# Regression analysis with time series data: Properties of the OLS estimator

Bernard Lejeune

Supplemental lecture notes IV
Advanced Econometrics
HEC-University of Liège
Academic year 2021-2022

- These lecture notes restate, in matrix form and with more details, the main results of Sections 10-3 and 11-2 of Wooldridge (2016).

## 1. Regression and time series data

- Regression analysis is not reserved to cross-sectional data. A multiple linear regression model such as:

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$$

$$\Leftrightarrow \quad y_t = X_t \beta + u_t$$

where $t = 1, ..., T$ is a time index, $X_t = (1, x_{t2}, ..., x_{tk})$ is a $1 \times k$ (row) vector of explanatory variables (including a constant) and $\beta = (\beta_1, \beta_2, ..., \beta_k)$ is a $k \times 1$ (column) vector, may likewise be used for analyzing time series data. As with cross-sectional data, stacking (with the correct temporal ordering) the observations of a sample of size $T$, we can write:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{T2} & \cdots & x_{Tk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix}$$

$$\Leftrightarrow \quad Y = X\beta + u$$

where $Y$ and $u$ are $T \times 1$ vectors, and $X$ is a $T \times k$ matrix, whose the $t$-th row is equal to $X_t$.

- Time series regression models may be used to evaluate the causal effect – i.e., the effect, the other factors being held constant – of some variables of interest on the dependent variable $y_t$, or more simply for forecasting. Depending on the case at hand and the question of interest, the set of independent variables $(x_{t2}, ..., x_{tk})$ may contain contemporaneous explanatory variables $z_t$ (static model), lagged explanatory variables $z_{t-1}$, $z_{t-2}, ...$ (distributed lag model), and/or lagged dependent variables $y_{t-1}$, $y_{t-2}, ...$ (autoregressive model).

- From a statistical point of view, the most distinctive feature of time series data is that the observations – which are viewed as realizations of a stochastic process – can almost never be assumed to be independent across time. In contrast, cross-sectional data are typically assumed to be obtained by random sampling in some population, which implies that they are independent across individuals. This independence property does no longer hold with time series data, and this is the basic reason why regression with time series data requires special attention when applying OLS.

# 2. Finite sample properties of OLS

- Following Wooldridge (2016), Section 10-3, the finite sample properties of the OLS estimator:

$$\hat{\beta} = (X'X)^{-1} X'Y = \left( \sum_{t=1}^{T} X_t' X_t \right)^{-1} \sum_{t=1}^{T} X_t' y_t$$

are exactly the same as in the cross-sectional case under the following assumptions:

  - TS.1 Linearity in parameters

    The available data are realizations of a stochastic process $\{(x_{t2}, ..., x_{tk}, y_t)$: $t = 1, ..., T\}$ following the linear model:

$$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$$

    where $(\beta_1, ..., \beta_k)$ are unknown parameters and $\{u_t$: $t = 1, ..., T\}$ is a sequence of error.

  - TS.2 No perfect collinearity

    In the sample (and thus in the underlying time series process), none of the explanatory variables $(x_{t2}, ..., x_{tk})$ is constant, and there is no exact linear relationship among them.

  - TS.3 Zero conditional mean

    For each $t$, the expected value of $u_t$ given any values of $X \equiv (X_1, ..., X_T)$ is equal to zero, which is equivalent to say that the expected value of $y_t$ given any values of $X \equiv (X_1, ..., X_T)$ is equal to $\beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} = X_t\beta$:

$$E(u_t|X) = 0 \;\Leftrightarrow\; E(y_t|X) = X_t\beta\,, \quad t = 1, ..., T$$

– TS.4 Homoskedasticity

For each $t$, the variance of $u_t$ given any values of $X \equiv (X_1, ..., X_T)$ is constant, which is equivalent to say that the variance of $y_t$ given any values of $X \equiv (X_1, ..., X_T)$ is constant:

$$Var(u_t|X) = \sigma^2 \;\Leftrightarrow\; Var(y_t|X) = \sigma^2\,, \quad t = 1, ..., T$$

where $\sigma^2$ is a unknown parameter.

– TS.5 No serial correlation

For all $t \neq s$, the covariance between of $u_t$ and $u_s$ given any values of $X \equiv (X_1, ..., X_T)$ is equal to zero, which is equivalent to say that the covariance between of $y_t$ and $y_s$ given any values of $X \equiv (X_1, ..., X_T)$ is equal to zero:

$$Cov(u_t, u_s|X) = 0 \;\Leftrightarrow\; Cov(y_t, y_s|X) = 0\,, \quad \text{for all } t \neq s$$

– TS.6 Normality

The distribution of the errors $u_t$ is the same given any values of $X \equiv (X_1, ..., X_T)$ – i.e., the errors $u_t$ are independent of $X$ – and they are independently[1] normally distributed with zero mean and variance equal to $\sigma^2$, which is equivalent to say that, given any value of $X \equiv (X_1, ..., X_T)$, the dependent variables $y_t$ are likewise independently normally distributed with mean equal to $X_t\beta$ and variance equal to $\sigma^2$:

$$u_t|X \sim N(0, \sigma^2) \;\Leftrightarrow\; y_t|X \sim N(X_t\beta, \sigma^2), \quad t = 1, ..., T$$

- The above set of assumptions deserves several comments:

  – Assumption TS.1 is the basically same as assumption MLR.1 for the cross-sectional case. Written in matrix form, it is the same as the linearity in parameters assumption E.1: $Y = X\beta + u$ stated in the supplemental lecture notes I (hereafter SNL-I).

  – Likewise, assumption TS.2 is the same as assumption MLR.3 for the cross-sectional case. Written in matrix form, it is likewise the same as the no perfect collinearity assumption E.2: $\text{rank}(X) = k$ stated in SNL-I.

  – Assumption TS.3 is a time series analog of assumption MLR.4 for the cross-sectional case. As assumption MLR.4, it says that the systematic part of the model in TS.1 is the conditional mean of $y_t$ given $X_t$, i.e., $E(y_t|X_t) = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} = X_t\beta$. But this assumption says more: it also requires that the conditional mean of $y_t$ given the explanatory variables of all time periods $X \equiv (X_1, ..., X_T)$ does actually only depend on the explanatory variables of the contemporaneous period $X_t$:

$$E(y_t|X) = E(y_t|X_1, ..., X_T) = E(y_t|X_t) = X_t\beta, \quad t = 1, ..., T$$

---

[1] Note that under (joint) normality, independence and zero correlation are equivalent.

When this assumption – which may equivalently be written as $E(u_t|X) = E(u_t|X_1, ..., X_T) = E(u_t|X_t) = 0,\ t = 1, ..., T$ – holds, it is said that the explanatory variables are 'strictly exogenous'. In the cross-sectional case, this strict exogeneity assumption was automatically satisfied, as a result of the random sampling assumption MLR.2, which implies independence across individuals. Assumption TS.3 may be viewed as retaining the strict exogeneity assumption implicit in the cross-sectional case – which is needed for OLS to have the same finite sample properties – without maintaining an independence assumption, i.e., while allowing for (arbitrary) dependence across time.
Written in matrix form, assumption TS.3 is the same as the zero conditional mean assumption E.3: $E(u|X) = 0 \Leftrightarrow E(Y|X) = X\beta$ stated in SNL-I.

- Similarly to assumption MLR.4 in the cross-sectional case, by the law of iterated expectations, the zero conditional mean assumption TS.3 implies that the unconditional mean of $u_t$ is zero (i.e., $E(u_t) = 0,\ t = 1, ..., T$), and that $u_t$ is uncorrelated (have zero covariance) with each explanatory variable, in each time period (i.e., $E(x_{sj}u_t) = 0,\ s, t = 1, ..T;\ j = 2, ..., k)$).

- The strict exogeneity assumption included in assumption TS.3 is much more restrictive than it might seem at first sight. On the one hand, it does not allow for the possibility of feedback from the current value of the dependent variable to the future values of the explanatory variables. For example, in the simple static model:

$$y_t = \beta_1 + \beta_2 z_t + u_t = X_t\beta + u_t, \quad \text{where } X_t = [\ 1 \quad z_t\ ]$$

we can not have that $z_{t+1}$ is partly determined by $y_t$. If it was the case, the strict exogeneity assumption:

$$E(y_t|X) = E(y_t|z_1, ...z_T) = E(y_t|z_t)$$

would certainly be violated[2]. On the other hand, it does not allow for models with lagged dependent variables. As a matter of fact, for the first order autoregressive model:

$$y_t = \beta_1 + \beta_2 y_{t-1} + u_t = X_t\beta + u_t, \quad \text{where } X_t = [\ 1 \quad y_{t-1}\ ]$$

we have:

$$\begin{aligned} & E(y_t|X) = E(y_t|y_0, ..., y_{T-1}) = y_t \\ \neq \quad & E(y_t|X_t) = E(y_t|y_{t-1}) = \beta_1 + \beta_2 y_{t-1} \end{aligned}$$

i.e., the strict exogeneity assumption is by construction violated.

- Assumption TS.4 is a time series analog of the homoskedasticity assumption MLR.5 for the cross-sectional case. Also, assumption TS.5 may be

[2] i.e., $y_t$ would certainly depends not only on $z_t$, but also on $z_{t+1}$. See Wooldridge (2016) p. 319 for a practical example. Note that, in this simple static model, the strict exogeneity assumption also forbids that $y_t$ depends on $z_{t-1}$, $z_{t-2}, ....$ But this is usually not really a concern: it can simply be circumvented by adding enough lags of $z_t$ in the original model, i.e., by estimating a distributed lag model.

viewed as a time series analog of the no correlation across individuals property which automatically holds in the cross-sectional case, as a result of the random sampling assumption MLR.2, which implies independence across individuals. Note that, similarly to assumption TS.3, both the constant variance and the no serial correlation assumption are assumed to hold not only for any value of the contemporaneous explanatory variables, but for any value of the explanatory variables of all time periods $X \equiv (X_1, ..., X_T)$. Note also that if assumption TS.5 restrict the temporal dependence of $u_t$, and equivalently of $y_t$, conditional on $X$, it assumes nothing – i.e., puts no restriction – about the temporal dependence of the explanatory variables. Written in matrix form, jointly considered, assumption TS.4 and assumption TS.5 are the same as the homoskedasticity & no correlation assumption E.4: $V(u|X) = \sigma^2 I \Leftrightarrow V(Y|X) = \sigma^2 I$ stated in SLN-I.

– Finally, assumption TS.6 is likewise a time series analog of the normality assumption MLR.6 for the cross-sectional case. Written in matrix form, it is likewise the same as the normality assumption E.5: $u|X \sim N(0, \sigma^2 I) \Leftrightarrow Y|X \sim N(X\beta, \sigma^2 I)$ stated in SLN-I.

- As outlined in the above discussion, assumptions TS.1 – TS.6 are the same as the so-called 'classical linear model' assumptions E.1 – E.5, i.e., the assumptions under which we actually established the finite sample properties of the OLS estimator in the cross-sectional case[3]. Under assumptions TS.1 – TS.6, the OLS estimator has thus the same finite sample properties as in the cross-sectional case. More specifically:

  – Under assumptions TS.1 – TS.3 (linearity in parameters, no perfect colinearity and zero conditional mean, which are the same as assumptions E.1 – E.3), from Property 1 in SNL-I, $\hat{\beta}$ is an unbiased estimator[4] of $\beta$.

  – If assumptions TS.4 – TS.5 (homoskedasticity and no serial correlation, which are the same as assumption E.4) are added to assumptions TS.1 – TS.3, from Property 2 and 3 in SNL-I, the (conditional) variance-covariance matrix of $\hat{\beta}$ is given[5] by $V(\hat{\beta}|X) = \sigma^2 (X'X)^{-1}$ and $\hat{\beta}$ is the best linear unbiased estimator[6] (BLUE) of $\beta$. Also, from Property 5 SNL-I, $\hat{s}^2 = \frac{1}{T-k} \sum_{t=1}^{T} \hat{u}_t^2 = \frac{\hat{u}'\hat{u}}{T-k}$ is an unbiased estimator[7] of $\sigma^2$.

  – If assumption TS.6 (normality, which is the same as assumption E.5) is added to assumptions TS.1 – TS.5, from Property 4 in SNL-I,, the (conditional) distribution of $\hat{\beta}$ is normal and given[8] by $\hat{\beta}|X \sim N(\beta, \sigma^2 (X'X)^{-1})$.

[3] In the cross-sectional case, we used these assumptions because, on the one hand, they are more convenient to work with, and on the other hand, they hold whenever the seminal assumptions MLR.1 – MLR.6 hold (i.e., the seminal assumptions MLR.1 – MLR.6 imply assumptions E.1 – E.5; note that the converse is not true).

[4] This property is the same as Theorem 10.1 in Wooldridge (2016), Section 10-3.

[5] This property is the same as Theorem 10.2 in Wooldridge (2016), Section 10-3.

[6] This property is the same as Theorem 10.4 in Wooldridge (2016), Section 10-3.

[7] This property is the same as Theorem 10.3 in Wooldridge (2016), Section 10-3.

[8] This property is the same as Theorem 10.5 in Wooldridge (2016), Section 10-3.

- Needless to say, as under assumptions TS.1 – TS.6 the finite sample properties of the OLS estimator are the same as in the cross-sectional case, all usual inference procedures derived in Section 4.1, 4.2 and 4.4.1 of SNL-I – i.e., the confidence intervals for $\beta_j$ or a single linear combination $R_0\beta$, the two-sided and one-sided $t$-tests of $\beta_j$ or a single linear combination $R_0\beta$, and the $F$-test of multiple linear restrictions – are of course valid and exact in finite sample.

# 3. Asymptotic properties of OLS

- The classical assumptions TS.1 – TS.6, which are equivalent to the matrix assumptions E.1 – E.5, are pretty restrictive for time series applications. This is in particular the case of the strict exogeneity assumption included in the zero conditional mean assumption TS.3 – which allows neither for the possibility of feedback from the current value of the dependent variable to the future values of the explanatory variables, nor for lagged dependent variables – and of the normality assumption TS.6.

- Fortunately, these somewhat restrictive assumptions are not needed for the OLS estimator to again actually have the same asymptotic properties as in the cross-sectional case. This however comes at a price : it requires the dependence of the observations across time not to be too strong, so that law of large numbers and central limit theorem for dependent data hold.

- Following Wooldridge (2016), Section 11-2, the asymptotic properties of the OLS estimator :
$$\hat{\beta} = (X'X)^{-1} X'Y = \left( \sum_{t=1}^{T} X_t' X_t \right)^{-1} \sum_{t=1}^{T} X_t' y_t$$
are again exactly the same as in the cross-sectional case under the following assumptions :

  - TS.1' Linearity in parameters and weak dependence

    The available data are realizations of a stochastic process $\{(x_{t2}, ..., x_{tk}, y_t)$: $t = 1, ..., T\}$ which is stationary, weakly dependent and follows the linear model :
    $$y_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} + u_t$$
    where $(\beta_1, ..., \beta_k)$ are unknown parameters and $\{u_t$: $t = 1, ..., T\}$ is a sequence of error.

  - TS.2' No perfect collinearity

    In the sample (and thus in the underlying time series process), none of the explanatory variables $(x_{t2}, ..., x_{tk})$ is constant, and there is no exact linear relationship among them.

  - TS.3' Zero conditional mean

    For each $t$, the expected value of $u_t$ given any values of $X_t = (1, x_{t2}, ..., x_{tk})$ is equal to zero, which is equivalent to say that the expected value of $y_t$

given any values of $X_t = (1, x_{t2}, ..., x_{tk})$ is equal to $\beta_1 + \beta_2 x_{t2} + ... + \beta_k x_{tk} = X_t\beta$ :

$$E(u_t|X_t) = 0 \Leftrightarrow E(y_t|X_t) = X_t\beta \,, \quad t = 1, ..., T$$

– TS.4' Homoskedasticity

For each $t$, the variance of $u_t$ given any values of $X_t = (1, x_{t2}, ..., x_{tk})$ is constant, which is equivalent to say that the variance of $y_t$ given any values of $X_t = (1, x_{t2}, ..., x_{tk})$ is constant :

$$Var(u_t|X_t) = \sigma^2 \Leftrightarrow Var(y_t|X_t) = \sigma^2 \,, \quad t = 1, ..., T$$

where $\sigma^2$ is a unknown parameter.

– TS.5' No serial correlation

For all $t \neq s$, the expected value of the cross-product between of $u_t$ and $u_s$ given any values of $(X_t, X_s) = (1, x_{t2}, ..., x_{tk}, x_{s2}, ..., x_{sk})$ is equal to zero :

$$E(u_t u_s|X_t, X_s) = 0 \,, \quad \text{for all } t \neq s$$

- The above set of assumptions deserves several comments :
  - Assumption TS.1' is the same as assumption TS.1 plus the assumption that the stochastic process $\{(x_{t2}, ..., x_{tk}, y_t) : t = 1, ..., T\}$ is stationary and weakly dependent. It may be viewed as the time series analogue of assumptions MLR.1 and MLR.2, where the random sampling assumption MLR.2 is replaced by an assumption of stationarity and weak dependence.
  - The stationarity assumption included in assumption TS.1' is mainly for convenience and is not crucial. As matter of fact, the asymptotic properties outlined hereafter also hold if the data are trended, provided appropriate time trends are included in the model. A similar statement holds for data with seasonality. See Wooldridge (2016) p. 348 for a short discussion.
  - In contrast, the weak dependence assumption included in assumption TS.1' is of primary importance : it is needed for law of large numbers and central limit theorem on which are based the asymptotic properties of OLS to hold. Intuitively, weak dependence requires that the correlation between observations at time $t$ and time $t + h$ goes to zero sufficiently quickly as $h \to \infty$. For a discussion and examples of weakly dependent time series, see Wooldridge (2016), Section 11-1. For a discussion and examples of time series which are not weakly dependent (they are called 'highly persistent' or 'strongly dependent'), and how such highly persistent time series can be transformed into a weakly dependent series by differencing, see Wooldridge (2016), Section 11-3.
  - Assumption TS.2' is exactly the same as assumption TS.2, as well as the same as assumption MLR.3 for the cross-sectional case.
  - Assumption TS.3' is the same as assumption TS.3, but without the strict

exogeneity assumption. It is the exact analogue of assumption MLR.4 for the cross-sectional case. As assumption MLR.4, by the law of iterated expectations, the zero conditional mean assumption TS.3' implies that the unconditional mean of $u_t$ is zero (i.e., $E(u_t) = 0$, $t = 1, ..., T$), and that $u_t$ is uncorrelated (have zero covariance) with each explanatory variable in the same (and only the same) time period (i.e., $E(x_{tj}u_t) = 0$, $t = 1, ..T$; $j = 2, ..., k$)).

– Assumption TS.4' is the same as assumption TS.4, except that the constant variance is assumed to hold only for any value of the contemporaneous explanatory variables $X_t$, and not for any value of the explanatory variables of all time periods $X \equiv (X_1, ..., X_T)$. It is the exact analogue of assumption MLR.5 for the cross-sectional case.

– Assumption TS.5' is a little bit more tricky to grasp. Essentially, it may be viewed as the same as assumption TS.5 – which under assumption TS.3 is equivalent[9] to $E(u_t u_s|X) = 0$, for all $t \neq s$ –, except that it entails conditioning only on the explanatory variables $X_t$ and $X_s$ in time periods coinciding with $u_t$ and $u_s$ rather than on the explanatory variables $X \equiv (X_1, ..., X_T)$ of all time periods. Likewise, it may be viewed as the time series analog of the no correlation across individuals property which automatically holds in the cross-sectional case, as a result of the random sampling assumption MLR.2, which implies independence across individuals. Formally, assumption TS.5' is however not equivalent to $Cov(u_t, u_s|X_t, X_s) = 0$, for all $t \neq s$: for this equivalence to hold, we would need that $E(u_t|X_t, X_s) = E(u_s|X_t, X_s) = 0$ holds[10], i.e., a stronger assumption (including strict exogeneity) than assumption TS.3'. We will see hereafter in Section 3.3 that assumption TS.5' may formally be interpreted in terms of 'dynamic completeness' of the model.

– In a nutshell, assumptions TS.1' – TS.5' may be viewed as a version of assumptions TS.1 – TS.5 where the strict exogeneity assumption has been traded for a (stationarity and) weak dependence assumption.

– On the other hand, assumptions TS.1' – TS.5' are the same as the seminal cross-sectional assumptions MLR.1 – MLR.5 with the random sampling assumption replaced by the weaker stationarity, weak dependence and no serial correlation assumptions. As matter of fact, assumptions MLR.1 – MLR.5 are just a special case[11] of assumptions TS.1' – TS.5'.

---

[9] By definition, $Cov(u_t, u_s|X) = E(u_t u_s|X) - E(u_t|X)E(u_s|X)$, so that under Assumption TS.3 $E(u_t|X) = E(u_s|X) = 0$, we have $Cov(u_t, u_s|X) = E(u_t u_s|X) = 0$.

[10] By definition, $Cov(u_t, u_s|X_t, X_s) = E(u_t u_s|X_t, X_s) - E(u_t|X_t, X_s)E(u_s|X_t, X_s)$, so that for having $Cov(u_t, u_s|X_t, X_s) = E(u_t u_s|X_t, X_s) = 0$ we would need that $E(u_t|X_t, X_s) = E(u_s|X_t, X_s) = 0$ holds, while assumption TS.3' only ensures that $E(u_t|X_t) = E(u_s|X_s) = 0$.

[11] In other words, assumptions MLR.1 – MLR.5 imply assumptions TS.1' – TS.5'.

## 3.1. Consistency of $\hat{\beta}$

- We have the following property[12] :

  Property 12 Consistency of $\hat{\beta}$
  Under assumptions TS.1' – TS.3', the OLS estimator $\hat{\beta}$ is a consistent estimator of $\beta$ :

$$\hat{\beta} \xrightarrow{p} \beta$$

  A sketch of the proof is as follows. It is basically the same as in the cross-sectional case. Under assumptions TS.1' – TS.2', we have :

$$\hat{\beta} = \left(\sum_{t=1}^{T} X_t'X_t\right)^{-1} \sum_{t=1}^{T} X_t'y_t = \beta + \left(\frac{1}{T}\sum_{t=1}^{T} X_t'X_t\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} X_t'u_t\right) \qquad (1)$$

  Under the stationarity and weak dependence assumptions, the observations $(X_t, y_t)$ are identically but non independently distributed (i.n.i.d.) across $t$, so that both $X_t'X_t$ and $X_t'u_t$ are likewise i.n.i.d. across $t$. If, as supposed by the weak dependence assumption, the dependence between observations at time $t$ and time $t+h$ goes to zero sufficiently quickly as $h \to \infty$, then, as in the cross-sectional case, a law of large numbers (LLN) can be applied to both sample average $\frac{1}{T}\sum_{t=1}^{T} X_t'X_t$ and $\frac{1}{T}\sum_{t=1}^{T} X_t'u_t$.
  If $\{Z_t\colon\ t = 1, ..., T\}$ is a stationary and weakly dependent stochastic process with $E(Z_t) = m$, then by the LLN we have[13] :

$$\bar{Z}_T = \frac{1}{T}\sum_{t=1}^{T} Z_t \xrightarrow{p} m$$

  Under the zero conditional mean assumption TS.3' $E(u_t|X_t) = 0$, by the law of iterated expectations, we have :

$$E\left(X_t'u_t\right) = E\left[E\left(X_t'u_t|X_t\right)\right] = E\left[X_t'E\left(u_t|X_t\right)\right] = E\left[X_t' \cdot 0\right] = 0 \qquad (2)$$

  Noting $E(X_t'X_t) = A$, from the LLN, we thus have :

$$\frac{1}{T}\sum_{t=1}^{T} X_t'X_t \xrightarrow{p} A \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} X_t'u_t \xrightarrow{p} 0$$

  so that, from (1), we finally have :

$$\hat{\beta} \xrightarrow{p} \beta + A^{-1} \cdot 0 = \beta$$

- As in the cross-sectional case, equation (2) in the above proof suggests that $\hat{\beta}$ would be consistent for $\beta$ under a weaker assumption than the zero conditional mean assumption TS.3'. It is actually sufficient to have $E\left(X_t'u_t\right) = 0$, i.e., the assumption TS.3" :

[12] This property is the same as Theorem 11.1 in Wooldridge (2016), Section 11-2. For a sketch of the proof similar to the one developed below, see Wooldridge (2016) p. 728-729. For a more detailed and rigorous treatment, see Hayashi (2000).

[13] This holds for $Z_t$ being a scalar, a vector or a matrix.

TS.3" Zero mean and zero correlation

For each $t$, the expected value of $u_t$ is zero and $u_t$ is uncorrelated with each explanatory variable $(x_{t2}, ..., x_{tk})$:

$$E(u_t) = 0 \text{ and } Cov(x_{tj}, u_t) = 0, \quad t = 1, ..T; j = 2, ..., k$$

Property 12' Consistency of $\hat{\beta}$ (bis)

Under assumptions TS.1' – TS.2' and assumption TS.3", the OLS estimator $\hat{\beta}$ is a consistent estimator of $\beta$:

$$\hat{\beta} \xrightarrow{p} \beta$$

## 3.2. Asymptotic normality of $\hat{\beta}$ and inference

- We have the following property[14]:

Property 13 Asymptotic normality of $\hat{\beta}$

Under assumptions TS.1' – TS.5', the OLS estimator $\hat{\beta}$ is asymptotically normally distributed:

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 A^{-1}), \quad \text{where } A = E(X_t' X_t) \tag{3}$$

A sketch of the proof is as follows. Besides additional complications due to the dependence of the observations across time, it is again basically the same as in the cross-sectional case. Under assumptions TS.1' – TS.2', from (1), we have:

$$\hat{\beta} = \beta + \left( \frac{1}{T} \sum_{t=1}^{T} X_t' X_t \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} X_t' u_t \right)$$

$$\Leftrightarrow \sqrt{T}(\hat{\beta} - \beta) = \left( \frac{1}{T} \sum_{t=1}^{T} X_t' X_t \right)^{-1} \left( T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t \right) \tag{4}$$

As already outlined, under the stationarity and weak dependence assumptions, both $X_t' X_t$ and $X_t' u_t$ are identically but non independently distributed (i.n.i.d.) across $t$, and the dependence between observations at time $t$ and time $t+h$ goes to zero sufficiently quickly as $h \to \infty$, so that, as in the cross-sectional case, a law of large numbers but also a central limit theorem can be applied to sample averages. From the law of large numbers (LLN), we have $\frac{1}{T} \sum_{t=1}^{T} X_t' X_t \xrightarrow{p} E(X_t' X_t) = A$, and we can write:

$$\sqrt{T}(\hat{\beta} - \beta) \overset{as}{=} A^{-1} \left( T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t \right) \tag{5}$$

where $\overset{as}{=}$ means 'asymptotically equivalent', so that $\sqrt{n}(\hat{\beta} - \beta)$ has asymptotically the same distribution as $A^{-1} \left( T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t \right)$.

[14] This property is the same as Theorem 11.2 in Wooldridge (2016), Section 11-2. For a sketch of the proof similar to the one developed below, see Wooldridge (2016) p. 729-730. For a more detailed and rigorous treatment, see Hayashi (2000).

Now, if $\{Z_t\colon\ t = 1, ..., T\}$ is a $(k \times 1)$ stationary and weakly dependent stochastic process with $E(Z_t) = m$, then by the central limit theorem (CLT) we have[15] :

$$\sqrt{T}(\bar{Z}_T - m) = T^{-\frac{1}{2}} \sum_{t=1}^{T} (Z_t - m) \xrightarrow{d} N(0, \Sigma), \text{ where } \Sigma = V\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} Z_t\right)$$

As already outlined, under the zero conditional mean assumption TS.3' $E(u_t|X_t) = 0$, from (2), we have $E(X_t' u_t) = 0$. From the CLT, we thus have:

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t \xrightarrow{d} N(0, B), \quad \text{where } B = V\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t\right)$$

so that[16] :

$$A^{-1}\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t\right) \xrightarrow{d} N(0, A^{-1} B A^{-1})$$

and thus, from (5):

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1}) \tag{6}$$

To complete the proof, it remains to show that, under the homoskedasticity and the no serial correlation assumptions TS.4' and TS.5', we have $B = \sigma^2 A$. The variance-covariance matrix $B$ can be written:

$$\begin{aligned} B &= V\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t\right) = E\left[\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t\right)\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t\right)'\right] \\ &= E\left[\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} X_t' u_t u_s X_s\right] = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E(u_t u_s X_t' X_s) \end{aligned} \tag{7}$$

Under the no serial correlation assumption TS.5' $E(u_t u_s|X_t, X_s) = 0$, for all $t \neq s$, by the law of iterated expectations, we have:

$$\begin{aligned} E(u_t u_s X_t' X_s) &= E\left[E(u_t u_s X_t' X_s|X_t, X_s)\right] \\ &= E\left[E(u_t u_s|X_t, X_s) X_t' X_s\right] \\ &= E\left[0 \cdot X_t' X_s\right] = 0 \end{aligned}$$

so that all terms with $t \neq s$ in (7) are equal to zero, and $B$ is thus equal to:

$$B = \frac{1}{T} \sum_{t=1}^{T} E(u_t^2 X_t' X_t)$$

Further, under the homoskedasticity assumption TS.4' $Var(u_t|X_t) = E(u_t^2|X_t) = \sigma^2$, for all $t$, by the law of iterated expectations, we have:

---

[15] Note that if $Z_t$ is identically and independently distributed (i.i.d.), then $\Sigma = V\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} Z_t\right) = \frac{1}{T} \sum_{t=1}^{T} V(Z_t) = V(Z_t)$.

[16] because a linear function of jointly normally distributed random variables is itself normally distributed, and $A^{-1}$ is a symmetric matrix.

$$\begin{aligned} E(u_t^2 X_t' X_t) &= E\left[E(u_t^2 X_t' X_t | X_t)\right] = E\left[E(u_t^2 | X_t) X_t' X_t\right] \\ &= E\left[\sigma^2 X_t' X_t\right] = \sigma^2 E(X_t' X_t) = \sigma^2 A \end{aligned}$$

so that $B$ is equal to:

$$B = \frac{1}{T} \sum_{t=1}^{T} \sigma^2 A = \sigma^2 A$$

and, from (6), we finally have:

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 A^{-1})$$

- The limiting distributional result (3) is the exact analogue of the limiting distributional result given by Property 7 in SNL-I for the cross-sectional case[17]. As in the cross-sectional case, it provides an approximate finite sample distribution for the OLS estimator $\hat{\beta}$:

$$\sqrt{n}(\hat{\beta} - \beta) \approx N(0, \sigma^2 A^{-1})$$

$$\Leftrightarrow \quad \hat{\beta} \approx N(\beta, \sigma^2 A^{-1}/T) \qquad (8)$$

  which can be used – when $n$ is sufficiently large – for performing inference (confidence interval, hypothesis testing) without having to rely on any other assumption than assumptions TS.1'–TS.5', i.e., in particular without having to rely on any normality assumption.

- As in the cross-section case, an estimator of the asymptotic variance $Avar(\hat{\beta}) = \sigma^2 A^{-1}/T$ is simply obtained by replacing $\sigma^2$ and $A$ by consistent estimators. From the LLN, a consistent estimator of $A = E(X_t' X_t)$ is given by $\frac{1}{T}\sum_{t=1}^{T} X_t' X_t = X'X/T$, and a consistent estimator of $\sigma^2$ is likewise still given[18] by $\hat{s}^2 = \frac{1}{T-k}\sum_{t=1}^{T} \hat{u}_t^2$, where $\hat{u}_t = y_t - X_t\hat{\beta}$:

  Property 14 Consistency of $\hat{s}^2$
  Under TS.1'–TS.5', $\hat{s}^2$ is a consistent estimator of $\sigma^2$:

$$\hat{s}^2 \xrightarrow{p} \sigma^2$$

  With $\hat{s}^2$ and $X'X/T$ as consistent estimators of $\sigma^2$ and $A$, an estimator of $Avar(\hat{\beta}) = \sigma^2 A^{-1}/T$ is given by:

$$\hat{V}(\hat{\beta}) = \hat{s}^2 \left(X'X\right)^{-1} \qquad (9)$$

- As the limiting distributional result (3), or equivalently the approximate distributional result (8), and the estimator (9) of the asymptotic variance $Avar(\hat{\beta})$ are the same as in the cross-sectional case, from the results derived in Section

[17] As a matter of fact, Property 7 in SNL-I is just a special case of Property 13.
[18] For a detailed proof, see Hayashi (2000) p. 115-116.

4.3 and Section 4.4.2 of SNL-I, it follows that all usual inference procedures – confidence interval for $\beta_j$ or a single linear combination $R_0\beta$, two-sided and one-sided $t$-tests of $\beta_j$ or a single linear combination $R_0\beta$, $F$-test (or Wald test) of multiple linear restrictions – are likewise asymptotically valid – i.e., approximately valid for $T$ sufficiently large – in the present time series context under assumptions TS.1'–TS.5'.

## 3.3. Heteroskedasticity and autocorrelation robust inference

- The outlined above asymptotic properties of the OLS estimator does not rely on any normality assumption, but they require both homoskedasticity and no serial correlation. If the homoskedasticity assumption TS.4' and/or the no serial correlation assumption TS.5' do not hold, then the OLS estimator is still consistent (only assumptions TS.1'–TS.3' are required for consistency), but all usual inference procedures are no longer valid.

- In time series, heteroskedasticity is usually a lower concern than with cross-sectional data[19]. On the other hand, serial correlation is often an issue, unless the model is dynamically complete : when a model is dynamically complete, then the no serial correlation assumption TS.5' is automatically satisfied. As matter of fact, a regression model :

$$y_t = X_t\beta + u_t, \quad t = 1, ..., T \tag{10}$$

is said to be dynamically complete if, for each $t$, we have :

$$E(y_t|X_t, y_{t-1}, X_{t-1}, ..., y_1, X_1) = E(y_t|X_t) = X_t\beta, \quad t = 1, ..., T \tag{11}$$

or equivalently :

$$E(u_t|X_t, y_{t-1}, X_{t-1}, ..., y_1, X_1) = E(u_t|X_t) = 0, \quad t = 1, ..., T \tag{12}$$

In words, the regression model (10) is dynamically complete if enough lags has been included in $X_t$ – which in all generality may contain contemporaneous explanatory variables $z_t$, lagged explanatory variables $z_{t-1}$, $z_{t-2}, ...$ and/or lagged dependent variables $y_{t-1}$, $y_{t-2}, ...$ –, so that adding further lags of $y$ and the explanatory variables do not matter for explaining $y_t$. For example[20], the autoregressive distributed lag model :

$$y_t = \beta_1 + \beta_2 z_t + +\beta_3 y_{t-1} + \beta_2 z_{t-1} + u_t = X_t\beta + u_t, \quad t = 1, ..., T$$

where $X_t = [1 \;\; z_t \;\; y_{t-1} \;\; z_{t-1}]$, is dynamically complete if we have :

$$\begin{aligned} E(y_t|z_t, y_{t-1}, z_{t-1}, ..., y_1, z_1) &= E(y_t|z_t, y_{t-1}, z_{t-1}), \quad t = 1, ..., T \\ &= \beta_1 + \beta_2 z_t + +\beta_3 y_{t-1} + \beta_2 z_{t-1} \end{aligned}$$

If (12) hold – i.e., if the model is dynamically complete –, then the no serial cor-

[19] Financial time series is a notable exception.

[20] For a discussion and other examples, see Wooldridge (2016), Section 11-4.

relation assumption TS.5' automatically hold. The proof is as follows. Suppose that $s < t$, by the law of iterated expectations, we have:

$$\begin{aligned} E(u_t u_s | X_t, X_s) &= E\left[E(u_t u_s | X_t, X_s, u_s) | X_t, X_s\right] \\ &= E\left[u_s E(u_t | X_t, X_s, u_s) | X_t, X_s\right] \end{aligned}$$

Because $s < t$ and $u_s$ is a function of $y_s$ and $X_s$, $(X_t, X_s, u_s)$ is a subset of the conditioning set in (12). Therefore, (12) imply that $E(u_t | X_t, X_s, u_s) = 0$, and thus:

$$E(u_t u_s | X_t, X_s) = E\left[u_s \cdot 0 | X_t, X_s\right] = 0$$

so that, for all $t \neq s$, we have:

$$E(u_t u_s | X_t, X_s) = 0$$

- When a model is not dynamically complete – it is most often the case of static and distributed lag models[21] –, then, along with the possible presence of heteroskedasticity, the no serial correlation assumption TS.5' is generally violated, and thus all usual inference procedures are no longer valid. However, as for heteroskedasticity (only) in the cross-sectional case, it is possible to derive inference procedures which are valid in the presence of both heteroskedasticity and serial correlation of unknown form, i.e., which are robust to arbitrary form of both heteroskedasticity and autocorrelation.

- Heteroskedasticity and autocorrelation robust inference procedures basically rely on the following property, which outlines the asymptotic properties of the OLS estimator $\hat{\beta}$ under the minimal set of assumptions TS.1' – TS.3' (linearity in parameters and weak dependence, no perfect colinearity and zero conditional mean):

  Property 14 Asymptotic properties of $\hat{\beta}$ without homoskedasticity and no serial correlation
  Under assumptions TS.1' – TS.3', the OLS estimator $\hat{\beta}$ is a consistent estimator of $\beta$:

$$\hat{\beta} \xrightarrow{p} \beta$$

  and is asymptotically normally distributed as:

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1}) \qquad (13)$$

  where:

$$A = E(X_t' X_t) \quad \text{and} \quad B = V\left(T^{-\frac{1}{2}} \sum_{t=1}^{T} X_t' u_t\right)$$

  The fact that $\hat{\beta}$ is consistent for $\beta$ under assumptions TS.1' – TS.3' was already outlined in Property 12. On the other hand, the limiting distribution result (13) has already be shown to hold likewise under assumptions TS.1' – TS.3' in the

[21] A time series regression model without any lagged dependent variable is quite unlikely to be dynamically complete.

sketch of the proof of Property 13 : see the intermediary result (6).

- The limiting distribution result (13) is the analogue of the limiting distributional result given by Property 9 in the supplemental lecture notes III (hereafter SNL-III) for the cross-sectional case[22]. As in the cross-sectional case, it provides an approximate finite sample distribution for the OLS estimator $\hat{\beta}$ :

$$\sqrt{T}(\hat{\beta}-\beta) \xrightarrow{d} N(0, A^{-1}BA^{-1})$$
$$\Leftrightarrow \quad \hat{\beta} \approx N(\beta, A^{-1}BA^{-1}/T) \tag{14}$$

which can be used – when $n$ is sufficiently large – for performing robust inference (confidence interval, hypothesis testing) without having to rely on any other assumption than assumptions TS.1' – TS.3', i.e., in particular without having to rely on the homoskedasticity and no serial correlation assumptions.

- For inference based on the limiting distributional result (13), or equivalently on the approximate distributional result (14), we need an estimator of the asymptotic variance $Avar(\hat{\beta}) = A^{-1}BA^{-1}/T$. This requires consistent estimators of $A$ and $B$. We already know that $X'X/T$ is a consistent estimator of $A$ (this directly follows from the LLN). It may be shown that a consistent estimator of $B = V\left(T^{-\frac{1}{2}}\sum_{t=1}^{T} X_t' u_t\right)$ is given by :

$$\hat{B} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 X_t' X_t + \frac{1}{T}\sum_{\tau=1}^{q}(1-\tfrac{\tau}{q+1})\sum_{t=\tau+1}^{T}(X_t'\hat{u}_t\hat{u}_{t-\tau}X_{t-\tau} + X_{t-\tau}'\hat{u}_{t-\tau}\hat{u}_t X_t)$$

where $\hat{u}_t = y_t - X_t\hat{\beta}$. This is formalized in the following property[23] :

Property 15 Consistent estimator of $B$
Under TS.1' – TS.3' (and some regularity conditions), $\hat{B}$ is a consistent estimator of $B$ :

$$\hat{B} \xrightarrow{p} B$$

To help intuitively understanding why $\hat{B}$ is a consistent estimator of $B$, it must be noted that $B = V\left(T^{-\frac{1}{2}}\sum_{t=1}^{T} X_t' u_t\right)$ can be written as :

$$B = \frac{1}{T}\sum_{t=1}^{T}E(u_t^2 X_t' X_t) + \frac{1}{T}\sum_{\tau=1}^{T-1}\sum_{t=\tau+1}^{T}E(X_t' u_t u_{t-\tau}X_{t-\tau} + X_{t-\tau}' u_{t-\tau}u_t X_t)$$

It then appears that $\hat{B}$ is basically the empirical counterpart of $B$, where (1) $u_t$ is replaced by its consistent estimator $\hat{u}_t$, (2) the averages of expectations are replaced by sample averages, and (3) the covariances between $X_t'u_t$ and $X_{t-\tau}'u_{t-\tau}$ are supposed to be equal to zero for all lags $\tau > q$ (i.e., only the covariances between observations with at most $q$ lags are considered). The integer $q$ controls

[22] As a matter of fact, Property 9 in SNL-III is just a special case of Property 14.
[23] For more details, see for example Hamilton (1994) p. 281-283.

how much autocorrelation is allowed. Theory states that $\hat{B}$ works for fairly arbitrary forms of autocorrelation, provided that $q$ grows with sample size $T$. For annual data, choosing a small $q$, such as $q = 1$ or $q = 2$, is likely to be enough. For quarterly or monthly data, $q$ should be larger ($q = 4$ or 8 for quarterly and $q = 12$ or 24 for monthly, provided enough data are available)[24].

- Using the outlined above consistent estimators of $A$ and $B$, an estimator of $Avar(\hat{\beta}) = A^{-1}BA^{-1}/T$ is given by:

$$\begin{aligned}\hat{V}_{HAC}(\hat{\beta}) &= (X'X)^{-1} \\ &\times \left(\sum_{t=1}^{T}\hat{u}_t^2X_t'X_t + \sum_{\tau=1}^{q}(1-\tfrac{\tau}{q+1})\sum_{t=\tau+1}^{T}(X_t'\hat{u}_t\hat{u}_{t-\tau}X_{t-\tau} + X_{t-\tau}'\hat{u}_{t-\tau}\hat{u}_tX_t)\right) \\ &\times (X'X)^{-1}\end{aligned} \tag{15}$$

where the subscript '$HAC$' stands for 'Heteroskedasticity and Autocorrelation Consistent'. As usual, the diagonal elements $V\hat{a}r_{HAC}(\hat{\beta}_j)$ of the $k \times k$ matrix estimator $\hat{V}_{HAC}(\hat{\beta})$ being the estimators of the variance $Avar(\hat{\beta}_j)$ of the estimator $\hat{\beta}_j$ of the different parameters $\beta_j$ ($j = 1, ..., k$), natural estimators of the asymptotic standard error $As.e.(\hat{\beta}_j) = \sqrt{Avar(\hat{\beta}_j)}$ of the estimator $\hat{\beta}_j$ of the different parameters $\beta_j$, as well as a natural estimator of the asymptotic standard error $As.e.(R_0\hat{\beta}) = \sqrt{Avar(R_0\hat{\beta})} = \sqrt{R_0Avar(\hat{\beta})R_0'}$ of the estimator $R_0\hat{\beta}$ of a single linear combination $R_0\beta$ of $\beta$, are likewise given by:

$$s.\hat{e}._{HAC}(\hat{\beta}_j) = \sqrt{V\hat{a}r_{HAC}(\hat{\beta}_j)}, \quad j = 1, ..., k \tag{16}$$

and:

$$s.\hat{e}._{HAC}(R_0\hat{\beta}) = \sqrt{R_0\hat{V}_{HAC}(\hat{\beta})R_0'} \tag{17}$$

where $R_0$ is a $1 \times k$ (row) vector of constants.

- As for heteroskedasticity robust inference in the cross-sectional case, the limiting distributional result (13), or equivalently the approximate distributional result (14), and the heteroskedasticity and autocorrelation consistent (or robust) estimators $\hat{V}_{HAC}(\hat{\beta})$, $s.\hat{e}._{HAC}(\hat{\beta}_j)$ and $s.\hat{e}._{HAC}(R_0\hat{\beta})$ given above in respectively (15), (16) and (17), provide all which is needed for performing heteroskedasticity and autocorrelation robust inference after OLS estimation. Following exactly the same reasoning as in Section 4.3 and Section 4.4.2 of SLN-I, it may readily be checked that if in all the usual inference procedures – confidence interval for $\beta_j$ or a single linear combination $R_0\beta$, two-sided and one-sided $t$-tests of $\beta_j$ or a single linear combination $R_0\beta$, $F$-test (or Wald test) of multiple linear restrictions – we replace the usual estimators $\hat{V}(\hat{\beta})$, $s.\hat{e}.(\hat{\beta}_j)$ and $s.\hat{e}.(R_0\hat{\beta})$ by their heteroskedasticity and autocorrelation consistent (or robust) versions $\hat{V}_{HAC}(\hat{\beta})$, $s.\hat{e}._{HAC}(\hat{\beta}_j)$ and $s.\hat{e}._{HAC}(R_0\hat{\beta})$, then we obtain inference procedures that are asymptotically valid – i.e., approximately valid for $T$ sufficiently large – under

[24] For more details, see Wooldridge (2016), Section 12-5.

only assumptions TS.1' – TS.3', i.e., without having to rely on neither the homoskedasticity assumption TS.4' nor the no serial correlation assumption TS.5' (as well as any normality assumption).

- Remarks :
  - All modern econometric software optionally provide heteroskedasticity and autocorrelation robust standard errors and allow to perform heteroskedasticity and autocorrelation robust tests. In applied works, when there is serial correlation, heteroskedasticity and autocorrelation robust standard errors are typically found larger than the usual OLS standard errors. This is because, in most cases, both data and errors are positively autocorrelated[25].
  - When there is substantial serial correlation and the sample size is small (where small can be as large as, say, 100), the heteroskedasticity and autocorrelation consistent estimator $\hat{V}_{HAC}(\hat{\beta})$ can be poorly behaved, so that the robust inference procedures may be poorly reliable. This partly explains why the use of heteroskedasticity and autocorrelation robust inference procedures are not as common in applied time series works as the use of heteroskedasticity robust inference procedures in cross-sectional empirical applications, where the available sample sizes are often much larger.
  - The presence of serial correlation may easily be checked by running an auxiliary regression of the form[26] :

$$\hat{u}_t = X_t b + \delta_1 \hat{u}_{t-1} + \delta_2 \hat{u}_{t-2} + ... + \delta_p \hat{u}_{t-p} + v_t \tag{18}$$

An usual $F$-test of $\mathrm{H}_0 : \delta_1 = ... = \delta_p = 0$ provides an asymptotically valid test of the null hypothesis of no serial correlation, which may also be viewed as a convenient test of the null hypothesis that the model is dynamically complete. In practice, choosing a number $p$ of lagged residuals equal to 1 or 2 should be enough for annual data. For quarterly or monthly data, $p$ should probably be larger. An asymptotically equivalent test may be performed using the $LM$ statistic[27] :

$$LM = T \cdot R^2_{\hat{u}^2}$$

where $R^2_{\hat{u}^2}$ is the $R$-squared from the auxiliary regression (18)[28]. This test is known as the Breush-Godfrey test for serial correlation.

The validity of the above tests requires that the homoskedasticity assumption TS.4' hold. If heteroskedasticity is suspected, this may be addressed

[25] For more details, see Wooldridge (2016), Section 12-1.

[26] Note that, unless the explanatory variables are assumed strictly exogenous, beside lagged residuals, to yield a valid test, the auxiliary regression must include all explanatory variables $X_t = (1, x_{t2}, ..., x_{tk})$ of the model. For more details, see Wooldrige (2016), Section 12-2.

[27] The decision rule of this $LM$ test is : reject $\mathrm{H}_0$ if $LM > \chi^2_{p;1-\alpha}$ and do not reject otherwise, where $\chi^2_{p;1-\alpha}$ is the quantile of order $1-\alpha$ of the $\chi^2(p)$ distribution. The $p$-value of the test, for a value $LM^*$ of the test statistic obtained in a particular sample, is given by : $p_{LM} = I\!P(v > LM^*)$, where $v \sim \chi^2(p)$.

[28] $T$ here refers to the actual number of observations in the auxiliary regression, i.e., the number of original observations minus the number of lagged residuals.

by using the heteroskedasticity robust $F$-test – as in the cross sectional case – rather than the usual $F$-test to test $\mathrm{H}_0: \delta_1 = ... = \delta_p = 0$ in the auxiliary regression (18). If this robust test does not reject the null hypothesis of no serial correlation, then homoskedasticity may validly be tested, as in the cross-sectional case, using the standard Breush-Pagan test for heteroskedasticity[29] outlined in SNL-III. If heteroskedasticity but no serial correlation is finally found, then inference procedures only robust to heteroskedasticity – as in the cross-sectional case – may be used instead of heteroskedasticity and autocorrelation robust inference procedures[30].

# Reference

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

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

Wooldridge J.M. (2016), *Introductory Econometrics : A Modern Approach*, 6th Edition, Cengage Learning.

[29] Just as the validity of the Breush-Godfrey test for serial correlation requires the homoskedasticity assumption TS.4' to hold, in the present time series context, the validity of the Breush-Pagan test for heteroskedasticity requires the no correlation assumption TS.5' to hold. For more details, see Wooldridge (2016), Section 12-6b.

[30] For more details, see Wooldridge (2016), Section 12-6a.