This note analyzes OLS estimation in a linear regression model for time series data. We first discuss the assumptions required on the data and on the error term, and we 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.

1 Time Series Regression Models

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

\[
y_t = x_t' \beta + \epsilon_t,
\]

for observations \( t = 1, 2, ..., T \), where \( \beta \) is a \( k \times 1 \) vector of parameters to be estimated and \( \epsilon_t \) is an error term. 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. \( x_t = z_t \). Then the linear regression

\[
y_t = z_t' \beta + \epsilon_t,
\]

(2)
is called a static regression. The dating of $y_t$ and $x_t$ are the same to model the contemporaneous relation, but no dynamic adjustment is modeled.

Next recall, that due to the temporal ordering of the time series observations, past events can be treated as given, or predetermined, 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, $y_{t-1}, y_{t-2}, \ldots$, 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

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

(3)

A model where the properties of $y_t$ are characterized as a function of only its own past is denoted as univariate time series model, and the specific model in (3), where $y_t$ depend only on the one period lagged value is denoted a first order autoregressive, or $AR(1)$, model.

The dynamic structure of the regression model can easily be more complex than (3) with lagged values of both the regressand, $y_t$, and the regressors, $x_t$. As an example, consider the dynamic regression model

$$y_t = \theta_1 y_{t-1} + x_t' \phi_0 + x_{t-1}' \phi_1 + \epsilon_t,$$  

(4)

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. It is the workhorse in the modelling of dynamic relations between variables.

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, $\partial y_t / \partial x_t, \partial y_{t+1} / \partial x_t, \ldots$, in different situations.

## 2 Properties of Linear Regression

In this section we discuss the statistical properties of ordinary least squares (OLS) applied to the linear time series regression model in (1), which requires that we specify the properties of the data $(y_t, x_t)$ and the error term $\epsilon_t$. We present a number of important asymptotic and finite sample results and compare the findings with the statistical properties of the cross-sectional regression. The material is not easy and we do not prove the results. 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).

In order to estimate the parameters in $\beta$ in the linear regression (1), the usual approach is to apply OLS estimation. One way to motivate OLS is to consider the so-called moment
condition
\[ E[x_t \epsilon_t] = 0, \]  
(5)

stating that the explanatory variables should be uncorrelated with the error term, see e.g. Wooldridge (2003, p. 27). The model implies that \( \epsilon_t = y_t - x_t' \beta \), and we can write the moment conditions as

\[
E[x_t (y_t - x_t' \beta)] = 0
\]

\[
E[x_t y_t] - E[x_t x_t'] \beta = 0.
\]

That suggests the population estimator
\[
\hat{\beta} = E[x_t x_t']^{-1} E[x_t y_t],
\]
which is defined if \( E[x_t x_t'] \) is non-singular.

From a given sample of \( y_t \) and \( x_t \), we cannot calculate the mathematical expectations in (6), but we may estimate the expectations by sample averages. For this to work, we need a law of large numbers (LLN) to apply, so that sample averages converge to the expectation, i.e.

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

Substituting sample averages we get the well known OLS estimator
\[
\hat{\beta} = \left( T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left( T^{-1} \sum_{t=1}^{T} x_t y_t \right).
\]

(7)

This way to motivate the OLS estimator is an example of a so-called method of moments (MM) estimation. We return to the analysis of the MM estimation principle later in the course.

To apply OLS is a regression model for time series data, 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 most cases we make the following assumption:

**Assumption 1 (Stationarity and weak dependence)** Consider a time series \( y_t \) and the \( k \times 1 \) vector time series \( x_t \). We assume that \( z_t = (y_t, x_t)' \) has a joint stationary distribution. We also assume that the process \( z_t \) is weakly dependent, so that \( z_t \) and \( z_{t+k} \) becomes approximately independent for \( k \rightarrow \infty \).

Under Assumption 1, most of the results for linear regression on random samples (i.e. cross-sectional data) carry over to the time series case. The idea of a regression analysis for time series data is to use observations from the past to characterize historical relationships. If we want to use the historical relationships to explain current and future developments we have to require that the future behaves like the past; and that is exactly the assumption of stationarity.
2.1 Consistency

A minimal 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.

**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 (5) holds, then the OLS estimator in (7) is a consistent estimator of the true value, i.e. $\hat{\beta} \rightarrow \beta$ as $T \rightarrow \infty$. The intuition for the result is straightforward. The OLS estimator is derived from the moment condition in (5). If this moment condition is true, then the derived estimator in (7) is consistent if a LLN apply to the sample averages; and that holds if Assumption 1 is satisfied.

In many cases it is natural to think of the model (1) as representing a conditional expectation, $E[y_t \mid x_t] = x_t'\beta$. For this to be true, we need that

$$E[\epsilon_t \mid x_t] = 0. \quad (8)$$

The conditional zero mean in (8) is stronger than and implies the no-contemporaneous-correlation assumption in (5). It follows that the OLS estimator is consistent whenever the regression model represents the conditional expectation of $y_t \mid x_t$.

We will not give a formal proof of Result 1. Instead we consider as an example the consistency of the OLS estimator in the first order autoregressive model (3).

**Example 1 (Consistency of OLS in an AR(1) model)** Consider the AR(1) model

$$y_t = \theta y_{t-1} + \epsilon_t, \quad t = 1, 2, ..., T.$$ We assume to have observed the time series for $y_0, y_1, ..., y_T$, where the first observation, $y_0$, is called the initial value. We write the OLS estimator as

$$\hat{\theta} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2},$$

and to illustrate the consistency we apply the usual manipulations:

$$\hat{\theta} = \frac{\sum_{t=1}^{T} (\alpha y_{t-1} + \epsilon_t) y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} = \theta + \frac{1}{T} \sum_{t=1}^{T} \epsilon_t y_{t-1} = \theta + \frac{1}{T} \sum_{t=1}^{T} \frac{\epsilon_t y_{t-1}}{y_{t-1}^2}.$$

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

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

$$\text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{y_{t-1}^2}{y_{t-1}} = q, \quad (9)$$

1 The assumption in (8) implies that $E[\epsilon_t \cdot f(x_t)] = 0$ for all functions $f(\cdot)$. More discussion on the relation between the assumptions can be found in Hayashi (2000).
for $0 < q < \infty$. This requirement says that the limit of the second moment should be positive and finite. In our example $E[\epsilon_t] = 0$ 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 variance is the same for all $t$, the requirement in (9) is automatically satisfied.

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

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \epsilon_t y_{t-1} = E[\epsilon_t y_{t-1}] = 0,$$

where the last equality follows from the assumption of predeterminedness in (5). Combining the results, we obtain

$$\lim_{T \to \infty} \hat{\theta} = \theta + \frac{\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \epsilon_t y_{t-1}}{\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2} = \theta + \frac{0}{q} = \theta,$$

which shows the consistency of the OLS estimator, $\hat{\theta}$.

### 2.2 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 cross-sectional regressions. As we will see, it is rarely possible to obtain unbiased estimators in dynamic models.

**Result 2 (Unbiasedness)** Again, let $y_t$ and $x_t$ obey Assumption 1. If the regressors, $x_t$, are strictly exogenous, so that

$$E[\epsilon_t | x_1, x_2, ..., x_t, ..., x_T] = 0,$$  

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

Note that whereas consistency is an asymptotic property, prevailing as $T \to \infty$, then 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 (10), implying a zero correlation between the error term, $\epsilon_t$, 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 again 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 $\epsilon_t$ is uncorrelated with lagged values of the explanatory variables, $y_{t-1}, y_{t-2}, ..., y_1$. But since $y_t$ is function of $\epsilon_t$ it is clear that $\epsilon_t$ 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 a dynamic regression model.

As an example, it can be shown that the OLS estimator of the autoregressive coefficient in the AR(1) model (3) is biased towards zero. The derivation of the bias is technically demanding and instead we use a Monte Carlo simulation to illustrate the idea.

EXAMPLE 2 (Small sample bias of OLS in an AR(1) model) As a data generating process (DGP) in the Monte Carlo simulation we use the specific AR(1) model
\[
y_t = \theta y_{t-1} + \epsilon_t, \quad t = 1, 2, \ldots, T, \tag{11}
\]
with an autoregressive parameter of \( \theta = 0.9 \) and \( \epsilon_t \sim N(0, 1) \). We use (11) to generate \( M = 5000 \) time series with a sample length \( T \), i.e.

\[
y^{(m)}_1, y^{(m)}_2, \ldots, y^{(m)}_T, \quad m = 1, 2, \ldots, M.
\]

For each time series we apply OLS to the regression model (3) and get the estimate \( \hat{\theta}_m \).

To illustrate the bias we calculate the average of the OLS estimates,

\[
\text{MEAN}(\hat{\theta}) = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_m, \tag{12}
\]

and also the bias, \( \text{BIAS}(\hat{\theta}) = \text{MEAN}(\hat{\theta}) - \theta \). The uncertainty of \( \hat{\theta} \) can be measured by the standard deviation of \( \hat{\theta}_m \) across the \( M \) replications. This is denoted the Monte Carlo standard deviation:

\[
\text{MCSD}(\hat{\theta}) = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (\hat{\theta}_m - \text{MEAN}(\hat{\theta}))^2}.
\]

Notice, that \( \text{MEAN}(\hat{\theta}) \) is itself an estimator, and the uncertainty related to the estimator can be measured by the Monte Carlo standard error, defined as

\[
\text{MCSE} = M^{-\frac{1}{2}} \cdot \text{MCSD}(\hat{\theta}).
\]

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

The results from \textit{PcNaive} are reported in Figure 1 (A) for different sample lengths \( T \in \{10, 15, \ldots, 100\} \). The mean is estimated as in (12), and the dotted lines are

\[
\text{MEAN}(\hat{\theta}) \pm 2 \cdot \text{MCSD}(\hat{\theta}).
\]

These lines are interpretable as the average 95% confidence band in each replication and not the 95% confidence band for the estimated \( \text{MEAN}(\hat{\theta}) \). The mean of the estimated
parameter 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 $\text{MEAN}(\hat{\theta}) = 0.7935$. The MCSD($\hat{\theta}$) = 0.2172 which implies that the Monte Carlo standard error of this estimate is $\text{MCSE} = 5000^{-\frac{1}{2}} \cdot 0.2172 = 0.0031$. A $t$-test for the hypothesis that the estimate is unbiased, i.e.

$$H_0 : \text{MEAN}(\hat{\theta}) = 0.9,$$

can be constructed as

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

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 as expected from the consistency of OLS.

To illustrate how the bias depends on the true autoregressive parameter, we redo the simulation with other values of the autoregressive parameter: $\theta \in \{0, 0.3, 0.5, 0.7, 0.9\}$. The obtained results for the bias are depicted in Figure 1 (B). If the true DGP is static, $\theta = 0$, the estimator is unbiased. If $\theta > 0$ the estimator is downward biased, with a bias that increases with $\theta$.

### 2.3 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 $\hat{\beta}$. To derive this we need to impose restrictions on the variance of the error term. For the linear regression in

$$y_t = x'_t \beta + \epsilon_t,$$

the following result holds.
RESULT 4 (Asymptotic distribution) Again, let $y_t$ and $x_t$ obey Assumption 1, and assume that the regressors are predetermined so that (5) holds. If the error term is homoskedastic, i.e.

$$E[\varepsilon_t^2 \mid x_t] = \sigma^2,$$

(14)

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

$$E[\varepsilon_t \varepsilon_s \mid x_t, x_s] = 0,$$

(15)

then the OLS estimator is asymptotically normally distributed, so that

$$\sqrt{T}(\hat{\beta} - \beta) \rightarrow N(0, \sigma^2 E[x_t x'_t]^{-1}),$$

(16)

for $T \rightarrow \infty$.

The result implies that we can test hypotheses on $\beta$. Inserting natural estimators for $\sigma^2$ and $E[x_t x'_t]$, the distributional result in (16) can be written as

$$\hat{\beta} \overset{d}{\sim} N\left( \beta, \hat{\sigma}^2 \left( \sum_{t=1}^{T} x_t x'_t \right)^{-1} \right),$$

(17)

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, $\varepsilon_t$.

The precise formulation of the condition in (15) is a little difficult to interpret, and often we ignore the conditioning on $x_t$ and $x_s$ and consider whether $\varepsilon_t$ and $\varepsilon_s$ are uncorrelated for $t \neq s$.

An alternative way to relate to the condition of no-serial-correlation in (15) 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

$$E[y_t \mid x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \ldots, y_1, x_1] = E[y_t \mid x_t] = x'_t \beta,$$

(18)

so that $x_t$ contains all relevant information in the available information set: $x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \ldots, y_1, x_1$, then we refer to the regression model in (13) as being a complete dynamic model. Assuming the complete dynamic model in (18) is practically the same as the no-serial-correlation assumption in (15); 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 (13).

If the considered regression model (13) is dynamic, such as the AR(1) model in (3) or the ADL model in (4), 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 $\varepsilon_t$. 

8
3 AUTOCORRELATION OF THE ERROR TERM

In this section we discuss the case where the no-serial-correlation assumption in (15) 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 that we have autocorrelation of the error term. In practice, autocorrelation is detected by looking at the estimated residuals, $\hat{\epsilon}_t$ ($t = 1, 2, ..., T$), and $Cov(\hat{\epsilon}_t, \hat{\epsilon}_s) \neq 0$ is referred to as residual autocorrelation.

3.1 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 (5). To see this, consider an AR(1) model like (3), and assume that the error term exhibit autocorrelation of first order, i.e. that $\epsilon_t$ follows a first order autoregressive model,

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

where $v_t$ is an IID error term with constant variance. Consistency requires that $E[\epsilon_t y_{t-1}] = 0$, but that is clearly not satisfied since both $y_{t-1}$ and $\epsilon_t$ depends on $\epsilon_{t-1}$. 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 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 (17) 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 (2004, section 4.10.2). A simpler discussion of the univariate case is given in Stock and Watson (2003, p. 504-507).

3.2 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 are truly autoregressive.
(2) The model is dynamically misspecified.
(3) The functional form is misspecified.
(4) The model is subject to a non-modelled structural break, e.g. a level shift.

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

**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 (13) and (19), then it is natural to use the information in both equations to derive an estimator.

If $\rho$ is known, the two equations can be combined to yield

$$\left( y_t - \rho y_{t-1} \right) = \left( x'_t - \rho x'_{t-1} \right) \beta + \left( \epsilon_t - \rho \epsilon_{t-1} \right),$$

or equivalently

$$y_t = \rho y_{t-1} + x'_t \beta - x'_{t-1} \rho \beta + v_t,$$

where the error term $v_t$ is now serially uncorrelated from (19). In practice, the parameter $\rho$ is unknown but can be consistently estimated by running the regression (19) on the estimated residuals from (13).

The transformation to equation (21) is chosen to remove the problem of the original system and is analogous to the idea of GLS estimation in the case of heteroskedasticity. Note, however, that (21) 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, $\rho$ and $\beta$. The GLS transformation therefore implies a so-called *common factor restriction*, and the estimation model in (21) is non-linear, see Verbeek (2004, p. 99-100) for details.

Consistent estimation of the parameters in (20) requires the usual moment condition

$$E \left[ (x'_t - \rho x'_{t-1}) (\epsilon_t - \rho \epsilon_{t-1}) \right] = 0.$$

This implies that $\epsilon_{t-1}$ should be uncorrelated with $x_t$, i.e. that $E[\epsilon_t x_{t+1}] = 0$. This shows that consistency of GLS requires a stronger assumption than consistency of OLS, see Wooldridge (2003, p. 406ff).

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.

**Dynamic Misspecification.** Residual autocorrelation indicates that the model is not dynamically complete. If the model in (13) 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

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

and the natural remedy is to extend the list of variables in $x_t$ 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 (21). 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

$$y_t = \alpha_0 y_{t-1} + x'_t \alpha_1 + x'_{t-1} \alpha_2 + \eta_t, \quad (22)$$

where $\eta_t$ is a new error term. Here we do not take the structure of (19) 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 (21) and (22).

**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.

**Non-Modelled Structural Shift.** Another reason for systematic variation in the estimated residuals could be that the true DGP changes, so that the level of $y_t$ changes at some point in time. If we do not account for the break in the regression, then the predicted values, $\hat{y}_t = x'_t \hat{\beta}$, will correspond to an average between the level before and after the shift. As a consequence, the residuals are positive (on average) before the break and negative (on average) after (or the opposite). That, again, results in residual autocorrelation.

Here 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 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 seems 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 $\hat{\beta}$ is consistent for any true value $\beta$. 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 $\hat{\beta}_i \rightarrow 0$. 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, that 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 dataset, and try to reformulate the model to explain the particular neglected features of the data.

Below we present a number of standard misspecification tests.

4.1 No-Autocorrelation

Recall that residual autocorrelation can indicate many types of misspecification of a model, and a 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 (13). This is done by running the auxiliary regression model

$$\hat{\varepsilon}_t = x_t'\delta + \gamma \hat{\varepsilon}_{t-1} + u_t,$$

where $\hat{\varepsilon}_t$ is the estimated residual from (13) and $u_t$ is a new error term. The original explanatory variables, $x_t$, are included in (23) to allow for the fact that $x_t$ is not necessarily strictly exogenous and may be correlated with $\varepsilon_{t-1}$. 

12
The null hypothesis of no autocorrelation corresponds to $\gamma = 0$, and can be tested by the $t$–ratio on $\gamma$. Alternatively we can compute $T \cdot R^2$, where $R^2$ is the coefficient of determination in the auxiliary regression. Note, that the residual $\hat{\epsilon}_t$ is orthogonal to the explanatory variables $x_t$, and any explanatory power in the auxiliary regression must be due to the included lagged residual, $\hat{\epsilon}_{t-1}$. Under the null hypothesis the statistic is asymptotically distributed as

$$ T \cdot R^2 \to \chi^2(1). $$

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. $\hat{\epsilon}_1 = 0$, and estimate the auxiliary regression for the same sample as the original model.

The test in (24) is valid asymptotically, i.e. for $T \to \infty$. An alternative finite sample test exists; 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 in (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.

### 4.2 No-Heteroskedasticity

To test the assumption of no-heteroskedasticity in (14), 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:

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

The null hypothesis is unconditional homoskedasticity, $\gamma_1 = \ldots = \gamma_k = \delta_1 = \ldots = \delta_k = 0$, and the alternative is that the variance of $\epsilon_t$ depends on $x_{it}$ or the squares $x_{it}^2$ for some $i = 1, 2, ..., k$. Again the test is based on the LM statistic $T \cdot R^2$, which is distributed as a $\chi^2$ 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 heteroskedasticity, which implies systematic variation of the variance over time, denoted autoregressive conditional heteroskedasticity (ARCH) will be covered later.

### 4.3 RESET Test for Correct Functional Form

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

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

where $\hat{y}_t = x_t'\hat{\beta}$ is the predicted value of the original regression. The null hypothesis of correct specification is $\gamma = 0$. The alternative is that the square of $\hat{y}_t = x_t'\hat{\beta}$ 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 is the $F$-test for $\gamma = 0$ and it is distributed as $F(1, T - k - 1)$.

## 4.4 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 for the focus on the normal distribution is that the convergence of $\hat{\beta}$ to the asymptotic normal distribution will be faster if $\epsilon_t$ is close to normal. Furthermore, under normality of the error terms, least squares estimation coincides with maximum likelihood estimation, which implies a number of nice asymptotic properties. We will return to maximum likelihood 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 $T_0$ correspond 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 (SK), 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 $u_t = (\hat{\epsilon}_t - \bar{\epsilon})/\hat{\sigma}$ be the standardized residuals, where $\bar{\epsilon} = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t$ and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(\hat{\epsilon}_t - \bar{\epsilon})^2$ denote the estimated mean and variance. Skewness (SK) and kurtosis ($K$) are defined as the estimated third and fourth moments:

$$SK = T^{-1}\sum_{i=1}^{T}u_i^3 \quad \text{and} \quad K = T^{-1}\sum_{i=1}^{T}u_i^4.$$  

The normal distribution is symmetric and has a skewness of $SK = 0$. The normal distribution has a kurtosis measure of $K = 3$ and $K - 3$ 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 $SK \rightarrow N(0, 6 \cdot T^{-1})$ and $K \rightarrow N(3, 24 \cdot T^{-1})$. For testing it is convenient to use the $\chi^2$-versions, i.e.

$$\xi_{SK} = \frac{T}{6} \cdot SK^2 \rightarrow \chi^2(1)$$

$$\xi_{K} = \frac{T}{24} (K - 3)^2 \rightarrow \chi^2(1).$$

14
Furthermore, it turns out that $\xi_{SK}$ and $\xi_K$ are asymptotically independent, which makes it easy to construct a joint test. One such test is the Jarque-Bera test, which is just

$$JB = \xi_{SK} + \xi_K \rightarrow \chi^2(2).$$

In the output of PcGive this is denoted the asymptotic test. Several refinements have been suggested. One version, that corrects for a correlation between $\xi_{SK}$ and $\xi_K$ in finite samples, is implemented in PcGive.

5 **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 aggregate consumption in the private sector, $\Delta c_t$, disposable income, $\Delta y_t$, and wealth including owner occupied housing, $\Delta w_t$. The three time series are depicted in Figure 2 (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 unreasonable to invoke Assumption 1.

Running a static regression from the full sample, 1973 : 2 – 2003 : 2, yields the
estimated equation
\[
\Delta c_t = 0.0002 + 0.196 \cdot \Delta y_t + 0.559 \cdot \Delta w_t,
\]
with \(t\)-ratios in parentheses. In this equation \(R^2 = 0.208\), indicating that 20% of the variation in consumption growth is explained by the regressors.

The estimated residuals from equation (25) are depicted in Figure 2 (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

\[
\hat{\epsilon}_t = 0.0002 + 0.049 \cdot \Delta y_t - 0.083 \cdot \Delta w_t - 0.313 \cdot \hat{\epsilon}_{t-1} + 0.014 \cdot \hat{\epsilon}_{t-2},
\]
where \(\hat{\epsilon}_t\) denote the estimated residuals. The first lag, \(\hat{\epsilon}_{t-1}\), is significantly negative, with a coefficient of \(-0.313\) and a \(t\)-ratio of \(-3.24\), while the second lag, \(\hat{\epsilon}_{t-2}\), 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 \(LM = T \cdot R^2 = 121 \cdot 0.092 = 11.176\), 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.

A natural solution to the first order autocorrelation is to formulate the first order ADL model, i.e. augmenting the regression (25) with the first lag of all variables. Estimating the dynamic model for the longest possible sample, 1973 : 3 – 2003 : 2, yields

\[
\Delta c_t = 0.0004 - 0.314 \cdot \Delta c_{t-1} + 0.249 \cdot \Delta y_t + 0.044 \cdot \Delta y_{t-1} + 0.513 \cdot \Delta w_t + 0.201 \cdot \Delta w_{t-1}.
\]
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 \(R^2 = 0.295\). For the dynamic model in (26) 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 dynamic model appears dynamically complete.

The residuals are depicted in Figure 2 (C). There still seems to be a number of large residuals, and the histogram has fatter tails than the normal distribution. Looking more formally at the residuals, we find \(SK = -0.391\) indicating a skewness to the left compared to the normal distribution. A test for symmetry, \(SK = 0\), can be constructed as

\[
\xi_{SK} = \frac{T}{6} \cdot SK^2 = \frac{120}{6} \cdot (-0.391)^2 = 3.053,
\]
which is not significant in a \(\chi^2(1)\) distribution. The measure of excess kurtosis is \(K - 3 = 2.074\), indicating that the distribution of the residuals has fat tails. A test for \(K = 3\) can be based on the statistic

\[
\xi_K = \frac{T}{24} \cdot (K - 3)^2 = \frac{120}{24} \cdot (5.079 - 3)^2 = 21.503,
\]
Null hypothesis:  | Statistic     | Distribution |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) No autocorrelation of order 1-2</td>
<td>1.360 [0.507]</td>
<td>$\chi^2(2)$</td>
</tr>
<tr>
<td>(b) Normality</td>
<td>1.290 [0.524]</td>
<td>$\chi^2(2)$</td>
</tr>
<tr>
<td>(c) No heteroskedasticity</td>
<td>12.033 [0.525]</td>
<td>$\chi^2(13)$</td>
</tr>
<tr>
<td>(d) Correct functional form (RESET)</td>
<td>0.00001 [0.994]</td>
<td>$F(1,110)$</td>
</tr>
</tbody>
</table>

Table 1: Misspecification testing for the preferred model (27).

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 $SK = K - 3 = 0$ can be constructed as the Jarque-Bera statistic:

$$JB = \xi_{SK} + \xi_{K} = 3.053 + 21.503 = 24.556,$$

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

$$\Delta c_t = 0.001 - 0.284 \cdot \Delta c_{t-1} + 0.140 \cdot \Delta y_t + 0.043 \cdot \Delta y_{t-1} + 0.327 \cdot \Delta w_t + 0.314 \cdot \Delta w_{t-1}$$
$$- 0.020 \cdot \text{Dum}743_t + 0.069 \cdot \text{Dum}754_t - 0.068 \cdot \text{Dum}774_t. \tag{27}$$

For this model $R^2 = 0.535$, indicating that the dummies remove much of the variation in the estimated residuals. Comparing (27) with (26), we see that the coefficient to income growth is lowered by the dummy variables and $\Delta y_t$ 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.

For this preferred model the results for the misspecification tests are reported in Table 1. Row (a) is the LM test for no autocorrelation. The $p-$value is 0.507 and we cannot reject the null hypothesis of a well specified model. (b) is the Jarque-Bera test for normality of the estimated residuals. Here the $p-$value is 0.524 accepting the null hypothesis of normal residuals. (c) is the LM test for no heteroskedasticity based on a regression with the regressors in (27) and their squares (in total 13 non-redundant variables). The null-hypothesis of a well specified model is again accepted with a $p-$value of 0.525. (d) is the RESET test for correct functional form. The $p-$value of 0.994 gives no evidence against the linearity of the regression line.
To conclude, the preferred model in (27) appears to be well specified, in the sense that we reject the types of misspecifications considered in Table 1. 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 in (27) is that all variables have been transformed to stationarity by first differences, so that all the information contained 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


