageable, finite data set and that is why it is called a "sample." This is purely theoretical and a near zero chance on exam day.