Econometrics 2, Fall 2004
Heino Bohn Nielsen
September 24, 2004

INTRODUCTION TO TIME SERIES

Abstract: This note introduces the concept of time series data. First we give some basic definitions and discuss the differences between cross-sectional data (analyzed in Econometrics 1) and time series data. We then say a few words on time dependence, which is a characteristic feature of time series; and illustrate how the basic results for the linear regression model are modified in a dynamic context, where the lagged dependent variable is introduced as a regressor. Finally, we discuss the problem of serial dependence in the error term.

1 Time Series Data

Most data in macroeconomics and finance come in the form of time series. A time series is a set of observations

$$y_1, y_2, ..., y_t, ..., y_T, \tag{1}$$

where the index t represents time such that the observations have a natural temporal ordering. In most time series models in economics, and throughout this course, the variable is observed with fixed intervals, such that the distances between successive time points are constant. As an example you could think of the gross domestic product (GDP) compiled at a quarterly frequency, or a money market interest rate recorded at a daily frequency.

The main assumption underlying *time series analysis* is that the observation at time t, y_t, is a realization of a random variable, y_t. Note that the standard notation does not distinguish between the random variable and a realization. Taken as a whole, the observed time series in (1) is a realization of a sequence of random variables, y_t, $t = 1, 2, ..., T$, often referred to as a *stochastic process*.

Here we notice an important difference between cross-section data and time series data. Recall, that in the cross section case we think of a data set, x_i, $i = 1, 2, ..., N$, as being sampled as N independent draws from a large population; and if N is sufficiently large we can characterize the distribution, e.g. by estimating the mean and variance. In the time series context, on the other hand, we are faced with T random variables, y_t, $t = 1, ..., T$, and only one realization from each. In general, therefore, we have no hope of characterizing the distributions corresponding to each of the random variables, unless we impose additional restrictions. Graph (A) in Figure 1 illustrates the idea of a general stochastic process, where the distributions differ from time to time. It is obvious that based on a single realized time series we cannot say much about the underlying stochastic process.

A realization of a stochastic process is just a sample path of T real numbers; and if history took a different course we would have observed a different sample path. If we could rerun history a number of times, M say, we would have M realized sample paths corresponding to different states of nature. Letting a superscript (m) denote the realizations, $m = 1, 2, ..., M$, we have M observed time series

$$\begin{array}{lcccccc} \text{Realization } 1: & y_1^{(1)}, & y_2^{(1)}, & ..., & \boxed{y_t^{(1)},} & ..., & y_T^{(1)} \\ & \vdots & \vdots & & \vdots & & \vdots \\ \text{Realization } m: & y_1^{(m)}, & y_2^{(m)}, & ..., & y_t^{(m)}, & ..., & y_T^{(m)} \\ & \vdots & \vdots & & \vdots & & \vdots \\ \text{Realization } M: & y_1^{(M)}, & y_2^{(M)}, & ..., & y_t^{(M)}, & ..., & y_T^{(M)}. \end{array} \tag{2}$$

For each point in time, t, we would then have a cross-section of M random draws, $y_t^{(1)}, ..., y_t^{(m)}, ..., y_t^{(M)}$, from the same distribution. This cross-section is not drawn from a fixed population, but is drawn from a hypothetical population of possible outcomes. This hypothetical population corresponds to a particular distribution in graph (A). Often we are interested in the *unconditional* mean, $E[y_t]$, which we would estimate with the sample average

$$\widehat{E}[y_t] = \frac{1}{M} \sum_{m=1}^{M} y_t^{(m)}. \tag{3}$$

The cross-sectional mean in (3) is sometimes referred to as the *ensemble mean* and it is the mean of a particular distribution in Figure 1. This is fundamentally different from the *time average* of a particular realized sample path, e.g.

$$\overline{y} = \frac{1}{T} \sum_{t=1}^{T} y_t^{(1)}, \tag{4}$$

Notice, that when we analyze a single realization, we normally ignore the superscript and use the notation $y_t = y_t^{(1)}$.

Of course, it is not possible in economics to generate more realizations of history. But if the distribution of the random variable y_t remains unchanged over time, then we can think of the T observations, $y_1, ..., y_T$, as drawn from the same distribution; and we can make inference on the underlying distribution of y_t based on observations from different

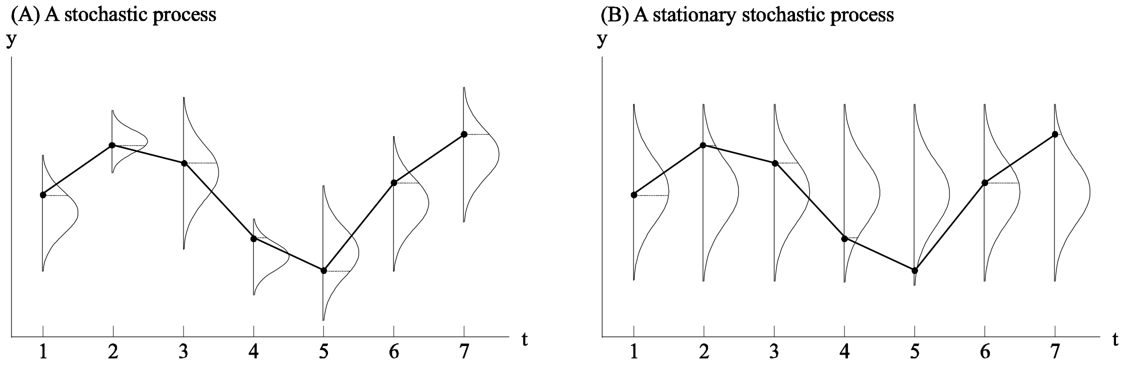

Figure 1: *Stochastic processes and realized time series.*

points in time. The property that the distribution of y_t is the same for all t is referred to as *stationarity*; and we will return to this issue several times during the course. Under stationarity the mean is constant, $E[y_t] = \mu$ for all t, and given an additional regularity condition the time average, $\overline{y}$, is a consistent estimator of $E[y_t]$.[1]

The idea of a stationary stochastic process is illustrated in graph (B) in Figure 1. Notice that the realized time series are identical in graph (A) and (B), and for a small number of observations it is often difficult to distinguish stationary from non-stationary time series.

2 Time Dependence

A characteristic feature of many economic time series is a clear dependence over time, and there are often non-zero correlations between observations at time t and $t-k$, for some lag k.

One way to characterize a stationary time series is by *the autocorrelation function* (ACF), defined as the correlation between y_t and y_{t-k}, i.e.

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{V}(y_t)}, \quad k = ..., -2, -1, 0, 1, 2, ...$$

The sample autocorrelations can be estimated by e.g.

$$\widetilde{\rho}_k = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T}(y_t - \overline{y})(y_{t-k} - \overline{y})}{\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2}$$

or

$$\widehat{\rho}_k = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T}(y_t - \overline{y})(y_{t-k} - \overline{y})}{\frac{1}{T-k}\sum_{t=k+1}^{T}(y_{t-k} - \overline{y})^2},$$

[1] The additional condition says that the observations y_t and y_{t+k} becomes approximately independent if we choose k large enough. This ensures that each new observation contains some new information on $E[y_t]$, and the estimator $\overline{y}$ is consistent as $T \to \infty$. A process with this property is called *ergodic*.

where $\overline{y}$ is the full-sample mean defined in (4)

The first estimator, $\widetilde{\rho}_k$, is the most efficient as it uses all the available observations in the denominator. For convenience it is sometimes preferred to discard the first k observations in the denominator and use the estimator $\widehat{\rho}_k$, which is just the OLS estimator in the regression model

$$y_t = a + \rho_k y_{t-k} + \text{residual}.$$

Positive autocorrelation is visible in a graph of a given time series as a tendency of a large observation, y_t, to be followed by another large observations, y_{t+1}, and vice versa. Negative autocorrelation is visible as a tendency of a large observation to be followed by a small observation.

2.1 Empirical Example

As an example of an actual time series, consider a quarterly time series for the natural log of the Danish GDP,

$$y_t = \log(\mathsf{GDP}_t), \quad t = 1971:2, ..., 2003:2,$$

in graph (A) in Figure 2. The time series has a clear positive linear trend, and by visual inspection it seems unlikely that the observations y_t could have been drawn from the same distribution for all t; and the underlying stochastic process is most certainly non-stationary. Graph (B) in Figure 2 shows the estimated autocorrelations, $\widehat{\rho}_k$, for the log of GDP. The ACF is declining very slowly, indicating a strong positive dependence between observations close to each other.

Next consider deviations of the log of Danish GDP from a linear trend. This is obtained as the residual, y_t^*, from a linear regression

$$y_t = \alpha_0 + \alpha_1 t + y_t^*.$$

The *detrended* series, $y_t^* = y_t - \widehat{\alpha}_0 - \widehat{\alpha}_1 t$, is given in graph (C). It is still doubtful whether the observations could correspond to the same underlying distribution, but to be sure we would have to apply a formal test for stationarity. We will return to the issue of stationarity tests later in the course. For the detrended series there is still a clear time dependence, and a high value of y_t^* is likely to be followed by a high value of y_{t+1}^*. This is the idea of positive autocorrelation, and the ACF in graph (D) illustrates a positive correlation between y_t^* and y_{t+k}^* for values of k up to 12.

Finally, consider the first difference of y_t, i.e.

$$\Delta y_t = y_t - y_{t-1} = \log\left(\frac{\mathsf{GDP}_t}{\mathsf{GDP}_{t-1}}\right) = \log\left(1 + \frac{\mathsf{GDP}_t - \mathsf{GDP}_{t-1}}{\mathsf{GDP}_{t-1}}\right) \simeq \frac{\mathsf{GDP}_t - \mathsf{GDP}_{t-1}}{\mathsf{GDP}_{t-1}},$$

where the last approximation is good if $\frac{\mathsf{GDP}_t - \mathsf{GDP}_{t-1}}{\mathsf{GDP}_{t-1}}$ is close to zero. The first difference of the log of a variable can therefore be interpreted as the relative growth rate; and changes in y_t are interpretable as percentage changes (divided by 100). Graph (E) shows the

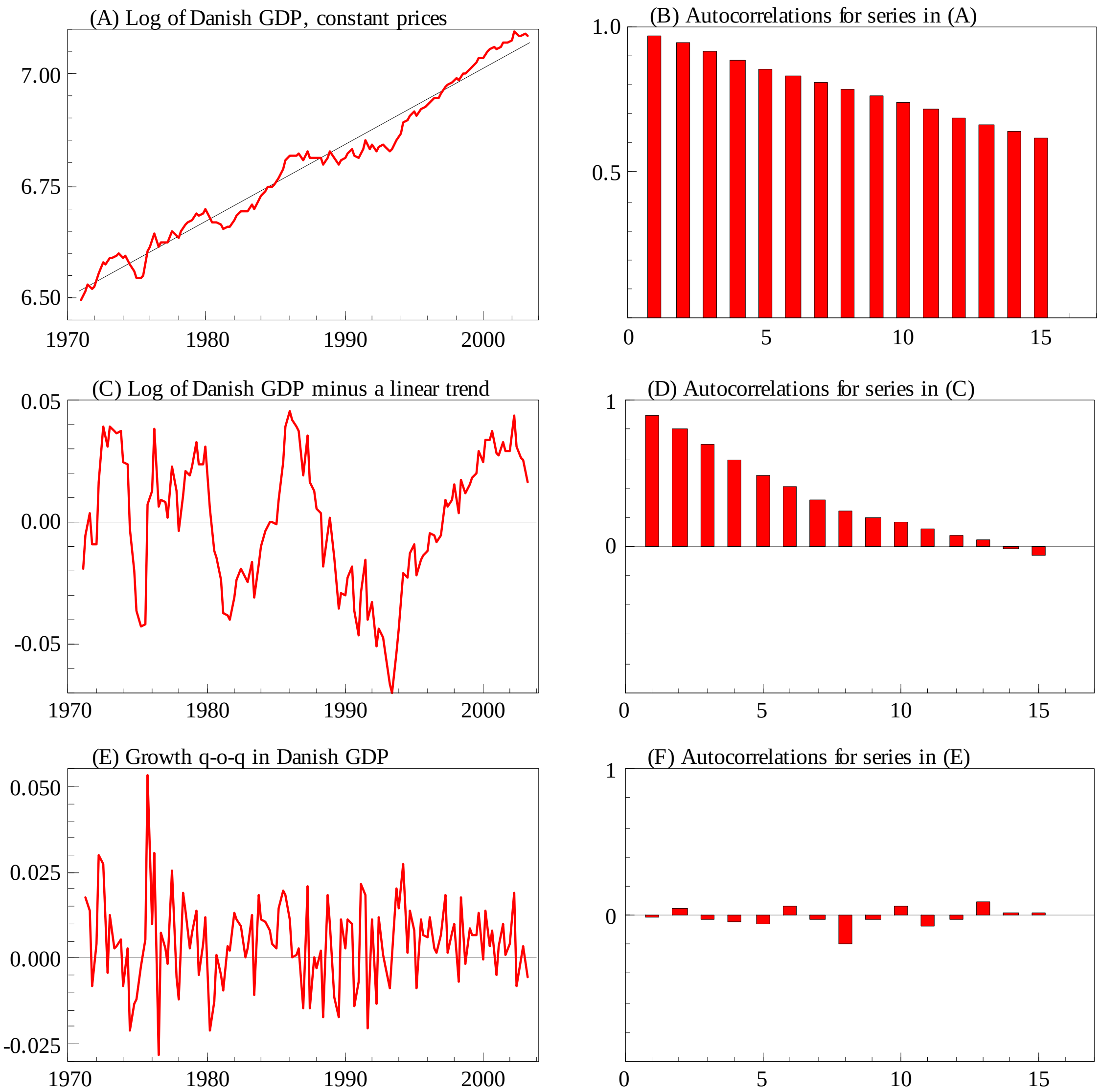

Figure 2: *The natural log of the Danish gross domestic product in levels, deviation from a linear trend, and in first differences. Actual time series and estimated autocorrelation functions.*

growth in GDP from quarter to quarter. For this time series it is not unlikely that the observations could have been drawn from the same distribution, i.e. that the underlying stochastic process is stationary. The time dependence is also much less pronounced, and the estimated autocorrelations are close to zero for all k, cf. the ACF in graph (F).

The different properties of the time series y_t, y_t^* and Δy_t demonstrate that although a given time series is non-stationary, some transformations of the series may be stationary. If the detrended series y_t^* is stationary while the original series is non-stationary the series y_t is called *trend stationary*. If, on the other hand, the first difference, Δy_t, is stationary while y_t is non-stationary, we refer to y_t as being *integrated of first order*, I(1), or *difference stationary*.

The example also serves to illustrate that due to the fixed ordering of the observations, graphical methods are often very useful in dealing with time series. A natural starting point for all time-series based empirical analysis is to look at time series graphs.

3 Linear Regression with Time Series Data

Since the observations in a time series have a temporal ordering, past events can be treated as given, or *predetermined*, in the analysis of current events. It is therefore natural to consider the properties of y_t given the past, e.g. the conditional expectation $E[y_t \mid y_1, y_2, ..., y_{t-1}]$. To formalize the time dependence consider the simple case of a first order dependence

$$E[y_t \mid y_1, y_2, ..., y_{t-1}] = \alpha y_{t-1}, \tag{5}$$

where the expectation of y_t conditional on the past is a function of y_{t-1} alone. The relation in (5) can be seen as a regression setup and we can consider the linear regression

$$y_t = \alpha y_{t-1} + \epsilon_t, \tag{6}$$

where ϵ_t is an identically and independently distributed (IID) error term. It follows by construction that the error term in (6) should satisfy the condition

$$E[\epsilon_t \mid y_1, y_2, ..., y_{t-1}] = 0, \tag{7}$$

which is a formal definition of predeterminedness of the explanatory variable in model (6).

The model (6) is referred to as a first order autoregressive, or AR(1), model, and we will analyze this model in more details later in the course. For the time being we will just consider (6) as a linear regression model to discuss how the dynamic structure of the time series modifies the properties of ordinary least squares (OLS) estimation.

3.1 Consistency

As discussed in Wooldridge (2003, p. 168–169) the assumption in (7) is sufficient for consistency of OLS, and it follows that OLS applied to the dynamic model (6) is consistent. To see this consider the usual manipulations of the OLS estimator

$$\widehat{\alpha} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \frac{\sum_{t=2}^{T} (\alpha y_{t-1} + \epsilon_t) y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \alpha + \frac{\frac{1}{T-1}\sum_{t=2}^{T} \epsilon_t y_{t-1}}{\frac{1}{T-1}\sum_{t=2}^{T} y_{t-1}^2}. \tag{8}$$

The sums run from $t = 2, ..., T$ because we have to condition on the first observation, y_1, the so-called *initial value*. We want to look at the behavior of the last term as $T \to \infty$, i.e. as the number of observations (in the time series dimension) increases. This is different from the cross-sectional case, where the asymptotic results were derived for increasing number of individuals, N.

First we assume that the denominator has a finite limit for $T \to \infty$, i.e.

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T-1} \sum_{t=2}^{T} y_{t-1}^2 = q, \tag{9}$$

for $0 < q < \infty$. This requirement says that the limit of the second moment should be positive and finite. In our example $E[y_t] = 0$ and the expression in (9) is the limit of the variance of y_t as $T \to \infty$. For a stationary process, i.e. where the distribution is the same for all t, the requirement in (9) is satisfied.

For the numerator, it follows from the law of large numbers that

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T-1} \sum_{t=2}^{T} \epsilon_t y_{t-1} = E[\epsilon_t y_{t-1}] = 0,$$

where the last equality follows from (7)[2]. Combining the results, we obtain

$$\operatorname*{plim}_{T\to\infty} \widehat{\alpha} = \alpha + \frac{\operatorname{plim} \frac{1}{T-1} \sum_{t=2}^{T} \epsilon_t y_{t-1}}{\operatorname{plim} \frac{1}{T-1} \sum_{t=2}^{T} y_{t-1}^2} = \alpha + \frac{0}{q} = \alpha,$$

which shows the consistency of OLS.

3.2 Small Sample Bias

Now we consider the finite sample properties of OLS in the dynamic model. It turns out, that we cannot show unbiasedness of OLS in (6). Remember from Econometrics 1 that in order to show unbiasedness we made the assumption of *strict exogeneity*, i.e.

$$E[\epsilon_t \mid y_1, y_2, ..., y_T] = 0, \tag{10}$$

which implies a zero correlation between ϵ_t and past, current and future values of y_t. This assumption cannot be made for dynamic models. We have that ϵ_t is uncorrelated with lagged values of the explanatory variables, $y_{t-1}, y_{t-2}, ..., y_1$, but since y_t is function of ϵ_t it is clear that ϵ_t cannot be uncorrelated with current and future values of the explanatory variables, i.e. $y_t, y_{t+1}, ..., y_T$.

It can be shown that the OLS estimator of the autoregressive coefficient in (6) is biased towards zero. The derivation of the bias is technically demanding and instead we use a Monte Carlo simulation to illustrate the idea.

As a data generating process (DGP) we use the specific AR(1) model

$$y_t = \alpha y_{t-1} + \epsilon_t, \quad \alpha = 0.9, \quad \epsilon_t \sim N(0,1), \quad t = 1, 2, ..., T, \tag{11}$$

[2] As you can see, the assumption needed for consistency is that of no contemporaneous correlation, $E[\epsilon_t y_{t-1}] = 0$. The assumption in (7) is a little stronger than that and implies that $E[\epsilon_t \cdot f(y_{t-1})] = 0$ for all functions $f(\cdot)$. More discussion on the differences between the assumptions can be found in Hayashi (2000).

and generate $M = 5000$ time series for the sample length T. The value $\alpha = 0.9$ of the autoregressive parameter corresponds to a positive autocorrelation in the same magnitude as for the detrended log of Danish GDP in graph (C) in Figure 2.

The Monte Carlo simulation is a formalization of the idea in (2), and we use the generating process in (11) to construct M realizations of the stochastic process. For each generated time series, $y_1^{(m)}, y_2^{(m)}, ..., y_T^{(m)}$, $m = 1, 2, ..., M$, we apply OLS to the regression model (6) (which is identical to the DGP) and get the OLS estimate $\widehat{\alpha}_m$, i.e.

$$\begin{array}{lccccccc}
\text{Realization } 1: & y_1^{(1)}, & ..., & y_t^{(1)}, & ..., & y_T^{(1)} & \text{gives estimate} & \widehat{\alpha}_1 \\
 & \vdots & & \vdots & & \vdots & & \\
\text{Realization } m: & y_1^{(m)}, & ..., & y_t^{(m)}, & ..., & y_T^{(m)} & \text{gives estimate} & \widehat{\alpha}_m \\
 & \vdots & & \vdots & & \vdots & & \\
\text{Realization } M: & y_1^{(M)}, & ..., & y_t^{(M)}, & ..., & y_T^{(M)} & \text{gives estimate} & \widehat{\alpha}_M.
\end{array}$$

To illustrate the bias we calculate the average estimate

$$\mathsf{MEAN}(\widehat{\alpha}) = \frac{1}{M}\sum_{m=1}^{M}\widehat{\alpha}_m, \tag{12}$$

and also the bias, $\mathsf{BIAS}(\widehat{\alpha}) = \mathsf{MEAN}(\widehat{\alpha}) - \alpha$. The uncertainty of $\widehat{\alpha}$ can be measured by the standard deviation of $\widehat{\alpha}_m$ across the M replications. This is denoted the *Monte Carlo standard deviation*,

$$\mathsf{MCSD}(\widehat{\alpha}) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\widehat{\alpha}_m - \mathsf{MEAN}(\widehat{\alpha}))^2}.$$

Notice, that $\mathsf{MEAN}(\widehat{\alpha})$ is itself an estimator, and the uncertainty related to the estimator can be measured by the *Monte Carlo standard error*, defined as

$$\mathsf{MCSE} = M^{-\frac{1}{2}} \cdot \mathsf{MCSD}(\widehat{\alpha}).$$

Be aware of the important difference between the $\mathsf{MCSD}(\widehat{\alpha})$, which is a measure of the uncertainty of $\widehat{\alpha}$, and the MCSE, which is a measure of the uncertainty of the estimator $\mathsf{MEAN}(\widehat{\alpha})$ in the simulation. The latter converges to zero for an increasing number of replications, $M \to \infty$.

The results from PcNaive are reported in graph (A) of Figure 3 for sample lengths $T \in \{10, 15, ..., 100\}$. The mean is estimated as in (12), and the dotted lines are

$$\mathsf{MEAN}(\widehat{\alpha}) \pm 2 \cdot \mathsf{MCSD}(\widehat{\alpha}).$$

These lines are interpretable as the average 95% confidence band in each replication and *not* the 95% confidence band for the estimated $\mathsf{MEAN}(\widehat{\alpha})$. The mean of the estimated parameter is lower than the true value for all sample lengths. For a very small sample

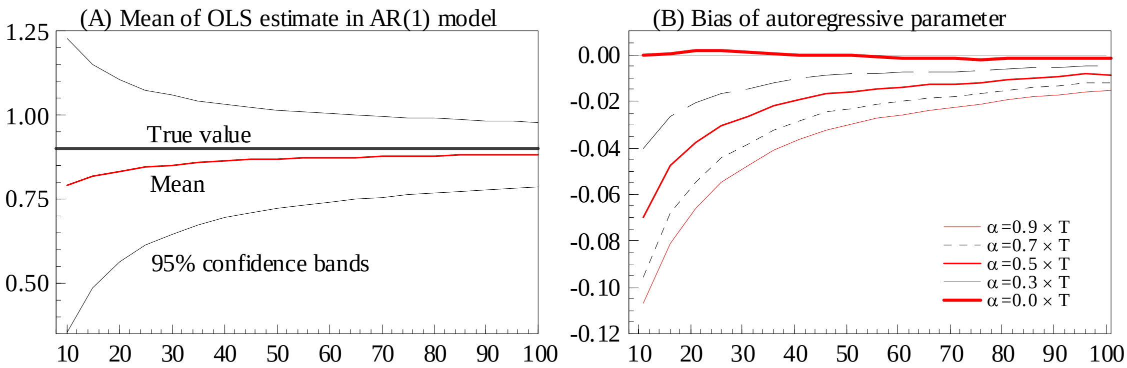

Figure 3: *Monte Carlo results. (A) illustrates the mean of the OLS estimate $\widehat{\alpha}$ of α in an AR(1) model. The true value in the DGP is $\alpha = 0.9$. (B) illustrates the bias $\widehat{\alpha} - \alpha$ for different values of α in the DGP. Based on 5000 replications.*

length of $T = 10$ the OLS estimator has a mean of $\mathsf{MEAN}(\widehat{\alpha}) = 0.7935$. The $\mathsf{MCSD}(\widehat{\alpha}) = 0.2172$ which implies that the Monte Carlo standard error of this estimate is $\mathsf{MCSE} = 5000^{-\frac{1}{2}} \cdot 0.2172 = 0.0031$. A $t-$test for the hypothesis that the estimate is unbiased, i.e.

$$\mathcal{H}_0 : \mathsf{MEAN}(\widehat{\alpha}) = 0.9,$$

can be constructed as

$$\tau = \frac{\mathsf{MEAN}(\widehat{\alpha}) - 0.9}{\mathsf{MCSE}} = \frac{0.7935 - 0.9}{0.0031} = -34.67,$$

which is clearly significant in the asymptotic $N(0, 1)$ distribution. We conclude that the bias is significantly different from zero. For larger samples the average converges to the true value, cf. the consistency of OLS in the dynamic model.

To illustrate how the bias depends on the true autoregressive parameter, we redo the exercise with other positive values $\alpha \in \{0, 0.3, 0.5, 0.7, 0.9\}$. The obtained results for the bias are depicted in graph (B). If the true DGP is static, $\alpha = 0$, the estimate is unbiased. If $\alpha > 0$ the estimator is downward biased, with a bias that increases with α.

4 Regression Models and Residual Autocorrelation

Until now we have considered time series models for a univariate time series, y_t. In this section we extend the framework to include regression models for time series data, e.g.

$$y_t = x_t'\beta + \epsilon_t, \quad t = 1, 2, ..., T, \tag{13}$$

where x_t is $K \times 1$ vector of explanatory variables and β is a $K \times 1$ vector of coefficients. Alternatively, we can stack the T equations in (13) using the matrix notation

$$Y = X\beta + \epsilon, \tag{14}$$

where $Y = (y_1, y_2, .., y_T)'$ is a $T \times 1$ vector, $X = (x_1, x_2, ..., x_T)'$ is the $T \times K$ stacked matrix of explanatory variables, and $\epsilon = (\epsilon_1, \epsilon_2, ..., \epsilon_T)'$ is a $T \times 1$ vector of error terms.

Using the matrix formulation in (14), the OLS estimator $\widehat{\beta}$ of β is given by

$$\widehat{\beta} = (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon,$$

and the variance of $\widehat{\beta}$ can be found as

$$\begin{aligned} \mathrm{V}\left[\widehat{\beta} \mid X\right] &= E\left[\left(\widehat{\beta} - \beta\right)\left(\widehat{\beta} - \beta\right)' \mid X\right] \\ &= E\left[(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1} \mid X\right] \\ &= (X'X)^{-1}X'\Omega X(X'X)^{-1}. \end{aligned} \tag{15}$$

Remember that a standard assumption used for deriving the variance of the OLS estimators is

$$E\left[\epsilon\epsilon' \mid X\right] = \sigma^2 I_T,$$

which implies that (15) reduces to the well-known expression

$$\mathrm{V}\left[\widehat{\beta} \mid X\right] = \sigma^2(X'X)^{-1}. \tag{16}$$

In models for cross-sectional data, an important issue is *heteroscedasticity.* That was discussed in Econometrics 1. Heteroscedasticity is the case where $E\left[\epsilon\epsilon' \mid X\right]$ is a diagonal matrix but where the diagonal elements are not identical, i.e.

$$E\left[\epsilon\epsilon' \mid X\right] = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_T^2 \end{pmatrix}.$$

Recall, that even in the case of heteroscedasticity, OLS is unbiased under the assumption in (10) and consistent under the assumption in (7). The OLS formula for the variance (16) is no longer correct, however, and inference in the OLS model should be based on the *heteroscedasticity consistent standard errors* constructed from an an estimate of (15). Also recall, that a more efficient estimator is given by generalized least squares (GLS) estimator that weights the observations according to the magnitude of the variances.

4.1 Autocorrelation of the Error Term

In time series models heteroscedasticity could still be an issue, but the problem is likely to be smaller. Instead we are often concerned with another type of misspecification denoted *autocorrelation* (of the error term). That is the case where consecutive error terms are correlated, e.g. $\mathrm{Cov}(\epsilon_t, \epsilon_{t-1}) \neq 0$. This implies that some off-diagonal elements in $E\left[\epsilon\epsilon' \mid X\right]$ are different from zero. The consequences of autocorrelation have similarities with the consequences of heteroscedasticity.

Firstly, OLS is still consistent under the assumption in (7). This implies that if the explanatory variables, x_t, are contemporaneously uncorrelated with the error term in the model (13), OLS is consistent even if the errors are autocorrelated. It is important to realize, however, that assumption (7) is not valid for a model with a lagged dependent variable and autocorrelated errors. To see this, consider an AR(1) model like (6), and assume that the error term is first order autocorrelated, i.e. that ϵ_t follows a first order autoregressive model

$$\epsilon_t = \rho\epsilon_{t-1} + v_t, \tag{17}$$

where v_t is IID with a constant variance. Consistency requires that $E[\epsilon_t y_{t-1}] = 0$. But this is clearly not satisfied since both y_{t-1} and ϵ_t depends on ϵ_{t-1}. As a result, *OLS is not consistent in a model with a lagged dependent variable and autocorrelated errors.*

Secondly, even if OLS is consistent, the OLS formula for the variance of the estimators (16) is no longer valid. In the spirit of White's heteroscedasticity robust standard errors, it is possible to find a consistent estimate of the matrix $X'\Omega X$ in (15) under autocorrelation, and that can be used to construct the so-called *heteroscedasticity-and-autocorrelation-consistent* (HAC) standard errors. This is discussed very briefly in Verbeek (2004, section 4.10.2). A simpler discussion of the univariate case is given in Stock and Watson (2003, p. 504-507).

Finally, the OLS estimator is no longer BLUE, and we can construct more efficient estimators. As an example consider the regression model (13) under first order autocorrelation in the error term, i.e. (17). If ρ is known the regression model can be transformed to

$$(y_t - \rho y_{t-1}) = (x_t - \rho x_{t-1})'\beta + v_t, \tag{18}$$

where the error term v_t is serially uncorrelated. This is an analog to the idea of the GLS transformation applied in the case of heteroscedasticity. Note, however, that the transformation cannot be applied to the first observation, and OLS applied directly to the transformed model (18) is only an approximation to the GLS estimator, see the discussion in Verbeek (2004, p. 99-100). In practice, the parameter ρ is unknown but can be consistently estimated by running the regression (17) on the estimated residuals from (13).

It should be stressed, that the GLS transformation to remove autocorrelation is rarely used in modern econometrics, see the discussion below.

4.2 Testing for Autocorrelation

In all time series models, tests for no autocorrelation should be routinely applied. In modern econometrics, the most commonly used test for the null hypothesis of no autocorrelation is a so-called lagrange multiplier (LM) test, based on an auxiliary regression. As an example we consider the test for no first order autocorrelation in the regression model

(13). This is done by running the auxiliary regression model

$$\widehat{\epsilon}_t = x_t'\delta + \gamma\widehat{\epsilon}_{t-1} + \xi_t,$$

where $\widehat{\epsilon}_t$ is the estimated residual from (13) and ξ_t is an error term. The null hypothesis of no autocorrelation corresponds to $\gamma = 0$, and can be tested by computing $T \cdot R^2$, where R^2 is the coefficient of determination in the auxiliary regression. Note, that the residual $\widehat{\epsilon}_t$ is orthogonal to the explanatory variables x_t, and any explanatory power in the auxiliary regression must be due to the included lagged residual, $\widehat{\epsilon}_{t-1}$. Under the null hypothesis the statistic is asymptotically distributed as

$$T \cdot R^2 \sim \chi^2(1). \tag{19}$$

Alternatively, an $F-$form of the test can be used as in PcGive.

Note, that the auxiliary regression needs one additional initial observation. It is customary to insert zeros in the beginning of the series of residuals, i.e. $\widehat{\epsilon}_1 = 0$, and estimate the auxiliary regression for the same sample that was used for the original estimation model.

While the test in (19) is valid asymptotically, i.e. for $T \to \infty$, an alternative test exists, which is in principal derived for finite samples; the so-called Durbin Watson (DW) test, see Verbeek (2004, section 4.7.2). This test is a part of the standard OLS output from most computer programs. The main problem with the DW test is that it is based on the assumption of strict exogeneity, i.e. assumption (10), which makes it invalid in most time series settings. And as autocorrelation is only a concern in time series models, the test is not particularly useful.

4.3 Autocorrelation and Dynamic Models

It is important to realize, that the error term in a regression model will automatically pick up the composite effect of everything not accounted for by the explanatory variables. Autocorrelation in the residuals can therefore indicate that the model is misspecified. Types of misspecifications that can produce residual autocorrelation include

(1) The model excludes an important and serially correlated variable.

(2) The true model is dynamic, but the lagged dependent variable is not included as a regressor.

(3) The functional form is misspecified, which produces a systematic behavior in the residuals.

(4) The model is subject to a structural break, e.g. a level shift.

A test for autocorrelation is therefore often interpreted as a broader test for misspecification of the regression model. The right thing to do if you find signs of residual autocorrelation is therefore not to make the GLS transformation but instead to reconsider and potentially change the model.

One starting point is to combine (13) and (17) to obtain the dynamic model

$$\begin{aligned} \epsilon_t &= \rho\epsilon_{t-1} + v_t \\ \left(y_t - x_t'\beta\right) &= \rho\left(y_{t-1} - x_{t-1}'\beta\right) + v_t \\ y_t &= \rho y_{t-1} + x_t'\beta - \rho x_{t-1}'\beta + v_t. \end{aligned}$$

The structure implies a number of restrictions on the parameters, the so-called *common factor restrictions*. These restrictions are implicitly imposed in the GLS estimation. In an empirical application, however, a finding of residual autocorrelation need not imply that the structure in (13) and (17) is correct, and the common factor restrictions are not necessarily valid. Instead of performing the GLS estimation, we can alternatively estimate the general dynamic model

$$y_t = \alpha_0 y_{t-1} + x_t'\alpha_1 + x_{t-1}'\alpha_2 + \eta_t, \tag{20}$$

where η_t is a new error term. Based on this model we can test the validity of the common factor restrictions. The model in (20) is denoted an *autoregressive distributed lag* (ADL) model, and we will return to the analysis of ADL models later in the course.

4.4 Empirical example

As an empirical example, consider a simple Phillips curve explaining the Danish inflation rate, INF, with the registered unemployment rate, UNR. The data are depicted in graph (A) of Figure 4 for the period 1971 : 2 − 2003 : 2. Running a simple regression yields

$$\mathsf{INF}_t = \underset{(0.00246)}{0.025441} - \underset{(0.03283)}{0.161705} \cdot \mathsf{UNR}_t,$$

with standard errors in parentheses. The actual values of inflation and the fitted values from the regression are depicted in graph (B). It is obvious from the graph that the unemployment rate is not able to explain all the features of inflation for the considered sample period. The residuals from this regression are depicted in graph (C). The residuals appear to be very systematic, indicating a positive autocorrelation. This is particularly true for the most recent years. Calculating the HAC standard errors yields 0.0041757 and 0.048617 respectively, which is much larger than the non-robust standard errors.

Denoting the residuals ϵ_t, we can make a LM test for no residual autocorrelation by running the regression

$$\epsilon_t = \underset{(0.002101)}{0.000344058} - \underset{(0.02808)}{0.00599805} \cdot \mathsf{UNR}_t + \underset{(0.07678)}{0.530869}\epsilon_{t-1}.$$

In this auxiliary regression, $R^2 = 0.275074$ and the LM statistic is given by $T \cdot R^2 = 129 \cdot 0.275074 = 35.485$, which is clearly rejected in a $\chi^2(1)$ distribution. We conclude that the residuals are positively autocorrelated.

Next consider the ADL model

$$\mathsf{INF}_t = \underset{(0.002869)}{0.0131526} + \underset{(0.08015)}{0.476504} \cdot \mathsf{INF}_{t-1} + \underset{(0.2469)}{0.356205} \cdot \mathsf{UNR}_t - \underset{(0.2478)}{0.441908} \cdot \mathsf{UNR}_{t-1}. \tag{21}$$

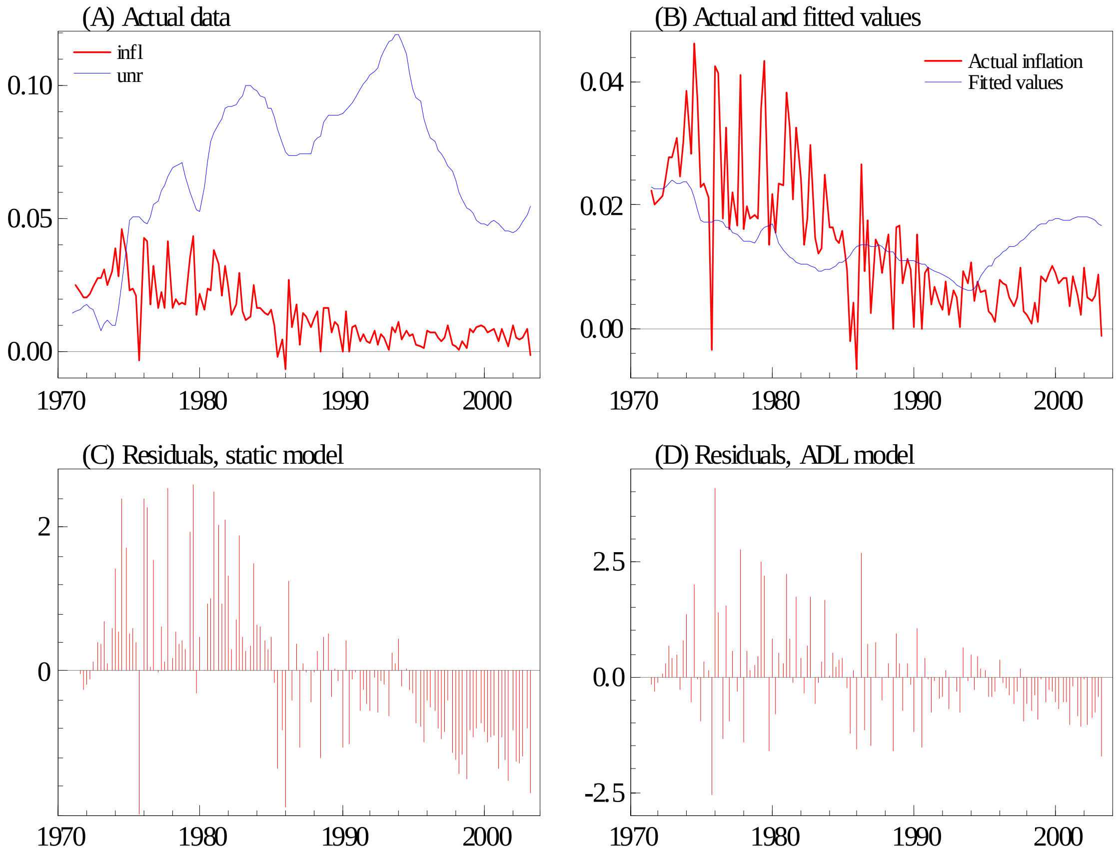

Figure 4: *Empirical example.*

The common factor restriction would imply that the product of the coefficients to INF_{t-1} and UNR_t should be identical to the coefficient to UNR_{t-1} with opposite sign. A Wald test for this hypothesis yields a test statistic of 3.494, corresponding to a $p-$value of 0.0640 in a F(1,124) distribution. In therefore seems to be a borderline case, whether the common factor restrictions are valid for the present case.

The estimated residuals from the ADL regression are illustrated in graph (C). Before 1995 the residuals appear non-systematic, but for the later years the residuals are systematically negative. This suggests, that the structure of the price determination or the labour marked has changed in recent years. An LM test for first order autocorrelation in the ADL regression yields a statistic of 14.125, which is still significant in a $\chi^2(1)$ distribution. This might suggest that the dynamics in the ADL model is not general enough and we could try to include more lags. Since the problems are related to the most recent period alone, however, a better solution would probably be to look for additional variables, that could explain the low inflation in the late 1990s in spite of a low unemployment rate. This last point illustrates that a rejected test for no autocorrelation is often an indication of general misspecification of the regression model and may suggest possible remedies.

References

[1] Hamilton, James D. (1994): *Time Series Analysis*, Princeton University Press.

[2] Hayashi, Fumio (2000): *Econometrics*, Princeton University Press.

[3] Stock, James H. and Mark W. Watson (2003): *Introduction to Econometrics*, Addison Wesley.

[4] Verbeek, Marno (2004): *A Guide to Modern Econometrics*, Second edition, John Wiley and Sons.

[5] Wooldridge, Jeffrey M. (2003): *Introductory Econometrics: A Modern Approach*, 2nd edition, South Western College Publishing.