Math Tools: Time Series Data

(Started: December 5, 2011; Revised: April 10, 2012)

We describe a couple ways to summarize the dynamic patterns evident in time series data: a sample of observations $(x_t, x_2, \ldots, x_T)$. What's different about time series data is that the order matters: x_3 is next to x_2 and x_4, which is typically relevant to how we think about them.

We develop two tools for describing the behavior of time series variables. The first is the *autocorrelation function*, a summary of the relation between x_t and x_{t-k} for various values of k. The second is the *cross-correlation function* a summary of the relation between x_t and y_{t-k}.

Autocovariances and autocorrelations

You may recall that the sample mean is

$$\bar{x} \;=\; T^{-1} \sum_{t=1}^{T} x_t$$

and the variance is

$$\gamma_x(0) \;=\; T^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^2.$$

The rational for the odd notation should be clear shortly.

Consider the covariance of x_t with x_{t-k}, for k a nonnegative integer. The sample covariance is computed

$$\gamma_x(k) \;=\; T^{-1} \sum_{t=k+1}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x}).$$

Since we only have the observations x_t for $t = 1, ..., T$, we need to start the sum at $t = k+1$. By longstanding convention, we nevertheless divide the sum by T rather than $T - k$. We could also consider negative values of k, but we'd have to adjust the range in the sum appropriately. We refer to $\gamma_x(k)$, a function of k, as the autocovariance function; that is, the covariances of x with itself, so to speak. When $k = 0$, we get the variance.

The shape of $\gamma_x(k)$ is useful in telling us about the dynamics of x, but it's more common to scale it by $\gamma_x(0)$ and convert it to a correlation. The autocorrelation function $\rho_x(k)$ is defined by

$$\rho_x(k) \;=\; \gamma_x(k)/\gamma_x(0).$$

Obviously $\rho_x(0) = 1$: x_t is perfectly correlated with x_t. But for other values of k it can take a variety of forms.

We see, for example, that autocorrelations of equity returns are very small: returns are virtually uncorrelated over time. Interest rates, however, are very persistent: the autocorrelations decline slowly with k. You can verify other patterns in the data we used in class.

Cross-covariances and cross-correlations

We can extend the idea to the relation between two variables, say x and y. The sample *cross-covariance function* (cross meaning across two variables) is defined by

$$\gamma_{xy}(k) \quad = \quad T^{-1} \sum_t (x_t - \bar{x})(y_{t-k} - \bar{y}),$$

where the sum is over the appropriate range. This is defined by integer values of k. If k is negative, we're looking at the covariance of x and future y. If k is positive, we're looking at the covariance of x and past y. Either way, we learn something about the dynamic association of x and y.

As before, it's conventional to report correlations rather than covariances. The *cross-correlation function* is

$$\gamma_{xy}(k) \quad = \quad \frac{\gamma_{xy}(k)}{\gamma_x(0)^{1/2}\gamma_y(0)^{1/2}},$$

the covariance divided by the product of standard deviations.