Descriptive Estimators

Overview

1. Moment estimators for μ, $\gamma(h)$, and correlations $\rho(h)$.
2. Simulate estimators using R.

See S&S, Appendix A, for further details on the properties of these estimators that we'll cover in the next class.

Moment estimators

Context, notation The general setting for estimation in this lecture is that we

- Observe $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ (n observations)
- Denote sequences by $\{X_1, X_2, \ldots, X_n\} = X_1^n$ and $\{x_1, x_2, \ldots, x_n\} = x_1^n$ and similarly for other symbols.
- Assume $\{X_t\}$ is weakly stationary and w_t is white noise. We will use the assumption of stationarity to allow averaging over time in place of averaging over the *ensemble* of processes.
- Covariances are $\text{Cov}(X_{t+h}, X_t) = \gamma_X(h)$ which are arranged in the $n \times n$ array $\Gamma = [\gamma(i-j)]$.

Reasons for studying moment estimators (a) They avoid the need for a specific model (other than stationarity); models ultimately provide more efficient estimators. Without specifying probability distributions for $\{X_t\}$ we cannot use *maximum likelihood.*

(b) Moment estimators provide an initial value for more efficient, iterative estimation algorithms such as ML. In such cases, it's important that the moment estimator be *consistent to order* $1/\sqrt{n}$ in order to obtain a one-step estimator.

Estimate of μ In general, the components of X_1^n are dependent, implying some sort of weighted estimator, such as the generalized least squares

(GLS) estimator ($\mathbf{1} = (1, 1, \ldots, 1)$)

$$\hat{\mu} = \frac{\mathbf{1}'\Gamma^{-1} x_1^n}{\mathbf{1}'\Gamma^{-1}\mathbf{1}} \tag{1}$$

(Think of $\hat{\mu}$ as the regression slope with correlated errors; all of this discussion extends to regression.) Because we don't know the covariances (we need an estimate of μ first), we begin with a more basic estimator,

$$\overline{X} = \frac{1}{n} \sum_{t=1}^{n} X_i$$

Some rationale: its often better in practice to have a simple estimator such as $\overline{X}$ which as known properties rather than a possibly better estimator such as $\hat{\mu}$ whose properties are unknown.

Sample covariances An *almost* unbiased estimator is

$$\tilde{\gamma}(h) = \frac{\sum_{t=1}^{n-|h|}(X_t - \overline{X})(X_{t+|h|} - \overline{X})}{n - |h|}$$

Many prefer the following estimator (for reasons explained below)

$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{n-|h|}(X_t - \overline{X})(X_{t+|h|} - \overline{X})}{n}$$

Note that $\gamma(h) = \gamma(-h)$ and $\hat{\gamma}(h) = \hat{\gamma}(-h)$. Neither estimator is unbiased because $\overline{X}$ appears in place of μ. Note the role of the bias in $\hat{\gamma}(h)$ is to shrink the estimator of the sequence to zero.

Sample correlations Frequently useful to have a scale-free estimator of the dependence, namely an estimator for

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

We will consider the plug-in estimator

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Positive definite Suppose $\mu = 0$ were known, then the biased estimator

$$g_h = \frac{\sum_{t=1}^{n-|h|} X_t X_{t+|h|}}{n}$$

produces a positive semi-definite sequence (*i.e.*, the resulting covariance matrix $G = [g_{ij}]$ is p.s.d., implying that quadratic forms such as $c'Gc \geq 0$) The unbiased estimator does not. In this sense, tolerating a little bias pointwise (at each lag h) provides a "better" estimator with a desired global property.

To see that G is p.s.d., consider the cross-products obtained by padding this series by zeros and then forming inner-products. (See exercise 1.24.) Other proofs rely on the Fourier transform. We will do these later.

Simulation results

Simulate the properties of estimates of the mean and covariances (and correlations) for several processes:

- White noise (Gaussian or Poisson), random walk
- One-sided moving average
- Autoregression

R commands that are useful in this case are `pairs` to see plots of the dependence of the estimates, in addition to `rnorm, rpois, filter` and commands to control looping (`for`). The function `acf` computes the autocovariances and autocorrelations.

Examples in the R code include:

- How well does $\overline{X}$ perform compared to the GLS estimator? The simulation compares the standard error and bias of $\overline{X}$ to those of the GLS estmator What are the properties of these estimators of the covariances? (Notice that it does this without having to compute the GLS estimator.)
- What are the sampling properties of $\hat{\rho}(h)$? The simulation considers the dependence among the estimates of $\gamma(h)$. In fact, for

large $|h|$, the estimated correlations look a lot like a stationary process themselves.