ARMA Processes

Non Normal residuals

By Olivier Croissant

As it is stated in Introduction to Time Series and Forcasting, by P.J. Brockwell and R.A. Davis 1996, Cap 5.2 pp 157:

The likelihood (5.2.1) of the vector $\mathbf{X}_n$ therefore reduces to

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n v_0 \cdots v_{n-1}}} \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(X_j - \hat{X}_j)^2/v_{j-1}\right\}. \qquad (5.2.6)$$

If Γ_n is expressible in terms of a finite number of unknown parameters $\beta_1, \ldots, \beta_r$ (as is the case when $\{X_t\}$ is an ARMA(p, q) process) the **maximum likelihood estimators** of the parameters are those values that maximize L for the given data set. When $X_1, X_2, \ldots, X_n$ are iid, it is known, under mild assumptions and for n large, that maximum likelihood estimators are approximately normally distributed with variances that are at least as small as those of other asymptotically normally distributed estimators (see e.g., Lehmann, 1983).

Even if $\{X_t\}$ is not Gaussian, it still makes sense to regard (5.2.6) as a measure of goodness of fit of the model to the data, and to choose the parameters $\beta_1, \ldots, \beta_r$ in such a way as to maximize (5.2.6). We shall always refer to the estimators, $\hat{\beta}_1, \ldots, \hat{\beta}_r$, so obtained as "maximum likelihood" estimators, even when $\{X_t\}$ is not Gaussian. Regardless of the joint distribution of $X_1, \ldots, X_n$, we shall refer to (5.2.1) and its algebraic equivalent (5.2.6) as the "likelihood" (or "Gaussian likelihood") of $X_1, \ldots, X_n$. A justification for using maximum Gaussian likelihood estimators of ARMA coefficients is that the large-sample distribution of the estimators is the same for $\{Z_t\} \sim \text{IID}(0, \sigma^2)$, regardless of whether or not $\{Z_t\}$ is Gaussian (see TSTM, Section 10.8).