LINEAR REGRESSION WITH TIME SERIES DATA

Econometrics C ⧫ Lecture Note 2
Heino Bohn Nielsen
February 6, 2012

This note analyzes OLS estimation in a linear regression model for time series data. We first discuss the assumptions required on the data and the error term, and present a number of important results for the OLS estimator; we focus on the interpretation and intuition for the results and no formal proofs are given. We then discuss the requirement that the error term should not be autocorrelated, which is an important design criteria for dynamic models. Finally, we discuss more broadly the issue of model formulation and misspecification testing and present an empirical example.

Outline

1 The Linear Regression Model

Let y_t be a variable of interest, and let x_t be a $k \times 1$ dimensional vector of explanatory variables. To model y_t as a function of x_t we consider the linear regression model

$$y_t = x_t'\beta + \epsilon_t, \tag{1}$$

for observations $t = 1, 2, ..., T$. Here β is a $k \times 1$ vector of parameters and ϵ_t is an error term, so that

$$y_t = x_t'\beta + \epsilon_t = x_{1t}\beta_1 + x_{2t}\beta_2 + ... + x_{kt}\beta_k + \epsilon_t.$$

In most applications the first explanatory variable is a constant term, $x_{1t} = 1$, in which case β_1 is the intercept of the regression. We use the index t to represent individual observations and in the next sections we will explicitly assume that y_t and x_t are time series. At this point, however, we have made no assumptions on the stochastic properties of y_t and x_t.

We will sometimes refer to the left hand side variable, y_t, as the *regressand*, the *dependent variable*, or the *endogenous variable*. The right hand side variables, x_t, are sometimes referred to as *explanatory variables*, *regressors*, *covariates*, and, under some specific assumptions, *exogenous variables*. Finally, we say that (1) is a regression of y_t on x_t.

1.1 Interpretation and the Ceteris Paribus Assumption

So far the regression model in (1) is a tautology, and it does not say anything on the relationship between y_t and x_t. For any set of observations $(y_t, x_t')'$ and any parameter value b the residual term can be chosen as $\epsilon_t = y_t - x_t'b$ to satisfy (1). To make the equation informative on the relationship between y_t and x_t we therefore have to impose restrictions on the behavior of ϵ_t that allow us to determine a unique value of β from equation (1). This is, loosely speaking, what is called *identification* in econometrics. And a condition for being able to estimate the parameter consistently, so that the estimator $\widehat{\beta}$ converges to the true value β as $T \to \infty$, is that the parameter is identified.

So, what are the identifying restrictions in a linear regression? In many cases it is natural to think of the model (1) as representing a conditional expectation, $E[y_t \mid x_t] = x_t'\beta$, such that

$$E[\epsilon_t \mid x_t] = 0. \tag{2}$$

Under this assumption we can think of a parameter β_j as the marginal effect of the expected value of y_t of a change in x_{jt}, i.e. as the partial derivative, $\frac{\partial}{\partial x_{jt}} E[y_t \mid x_t] = \beta_j$. We therefore interpret β_j as the effect of a marginal change in the variable x_{jt} holding the remaining variables in x_t constant; this is known as the *ceteris paribus* assumption. If (2) is fulfilled, we refer to the regressor as being *predetermined*. As we will see in §2 below, predeterminedness is a sufficient condition for identification of β in a linear regression model.

Note that the assumption in (2) is not an innocuous technicality. It states that all information that is relevant for the relationship between x_t and y_t has been included in the model. Firstly, this excludes that a variable in x_t depends on y_t through some feedback mechanism operating at time t. If this is the case there exists two equations linking y_t and x_t and in our simple regression there is no way to say which one we would actually obtain. Secondly, remember that if a relevant variable is excluded from the regression model then it will be picked up by the error term. For (2) to be true we need that any such variable is unrelated to x_t.

1.2 Properties of Conditional Expectations

The interpretation of the linear regression is intimately linked to the conditional expectation. First note that it is always possible to decompose a stochastic variable, y_t, into a conditional expectation and an error term with conditional expectation zero, i.e. for any vector x_t,

$$y_t = E[y_t \mid x_t] + \epsilon_t,$$

where $E[\epsilon_t \mid x_t] = 0$. In general $E[y_t \mid x_t]$ is some nonlinear function of x_t. The central assumptions in the linear regression model are that x_t includes all the relevant conditioning information and that the functional form of the conditional expectation is linear, $E[y_t \mid x_t] = x_t'\beta$. At this point we emphasize three important properties of the conditional expectation.

Firstly, we have the well-known result that

$$E[g(x_t) \mid x_t] = g(x_t),$$

meaning that if we condition on x_t then we can treat the stochastic variable as non-random.

Secondly, the condition (2) implies that ϵ_t is uncorrelated with any function of x_t, an therefore also uncorrelated with the individual variables, $x_{1t}, x_{2t}, ..., x_{kt}$. The condition (2) therefore also states that the functional form of the regression model has been correctly specified; no non-linear effects have been neglected.

Thirdly, $E[\epsilon_t \mid x_t] = 0$ implies an unconditional zero expectation, $E[\epsilon_t] = 0$. This is an example of a results called the *law of iterated expectations*, which states that

$$E\left[E\left[y_t \mid x_t \quad\right] \mid x_t, z_t\right] \;=\; E\left[y_t \mid x_t\right] \qquad (3)$$

$$E\left[E\left[y_t \mid x_t, z_t\right] \mid x_t \quad\right] \;=\; E\left[y_t \mid x_t\right]. \qquad (4)$$

It is easy to follow the intuition in the result (3): Recall that $E\left[y_t \mid x_t\right] = g(x_t)$ is some general function of x_t and since all the information in x_t is also contained in the larger information set $w_t = (x_t', z_t')'$, the conditional expectation of $g(x_t)$ is the function itself. To understand the result in (4) it is informative to think of the conditional expectation as a prediction. The result states that we can not improve the prediction given the small information set, $E\left[y_t \mid x_t\right]$, by first imagining the prediction using a larger information set, $E\left[y_t \mid x_t, z_t\right]$, and then try to forecast that best prediction using only x_t. The general result is that it is always *the smallest information set that dominates.*

1.3 Examples of Time Series Regressions

Depending on the variables included in the vector of regressors, x_t, the interpretation of the linear regression in (1) changes.

As a first example, let the vector of regressors contain k explanatory variables dated at the same point in time as the left hand variable, i.e. the equation (1). Then the linear regression is called a *static regression*.

Next recall, that due to the temporal ordering of the time series observations, past events can be treated as given in the analysis of current events. Since many economic time series seem to depend on their own past it is natural to include the *lagged values*, $y_{t-1}, y_{t-2}, ...$, in the explanation of the current value. As an example we can let, $x_t = y_{t-1}$, and the regression model is given by

$$y_t = \theta y_{t-1} + \epsilon_t. \tag{5}$$

A model where the properties of y_t are characterized as a function of only its own past is denoted a *univariate time series model*, and the specific model in (5), where y_t depend only on y_{t-1}, is denoted a *first order autoregressive*, or AR(1), *model*.

The dynamic structure of the regression model can easily be more complex than (5) with lagged values of both the regressand, y_t, and the regressors, x_t. As an example, consider the dynamic regression model

$$y_t = \theta_1 y_{t-1} + x_t'\varphi_0 + x_{t-1}'\varphi_1 + \epsilon_t, \tag{6}$$

where y_t is modelled as a function of y_{t-1}, x_t, and x_{t-1}. This model is denoted an *autoregressive distributed lag*, or ADL, *model*.

The above models are useful in different contexts and later in the course we go into more details with the interpretations of the models. In particular we want to characterize the dynamic properties such as the dynamic impacts, $\partial y_t/\partial x_t, \partial y_{t+1}/\partial x_t, ...$

2 Method of Moments Estimation

One way to derive the ordinary least squares (OLS) estimator of β in the linear regression (1) is by appealing to the so-called *method of moments* (MM) estimation principle, see also Wooldridge (2006). Here we briefly review the MM estimation principle and the relation to the identification of the parameters. We return to applications of MM estimation in more complicated situations later in the course.

The conditional zero mean in (2) states that x_t does not contain information on the expected value of ϵ_t. This implies in particular that x_t and ϵ_t are uncorrelated, i.e. that

$$E[x_t\epsilon_t] = 0, \tag{7}$$

since $E[x_t\epsilon_t] = E[E[x_t\epsilon_t \mid x_t]] = E[x_t E[\epsilon_t \mid x_t]] = 0$ from the properties of the conditional expectation. We will refer to (7) as a set of *moment conditions*. Inserting the expression for the error terms, $\epsilon_t = y_t - x_t'\beta$, yields the system of k equations to determine the k parameters in β, and if there is a unique solution we say that the system identifies the parameter. In particular we have that

$$\begin{aligned} E[x_t(y_t - x_t'\beta)] &= 0 \quad \text{or,} \\ E[x_t y_t] - E[x_t x_t']\beta &= 0. \end{aligned}$$

If the $k \times k$ matrix $E[x_t x_t']$ is non-singular it can inverted to give the solution:

$$\beta = E[x_t x_t']^{-1} E[x_t y_t], \tag{8}$$

which is the population parameter.

From a given finite sample of y_t and x_t $(t = 1, 2, ..., T)$, we cannot compute the expectations and (8) is infeasible. The idea of the MM estimation principle is to replace the expectations by sample averages, which defines the well known OLS sample estimator,

$$\widehat{\beta} = \left(T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} x_t y_t \right). \tag{9}$$

For the step from (8) to (9) to work, we need a law of large numbers (LLN) to apply, so that sample averages converge to the expectations, i.e.

$$T^{-1} \sum_{t=1}^{T} x_t y_t \to E[x_t y_t] \quad \text{and} \quad T^{-1} \sum_{t=1}^{T} x_t x_t' \to E[x_t x_t'].$$

Under this assumption it holds that $\widehat{\beta}$ defined in (9) converges to the true value in (8).

Note that three distinct conditions are needed to derive the estimator: (i) The explanatory variables should be predetermined as stated by the moment condition in (7). This is a property of the model and states that there is a unique value of β that satisfies the equations; i.e that the parameters are identified. (ii) A LLN should apply to the data, which is related to the stochastic properties of y_t and x_t. A central part of any econometric analysis is to ensure that these two conditions are fulfilled. (iii) The matrices $T^{-1} \sum_{t=1}^{T} x_t x_t'$ and $E[x_t x_t']$ should be non-singular. Here they are quadratic by construction, and the only assumption is we need is that the regressors are not linearly dependent. This is the usual assumption of no perfect multicollinearity.

3 Properties of Time Series Regressions

In this section we assume that y_t and x_t are time series. We discuss the statistical properties of the OLS estimator in the linear regression model in (1), and present a number of important asymptotic and finite sample results. The results are very similar to the results known from cross-sectional regression, but the implications and interpretations differ somewhat due to different sampling of the data.

The statistical analysis underlying the presented results is not easy and we do not give proofs. Instead we discuss the intuition and look at some simple examples. A more rigorous and technically demanding coverage, including proofs of the theorems, can be found in Davidson (2001). Hamilton (1994) also go through the calculations for many relevant time series regressions, and a very informative discussion on time series regressions is found in Hayashi (2000).

As discussed above we need to impose assumptions to ensure that a LLN applies to the sample averages. There are several ways to formulate the requirements, see Davidson (2001), but in this course we use the following formulation:

Assumption 1 (stationarity and weak dependence): *Consider a time series y_t and the $k \times 1$ vector time series x_t. For the process $z_t = (y_t, x_t')'$ make the following assumptions: (i) z_t has a stationary distribution. And (ii) the process z_t is weakly dependent, so that z_t and z_{t-h} become approximately independent for $h \to \infty$.*[1]

For analysis of cross-sectional data it is common to assume that the observations are *identically* and *independently* distributed (IID). That assumption is typically too restrictive for time series observations, but most of the results for the regression model continue to hold for time series satisfying Assumption 1. The idea is that 'identical distributions' is replaced by the assumption of stationarity, while 'independence' is replaced by the assumption of weak dependence.

3.1 Consistency

A first requirement for an estimator is that it is consistent, so that the estimator converges to the true value as we get more and more observations (i.e. for $T \to \infty$):

Result 1 (consistency): *Consider a data set, y_t and x_t, that obeys Assumption 1. If the regressors, x_t, are predetermined so that the moment condition (7) holds, then the OLS estimator in (9) is a consistent estimator of the true value, i.e. $\widehat{\beta} \to \beta$ as $T \to \infty$.*

An alternative formulation is that $\operatorname{plim} \widehat{\beta} = \beta$. The intuition for the result is straightforward and follows the steps in the derivation in §2. First, we need the moment conditions to be satisfied. Next we need a LLN to apply to the sample averages so that $\widehat{\beta} \to \beta$. And Assumption 1 is sufficient to ensure that. A sketch of the proof for consistency of OLS in a simple regression model in presented in Box 1.

Note again that the moment condition in (7) is implied by the zero conditional expectation, $E[\epsilon_t \mid x_t] = 0$, i.e. predeterminedness. This implies that if the regression model represents the conditional expectation, $E[y_t \mid x_t] = x_t'\beta$, then the OLS estimator, $\widehat{\beta}$, is consistent under Assumption 1.

[1] The formulation here is not precise, but is meant to capture the ida of the *mixing* concept from probability theory, see e.g. Davidson (2001, p. 70).

Box 1: Illustration of Consistency

Consider a regression model

$$y_t = x_t\beta + \epsilon_t, \quad t = 1, 2, ..., T, \tag{B1-1}$$

where the single explanatory variable is predetermined, $E[\epsilon_t x_t] = 0$. To illustrate the consistency of the OLS estimator in (9) we insert the expression for y_t in the formula to obtain

$$\widehat{\beta} = \frac{\frac{1}{T}\sum_{t=1}^T y_t x_t}{\frac{1}{T}\sum_{t=1}^T x_t^2} = \frac{\frac{1}{T}\sum_{t=1}^T (x_t\beta + \epsilon_t) x_t}{\frac{1}{T}\sum_{t=1}^T x_t^2} = \beta + \frac{\frac{1}{T}\sum_{t=1}^T \epsilon_t x_t}{\frac{1}{T}\sum_{t=1}^T x_t^2}. \tag{B1–2}$$

We look at the behavior of the last term as $T \to \infty$.

First, assume that the denominator has a non-zero limit for $T \to \infty$, i.e.

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^T x_t^2 = \sigma_x^2, \tag{B1-3}$$

for $0 < \sigma_x^2 < \infty$. This requirement says that the limit of the second moment should be positive and finite. A stationary process has the same moments for all t, so the average second moment will converge to a constant as stated in (B1-3).

If a LLN applies, i.e. under Assumption 1, the limit for the numerator is

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

where the last equality follows from the assumption of predeterminedness. Combining the results, we obtain

$$\operatorname*{plim}_{T\to\infty} \widehat{\beta} = \beta + \frac{\operatorname{plim}\frac{1}{T}\sum_{t=1}^T \epsilon_t x_t}{\operatorname{plim}\frac{1}{T}\sum_{t=1}^T x_t^2} = \beta + \frac{0}{\sigma_x^2} = \beta,$$

which shows the consistency of the OLS estimator, $\widehat{\beta}$.

It should be emphasized that Result 1 gives *sufficient* conditions for consistency. The conditions are not *necessary*, however, and in the analysis of non-stationary variables later in the course we will see examples of estimators that are consistent by other arguments than the one used to obtain Result 1. As an example, it turns out that the estimator $\widehat{\theta}$ of θ in the first order autoregressive model (5) is consistent even if y_t does not fulfill Assumption 1.

3.2 What is OLS Actually Estimating?

There is a subtle point related to the interpretation of the potential OLS inconsistency, which is worth noting. To illustrate this consider a linear equation

$$y_t = \beta_1 + x_t\beta_2 + w_t\beta_3 + u_t, \tag{10}$$

where y_t depends on two explanatory variables. We assume that the requirements for OLS are fulfilled, in particular that $E[u_t \mid x_t, w_t] = 0$. Now suppose that we incorrectly omit the variable w_t and consider the linear regression

$$y_t = \beta_1 + x_t \beta_2 + \epsilon_t, \tag{11}$$

which is *misspecified* due to the *omitted variable*, w_t. The regression error term is given by $\epsilon_t = w_t \beta_3 + u_t$, and using the result in Box 1 the OLS estimator in the misspecified model, $\breve{\beta}_2$ say, is inconsistent when $E[\epsilon_t x_t] = E[(w_t \beta_3 + u_t) x_t] \neq 0$, i.e. when x_t and w_t are correlated.

To gain insight in the mechanics of the linear regression it is natural to ask what OLS in equation (11) is estimating. To answer this questions, remember that for a stochastic variable of interest, y_t, and a set of conditioning variables, x_t, we may always decompose y_t into a conditional expectation given x_t and a remainder term with conditional expectation zero. For the misspecified regression model above we may write

$$y_t = E[y_t \mid x_t] + \epsilon_t, \quad \text{where} \quad E[\epsilon_t \mid x_t] = 0.$$

Assuming that the conditional expectation is a linear function, $E[y_t \mid x_t] = b_1 + x_t b_2$, we get the model

$$y_t = b_1 + b_2 x_t + \epsilon_t, \tag{12}$$

The OLS moment condition, $E[\epsilon_t \mid x_t] = 0$, is fulfilled per construction for (12) and the OLS estimator, $\hat{b}_2$, which equals to $\breve{\beta}_2$ from (11), is a consistent estimator of b_2. This parameter has the interpretation of a partial derivative,

$$\operatorname{plim} \breve{\beta}_2 = b_2 = \frac{\partial E[y_t \mid x_t]}{\partial x_t} \neq \frac{\partial E[y_t \mid x_t, w_t]}{\partial x_t} = \beta_2. \tag{13}$$

The intuitive point is that OLS consistently estimates the parameters in the *linear conditional expectation*. The problem is that the structural economic equation stated in (10) does *not* correspond to the conditional expectation in (12) and the model will not identify the economically relevant parameter β_2. To put it differently, the problem is not that the mechanics of the OLS estimator are invalid, the problem is that it consistently estimates an irrelevant quantity. OLS is essentially estimating a correlation coefficient and whether that can be given a structural economic interpretation in terms of economic theory depends on the setting, and in practice the moment condition for OLS always has to be motivated.

3.3 Unbiasedness and Finite Sample Bias

Consistency is a minimal requirement for an estimator. A more ambitious requirement is that of unbiasedness, which is often quoted for OLS in regressions for IID data. As we will see, it is rarely possible to obtain unbiased estimators in dynamic models.

Result 2 (unbiasedness): *Let y_t and x_t obey Assumption 1. If the regressors, x_t, are strictly exogenous, so that*

$$E[\epsilon_t \mid x_1, x_2, ..., x_t, ..., x_T] = 0, \tag{14}$$

then the OLS estimator is unbiased, i.e. $E[\widehat{\beta} \mid x_1, x_2, ..., x_T] = \beta$.

To show this result consider the last term of (B1–2) in Box 1. Recall that the unconditional expectation of the ratio is not the ratio of the expectations. Instead take expectations conditional on $x_1, ..., x_t, ..., x_T$ and assume (14) to get

$$E\left[\widehat{\beta} \mid x_1, x_2, ..., x_t, ..., x_T\right] = \beta + \frac{\frac{1}{T}\sum_{t=1}^{T} x_t E\left[\epsilon_t \mid x_1, x_2, ..., x_t, ..., x_T\right]}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} = \beta.$$

From the rules for conditional expectations we may also write the result unconditionally: $E[\widehat{\beta}] = E[E[\widehat{\beta} \mid x_1, x_2, ..., x_t, ..., x_T]] = \beta$.

Whereas consistency is an asymptotic property, prevailing as $T \to \infty$, unbiasedness is a finite sample property stating that the expectation of the estimator equals the true value for all sample lengths.

Note, however, that in order to prove unbiasedness we have to invoke the assumption of *strict exogeneity* in (14), implying a zero correlation between the error term, ϵ_t, and both past, current and future values of x_t. Strict exogeneity is often a reasonable assumption for cross-sectional data where the randomly sampled cross-sectional units are independent, but for time series data the assumption is in most cases too strong. As an example, consider the first order autoregressive model

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

Due to the structure of the time series it might be reasonable to assume 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. $y_t, y_{t+1}, ..., y_T$.

As a consequence we have the following auxiliary result:

Result 3 (estimation bias in dynamic models): *In general, the OLS estimator is not unbiased in regression models with lagged dependent variables.*

As an example, it can be shown that the OLS estimator of the autoregressive coefficient in the AR(1) model (5) is biased towards zero. The derivation of the bias is technically demanding and instead Box 2 presents a Monte Carlo simulation to illustrate the idea.

3.4 Complete Dynamic Models and Asymptotic Distribution

To make inference on the estimator, i.e. to test hypotheses on the parameter, we need a way to approximate the distribution of $\widehat{\beta}$. To derive this we need to impose restrictions on the variance of the error term.

Box 2: Bias of OLS in an AR(1) Model

To illustrate the bias of OLS we use a Monte Carlo simulation. As a data generating process (DGP) we use the AR(1) model

$$y_t = \theta y_{t-1} + \epsilon_t, \quad t = 1, 2, ..., T, \tag{B2-1}$$

with an autoregressive parameter of $\theta = 0.9$ and $\epsilon_t \sim N(0,1)$. We generate $M = 5000$ time series with a sample length T, i.e. $y_1^{(m)}, y_2^{(m)}, ..., y_T^{(m)}$ for $m = 1, 2, ..., M$. For each time series we apply OLS to the regression model (B2-1) and get the estimate $\widehat{\theta}_m$.

To characterize the estimator we calculate the average and standard deviation:

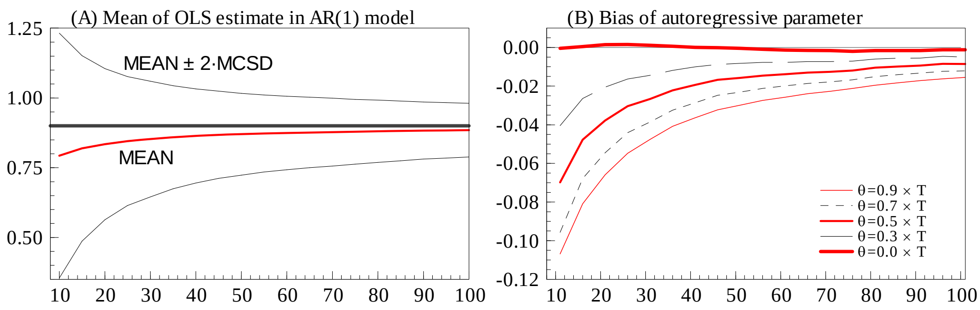
$$\mathsf{MEAN}(\widehat{\theta}) = \frac{1}{M}\sum_{m=1}^{M}\widehat{\theta}_m \quad \text{and} \quad \mathsf{MCSD}(\widehat{\theta}) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\widehat{\theta}_m - \mathsf{MEAN}(\widehat{\theta}))^2},$$

and also the bias, $\mathsf{BIAS}(\widehat{\theta}) = \mathsf{MEAN}(\widehat{\theta}) - \theta$. Notice, that $\mathsf{MEAN}(\widehat{\theta})$ is itself an estimator, and the uncertainty related to the estimator can be measured by the *Monte Carlo standard error*, defined as $\mathsf{MCSE} = M^{-\frac{1}{2}} \cdot \mathsf{MCSD}(\widehat{\theta})$. Be aware of the important difference between the $\mathsf{MCSD}(\widehat{\theta})$, which measures of the uncertainty of $\widehat{\theta}$ between realization, and the MCSE, which measures the uncertainty of $\mathsf{MEAN}(\widehat{\theta})$ between simulation experiments. The latter converges to zero for an increasing number of replications, $M \to \infty$.

The results from $\mathsf{PcNaive}$ are reported in Figure (A) for sample lengths $T \in \{10, 15, ..., 100\}$. The confidence bands, $\mathsf{MEAN}(\widehat{\theta}) \pm 2 \cdot \mathsf{MCSD}(\widehat{\theta})$, measures the uncertainty of $\widehat{\theta}$ in each replication. The mean is lower than the true value for all sample lengths. For a very small sample length of $T = 10$ the OLS estimator has a mean of $\mathsf{MEAN}(\widehat{\theta}) = 0.7935$. The $\mathsf{MCSD}(\widehat{\theta}) = 0.2172$ 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 unbiasedness, $\mathcal{H}_0 : \mathsf{MEAN}(\widehat{\theta}) = 0.9$, can be constructed as

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

which is clearly significant in the asymptotic $N(0,1)$ distribution. For larger samples the average converges to the true value as expected from the consistency of OLS.



To illustrate how the bias depends on the autoregressive parameter, Figure (B) reports the bias for other values, $\theta \in \{0, 0.3, 0.5, 0.7, 0.9\}$. If the DGP is static, $\theta = 0$, the estimator is unbiased. If $\theta > 0$ the estimator is downward biased, with a bias that increases with θ.

RESULT 4 (ASYMPTOTIC DISTRIBUTION): *Let y_t and x_t obey Assumption 1, and assume that the regressors are predetermined so that (7) holds. If ϵ_t is homoskedastic, i.e.*

$$E[\epsilon_t^2 \mid x_t] = \sigma^2, \tag{15}$$

with no serial correlation, i.e. for all $t \neq s$,

$$E[\epsilon_t \epsilon_s \mid x_t, x_s] = 0, \tag{16}$$

then the OLS estimator is asymptotically normally distributed, so that

$$\sqrt{T}\left(\widehat{\beta} - \beta\right) \to N(0, \sigma^2 E[x_t x_t']^{-1}), \tag{17}$$

as $T \to \infty$.

To interpret the statement in (17) we note that $\sqrt{T}(\widehat{\beta} - \beta) \to N(0, V)$, where V is called the asymptotic variance. Intuition for the expression for the asymptotic variance can be obtained for the case of a single regressor with mean zero, in which case

$$V = \frac{\sigma^2}{E[x_t^2]} = \frac{V(\epsilon_t)}{V(x_t)}.$$

This is a measure of the noise-to-information ratio. Note that the variance of the estimator is $V(\widehat{\beta}) = T^{-1}V$, which collapses with the rate of T. The result implies that we can test hypotheses on β. Inserting natural estimators for σ^2 and $E[x_t x_t']$, the distributional result in (17) can be written as

$$\widehat{\beta} \overset{a}{\sim} N\left(\beta, \widehat{\sigma}^2 \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}\right), \tag{18}$$

which is again similar to the formula for the cross-sectional case. It is worth emphasizing that the asymptotic normality is the result of a central limit theorem (CLT) and it does not require normality of the error term, ϵ_t.

The precise formulation of the condition in (16) is a little difficult to interpret, and often we ignore the conditioning on x_t and x_s and consider whether ϵ_t and ϵ_s are uncorrelated for $t \neq s$. An alternative way to relate to the condition of no-serial-correlation in (16) is to think of a model for the conditional expectation of y_t given the entire joint history of y_t and x_t. If it holds that

$$E[y_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ..., y_1, x_1] = E[y_t \mid x_t] = x_t'\beta, \tag{19}$$

so that x_t contains all relevant information in the available information set: $x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ..., y_1, x_1$, then we refer to the regression model as being a *complete dynamic model*. Assuming the complete dynamic model in (19) is practically the same as the no-serial-correlation assumption in (16); and the idea is that there is no systematic information in the past of y_t and x_t which has not been used in the construction of the regression model.

If the considered regression model is dynamic, such as the AR(1) model in (5) or the ADL model in (6), then most people would have as a design criteria that the model should be dynamically complete, i.e. free of serial correlation of the error term. The reason is that the variables in x_t have been chosen to represent the systematic variation of y_t over time; and for the econometric model to be successful in that respect we require that no systematic variation is left in ϵ_t.

4 Autocorrelation of the Error Term

In this section we discuss the case where the no-serial-correlation assumption in (16) is violated. This is the case if consecutive error terms are correlated, e.g. if $Cov(\epsilon_t, \epsilon_s) \neq 0$ for some $t \neq s$. In this case we say there is *autocorrelation of the error term*. In practice, autocorrelation is detected by looking at the estimated residuals, $\widehat{\epsilon}_t$ $(t = 1, 2, ..., T)$, and $Cov(\widehat{\epsilon}_t, \widehat{\epsilon}_s) \neq 0$ is referred to as *residual autocorrelation*.

4.1 Consequences of Autocorrelation

Note that autocorrelation will not in general violate the assumptions for Result 1, and OLS is consistent if the explanatory variables, x_t, are contemporaneously uncorrelated with the error term. If the model includes a lagged dependent variable, however, autocorrelation of the error term *will* violate the assumption in (7). To see this, consider an AR(1) model like (5), and assume that the error term exhibit autocorrelation of first order, i.e. that ϵ_t follows a first order autoregressive model,

$$\epsilon_t = \rho\epsilon_{t-1} + v_t, \tag{20}$$

where v_t is an IID error term. Consistency requires that $E[\epsilon_t y_{t-1}] = 0$, but that is clearly not satisfied since both y_{t-1} and ϵ_t depends on ϵ_{t-1}. We have the following result:

Result 5 (inconsistency of ols with autocorrelation and lagged regressand): *In a regression model including the lagged dependent variable, the OLS estimator is in general not consistent in the presence of autocorrelation of the error term.*

This is an additional motivation for the fact that no-autocorrelation is an important design criteria for dynamic regression models.

Next, even if OLS is consistent, the standard formula for the variance in (18) is no longer valid. The asymptotic normality still holds, and in the spirit of White's heteroskedasticity robust standard errors, it is possible to find a consistent estimate of the correct covariance matrix under autocorrelation, the so-called *heteroskedasticity-and-autocorrelation-consistent* (HAC) standard errors. This is discussed very briefly in Verbeek (2008, Section 4.10.2), and a simpler discussion of the univariate case is given in Stock and Watson (2003, p. 504-507). HAC covariance matrices have a natural role in method of moments estimation and we return to the issue later in the course.

4.2 Interpretation of Residual Autocorrelation

It is important to realize, that the residuals of a regression model pick up the composite effect of everything not accounted for by the explanatory variables. The interpretation of residual autocorrelation therefore depends on the likely reason for the indications of autocorrelation. Possible sources of residual autocorrelation include:

(1) The error terms of the DGP are autoregressive.
(2) The estimation model is dynamically misspecified.
(3) The estimation model wrongly omits a persistent variable.
(4) The functional form is misspecified.

Autocorrelation is often interpreted as a sign of misspecification of the model, and the relevant solution depends on the interpretation of the residual autocorrelation.

4.2.1 Autoregressive Errors in the DGP

If it holds that the error term in the regression model is truly autoregressive, i.e. that the DGP is given by the two equations (1) and (20):

$$y_t = x_t'\beta + \epsilon_t \quad \text{where} \quad \epsilon_t = \rho\epsilon_{t-1} + v_t,$$

then it is natural to use the information in both equations to derive an estimator of β.

If ρ is known, the two equations can be combined to yield

$$(y_t - \rho y_{t-1}) = \left(x_t' - \rho x_{t-1}'\right)\beta + (\epsilon_t - \rho\epsilon_{t-1}), \tag{21}$$

or equivalently

$$y_t = \rho y_{t-1} + x_t'\beta - x_{t-1}'\rho\beta + v_t, \tag{22}$$

where the error term $v_t = \epsilon_t - \rho\epsilon_{t-1}$ is now serially uncorrelated. In practice, the parameter ρ is unknown but can be consistently estimated by running the regression (20) on the estimated residuals from (1).

The transformation to equation (22) is chosen to remove the problem of the original system and it is analog to the idea of GLS estimation in the case of heteroskedasticity. Note, however, that (22) is subject to a restriction on the parameters: There are three regressors in the equation, but the parameters are made up of only two free parameters, ρ and β. The GLS transformation therefore implies a so-called *common factor restriction*, and there are no closed form solution for the estimators of of the parameters in (22); see Verbeek (2008, p. 107-108) for details.

For given ρ, consistent estimation of β in (21) requires the usual moment condition

$$E[(x_t - \rho x_{t-1})(\epsilon_t - \rho\epsilon_{t-1})] = 0.$$

This implies that ϵ_{t-1} should be uncorrelated with x_t, $E[\epsilon_t x_{t+1}] = 0$. This is in the direction of strict exogeneity and consistency of GLS requires stronger assumptions than consistency of OLS; see also Wooldridge (2006).

It should be emphasized that *the GLS transformation to remove autocorrelation is rarely used in modern econometrics*. The most important reason is (as discussed above) that the finding of residual autocorrelation for a regression model does not imply that the error term of the DGP is autoregressive. The second reason is the strong assumption needed for consistency of GLS.

4.2.2 Dynamic Misspecification

Residual autocorrelation indicates that the model is not dynamically complete. If the model in (1) is dynamic that is normally interpreted as a violation of a design criteria. The econometric model is therefore misspecified and should be reformulated. Autocorrelation implies that

$$E[y_t \mid x_t] \neq E[y_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ..., y_1, x_1],$$

and the natural remedy is to extend the list of regressors in order to capture all the systematic variation.

If the estimated residuals seem to exhibit first order autocorrelation, then a starting point is the transformed model in (22). The common factor restrictions are imposed by an assumed structure of the DGP, which is not necessarily valid. Instead of the non-linear GLS equation, we can alternatively estimate the unrestricted ADL model

$$y_t = \alpha_0 y_{t-1} + x_t' \alpha_1 + x_{t-1}' \alpha_2 + \eta_t, \tag{23}$$

where η_t is a new error term. Here we do not take the structure of (20) at face value and we use it only indicative to extend the list of regressors to obtain a dynamically complete model. If we are interested, we can test the validity of the common factor restrictions by comparing (22) and (23).

4.2.3 Omitted Variables

Omitted variables in general can also produce residual autocorrelation. Consider a simple DGP given by

$$y_t = x_{1t} \cdot \beta_1 + x_{2t} \cdot \beta_2 + \epsilon_t,$$

and consider an estimation model that omits the variable x_{2t}:

$$y_t = x_{1t} \cdot \beta_1 + u_t.$$

Then the new error term is $u_t = x_{2t} \cdot \beta_2 + \epsilon_t$, which is autocorrelated if there are persistent movements in x_{2t}.

One example is if the level of y_t changes at some point in time, so that the DGP contains a level shift. If we do not account for the break in the regression, then the predicted values, $\widehat{y}_t = x_t' \widehat{\beta}$, will correspond to an average between the level before and after the shift. As a consequence, the residuals are mainly positive before the break and mainly negative after the break (or the opposite). That, again, results in residual autocorrelation.

In this case the solution is to try to identify the shift and to account for it in the regression model. If there is a shift in the level of y_t at time T_0, then we could extend the list of regressors with a dummy variable taking the value 1 for $t \geq T_0$ and 0 otherwise. That would allow the level after T_0 to be different from the level before. This is mainly a technical solution, and from an economic point of view it is most reasonable if the dating of the break, T_0, can be interpreted; e.g. the time of the German reunification. In any case a preferable solution would be to find a variable that *explains* the shift, but that is often extremely difficult.

4.2.4 Misspecified Functional Form

If the true relationship between y_t and x_t is non-linear, then the residuals from a linear regression will typically be autocorrelated. Think of a true relationship being a parabola and a linear regression line. Then the residuals will be systematic over time–reflecting the systematic differences between a parabola and a straight line; and that translates into residual autocorrelation.

In this case the obvious solution is to try to reformulate the functional form of the regression line.

5 Model Formulation and Misspecification Testing

In this section we briefly outline an empirical strategy for dynamic modelling; with the explicit goal to find a model representation that is dynamically complete.

So far we have assumed knowledge of the list of relevant regressors, x_t. In reality we need a way to choose these variables; and often economic theory is helpful in pointing out potential explanations for the variable of interest y_t. From Result 1 we know that the estimator $\widehat{\beta}$ is consistent for any true value β. So if we include a redundant regressor (i.e. a variable with true parameter $\beta_i = 0$) then we will be able to detect it as $\widehat{\beta}_i \to 0$ for $T \to \infty$. If, on the other hand, we leave out an important variable (with $\beta_i \neq 0$), then the estimators will not be consistent in general. This asymmetry suggests that it is generally recommendable to start with a larger model and then to simplify it by removing insignificant variables. This is the so-called *general-to-specific* principle. The opposite *specific-to-general* principle is dangerous because if the initial model is too restricted (by leaving out an important variable) then estimation and inference will be invalid.

We can never prove that a model is well specified; but we can estimate a model and test for indications of misspecifications in known directions: E.g. autocorrelation, heteroskedasticity, wrong functional form, etc. If the model passes all the tests, then we have no indications that the model is misspecified and we may think of the model as representing the main features of the data.

Above, we discussed that the finding of autocorrelation allows different interpretations. This is true more generally, and if we cannot reject a certain type of misspecification for the model, we do not necessarily know why, and it is difficult to use the misspecification in a mechanical manner to improve the model. Whenever the model is rejected we have to reconsider the data and the problem at hand, and try to reformulate the model to explain the particular neglected features of the data.

Below we present a number of standard misspecification tests; and in §6 we consider an empirical example.

5.1 Test for No-Autocorrelation

Recall that residual autocorrelation can indicate many types of misspecification of a model, and *the test for no autocorrelation should be routinely applied in all time series regressions.* In modern econometrics, the most commonly used test for the null hypothesis of no autocorrelation is a so-called Breusch-Godfrey Lagrange Multiplier (LM) test. As an example we consider the test for no first order autocorrelation in the regression model (1). This is done by running the auxiliary regression model

$$\widehat{\epsilon}_t = x_t'\delta + \gamma\widehat{\epsilon}_{t-1} + u_t, \tag{24}$$

where $\widehat{\epsilon}_t$ is the estimated residual from (1) and u_t is a new error term. The original explanatory variables, x_t, are included in (24) to allow for the fact that x_t is not necessarily strictly exogenous and may be correlated with ϵ_{t-1}.

The null hypothesis of no autocorrelation corresponds to $\gamma = 0$, and can be tested by the $t-$ratio on γ, which follows an $N(0,1)$ asymptotically. Alternatively we can compute the LM test statistic, $\xi_{AR} = T \cdot R^2$, where R^2 is the coefficient of determination in the auxiliary regression (24). 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

$$\xi_{AR} = T \cdot R^2 \to \chi^2(1). \tag{25}$$

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. $\widehat{\epsilon}_0 = 0$, and estimate the auxiliary regression for the same sample as the original model.

Sometimes a small-sample adjustment is applied and the statstic $(T-k)\,R^2$ is considered, but that is in most cases not important for the conclusions. An often preferred alternative is to consider an $F-$form of the test; this is the one implemented in PcGive.

The test in (25) is valid asymptotically, i.e. for $T \to \infty$. An alternative finite sample test exists: The so-called Durbin Watson (DW) test, see Verbeek (2008, 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 in (14), which makes it invalid in most time series settings.

5.2 Test for No-Heteroskedasticity

To test the assumption of no-heteroskedasticity in (15), we use an LM test against heteroskedasticity of an unknown form (due to White). It involves the auxiliary regression of the squared residuals on the original regressors and their squares:

$$\widehat{\epsilon}_t^2 = \gamma_0 + x_{1t}\gamma_1 + ... + x_{kt}\gamma_k + x_{1t}^2\delta_1 + ... + x_{kt}^2\delta_k + u_t.$$

The null hypothesis is unconditional homoskedasticity, $\gamma_1 = ... = \gamma_k = \delta_1 = ... = \delta_k = 0$, and the alternative is that the variance of ϵ_t depends on x_{it} or the squares x_{it}^2 for some $i = 1, 2, ..., k$, i.e. at least one of the parameters $\gamma_1, ..., \gamma_k, \delta_1, ..., \delta_k$ non-zero. Again the test is based on the LM statistic $\xi_{HET} = T \cdot R^2$, which is distributed as a $\chi^2(2k)$ under the null. Sometimes a more general test is also considered, in which the auxiliary regression is augmented with all the non-redundant cross terms, $x_{it} \cdot x_{jt}$.

A particular dynamic form of heteroskedacity, which implies systematic variation of the variance over time, is often relevant in time series models. This is denoted autoregressive conditional heteroskedasticity (ARCH) and will be discussed in some detail later in the course.

5.3 Test for Correct Functional Form: RESET

To test for the functional form of the regression, the so-called RESET test can be used. The idea is to consider the auxiliary regression model

$$\widehat{\epsilon}_t = x_t'\delta + \gamma\widehat{y}_t^2 + u_t,$$

where $\widehat{y}_t = x_t'\widehat{\beta}$ is the predicted value of the original regression. The null hypothesis of correct specification is $\gamma = 0$. The alternative is that the square of $\widehat{y}_t = x_t'\widehat{\beta}$ has been omitted. This indicates that the original functional form is incorrect and could be improved by powers of linear combinations of the explanatory variables, x_t. The RESET test statistic, ξ_{RESET}, is the $F-$test for $\gamma = 0$ and it is distributed as $F(1, T - k - 1)$.

5.4 Test for Normality of the Error Term

The derived results for the regression model hold without assuming normality of the error term. It is still a good idea, however, to thoroughly examine the residuals from a regression model, and the normal distribution is a natural benchmark for comparison. The main reasons to focus on the normal distribution is that the convergence of $\widehat{\beta}$ to the asymptotic normal distribution is faster if ϵ_t is close to normal. Furthermore, under normality of the error terms, least squares estimation coincides with *maximum likelihood* (ML) estimation, which implies a number of nice asymptotic properties. We will return to ML estimation later in the course.

It is always a good starting point to plot the residual to get a first visual impression of possible deviations from normality. If some of the residuals fall outside the interval of say three standard errors, it might be an indication that an extraordinary event has taken place. If a big residual at time T_0 corresponds to a known shock, the observation may be accounted for by including a dummy variable with the value 1 at T_0 and zero otherwise. Similarly, it is also useful to plot a histogram of the estimated residuals and compare with the normal distribution.

A formal way of comparing a distribution with the normal is to calculate *skewness* (S), which measures the asymmetry of the distribution, and *kurtosis* (K), which measures the proportion of probability mass located in the tails of the distribution. Let $u_t = (\widehat{\epsilon}_t - \overline{\epsilon})/\widehat{\sigma}$ be the standardized residuals, where $\overline{\epsilon} = T^{-1}\sum_{t=1}^{T}\widehat{\epsilon}_t$ and $\widehat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(\widehat{\epsilon}_t - \overline{\epsilon})^2$ denote the estimated mean and variance. Skewness (S) and kurtosis (K) are defined as the estimated third and fourth central moments:

$$\mathsf{S} = T^{-1}\sum_{t=1}^{T} u_t^3 \quad \text{and} \quad \mathsf{K} = T^{-1}\sum_{t=1}^{T} u_t^4.$$

The normal distribution is symmetric and has a skewness of $\mathsf{S} = 0$. The normal distribution has a kurtosis measure of $\mathsf{K} = 3$ and $\mathsf{K} - 3$ is often referred to as *excess kurtosis*. If K is larger than three the distribution has 'fat' tails in the sense that more probability mass is located in the tails.

Under the assumption of normality, it holds that the estimated skewness and kurtosis are asymptotically normal (due to a central limit theorem), so that $\mathsf{S} \to N(0, 6 \cdot T^{-1})$ and $\mathsf{K} \to N(3, 24 \cdot T^{-1})$. For testing it is convenient to use the χ^2-versions, i.e.

$$\begin{aligned} \xi_S &= \frac{T}{6} \cdot \mathsf{S}^2 &&\to \chi^2(1) \\ \xi_K &= \frac{T}{24} \cdot (\mathsf{K} - 3)^2 &&\to \chi^2(1). \end{aligned}$$

Furthermore, it turns out that ξ_S and ξ_K are asymptotically independent, which makes is easy to construct a joint test. One such test is the *Jarque-Bera* test, which is just

$$\xi_{JB} = \xi_S + \xi_K \to \chi^2(2).$$

In the output of PcGive this is denoted the asymptotic test. Several refinements have been suggested. One version, that corrects for a correlation between ξ_S and ξ_K in finite samples, is implemented in PcGive.

6 Empirical Example

As an empirical example we estimate a simple consumption function for aggregate Danish consumption. We consider a data set including the first differences of the logarithm of aggregate consumption in the private sector, Δc_t, disposable income, Δy_t, and wealth including owner occupied housing, Δw_t. The three time series are depicted in Figure 1 (A). All time series are relatively volatile; but income appears more erratic than consumption. Importantly, all three time series in first differences look stationary, and it does not seem

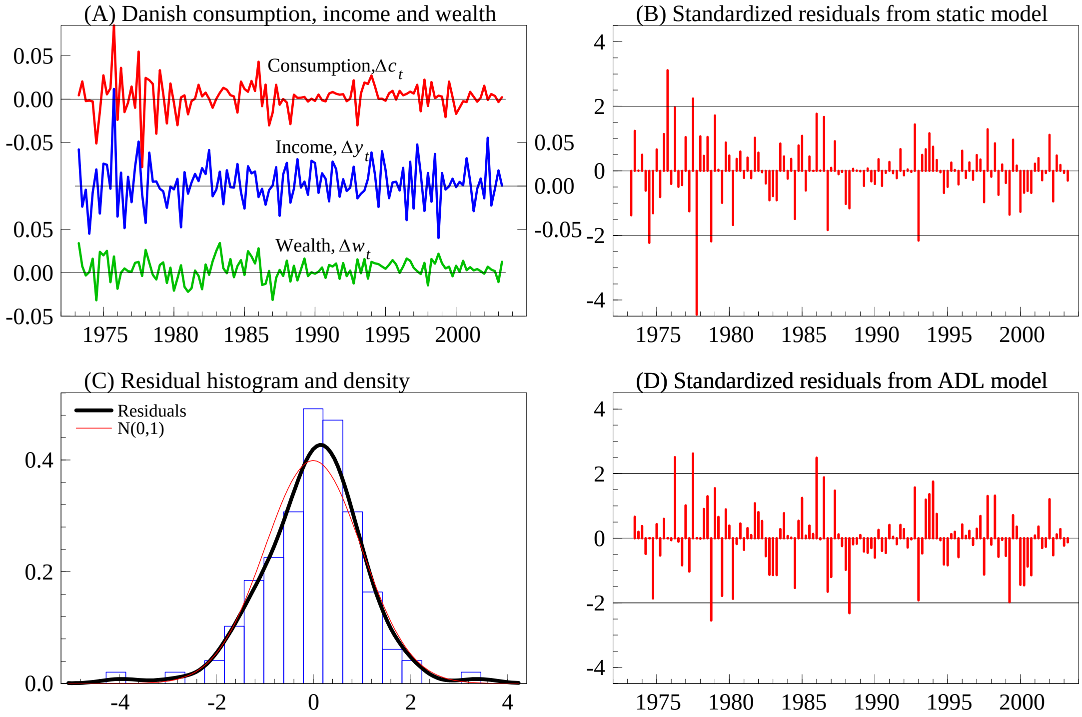


Figure 1: *Danish consumption data and estimated residuals.*

unreasonable to invoke Assumption 1. Later in the course we turn to testing for non-stationarity; here we make the judgement based on a graphical inspection.

If we believe that income and wealth are predetermined relative to consumption then we may consider the linear regression model

$$\Delta c_t = \beta_0 + \beta_1 \cdot \Delta y_t + \beta_2 \cdot \Delta w_t + \epsilon_t.$$

The predeterminedness assumption is not testable and it requires, for example, no-reverse causality, so that consumption does not affect income and wealth at time t. This is not trivial and there could be effects e.g. via the national accounts construction of disposable income. Also it requires that any omitted variable is uncorrelated with income and wealth. Again this has be defended in each case. Here a candidate for an omitted variable could be the interest rate, and if so we require that the interest rate is not correlated with income. We do not want to dwell anymore with the moment conditions, just emphasize the point that the assumptions are not trivial and should be discussed.

Running this regression for the full sample, $1973:2-2003:2$, yields the estimated equation

$$\Delta c_t = \underset{(0.11)}{0.0002} + \underset{(3.13)}{0.196} \cdot \Delta y_t + \underset{(4.45)}{0.559} \cdot \Delta w_t + \widehat{\epsilon}_t, \tag{26}$$

where the numbers in parentheses are $t-$ratios for the hypothesis that $\beta_i = 0$. In this equation $R^2 = 0.208$, indicating that 20% of the variation in consumption growth is explained by the regressors. Although it is tempting to discuss significance of the coefficients, we cannot attach much weight to the $t-$ratios at this point because we do not know whether the assumptions in Result 4 are fulfilled.

The estimated residuals from equation (26) are depicted in Figure 1 (B); we note some very large residuals in the mid 70'ties. To test the null hypothesis of no-autocorrelation we construct a Breusch-Godfrey test. We focus on autocorrelation of order one and two and consider the auxiliary regression

$$\widehat{\epsilon}_t = \underset{(0.13)}{0.0002} + \underset{(0.78)}{0.049} \cdot \Delta y_t - \underset{(-0.66)}{0.083} \cdot \Delta w_t - \underset{(-3.24)}{0.313} \cdot \widehat{\epsilon}_{t-1} + \underset{(0.15)}{0.014} \cdot \widehat{\epsilon}_{t-2},$$

where $\widehat{\epsilon}_t$ denote the estimated residuals. The first lag, $\widehat{\epsilon}_{t-1}$, is significantly negative, with a coefficient of -0.313 and a $t-$ratio of -3.24, while the second lag, $\widehat{\epsilon}_{t-2}$, is insignificant. That indicates a negative first order autocorrelation of the residuals, so that a large positive residual is often followed by a large negative residual. The coefficient of determination in the auxiliary regression is $R^2 = 0.092$ and the LM statistic is given by $\xi_{LM} = T \cdot R^2 = 121 \cdot 0.092 = 11.176$, which is again clearly rejected in a $\chi^2(2)$ distribution with a $p-$value of 0.004. We conclude that the residuals are negatively autocorrelated, and we note that this finding invalidates the $t-$ratios in (26).

A natural solution to the first order autocorrelation is to formulate the first order ADL model, i.e. augmenting the regression (26) with the first lag of all variables. Estimating the dynamic model for the longest possible sample, $1973:3-2003:2$, yields

$$\begin{aligned} \Delta c_t &= \underset{(0.25)}{0.0004} - \underset{(-3.39)}{0.314} \cdot \Delta c_{t-1} + \underset{(3.78)}{0.249} \cdot \Delta y_t + \underset{(0.66)}{0.044} \cdot \Delta y_{t-1} \\ &\quad + \underset{(4.07)}{0.513} \cdot \Delta w_t + \underset{(1.53)}{0.201} \cdot \Delta w_{t-1} + \widehat{\epsilon}_t. \end{aligned} \tag{27}$$

For the dynamic model in (27) the LM test for no second order autocorrelation is 1.092, corresponding to a $p-$value of 0.579 in a $\chi^2(2)$. We conclude that the model in (27) appears dynamically complete. The LM test for no heteroskedasticity is also accepted and we conclude that the $t-$ratios follow standard normal distributions, assymptotically. We note that the lag of consumption growth is significant ($t-$ratio of -3.39) while the lags of the changes in income and wealth are not significantly different from zero. In this equation $R^2 = 0.295$.

In the residuals for this model there still seems to be a number of large residuals, and the histogram in Figure 1 (C) has fatter tails than the normal distribution. Looking more formally at the residuals, we find $\mathsf{S} = -0.391$ indicating a skewness to the left compared to the normal distribution. A test for symmetry, $\mathsf{S} = 0$, can be constructed as

$$\xi_S = \frac{T}{6} \cdot \mathsf{S}^2 = \frac{120}{6} \cdot (-0.391)^2 = 3.053,$$

which is not significant in a $\chi^2\,(1)$ distribution. The measure of excess kurtosis is $\mathsf{K} - 3 = 2.074$, indicating that the distribution of the residuals has fat tails. A test for $\mathsf{K} = 3$ can be based on the statistic

$$\xi_K = \frac{T}{24} \cdot (\mathsf{K} - 3)^2 = \frac{120}{24} \cdot (5.079 - 3)^2 = 21.503,$$

	Coefficient	*Std.Error*	*t−value*
Δc_{t-1}	-0.284	0.079	-3.62
Δy_t	0.140	0.060	2.34
Δy_{t-1}	0.043	0.056	0.77
Δw_t	0.327	0.110	2.98
Δw_{t-1}	0.314	0.110	2.84
Dum743_t	-0.050	0.014	-3.57
Dum754_t	0.069	0.015	4.60
Dum774_t	-0.068	0.014	-4.89
Constant	0.0013	0.0014	0.95
$\widehat{\sigma}$	0.0132	log-likelihood	353.30
R^2	0.535	T	120
	Statistic	$[p-val]$	*Distribution*
No autocorrelation of order 1-2	1.360	[0.51]	$\chi^2(2)$
Normality	1.290	[0.52]	$\chi^2(2)$
No heteroskedasticity	12.033	[0.53]	$\chi^2(13)$
Correct functional form (RESET)	0.00001	[0.99]	$F(1,110)$

Table 1: *Modelling Δc_t by OLS for $t = 1973:2-2003:2$.*

which is much larger than the 5% critical value of 3.84 in a $\chi^2(1)$ distribution. The distribution is close to symmetric, but with a marked excess kurtosis. A combined test for $\mathsf{S} = \mathsf{K} - 3 = 0$ can be constructed as the Jarque-Bera statistic:

$$\xi_{JB} = \xi_S + \xi_K = 3.053 + 21.503 = 24.556,$$

which is clearly significant in a $\chi^2(2)$ distribution.

The deviation from normality was only due to excess kurtosis. Kurtosis often reflects outliers, i.e. a few large residuals. In the present case the residuals for three observations, $1974:3$, $1975:4$ and $1977:4$, look extreme, and we might want to condition on these observations by inserting dummy variables. Looking at an economic calendar we note that $1975:4$ corresponds to a temporary VAT reduction, while $1974:3$ and $1977:4$ correspond to announced contractive policy measures. Inserting dummy variables (of the form $0,...,0,1,0,...,0$) for these observations yields the estimated model reported in Table 1. For this model $R^2 = 0.535$, indicating that the dummies account for much of the variation. Comparing with the results in (27), we see that the coefficient to income growth is lowered by the dummy variables and Δy_t is less significant. The reason is that the policy measures affected both income and consumption and the dummies take out the effects from these special events in the estimation of the parameters.

The estimated residuals for this preferred model are reported in Figure 1 (D). From the results in Table 1 we note that the LM test for no autocorrelation has a p−value of

0.51 and we cannot reject the null hypothesis of a well specified model. The Jarque-Bera test for normality of the estimated residuals has a $p-$value is 0.52 and we accept the null hypothesis of normal residuals. The LM test for no-heteroskedastiticy based on a regression with the regressors in Table 1 and their squares (in total 13 non-redundant variables) is also reported. The null-hypothesis of a well specified model is again accepted with a $p-$value of 0.53. Finally, the RESET test for correct functional form gives a $p-$value of 0.99 with no evidence against the linearity of the regression line.

To conclude, the preferred model in Table 1 appears to be well specified, in the sense that we reject the types of misspecifications considered. This does not imply, however, that we have reached a final model; and we return to the analysis of Danish consumption later in the course. A main drawback of the model is that all variables have been transformed to stationarity by first differences and all information in the levels of the variables is eliminated. Later we will use cointegration techniques to recover the information in the levels of consumption, income and wealth.

References

Davidson, J. (2001): *Econometric Theory.* Blackwell, Oxford.

Hamilton, J. D. (1994): *Time Series Analysis.* Princeton University Press, Princeton.

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

Stock, J. H., and M. W. Watson (2003): *Introduction to Econometrics.* Addison Wesley.

Verbeek, M. (2008): *A Guide to Modern Econometrics.* John Wiley & Sons, 3rd edn.

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