A SEMIPARAMETRIC LONG MEMORY MODEL

The spectral density of the fractional ARIMA(p,d,q) model can be written as the product

\[ f(\lambda) = \frac{\sigma^2}{2\pi} |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) , \]

where \( f^*(\lambda) \) is the spectral density of an ARMA(p,q) process. Even if this model holds exactly, there is the possibility that we will misspecify \( p \) and \( q \) (i.e., use the wrong values). As a result, the maximum likelihood or Whittle estimate of \( d \) will be asymptotically biased. Another type of misspecification occurs if \( f^*(\lambda) \) is not in fact ARMA but we assume that it is. Here again, the standard estimates of \( d \) will be asymptotically biased. This asymptotic bias is a serious problem, since if our focus is on the long memory aspects of the series, then the only parameter of any real interest is \( d \). Clearly, then, it would be nice to have a way of estimating \( d \), even if we are not able to specify a fully parametric model for the short memory aspects of the process, \( f^* \).

We start by generalizing the model (1): We now suppose simply that \( f^*(\lambda) \) is continuous at \( \lambda=0 \), and that there exist finite positive constants \( C_1, C_2 \) such that \( C_1 \leq f^*(\lambda) \leq C_2 \) for all \( \lambda \in [0, \pi] \). (This is indeed a generalization of the fractional ARIMA(p,d,q) model, since the above assumptions are satisfied if \( f^* \) corresponds to a stationary invertible ARMA process. The point of making these assumptions is to ensure that \( f^* \) is not itself the spectral density of a long memory process.) Then (1) becomes a semi-parametric model, in the sense that the long memory aspects of the series are parametrically specified (by \( d \)), but the short memory aspects \( (f^*) \) are not required to obey any parametric model.

Geweke and Porter-Hudak proposed an estimator of \( d \) in the semiparametric long memory model described above, based on regression of the log periodogram. They assumed that asymptotically (as \( n \to \infty \)), the first \( M \) normalized periodogram ordinates \( \{I_j/\hat{f}_j\}_{j=1}^M \) are \( iid \) \( \frac{1}{2}\chi^2_2 \). In fact, Hurvich and Beltrao have shown that this assumption is incorrect in the long-memory case \( (d \neq 0) \), but it will be a reasonable approximation except at the first few Fourier frequencies. We will assume here that it holds exactly. Then it can be shown that \( E[\log(I_j/\hat{f}_j)] = -C \), where \( C = .577216 \ldots \) is Euler’s constant, and \( \text{var}[\log(I_j/\hat{f}_j)] = \pi^2/6 \). Therefore, \( \varepsilon_j = \log(I_j/\hat{f}_j) + C \) are \( iid \) with \( E[\varepsilon_j] = 0, \ \text{var}[\varepsilon_j] = \pi^2/6 \) for \( j = 1, \ldots, M \). Using (1), we obtain
\[
\log I_j = \log f_j + \log(I_j/f_j) \\
= \log f_j - C + \varepsilon_j \\
= [\log \left( \frac{\sigma^2}{2\pi f_j} \right) - C] - 2d \log |1 - e^{-i\omega_j}| + \varepsilon_j.
\]

If \( M \) is small compared to \( n \), then the continuity of \( f^*(\lambda) \) at \( \lambda = 0 \) implies that \( f^*(\omega) \) is essentially constant at \( f_0^* \) on the interval \( \omega \in [\omega_1, \omega_M] \). Therefore, we have, to a good approximation,

\[
\log I_j = [\log \left( \frac{\sigma^2}{2\pi f_0^*} \right) - C] - 2d \log |1 - e^{-i\omega_j}| + \varepsilon_j, \quad j = 1, \ldots, M. \tag{2}
\]

Using (2) as motivation, Geweke and Porter-Hudak proposed to estimate \( d \) using the least-squares estimator of the slope parameter, in a linear regression of \( \{\log I_j\}_{j=1}^M \) on \( \{-2\log |1 - e^{-i\omega_j}|\}_{j=1}^M \). We will denote the resulting estimate by \( \hat{d}^{(GPH)} \).

Assuming that (2) holds exactly, it can be shown that \( E[\hat{d}^{(GPH)}] = \hat{d} \), and \( \text{var}[\hat{d}^{(GPH)}] \propto \pi^2/(6M) \). Thus, if we let \( M \to \infty \) (slowly), then \( \hat{d}^{(GPH)} \) will be a consistent estimate of \( d \).

If \( f^* \) is actually ARMA(\( p,q \)) with \( p \) and \( q \) known, then the maximum likelihood estimator of \( d \) will have variance proportional to \( 1/n \), and therefore will be more efficient than \( \hat{d}^{(GPH)} \). This is not so important, however, once we admit that the ARMA assumption on \( f^* \) will rarely be satisfied in practice. The GPH estimate of \( d \) is useful since it does not require the user to specify a parametric model for \( f^* \).

The main practical problem we encounter in using \( \hat{d}^{(GPH)} \) is the selection of the number of frequencies, \( M \). If we take \( M \) too small, the variance of \( \hat{d}^{(GPH)} \) will be unacceptably high. If we take \( M \) too large, then the assumption that \( f^*(\omega) \) is approximately constant for \( \omega \in [\omega_1, \omega_M] \) will break down, and \( \hat{d}^{(GPH)} \) will be biased. Thus, the choice of \( M \) involves a tradeoff between bias and variance. Hurvich and Beltrao have proposed a data-driven method of selecting \( M \), using frequency domain cross validation.