Chapter 6

MODEL ESTIMATION, SELECTION, AND CHECKING

Model estimation, selection, and diagnostic checking are three interwoven components of time series analysis. If, within a specified class of nonlinear models, a particular linearity test statistics indicates that the DGP underlying an observed time series is indeed a nonlinear process, one would ideally like to be able to select the correct lag structure and estimate the parameters of the model. In addition, one would like to know the asymptotic properties of the estimators in order to make statistical inference. Moreover, it is evident that a good, perhaps automatic, order selection procedure (or criterion) helps to identify the most appropriate model for the purpose at hand. Finally, it is common practice to test the series of standardized residuals for white noise via a residual-based diagnostic test statistic.

In this chapter, we focus on these three themes within the context of parametric nonlinear modeling. Specifically, we consider the class of identifiable parametric stochastic models

\[ Y_t = g(Y_{t-1}, \ldots, Y_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}; \theta_g) + \eta_t \] (6.1)

where

\[ \eta_t = h(Y_{t-1}, \ldots, Y_{t-u}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-v}; \theta_h)^{1/2} \varepsilon_t. \]

Here \( \{Y_t, t \in \mathbb{Z}\} \) is a strictly stationary and ergodic univariate stochastic process; \( g(\cdot; \theta_g) \) and \( h(\cdot; \theta_h) \) are two real-valued measurable (known) functions on \( \mathbb{R}^{p+q} \) and \( \mathbb{R}^{u+v} (u \leq p) \), respectively; and \( \theta = (\theta'_g, \theta'_{h})' \) is a vector of unknown parameters that we wish to estimate, and we have available a set of observations \( \{Y_t\}_{t=1}^T \) with which to do so. Further, we assume that \( h(\cdot; \theta) \) is a non-negative function of past \( Y_t \)'s and \( \varepsilon_t \)'s.

The class of models (6.1) covers a wide range of nonlinear models, including many models introduced earlier in this book. Numerous methods have been proposed for estimating models contained within this class. Here, we do not provide a full
technical treatment of the subject. Rather we elaborate on some commonly used estimation methods and, in some cases, their practical implementation. Throughout the discussion, we assume that (6.1) is completely known. In practice, however, this is seldom the case and the model structure needs to be specified first. This is a model selection problem, and there are several ways to approach it. One is to develop model selection criteria on the basis of the asymptotic properties of the estimated parameters, and we will therefore spend some time discussing these criteria here. Alternatively, model selection criteria have been suggested on the basis of sample reuse such as cross-validation (CV). Since several of the latter criteria are (asymptotically) linked to criteria in the first group, we include them as well in this chapter. Similarly, the effect of parameter estimation errors becomes relevant when checking for model adequacy.

Given the above themes, the chapter consists of three interrelated parts. First, in Section 6.1.1, we discuss the method of quasi maximum likelihood (QML) estimation and, in particular, nonlinear least squares (NLS) estimation within the general framework of model (6.1). In Section 6.1.2, we consider the method of conditional least squares (CLS) estimation tailor-made for SETARMA, subset SETARMA, STAR, and BL models. In Section 6.1.3, we present an iteratively weighted least squares algorithm for QML estimation of double threshold ARCH models.

In the second part, we concentrate on model selection rules that are associated with the QML and NLS estimation methods. Both estimation methods are likely the most commonly used in practice. Consequently, the associated order selection rules are of quite general interest. In the third part, we discuss a general class of standardized-residuals-based correlation test statistics. The proposed tests avoid potential “size distortion” problems due to estimation uncertainty. Finally, in Section 6.4, we bring together elements of (subset) TARSO model estimation, TARSO model selection and checking, to analyze an important nonlinear time series problem from the area of hydrology.

6.1 Model Estimation

6.1.1 Quasi maximum likelihood estimator

Consider model (6.1). Let \( p^* = p \lor u, q^* = q \lor v, \) \( Y_0 = (Y_0, \ldots, Y_{1-p^*})' \) be the initial starting values of the process \( \{Y_t, t \in \mathbb{Z}\} \), and \( \varepsilon_0 = (\varepsilon_0, \ldots, \varepsilon_{1-q^*})' \) be the starting innovations. In addition, let \( \theta_0 = (\theta_0^g, \theta_0^h)' \) denote the true value of the parameter vector \( \theta \), and \( Y_t = (Y_1, \ldots, Y_t)' \). We assume that \( \theta_0 \) belongs to \( \Theta = \Theta_g \times \Theta_h \subset \mathbb{R}^{p+q} \times \mathbb{R}^{u+v} \).

Under the above assumptions, it is easily seen that the conditional mean and variance of \( \{Y_t, t \in \mathbb{Z}\} \) given \( Y_{t-1} \) and \( \Theta \) are

\[
\mathbb{E}(Y_t|Y_{t-1}, \Theta) = g(Y_{t-1}, \ldots, Y_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}; \theta_0^g) \equiv \mu_t(\theta_0^g)
\]

\[
\text{Var}(Y_t|Y_{t-1}, \Theta) = h(Y_{t-1}, \ldots, Y_{t-u}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-v}; \theta_0^h) \varepsilon_t \equiv \sigma_t^2(\theta_0^h).
\]