Contents

Chapter 1

Introduction into Time Series

1.1 The Nature of the Time Series Model

In this section we shall outline some of the basic issues in time series analysis. The discussion will be sketchy, but our intent is just to get an idea of where we are going. A *pure time series model* is a statistical model that explains a time series (or a vector process) by its *past history alone*. No other variables serve as explanatory variables. As such, these models are atheoretical, and are often employed as forecasting tools. The goal is to extract the **time dependence** in the data.
The emphasis here is on the fact that time series are **time dependent processes.**

For example, consider the $1st$ order autoregressive process, $AR(1)$, with drift:

$$y_t - \mu \;=\; \phi(y_{t-1} - \mu) + \epsilon_t \tag{1.1}$$

This is a simple example of a stochastic process where the uncertainty is due to ϵ_t. We model y_t solely as a function of its lag value. Further, if we append the following assumption: $\epsilon_t \sim \text{NID}(0, \sigma^2)$ then if $\mid \phi \mid < 1$ the observations fluctuate around the uncodntional mean μ. Under these assumptions we may estimate *consistently* μ and ϕ by ordinary least squares (OLS).

1.1.1 Notes

1. **Stationarity**: By constraining ϕ in the interval $[-1, 1] \Rightarrow$ process is **stationary**$\Rightarrow$no tendency for the spread (variance) to increase or decrease

2. **Autoregressive models:** Could increase the order of the dependency $(y_{t-2}, y_{t-3}, \ldots)$. A common modeling strategy is simply to model y_t by adding more and more lag values **.**

3. **Forecasting** can easily be accomplished for this kind of time series model

   $$\tilde{y}_{T+1/T} = \tilde{\mu} + \tilde{\phi}(y_T - \tilde{\mu}) \tag{1.2}$$

   where $\tilde{y}_{T+1/T}$ is the forecast of y_{T+1} with information at time T (i.e. a conditional forecast) as $\tilde{\phi} \to 1$ more weight given to y_T in forecasting

(a) **Moving-Average Error (MA) models** If we allow dependencies to enter the time series model from the ϵ_t

$$y_t - \mu = \epsilon_t + \theta\epsilon_{t-1} \qquad MA(1) \tag{1.3}$$

(b) **ARMA**: Combining autoregressive and moving average error models

$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t + \theta\epsilon_{t-1} \qquad \text{ARMA}(1,1)$$

4. **Unit Roots** (Chapter 10) : Stationarity is something that is crucial in the modeling exercise, if $\phi = 1$ we say the variable y_t has a **unit root** or is **integrated of order 1, I(1)**

$$y_t = y_{t-1} + \epsilon_t$$

5. **Trends:** Currently there is some discussion in the macro econometric literature over whether economic time series are best modeled as stationary in differences (*i.e.* there is a unit root of some order in the process) or whether the variables are stationary about some trend

$$y_t = \alpha + \beta t + \epsilon_t \qquad t = 1, \ldots T$$

6. Unit Root and Trend Stationary Variables: of course, there is nothing stopping us from combining the two models:

$$y_t = \alpha + \beta t + y_{t-1} + \epsilon_t \tag{1.4}$$

7. **ARIMA:** We may in general incorporate the integrated nature of time series in the (non-seasonal) autoregressive moving average models $\Rightarrow$ ARIMA(p, d, q)

   - p is the order of the autoregressive process
   - d is the amount of differencing to achieve stationarity
   - q is the order of the moving average process

8. **Cointegration (Chapter 11)**: Take I(1) processes y_t and $(1 \times k)$ x_t^T such that

$$y_t = x_t^T\theta + u_t$$

and u_t is **stationary** $ARMA$. This is said to describe a cointegrated relation.

1.2 Univariate Analysis

The analysis of time series as $ARIMA$ models is identified with Box-Jenkins analysis (Chapter 3). We can have seasonal $ARIMA$ and also incorporate dummies, trends and other deterministic variables.

Three parts to traditional Box-Jenkins analysis:

1. **Identification** (Finding a candidate model)
   (a) determining the order of differencing required for stationarity (d)
   (b) determining the order of $ARMA(p, q)$ on the differenced process
2. **Estimation** (Chapter 4)
   (a) typically by maximizing a likelihood function (the moving average part implies that this will be a nonlinear operation).
3. **Diagnostic Testing and Forecasting** (Chapters 5 and 6)

Once the model is identified several diagnostic checks and model selection methods are applied. Often the intent in these kinds of exercises is ultimately forecasting (indeed sometimes the absence of economic model building is justified on the basis of a forecasting motive).

We will see that the concepts and techniques developed for the pure time series model are often applied in our more conventional econometric models. In fact to encompass these kinds of models, there are the so-called $ARMAX$ models -where the X part refers to the economic variables that we might wish to include (or $ARIMAX$ to include integrated processes). This year we will also study fractional process that allow long memory without the variable being nonstationary.

1.3 Multivariate Time Series

We may model vector processes in much the same way; we introduce not only own lag variables on the right hand side of the equation but also lags of other **related** variables. Hence there is a **system** of equations called **vector ARMA** models and are discussed at length in Chapter 7. These are usually quite difficult to estimate and only special cases are tackled in the applied literature.

The most popular being **Vector Autoregressive models** (VAR): Y_t is now an $(N \times 1)$ vector and form a system of equations, say a $VAR(1)$,

$$Y_t = \Phi Y_{t-1} + \varepsilon_t \tag{1.5}$$

$$E\left[\varepsilon_t \varepsilon_t^T\right] = \Omega \tag{1.6}$$

Multivariate time series models are really the domain of economists and programs like STATA have only recently got into the game.

1.4 Estimation

Estimation of the times series models in STATA/RATS/TSP is typically by **maximum likelihood.**

1. We shall see for strictly AR processes, this can be equivalent to OLS estimation

2. Estimation of the MA part is nonlinear and requires more sophisticated programming using numerical optimization (procedure in STATA has many nonlinear options whereas RATS/TSP is Gauss- Newton).

1.4.1 Maximum Likelihood Estimation (MLE)

The joint density function for observations $y_1, \ldots, y_T$ is assumed to depend upon n unknown parameters in the vector $\Psi = (\psi_1, \ldots, \psi_n)^T$. Note that T is transpose which should be obvious from the context to separate it from the number of observations also denoted by T.

Denote the *joint density* in T observations:

$$L(y_1, \ldots y_T; \Psi) \tag{1.7}$$

1. The *likelihood* is a reinterpretation of the density, as a function of the parameters Ψ given the observed or drawn sample . i.e. given the data $y_1 \ldots y_T$, we consider different values for the Ψ

$$L(\Psi; y_1, \ldots, y_T) \text{ or } L(\Psi)$$

2. Object of MLE is to maximize $L(\Psi)$ or more commonly the log likelihood (a monotonic transform of the likelihood): $\log(L)$ w.r.t. to choice over Ψ.

3. The maximum likelihood estimator is obtained by setting the **score** $\frac{\partial logL}{\partial \Psi}$ to zero:

$$\frac{\partial logL}{\partial \Psi} = 0 \qquad \} \text{likelihood equations} \tag{1.8}$$

4. Likelihood equations are typically nonlinear in the parameters which will require an iterative procedure like Gauss-Newton.

5. Even though the models we have dealt with previously have been the fixed X with iid support (or some form of heteroskedasticity or autoregressive errors), all the desirable properties of MLE obtain for these kinds of time series models given our assumptions.

1.4.2 Properties for MLE

Let $\tilde{\Psi} = \{\tilde{\Psi}_1, \ldots, \tilde{\Psi}_n\}$ be the MLE estimate

$$\sqrt{T}(\tilde{\Psi} - \Psi) \underset{\sim}{d} N(0, T\ Avar) \tag{1.9}$$

where **AVar** is the asymptotic variance-covariance matrix obtained in the usual way from the probability limit of the inverse **information matrix (IA)**:

$$\mathbf{IA}(\Psi) = p\lim T^{-1}\{-\frac{\partial^2 log(L)}{\partial \mathbf{\Psi} \partial \mathbf{\Psi}^T}\}. \tag{1.10}$$

that is

$$\mathbf{Avar}(\tilde{\Psi}) = T^{-1}\mathbf{IA}(\Psi)^{-1} \tag{1.11}$$

In practice, we evaluate the **Avar** at the MLE : $\tilde{\Psi}$. This is the estimates of the **Avar** produced by the program STATA/RATS/TSP (there are options to use numerical or analytical derivatives).

1.4.3 Notes on MLE

1. Review maximum likelihood estimation in the classical linear regression model. There are many good references for this (Davidson and MacKinnon or Greene will do).

2. With an autoregressive error terms u in

   $$y = X\beta + u$$

   the estimation procedure in the generalized regression model is generalized least squares (GLS). The same basic principle is followed when the error structure is MA.

3. Suppose we have normal support $u \sim N(\mathbf{0}, \Omega)$, then the likelihood is:

   $$logL = -\frac{T}{2}log(2\pi) - \frac{T}{2}log \mid \Omega \mid -\frac{1}{2}u^T\Omega^{-1}u \tag{1.12}$$

4. Maximization of the likelihood is done by an iterative procedure. We will study the numerical optimization problem (usually some gradient method) in some detail in Chapter 4. The basic iterative procedure is in terms of the parameter updates is:

   $$\Psi^{i+1} = \Psi^i + \underbrace{\left[I^i(\Psi^i)\right]^{-1}\{\frac{\partial logL}{\partial \Psi}\}} \tag{1.13}$$

   where all the right hand-side of equation (1.13) is evaluated at Ψ^i. Note that $I^i(\Psi)$ is the information matrix and $\frac{\partial logL}{\partial \mathbf{\Psi}}$ is the *gradient (score)* also evaluated at $\mathbf{\Psi}^i$

5. Sometimes minus the Hessian (the second derivative of the log likelihood with no restrictions imposed; the plim of this is the information matrix) is used.

6. Most optimization schemes are based on this or some similar algorithm and iteration proceeds until some form of convergence (which may be stated in terms of a number of things).

7. To start the algorithm we need to supply initial estimates (STATA/RATS/TSP) have default values and the programmer can change these).

8. As we shall see, many efficient optimization routines have variable step procedures.