INTRODUCTION TO ECONOMIC TIME SERIES

ECONOMETRICS C ♦ LECTURE NOTE 1
HEINO BOHN NIELSEN
FEBRUARY 6, 2012

Τhis note introduces some key concepts in time series econometrics. First, we present by means of examples some characteristic features of macroeconomic and financial time series. We then present some underlying assumptions on the way time series data are sampled and give a heuristic introduction to stochastic processes. We then present the definition of stationarity and discuss how a time series which is not stationary can sometimes be made stationary by means of simple transformations. Stationarity turns out to be an important requirement for the usual results from (cross-sectional) regression analysis to carry over to the time series case; and the distinction between stationary and non-stationary time series will be central throughout the course.

OUTLINE

1 FEATURES OF ECONOMIC TIME SERIES

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. Most time series in economics–and all time series considered in this course–are observed with fixed intervals, such that the distances between successive time points, t and $t + 1$, are constant. This section presents some characteristic features of economic time series.

Time Dependence. One characteristic feature of many economic time series is a clear dependence over time, and there is often a non-zero correlation between observations at time t and $t - h$ for small values of h. As an example, Figure 1 (A) shows monthly data for the US unemployment rate in percent of the labour force, $1948 : 1 - 2005 : 7$. The development in unemployment is relatively sluggish, and a reasonable conjecture on the unemployment rate a given month (for example the most recent, $u_{2005:7}$) would be a function of the unemployment rate the previous month, $u_{2005:6}$. Often we say that there is a high degree of *persistence* in the time series.

The temporal ordering of the observations implies that the observation for $2005 : 6$ always precedes the observation for $2005 : 7$, and technically u_{t-1} is *predetermined* when u_t is generated. This suggests that the past of u_t can be included in the information set in the analysis of u_t, and a natural object of interest in a time series analysis would be a model for u_t given the past, e.g. the *conditional expectation*, $E[u_t \mid u_1, u_2, ..., u_{t-1}]$. Throughout it will be important to distinguish the conditional expectation and the *unconditional expectation*, $E[u_t]$.

Trends. Another characteristic feature shared by many time series is a tendency to be trending. As an example, consider in Figure 1 (B) the quarterly Danish productivity, $1971 : 1 - 2005 : 2$, compiled as the log of real output per hour worked. In this case the trend represents the underlying economic growth, and the slope of the trend is by and large constant and approximately equal to 0.5% from quarter to quarter.

Co-Movements. A third interesting feature is a tendency for some time series to move together. Graph (C) shows time series for the log of real disposable income, y_t, and the log of real consumption, c_t, for the private sector in Denmark, $1971 : 1 - 2005 : 2$. The movements from quarter to quarter sometimes differ markedly, but the slow underlying movements seem to be related. This suggests that there could be some sort of equilibrium relation governing the movements of consumption and income in the long run.

A stylized economic theory may suggest that while both y_t and c_t have an upward drift due to the growth in productivity and real wages, the savings rate, $y_t - c_t$, should
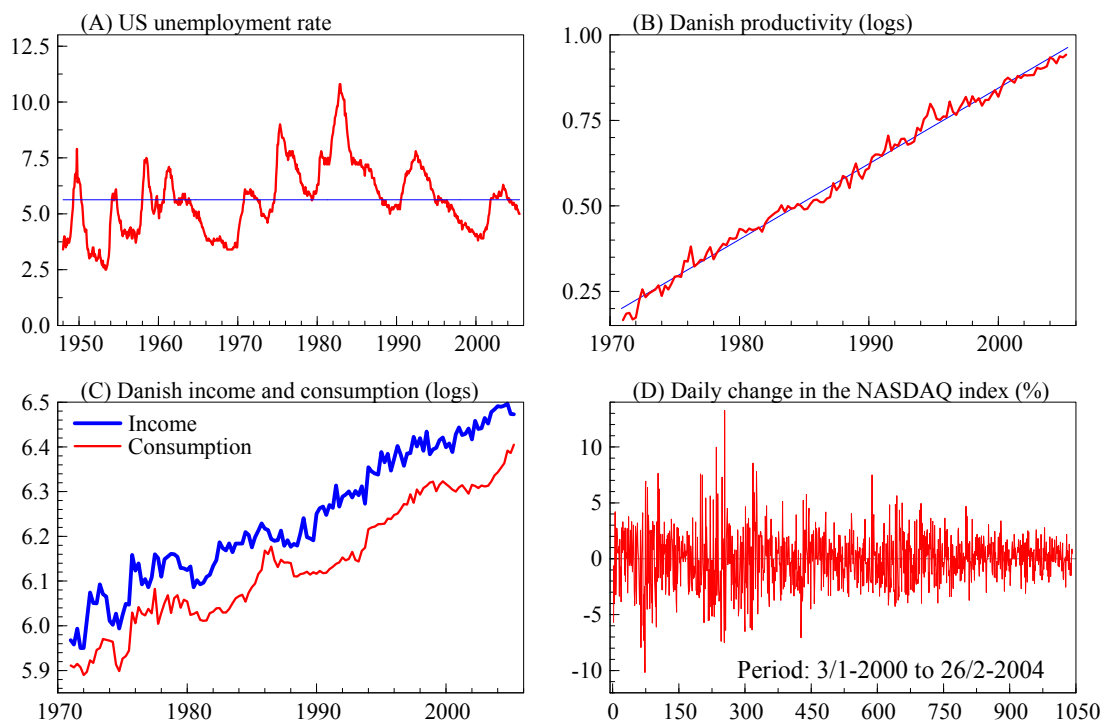
Figure 1: *Examples of macro-economic and financial time series.*

be stable. That implies that the short-term movements in the consumption and income series in graph (C) may differ but they are tied together in the longer run and do not move too far apart.

Volatility Clustering. For many financial time series, not only the mean but also the variance is of interest. This is because the variance is related to measures of the uncertainty or risk of a particular variable. A characteristic feature of this type of data is a tendency for the variance or *volatility* to change over time reflecting periods of rest and unrest on financial markets. Graph (D) shows percentage changes from day to day in the NASDAQ stock market index for the period January 3, 2000 to February 26, 2004. A visual inspection suggests that the volatility is much larger in some period than in others. This phenomenon is known as *volatility clustering*.

One primary goal of this course is to introduce the tools for analyzing time series with the characteristics presented in Figure 1. We want to construct measures of the dependence between successive observations, y_t and y_{t-1}, and to develop simple econometric models for the time dependence. We also want tools to analyze trending variables; in particular we want to address the dynamic interaction and the existence of equilibrium relations.

3

2 STOCHASTIC PROCESSES AND STATIONARITY

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 by the sample moments: e.g. the mean and variance. In the time series context, on the other hand, we are faced with T random variables, y_t $(t = 1, 2, ..., T)$, 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. Figure 2 (A) 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 would have M observed time series:

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

For each point in time, t, we would then have a cross-section of M random draws, $y_t^{(1)}, y_t^{(2)}, ..., y_t^{(m)}, ..., y_t^{(M)}$, from the same distribution. This cross-section is not drawn from a fixed population, but is drawn from a hypothetical population of possible outcomes, corresponding to a particular distribution in Figure 2 (A). Often we are interested in the *unconditional* mean, $E[y_t]$, which we could 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 2. This concept is fundamentally different
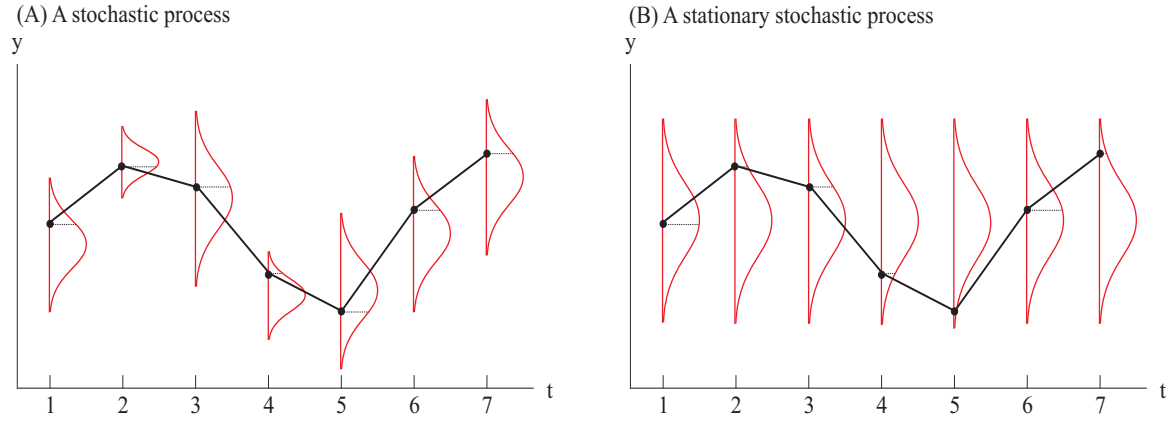
(A) A stochastic process

(B) A stationary stochastic process

Figure 2: *Stochastic processes and realized time series.*

from the *time average* of a particular realized sample path, e.g.

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

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

2.1 STATIONARITY

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 points in time. The property that the distribution of y_t is the same for all t is referred to as *stationarity*. Formally, the definition is given as follows:

DEFINITION 1 (STRICT STATIONARITY): *A time series, $y_1, y_2, ..., y_t, ..., y_T$, is strictly stationary if the joint distributions of s observations*

$$\left(y_t, y_{t+1}, ..., y_{t+s}\right) \quad and \quad \left(y_{t+h}, y_{t+1+h}, ..., y_{t+s+h}\right)$$

are the same for all integers h.

Strict stationarity implies that all structures and characteristics do not depend on the location on the time axis, and all moments are constant over time. Another concept focuses on the first two moments and requires only constancy of the mean and variance:

5

DEFINITION 2 (WEAK STATIONARITY): *A time series, $y_1, y_2, ..., y_t, ..., y_T$, is weakly stationary if*

$$
\begin{aligned}
E[y_t] &= & & \mu \\
V[y_t] &= & E[(y_t - \mu)^2] & = & \gamma_0 \\
Cov[y_t, y_{t-h}] &= & E[(y_t - \mu)(y_{t-h} - \mu)] & = & \gamma_h & \text{for } h = 1, 2, ...
\end{aligned}
$$

for all values of t.

Note that the mean, μ, and the variance, γ_0, are the same for all t and the covariance between y_t and y_{t-h}, denoted γ_h, only depends on the distance between the observations, h. Since the definition is related only to mean and covariances, a *weakly stationary* time series is sometimes referred to as *covariance stationary* or *second-order stationary*. In most cases it is also assumed that μ, γ_0, and γ_h are finite. When we assume stationarity in the rest of this course, we will take that to mean weak stationarity.

The idea of a stationary stochastic process is illustrated in Figure 2 (B). Here each new observation, y_t, contains information on the same distribution, and we can use all observations to estimate the common mean, μ. 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.

2.2   WEAK DEPENDENCE

Recall, that to show consistency of the estimators in a regression model we use the law of large numbers (LLN), stating that the sample average converges (in probability) to the population mean. And to derive the asymptotic distribution of the estimator, so that we can test hypotheses, we use the central limit theorem (CLT) stating that the appropriately normalized sample average converges (in distribution) to a normal distribution. In regression models for identically and independently distributed (IID) observations this is reasonably straightforward and the simplest versions of the LLN and CLT apply, see Wooldridge (2006, p. 774 ff.).

In a time series setting, where the IID assumption is unreasonable, things are more complicated. There exists more advanced versions of the LLN and the CLT, however, that allow the analysis of dependent observations, and in cases where such results apply, most of the results derived for regression models for IID data carry over to the analysis of time series. Two main assumptions are needed: The first important assumption is stationarity, which replaces the cross-sectional assumption of identical distributions. That assumption ensures that observations origin from the same distribution.

The second assumption is comparable to the assumption of independence, but is less stringent. In particular, the LLN and CLT can be extended to allow y_t to depend on y_{t-h}, but the dependence cannot be too strong. In particular, the additional condition says that the observations y_t and y_{t-h} becomes approximately independent if we choose h large enough. This is known as the assumption of *weak dependence* and it ensures that

each new observation contains some *new* information on $E[y_t]$. The interested reader is referred to Hayashi (2000, Section 2.2) for a more detailed discussion. Throughout this course we consider the assumption of weak dependence as a technical regularity condition, and all the stationary processes you will see in the course are also weakly dependent.

Under the assumption of stationarity and weak dependence the time average, $\overline{y}_T$, is a consistent estimator of the ensemble mean, $E[y_t] = \mu$; and most of the results from IID regression carry over to the time series case. If the assumptions fail to hold, the standard results from regression analysis cannot be applied; and *the most important distinction in time series econometrics is whether the time series of interest are stationary or not.* For the first part of this course we focus on stationary time series; but later we consider how the tools should be modified to allow for the analysis of non-stationary time series.

To illustrate the kinds of mechanisms that are at play, Figure 3 illustrates one realization of 200 IID observations in (A), a stationary stochastic process in (C), and a non-stationary stochastic process in (E). The right hand side column illustrates the LLN by showing the average of the first T observations, $\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t = \frac{1}{T}(y_1 + y_2 + ... + y_T)$, for increasing values of T. For IID observations the average clearly converges to the true zero mean. The same is the case for the dependent but stationary process, although the fluctuations are larger. Note, however, that the LLN does not apply to the non-stationary case. Here the time dependence is too strong and the average, $\overline{y}_T$, has no tendency to converge to zero.

3   STATIONARITY AND TIME DEPENDENCE

It follows from the definition, that a stationary time series should fluctuate around a constant level. For a stationary time series the level can be seen as the equilibrium value of y_t, while deviations from the mean, $y_t - \mu$, can be interpreted as deviations from equilibrium. In terms of economics, the existence of an equilibrium requires that there are some forces in the economy that pull y_t towards the equilibrium level. We say that the variable is *mean reverting* or *equilibrium correcting.* If the variable, y_t, is hit by a shock in period t so that y_t is pushed out of equilibrium, stationarity implies that the variable should revert to equilibrium, and a heuristic characterization of a stationary process is that a shock to the process y_t has only transitory effects.

The fact that a time series adjusts back to the mean, does not imply that the deviations from equilibrium cannot be systematic. Stationarity requires that the unconditional distribution of y_t is constant, but the distribution given the past, e.g. the distribution of $y_t \mid y_{t-1}$ may depend on y_{t-1} so that y_t and y_{t-1} are correlated. Another way to phrase this is in terms of forecasts: You may think of the expectation $E[y_t \mid I_t]$ as the best forecast of y_t given the information set, I_t. In this respect the mean of an unconditional distribution in Figure 2 (B) corresponds to the best forecast based on an empty information set, $I_t = \varnothing$, i.e. in a situation where we have not observed the history of the process, $y_{t-1}, y_{t-2}, ..., y_1$. If y_t has a tendency to fluctuate in a systematic manner and we
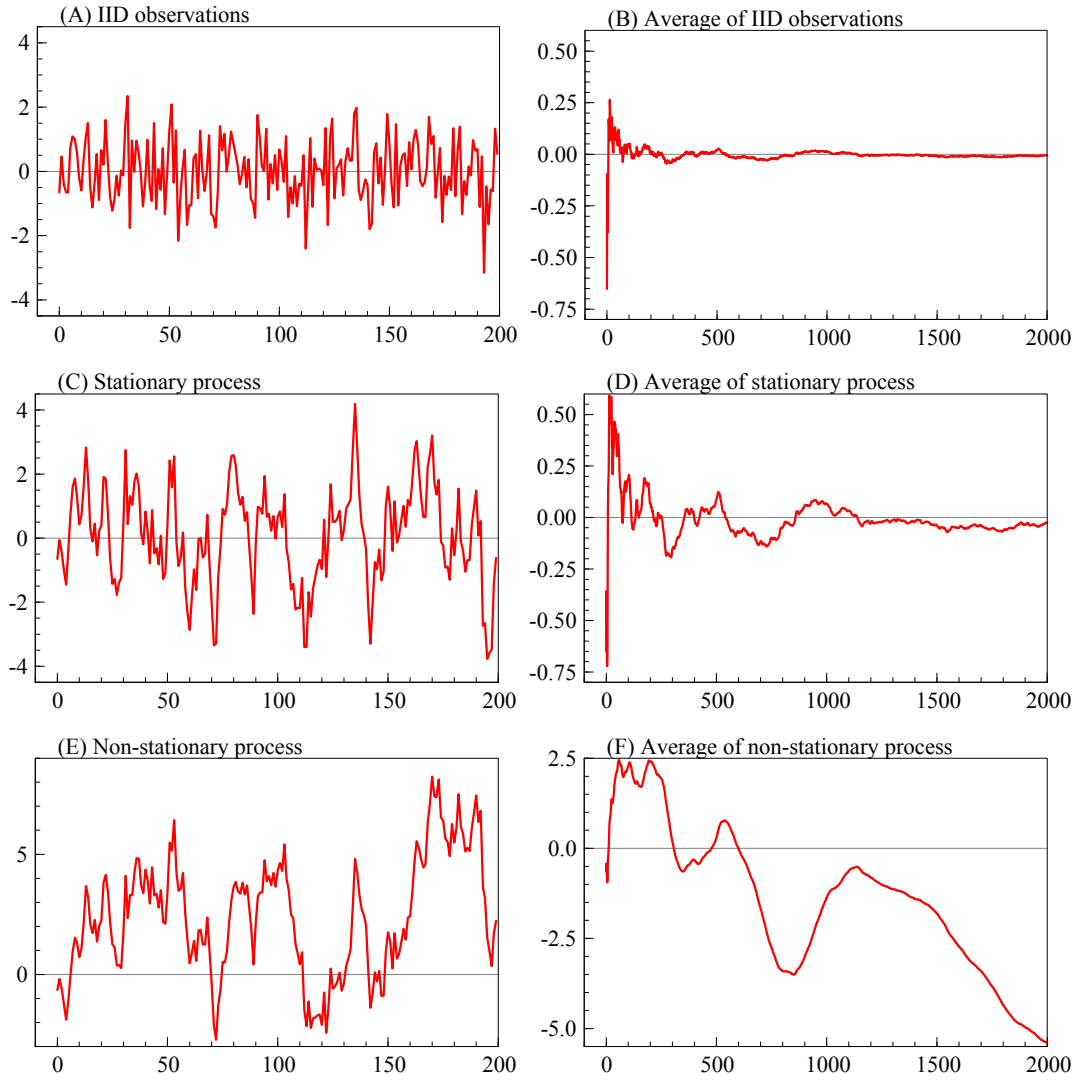
Figure 3: *Time series of* 200 *observations for (A): An IID process. (C): A dependent but stationary process. And (E): A non-stationary process. The graphs (B), (D), and (F) show the sample averages of $y_1, ..., y_T$, that is $\overline{y}_T = \frac{1}{T} \sum_{t=1}^{T} y_t$, as a function of $T = 1, 2, ..., 2000$.*

have observed that observation four is below average, $y_4 < \mu$, that could suggest that the best forecast of the next observation, the conditional expectation $E[y_5 \mid y_4]$, is also likely to be smaller than μ. This highlights the important difference between conditional and unconditional expectations.

In terms of economics the deviations from equilibrium could reflect business cycles; we expect business cycles to be systematic but to revert at some point in time.

3.1 MEASURES OF TIME DEPENDENCE

One way to characterize the time dependence in a time series, y_t, also known as the *persistence*, is by the correlation between y_t and y_{t-h}, defined as

$$Corr(y_t, y_{t-h}) = \frac{Cov(y_t, y_{t-h})}{\sqrt{V(y_t) \cdot V(y_{t-h})}}.$$

If the correlation is positive we denote it positive autocorrelation, and in a graph of the time series that is visible 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.

Under stationarity it holds that $Cov(y_t, y_{t-h}) = \gamma_h$ only depends on h, and the variance is constant, $V(y_t) = V(y_{t-h}) = \gamma_0$. For a stationary process the formula can therefore be simplified, and we define *the autocorrelation function* (ACF) as

$$\rho_h = \frac{Cov(y_t, y_{t-h})}{V(y_t)} = \frac{\gamma_h}{\gamma_0}, \quad h = ..., -2, -1, 0, 1, 2, ...$$

The term autocorrelation function refers to the fact that we consider ρ_h as a function of the lag length h. It follows from the definition that $\rho_0 = 1$, the autocorrelations are symmetric, $\rho_h = \rho_{-h}$, and they are bounded, $-1 \leq \rho_h \leq 1$. For a stationary time series the autocorrelations could be non-zero for a number of lags, but we expect ρ_h to approach zero as h increases. In most cases the convergence to zero is relatively fast.

For a given data set, the sample autocorrelations can be estimated by e.g.

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

or alternatively, by

$$\widehat{\rho}_h = \frac{\frac{1}{T-h} \sum_{t=h+1}^{T} (y_t - \overline{y})(y_{t-h} - \overline{y})}{\frac{1}{T-h} \sum_{t=h+1}^{T} (y_{t-h} - \overline{y})^2}.$$

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

$$y_t = c + \rho_h y_{t-h} + \text{residual}.$$

There are several formulations of the variance of the estimated autocorrelation function. The simplest result is that if the true correlations are all zero, $\rho_1 = \rho_2 = ... = 0$, then the asymptotic distribution of $\widehat{\rho}_h$ is normal with variance $V(\widehat{\rho}_h) = T^{-1}$. A 95% confidence band for $\widehat{\rho}_h$ is therefore given by $\pm 2/\sqrt{T}$.

An complementary measure of the time dependence is the so-called *partial autocorrelation function* (PACF), which is the correlation between y_t and y_{t-h}, conditional on the

intermediate values, i.e.

$$\theta_h = Corr(y_t, y_{t-h} \mid y_{t-1}, ..., y_{t-h+1}).$$

One way to understand the PACF is the following: If y_t is correlated with y_{t-1}, then y_{t-1} is correlated with y_{t-2}. This implies that y_t and y_{t-2} are correlated by construction, but some of the effect is indirect (it goes through y_{t-1}). The PACF measures the direct relation between y_t and y_{t-2}, after the indirect effect via y_{t-1} is removed. The PACF can be estimated as the OLS estimator $\widehat{\theta}_h$ in the regression

$$y_t = c + \theta_1 y_{t-1} + ... + \theta_h y_{t-h} + \text{residual},$$

where the intermediate lags are included. If $\theta_1 = \theta_2 = ... = 0$, it again holds that $V(\widehat{\theta}_h) = T^{-1}$, and a 95% confidence band is given by $\pm 2/\sqrt{T}$. For a stationary time series the partial autocorrelation function, θ_h, should also approach zero as h increases.

In the next section we consider examples of the estimated ACF, and later in the course we discuss more precisely how the ACF and PACF convey information on the properties of the data.

4 TRANSFORMATIONS TO STATIONARITY

A brief look at the time series in Figure 1 suggests that not all economic time series are stationary. The US unemployment rate in Figure 4 (A) seems to fluctuate around a constant level. The deviations are extremely persistent, however, and the unemployment rate is above the constant level for most of the 20 year period $1975 - 1995$. If the time series is considered stationary, the speed of adjustment towards the equilibrium level is very low; and it is probably a more reasonable assumption that the expected value of the unemployment rate is not constant. An economic interpretation of this finding could be that there have been systematic changes in the natural rate of unemployment over the decades, and that the actual unemployment rate is not adjusting towards a constant level but towards a time varying NAIRU.

The persistence of the time series itself translates into very large and significant autocorrelations. Graph (B) indicates that the ACF remains positive for a very long sample length, and the correlation between u_t and u_{t-15} is around 0.6. Whether the unemployment rate actually corresponds to a sample path from a stationary stochastic process is an empirical question. We will discuss formal testing of this hypothesis at a later stage.

4.1 TREND-STATIONARITY

The Danish productivity level in Figure 1 (B) has a clear positive drift and is hence non-stationary. The trend appears to be very systematic, however, and it seems that productivity fluctuates around the linear trend with a constant variance. One reason could be that there is a constant autonomous increase in productivity of around 2% per year;
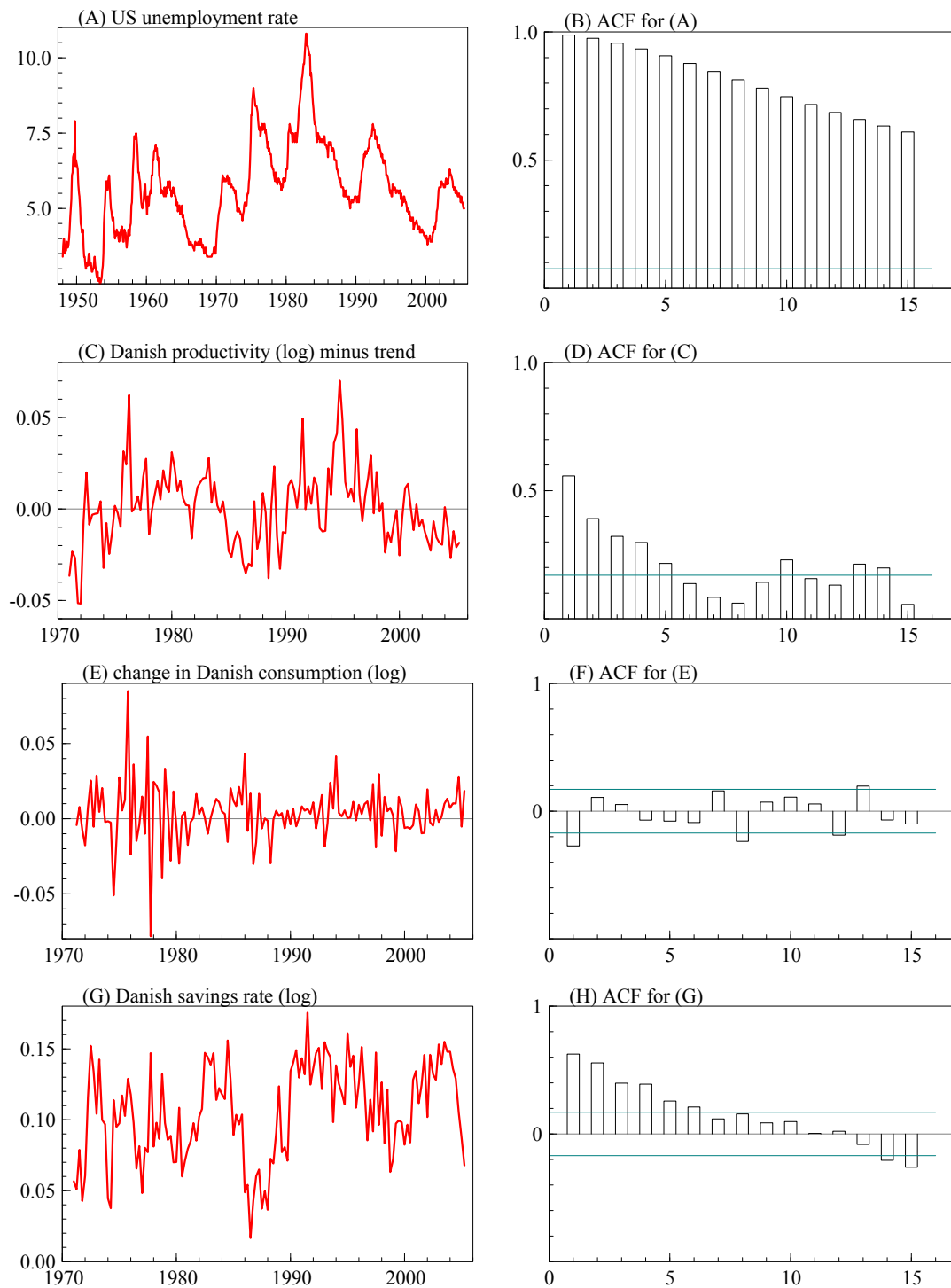
Figure 4: *Examples of time series transformations to stationarity.*

and equilibrium is defined in terms on this trend, so that the deviations of productivity from the underlying trend are stationary.

Figure 4 (C) shows the deviation of productivity from a linear trend. The deviations are calculated as the estimated residual, $\widehat{q}_t^*$, in the linear regression on a constant and a trend,

$$q_t = \delta + \gamma t + q_t^*, \tag{5}$$

where q_t is the log of measured productivity. The deviations from the trend, $\widehat{q}_t^* = q_t - \widehat{\delta} - \widehat{\gamma}t$, are still systematic, reflecting labour hoarding and other business cycle effects, but there seems to be a clear reversion of productivity to the trending mean. This is also reflected in the ACF of the deviations, which dies out relatively fast.

A non-stationary time series, q_t, that becomes stationary after a deterministic linear trend has been removed is denoted *trend stationary*. One way to think about a trend-stationary time series is that the stochastic part of the process is stationary, but the stochastic fluctuations appear around a trending deterministic component, $\delta + \gamma t$. Deterministic de-trending, like in the regression (5), is one possible way to transform a non-stationary time series to stationarity.

4.2 DIFFERENCE-STATIONARITY

Also the aggregate Danish consumption in Figure 1 (C) is clearly non-stationary due to the upward drift. In this case, the deviations from a deterministic trend are still very persistent, and the de-trended consumption also looks non-stationary. An alternative way to remove the non-stationarity is to consider the first difference, $\Delta c_t = c_t - c_{t-1}$. The changes in Danish consumption, Δc_t, depicted in Figure 4 (D) fluctuates around a constant level and could look stationary. The persistence is not very strong, which is also reflected in the ACF in graph (E). A variable, c_t, which itself is non-stationary but where the first difference, Δc_t, is stationary is denoted *difference stationary*. Often it is also referred to as *integrated of first order* or *I(1)* because it behaves like a stationary variable that has been integrated (or cumulated) once.

Whether it is most appropriate to transform a variable to stationarity using deterministic de-trending or by first differencing has been subject to much debate in the economics and econometrics literature. The main difference is whether the stochastic component of the variable is stationary or not, that is whether the stochastic shocks to the process have transitory effects only (the trend-stationary case), or whether shock can have permanent effects (the difference stationary case). We return to the discussion later, where we present formal tests between the two cases.

We may note that the first difference of variable in logs have a special interpretation. As an example, let $c_t = \log(\mathsf{CONS}_t)$ measures the log of consumption. It holds that

$$\Delta c_t = c_t - c_{t-1} = \log\left(\frac{\mathsf{CONS}_t}{\mathsf{CONS}_{t-1}}\right) = \log\left(1 + \frac{\mathsf{CONS}_t - \mathsf{CONS}_{t-1}}{\mathsf{CONS}_{t-1}}\right) \simeq \frac{\mathsf{CONS}_t - \mathsf{CONS}_{t-1}}{\mathsf{CONS}_{t-1}},$$

where the last approximation is good if the growth rate is close to zero. Therefore, Δc_t can be interpreted as the relative growth rate of CONS_t; and changes in c_t are interpretable as percentage changes (divided by 100).

4.3   Cointegration

A third way to remove non-stationarity is by taking linear combinations of several variables. In Figure 1 (C) the development of income and consumption have many similarities and the savings rate, $s_t = y_t - c_t$, depicted in Figure 4 (G) is much more stable. It seems to fluctuate around a constant level with peaks corresponding to known business cycle episodes in the Danish economy. The time dependence dies out relatively fast, cf. graph (H), and the savings rate may correspond to a stationary process.

The property that a combination of non-stationary variables become stationary is denoted *cointegration*. The interpretation is that the variables themselves, y_t and c_t, move in a non-stationary manner, but they are tied together in an equilibrium relation, and the dynamics of the economy ensures that the savings rate only shows transitory deviations from the equilibrium level. Cointegration is the main tool in time series analysis of non-stationary variables, and we will return to the cointegration analysis later in the course.

References

HAYASHI, F. (2000): *Econometrics*. Princeton University Press, Princeton.

VERBEEK, M. (2004): *A Guide to Modern Economtrics*. John Wiley & Sons, 2nd edn.

WOOLDRIDGE, J. M. (2006): *Introductory Econometrics: A Modern Approach*. Thomson, South-Western Publishing, 3rd edn.