Instrumental Variables Estimators
for State Space Models of Time Series
by
M. Aoki

Introduction

The instrumental variable method, also known as the method of moments, has been used extensively both in the engineering and the econometric literature. For example, $\hat{S}$öderstrom and Stoica (1983), and Ljung (1987); and Bowden and Turkington (1984), White (1984), and Sagan (1988) are books, respectively, in engineering and econometrics, which discuss the method. There are also numerous journal articles cited in these books. In the econometric literatures, the instrumental variable method is usually .splied to statis regression models. In the engineering literature, the metod is .splied to (scalar-valued) dynamic processes in ARMA or related representations. When .splied to state space models, one usually converts them into equivalent ARMA or regression model forms before the method is .splied.

Estimation of state space models in innovation representation is discussed in Aoki (1983, 1987) in which an algorithm, which adapts system identification algorithm considered in realization theory in system theory, has been proposed to estimate state space model in innovation representation. This algorithm has been extensively tested by him and some of his colleagues, and has been found to be computationally efficient, robust, and produce good prediction models, even though the estimators are not maximum likelihood estimators. Some of the .splications are found in Aoki (1987); Cerchi and Havenner (1988). See also the references cited in Aoki (1987, 1990).

Later, Havenner and Aoki (1988) provided an instrumental variable interpretation of the state space modeling algorithm proposed in Aoki (1987). In particular, the estimator in Aoki (1987) for matrix C in the observation equation in (1) below has been shown to be identical with the instrumental variable estimator in which the stacked current and past data vector is used as instruments. This paper provides further detail of this interpretation and shows that the state vectors of the forward and backward innovation models are asymptotically the most efficient instrumental variables in estimating system matrices C and M respectively, and both are used to obtain asymptotically the most efficient estimator of the matrix A, where C is the observation matrix, M is the cross-covariance matrix of the state vector of the forward innovation model and the data vector which is needed in estimating matrix B and A is the dynamic matrix, all introduced in (1) below.

Estimation of System Matrices

This section describes two types of estimators for system matrices of state space models from given sets of time series data, and examines statistical properties of the proposed estimators. The estimators of the first type are based on the stochastic realization

G. Conte et al., *New Trends in Systems Theory*

theory, and are suggested by the systems literature on the deterministic realization theory. The estimators of the second type are instrumental variables estimators. We describe an important relation between these two estimators. The first type of estimators are computationally more efficient than the second, but the latter is asymptotically more efficient than the former. There is an important class of VAR data generating processes for which the two types of estimators are equally asymptotically efficient.

Two Classes of Estimators of System Matrices

The Hankel matrix built up from the sample covariance matrices of the data set can be factored as the produce of the observability matrix O and another matrix Ω which has the same structure as the reachability matrix. This factored expression suggests a way of computing the system matrices A and C in the innovation representation given below. The matrix B and the noise covariance matrix are then computed after solving the matrix Riccati equation, see Aoki [1987].

The basic steps in constructing innovation models are collected here for easy reference. Consider building an innovation model

$$z_{t+1} = Az_t + Be_t, \text{ and } \quad y_t = Cz_t + e_t, \tag{1}$$

where e_t is a mean-zero, serially uncorrelated covariance-stationary process with $cov \;\; e_t = \Delta$. The covariances matrices of $\{y_t\}$ are related to the system matrices and the noise covariance matrix by the relations

$$\Lambda_k = E(y_{t+k}y_t') = CA^{k-1}M, \, k \geq 1, \text{ and } \Lambda_0 = C\Pi C' + \Delta, \tag{2}$$

where the matrix Π is the covariance matrix of the state vector $\Pi = E(z_t z_t')$, and the matrix M denotes the cross-covariance matrix between the vectors z_{t+1} and y_t, and is related to the other system matrices by $M = A\ \Pi C' + B\ \Delta$. In this chapter, the matrix A is assumed to be stable so that $\{z_t\}$ is covariance stationary, i.e., the matrix Π is constant. From weak stationarity, Π satisfies the matrix algebraic equation, called the Riccati equation $\Pi = A\ \Pi A' + B\ \Delta B'$. Both matrices B and Δ depend on Π, since $\Delta = \Lambda_0 - C\ \Pi C'$, and $B = (M - A\ \Pi C')\Delta^{-1}$. The matrix Δ must be positive definite for its inverse to exist. This condition is called the full rank or regularity condition. We require that none of the components of the data vector $\{\ y_t\ \}$ be collinear for all times so that the matrix Λ_0 is nonsingular.

The theoretical Hankel matrix is constructed from the covariance matrices for the data vector process given in (2). The sample Hankel matrix is similarly made using sample covariance matrices. They are denoted by a hat over the theoretical ones. The north-west section of size J sub-matrix rows by K sub-matrix columns of the Hankel matrix is the product of the observability matrix $O_J' = [C'A'C' \cdots (A')^{J-1}C']$ with another matrix called Ω_K $\Omega_K = [M\ AM\ A^2M \cdots A^{K-1}M]$, because the covariance matrices of the weakly stationary mean zero data generating stochastic processes with rational spectral densities have the special structure shown in (2).