Lecture 16

Modeling with Time Series Data. Part 1.

Until now we have analyzed cross-sectional data within the framework of Model A and then Model B, relaxing the assumption of non-stochastic regressors. The common feature of such type of models is that in the data generating process (DGP) – when observations on a number of units are all taken at the same (one) point in time – the ordering of the observations does not influence regression results. In other words, this ordering is arbitrary.

In time series analysis, observations are usually collected at fixed intervals in time (regular spans). As a consequence, there is a so called natural ordering that is reflected by a particular time index. Moreover, there is a certain degree of regularity (persistence) in models with time series data: successive observations are correlated. This describes the behavior of many macroeconomic variables.

In this lecture we will switch to time series models, looking at the analysis within the framework of Model C.

Assumptions of Model C:

C.1 *The model is linear in parameters and correctly specified:* $Y = \beta_1 + \beta_2 \cdot X_2 + \ldots + \beta_k \cdot X_k + u$;

C.2 *The time series for the regressors are (at most) weakly persistent;*

C.3 *There does not exist an exact linear relationship among the regressors;*

C.4 Linearity: *The disturbance term has zero expectation*. $\mathrm{E}(u_t) = 0$ for all *t*;

C.5 Homogeneity: *The disturbance term is homoscedastic.* $\sigma_{u_t}^2 = \sigma_u^2$ for all *t*;

C.6 Independence: *The values of the disturbance term have independent distributions*. u_t is distributed independently of $u_{t'}$ for all $t \neq t'$;

C.7 *The disturbance term is distributed independently of the regressors: u_i is distributed independently of $X_{ji'}$ for all i' and all j;*

C.8 *The disturbance term has a normal distribution:* $u_t \sim N(0, \sigma_u^2)$

Note that C.2 is a new assumption. Let's look at it in a greater detail:

C.2: Advanced technical concepts are used in the precise definition of weakly persistence. Therefore, it goes beyond the scope of the introduction to econometrics course. Intuitively, highly persistence is connected with strongly dependence. For example, C.2 is violated in the random walk model $X_t = X_{t-1} + u_t$, where $t = 1,2, \ldots$ and u_t is an i.i.d. with zero mean and constant variance. For period $t + h > 0$ the relationship becomes: $X_{t+h} = X_{t+h-1} + u_{t+h}$, where $h > 0$. This can be rewritten as $X_{t+h} = X_t + u_{t+1} + u_{t+2} + \ldots + u_{t+h}$.

Note that $\mathrm{E}(X_{t+h}|X_t) = \mathrm{E}(X_t + u_{t+1} + u_{t+2} + \ldots + u_{t+h}) = X_t$. It means that the value of X_t today affects all the future values of X_{t+h} – indication of highly persistence (strong dependence). Models of the type $X_t = \rho \cdot X_{t-1} + u_t$ with $|\rho| < 1$ are weakly persistent. In fact, $\mathrm{E}(X_{t+h}|X_t) = \rho^h \cdot X_t$ => not so strong dependence (the expectation approaches to zero for $h \to \infty$).

As a first approximation, we could use stationarity as a definition, and this criterion seems to be widely adopted in practice (stationarity will be considered in next lectures). But in theory some non-stationary models can be weakly persistent, while not all weakly persistent models are stationary.

All other Model C assumptions and the consequences of their violations are similar to those of the Model B. However, assumptions C.6 and C.7 have a greater relative importance in the context of time series models.

C.6: Due to a certain degree of persistence inherited in time series data, it frequently happens that subsequent values of disturbance terms are correlated. For cross-sectional data the violation of C.6 is a rare situation as observations are generated randomly (even if it happens, we can always rearrange a sample to get rid of this dependence).

C7: This assumption can be divided into 2 parts:

Part (1): The disturbance term in any observation is distributed independently of the values of the regressors in the same observation, and

Part (2): The disturbance term in any observation is distributed independently of the values of the regressors in the other observations.

If both parts hold, then estimates are unbiased. At the same time, (1) is necessary for consistency (but not sufficient). For cross-sectional models part (2) is usually not violated since observations are generated randomly. Hence, unbiasedness depends on part (1) that can be violated by measurement errors and estimation of simultaneous equations. However, for time series data part (2) becomes a major concern. To understand it, let's look at the simple linear regression model:

$$Y = \beta_1 + \beta_2 \cdot X_2 + u$$

Remember that the OLS slope coefficient can be decomposed into the true value and an error term:

$$b_2^{\mathrm{OLS}} = \beta_2 + \sum a_i u_i, \qquad \text{where} \qquad a_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}$$

Taking expectations: $\mathrm{E}\left(b_2^{\mathrm{OLS}}\right) = \mathrm{E}(\beta_2) + \mathrm{E}(\sum a_i u_i) = \beta_2 + \sum \mathrm{E}(a_i u_i)$.

Assuming that u_i is independent of a_i, where a_i is a function of all X_i, the decomposition becomes $\mathrm{E}\left(b_2^{\mathrm{OLS}}\right) = \beta_2 + \sum \mathrm{E}(a_i u_i) = \beta_2 + \sum \mathrm{E}(a_i) \cdot \mathrm{E}(u_i) = \beta_2 + \sum \mathrm{E}(a_i) \cdot 0 = \beta_2$. Therefore, for unbiasedness part (1) is insufficient and part (2) assumption is needed.

C.7 assumption for unbiasedness

	Cross-sectional	**Time series**
Part (1)	Required (main concern)	Required
Part (2)	Usually holds by default due to DGP	Required

Consider a model with a lagged dependent variable where the dependent variable with some lag is a part of regressors. The simplest case is the considered random walk model of the form: $Y_t = Y_{t-1} + u_t$. In the next observation: $Y_{t+1} = Y_t + u_{t+1}$

OLS gives a biased result as u_t affects Y_t that is a regressor for the next observation – part (2) condition is violated. The value of bias can be determined implementing Monte Carlo simulation. There is no analytical expression for the bias. At the same time, it can be noted that when the number of observations increases in the sample, the bias seems to disappear – evidence of consistency. Analytically, for more general case of the model:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t$$

$$b_2^{\mathrm{OLS}} = \frac{\widehat{\mathrm{Cov}}(Y_t, Y_{t-1})}{\widehat{\mathrm{Var}}(Y_{t-1})}$$

By taking probability limits:

$$\text{plim}\left(b_2^{\text{OLS}}\right) = \text{plim}\left(\frac{\widehat{\text{Cov}}(Y_t, Y_{t-1})}{\widehat{\text{Var}}(Y_{t-1})}\right) = \frac{\text{plim}\left(\widehat{\text{Cov}}(Y_t, Y_{t-1})\right)}{\text{plim}\left(\widehat{\text{Var}}(Y_{t-1})\right)} = \frac{\text{Cov}(Y_t, Y_{t-1})}{\text{Var}(Y_{t-1})} =$$
$$= \frac{\text{Cov}((\beta_1 + \beta_2 Y_{t-1} + u_t), Y_{t-1})}{\text{Var}(Y_{t-1})} = \beta_2 \cdot \frac{\text{Cov}(Y_{t-1}, Y_{t-1})}{\text{Var}(Y_{t-1})} + \frac{\text{Cov}(u_t, Y_{t-1})}{\text{Var}(Y_{t-1})} =$$
$$= \beta_2 + \frac{\sigma_{Y_{t-1}, u_t}}{\sigma^2_{Y_{t-1}}} = \beta_2 + \frac{0}{\sigma^2_{Y_{t-1}}} = \beta_2$$

Hence, the estimate is consistent. We used part (1) condition $\sigma_{Y_{t-1}, u_t} = 0$ by showing consistency. It is quite reasonable to be valid as Y_{i-1} is determined before u_i is generated => they are independent.

C.7 assumption for consistency

	Cross-sectional	**Time series**
Part (1)	Required (main concern)	Required (but not sufficient – technical issue)
Part (2)	Usually holds by default due to DGP	Not Required

Static models and models with lags:

The values of economic variables can depend not only on current values of other (explanatory) variables, but also on their preceding values. When such a relationship with the preceding values of explanatory variables exists, we can speak about a model with a **distributed lag**. This characteristic is notable for time series data. Suppose, we need to estimate the parameters of the following model: $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \ldots + u_t$. The following problems can arise:

1) ***Multicollinearity.*** Lagged values of an explanatory variable can be strongly correlated with each other.
2) ***Decrease in the sample size:*** To include an observation into the sample, the data on the values of regressors in the preceding observations are necessary. For a number of the earliest observations such data are unavailable, therefore, they are excluded from the sample, which thus decreases in size.
3) ***Reduction in the number of degrees of freedom:*** A large number of parameters are estimated and, therefore, the number of degrees of freedom is reduced. The solution can be to make an assumption concerning the distribution of β_i, which is assumed to be characterized with small number of parameters.

Illustration of multicollinearity:

Let's consider a constant elasticity function in which consumer expenditure on housing services depends on aggregate disposable personal income and its relative price index. It takes the form: $HOUS = \beta_1 DPI^{\beta_2} PRELHOUS^{\beta_3} \cdot v$. Here β_2 is the income elasticity and β_3 is the price elasticity for expenditure on housing services. By taking logarithms, the model can be viewed as a linear regression model:

$$LGHOUS = \log\beta_1 + \beta_2 LGDPI + \beta_3 LGPRHOUS + \log v.$$

Then, slope coefficients are estimated using OLS procedure. This corresponds to the cross-sectional analysis.

Time series analysis allows to introduce some simple dynamics. In fact, changes in income and price level at some period t are continuing to be reflected in expenditure on housing for subsequent time periods. There is a so called inertia in the response of housing expenditure.

Therefore, new specifications are considered in which lagged values of income and price level become regressors. They are described in the table below.

Alternative dynamic specifications, housing services
(*LGHOUS* is the dependent variable, USA, 1959-2003)

Variable	(1)	(2)	(3)	(4)	(5)
LGDPI	1.03 (0.01)	–	–	0.33 (0.15)	0.29 (0.14)
LGDPI(–1)	–	1.01 (0.01)	–	0.68 (0.15)	0.22 (0.20)
LGDPI(–2)	–	–	0.98 (0.01)	–	0.49 (0.13)
LGPRHOUS	–0.48 (0.04)	–	–	–0.09 (0.17)	–0.28 (0.17)
LGPRHOUS(–1)	–	–0.43 (0.04)	–	–0.36 (0.17)	0.23 (0.30)
LGPRHOUS(–2)	–	–	–0.38 (0.04)	–	–0.38 (0.18)
R^2	0.9985	0.9989	0.9988	0.9990	0.9993

It is evident that direct inclusion of lagged explanatory variables results in severe multicollinearity. It is caused by the high correlation between current and lagged values. Moreover, long lag structure is expected for housing expenditure (2 lags is not always enough), therefore, some other approaches are needed to construct a dynamic model. To alleviate the problem, one can impose some restrictions on the regression coefficients β_i (for example, by assuming that they follow a certain distribution) or employ autoregressive distributed lag model ADL(p, q). The parameter p stands for the maximum number of lags of the dependent variable, and q is the maximum lag of explanatory variables. General expression:

$$\boxed{Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \ldots \beta_{q+2} X_{t-q} + \alpha_3 Y_{t-1} + \alpha_4 Y_{t-2} + \alpha_{p+2} Y_{t-p} + u_t - \ ADL(p,q)}$$

Nevertheless, long-run estimates of elasticities can be obtained by looking at equilibrium relationship. For example, if the process ADL(0,2) $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + u_t$ ever reached equilibrium, the following holds:

$$\bar{Y} = \beta_1 + \beta_2 \bar{X} + \beta_3 \bar{X} + \beta_4 \bar{X} = \beta_1 + (\beta_2 + \beta_3 + \beta_4)\bar{X}, \qquad \text{where } \bar{X} \text{ and } \bar{Y} \text{ are equilibrium values.}$$

Hence, $\beta_2 + \beta_3 + \beta_4$ is a measure of the long-run effect. While the impact of current X_t on Y_t is measured by β_2 – short-run effect. For instance, in specification (5) the long-run income elasticity is equal to $0.29 + 0.22 + 0.49 = 1$. The long-run price elasticity equals $-0.28 + 0.23 - 0.38 = -0.43$. Corresponding short-run elasticities are 0.29 and -0.28, respectively.

Furthermore, in order to test significance of obtained estimate for the long-run effect, one needs its standard error. It is derived by the following reparametrization of the model. Adding and subtracting $\beta_3 X_t$ and $\beta_4 X_t$ from the right side equation of $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + u_t$, one can get:

$$\boxed{Y_t = \beta_1 + (\beta_2 + \beta_3 + \beta_4) X_t - \beta_3 (X_t - X_{t-1}) - \beta_4 (X_t - X_{t-2}) + u_t}$$

Multicollinearity is unlikely to be an issue here because X_t may not be highly correlated with $X_t - X_{t-1}$ and $X_t - X_{t-2}$.

Estimation of distributed lag models:

As was discussed, one way to deal with the problem 3) is to choose a particular time structure of the model by specifying certain behavior of coefficients over time. The most frequently used such distributions are the following:

1) Geometric distribution (Koyck distribution): coefficients of the explanatory variables have geometrically declining weights, i.e.

$\beta_1; \beta_2 = \beta_1 \cdot \rho; \beta_3 = \beta_1 \cdot \rho^2 \ldots \beta_k = \beta_1 \cdot \rho^{k-1}$, where $0 < \rho < 1$.

2) Polynomial distribution:

$\beta_s = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot s + \tilde{\beta}_2 \cdot s^2$, where s are integers.

It is also possible to impose restrictions on the parameters of uniform, linear, "triangular" and other distributions.

Geometrically Distributed lag (Koyck model):

General specification of the model is described as:

$$\boxed{Y_t = \beta_1 + \beta_2 X_t + \beta_2 \rho X_{t-1} + \beta_2 \rho^2 X_{t-2} + \ldots + u_t} \quad (*)$$

The problem of inclusion of infinite number of lags into the model is resolved by Koyck transformation. Multiplying both parts of one time period lagged (*) by ρ:

$$\rho Y_{t-1} = \beta_1 \rho + \beta_2 \rho X_{t-1} + \beta_2 \rho^2 X_{t-2} + \beta_2 \rho^3 X_{t-3} + \ldots + \rho u_{t-1} \quad (**)$$

Subtracting (**) from (*), note that terms with X from the right side of both expressions are cancelled starting from the first observation's period till $t-1$:

(*) – (**):$\Rightarrow Y_t - \rho \cdot Y_{t-1} = \beta_1(1-\rho) + \beta_2 \cdot X_t + u_t - \rho \cdot u_{t-1}$.

This can be rewritten as:

$$\boxed{Y_t = \beta_1(1-\rho) + \rho \cdot Y_{t-1} + \beta_2 \cdot X_t + v_t}, \quad \text{where } v_t = u_t - \rho \cdot u_{t-1}$$

However, part (1) of the assumption C.7 that is necessary for both consistency and unbiasedness does not hold as v_t is related to Y_{t-1} because u_{t-1} is a component of v_t and u_{t-1} influences Y_{t-1}. Thus, <u>direct implementation of OLS gives biased and inconsistent estimates</u>. Instead, non-linear procedures of estimation are used. For instance, we can stop increasing the number of lags in (*) when the next increment in lags does not change the values of estimates (achieving stability of coefficients).

<u>Short-run and long-run effects:</u>

Short-run influence of X on Y: β_2

$\bar{Y} = \beta_1 + \beta_2 \bar{X} + \beta_2 \rho \bar{X} + \beta_2 \rho^2 \bar{X} + \ldots + u_t = \beta_1 + \beta_2 \bar{X}(1 + \rho + \rho^2 + \ldots) + u_t = \beta_1 + \frac{\beta_2 \bar{X}}{1-\rho} + u_t$.

Long-run influence of X on Y: $\frac{\beta_2}{1-\rho}$

Polynomial distributed lag (Almon):

The problem of Koyck procedure is that it is very restrictive where the values of coefficients decline in geometric proportions. However impact of economic variables may be better explained by a quadratic, cubic or higher order polynomial of the form:

$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + \ldots + \beta_{n+2} X_{t-n} + u_t$, where m-order polynomial lag structure defined as:

$$\beta_s = \gamma_0 + \gamma_1 \cdot s + \gamma_2 \cdot s^2 + \ldots + \gamma_m \cdot s^m; \quad s = 0,1,2, \ldots$$

The model becomes more flexible. It can incorporate variety of lags.

Calculating for:

$s = 0 \qquad \beta_0 = \gamma_0;$

$s = 1 \qquad \beta_1 = \gamma_0 + \gamma_1 + \gamma_2 + \ldots + \gamma_m;$

$s = 2 \qquad \beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2 + \ldots + 2^m \gamma_m;$

…………………………………………………………

Setting n = 3 and m = 2, we get:

$$Y_t = \beta_1 + \gamma_0 X_t + (\gamma_0 + \gamma_1 + \gamma_2) X_{t-1} + (\gamma_0 + 2\gamma_1 + 4\gamma_2) X_{t-2} + (\gamma_0 + 3\gamma_1 + 9\gamma_2) X_{t-3} + u_t =$$
$$= \beta_1 + \gamma_0 \underbrace{(X_t + X_{t-1} + X_{t-2} + X_{t-3})}_{Z_0} + \gamma_1 \underbrace{(X_{t-1} + 2X_{t-2} + 3X_{t-3})}_{Z_1} + \gamma_2 \underbrace{(X_{t-1} + 4X_{t-2} + 9X_{t-3})}_{Z_2}$$

Therefore, from estimation of $\gamma' s$ we can derive $\beta' s$, substituting $\gamma' s$ into the expression for β_s.