2.1 Introduction

This chapter deals with the first step in the preliminary analysis of the data in order to find out whether these can be analyzed usefully in the context of a VAR model. The main aim of the VAR model in this study is to discover meaningful linear lead-lag relationships between the variables, without using too much a priori information. The bulk of the information should therefore be extracted from the data.

Since an empirical VAR model is only based on the estimates of variances and covariances of the included variables, it is necessary to check whether one can interpret these estimates in a sound way. In the theory and practice of time series analysis of macroeconomic variables a number of interesting techniques exist that supply applied researchers with different interpretations of the variances and covariances, such as the spectrum and the "random shock" or impulse response representation. Econometric analysis of the rational distributed lag model has provided other interpretations and a statistical toolbox with slightly different contents, i.a. checks on model adequacy developed for relatively small samples.

Regression equations are the natural building blocks of a VAR model. In the context of regression analysis it has been argued that statistical analysis can only be meaningful within the framework of a flexible statistical model (Spanos (1986)) which has to be general enough to capture the major properties of the data. It is statistically unsound to attach meaning to outcomes of tests of specific hypotheses against a more general one if important aspects of the general model can be rejected a priori. Since we would rather not start with a disabled model, blind to the insights we can get from the statistical toolbox of the regression equation, we begin by setting up a slightly modified VAR model. It should serve as a useful maintained hypothesis for statistical analysis and thus be able to describe the well known features of interest of real quarterly macroeconomic data.

In the following section we introduce a formal representation of the modified VAR model. From this set-up we derive conditions on the univariate properties of the included variables which they have to fulfill in order to meaningfully relate them to each other in the multivariate model. In the third
section we present a derivation of the marginal univariate processes in the presence of unit root nonstationarity.

In the fourth section we give our definition of unit root processes. We then discuss the properties of these processes in more detail. In the last section we introduce alternative models which can describe nonstationarity, long memory and persistence. We advocate the use of data analytic tools geared towards detection of these alternatives. In appendix A2.2 we discuss statistical tests for unit root nonstationary. We apply these techniques in §7.4.

2.2 The model

The basic model reads:

\[ \Phi(L)\gamma_t = \omega_t, \quad t \in \mathbb{Z} \quad (2.1) \]

\[ y_t = S(L)(w(x_t) - g(t) - h_t(L)\eta_t), \quad (2.2) \]

\[ \Phi(L) = V(L)M(L)U(L), \quad (2.3) \]

where

- \( x_t \) is an \( n \)-column vector of observed variables of interest,
- \( y_t \) is a real, purely stochastic, zero mean \( n \)-vector, a transformation of interest of \( x_t \), and
- \( \Phi(L) = \Phi_0 + \Phi_1L + \Phi_2L^2 + \ldots + \Phi_pL^p \) is an real \( n \times n \)-matrix lag polynomial, with all roots of \( \text{det}(\Phi(z)) = 0 \) on or outside the unit circle, and with \( \Phi_0 = I_n \), the identity matrix of order \( n \),
- \( \omega_t \) is an \( n \)-vector of serially uncorrelated disturbances, \( \omega_t \sim N(0, \Sigma) \),
- \( t = p, p+1, \ldots, T \). Normality is not essential. \( \text{det}(\Sigma) \neq 0 \), so that identities and strict linear dependencies among the equations are excluded in (2.1).
- \( S(L) \) is a filter that preserves only the variation of interest in \( (w(x_t) - g(t) - h_t(L)\eta_t) \),
- \( w(x_t) \) is a nonlinear function of \( x_t \),
- \( g(t) \) is a deterministic or perfectly predictable function of time,
- \( V(L) \) and \( U(L) \) are real matrix lag polynomials of orders \( p_v \) and \( p_u \) with the roots of \( \text{det}(V(z)) = 0 \) and \( \text{det}(U(z)) = 0 \) outside the unit circle,
- \( M(L) \) is a real diagonal matrix with the roots of \( \text{det}(M(z)) = 0 \) on the unit circle,
- \( L \) is the lag operator: \( L^i y_t = y_{t-i}, \quad i \in \mathbb{Z} \), and
- \( \eta_t \) is an \( n \)-vector of disturbances different from \( \omega_t \), influencing \( y_t \) via \( h_t(L) \).