Introductory Econometrics

Based on the textbook by Wooldridge: *Introductory Econometrics: A Modern Approach*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

December 17, 2012

Outline

What is so special about time-series data?

Time-series data typically violate the random-sampling assumption. Observations (often indexed by t rather than i) are usually related to each other, and often X_t and X_{t+1} are related more strongly than X_t and X_{t+2}.

The ordering of time plays a role. Observations at t can affect observations at $t+1$ but not vice versa (no StarTrek).

Static and dynamic regressions

Regressions such as

$$y_t = \beta_0 + \beta_1 x_{t,1} + \ldots + \beta_k x_{t,k} + u_t$$

are called *static regressions*. If u_t are independent, they can be handled like other regressions.

Regressions with lags among regressors, such as:

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t-1,1} + \beta_3 x_{t-2,1} + \beta_4 x_{t,2} + u_t,$$

are called *regressions with distributed lags*.

Regressions with lagged dependent variables as regressors, such as:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 x_{t,1} + \beta_4 x_{t,2} + u_t,$$

are called *dynamic regressions*.

Impact and long run

In models such as

$$y_t = \beta_0 + \beta_1 x_{t,1} + \beta_2 x_{t-1,1} + \beta_3 x_{t-2,1} + \beta_4 x_{t,2} + u_t,$$

a change in $x_{t,1}$ immediately incurs a change in y_t of β_1, this is the 'impact multiplier'. It will also affect y_{t+1} and y_{t+2}. The sum $\beta_1 + \beta_2 + \beta_3$ can be seen as the total effect of the impulse, it is called the 'long-run multiplier'.

Assumptions with time-series data

The assumption of a linear reaction **[TS.1]**

$$y_t = \beta_0 + \beta_1 x_{t,1} + \ldots + \beta_k x_{t,k} + u_t$$

remains as is. Some of the regressors could be lags of other regressors, but preferably not leads.

The assumption of non-multicollinearity **[TS.2]** remains as is.

The assumption of zero-mean errors **[TS.3]**

$$\mathrm{E}(u_t|X) = 0, t = 1, \ldots, n,$$

essentially remains as is. Note that X includes all leads and lags of the regressors. TS.3 is violated in dynamic regression.

Unbiasedness of OLS in time-series regressions

Theorem

Under assumptions TS.1, TS.2, TS.3, the OLS estimator is unbiased conditional on X and therefore unconditionally as well: $\mathrm{E}\hat{\beta}_j = \beta_j, j = 0, 1, \ldots, k$.

Further assumptions with time-series data

The homoskedasticity assumption **TS.4**

$$\mathrm{var}(u_t|X) = \mathrm{var}\, u_t \equiv \sigma^2, t = 1, \ldots, n$$

requires equal variances conditionally and thus also unconditionally.

The no-serial-correlation assumption **TS.5**

$$\mathrm{cov}(u_t, u_s|X) = 0 \Rightarrow \mathrm{corr}(u_t, u_s|X) = 0 \Rightarrow \mathrm{corr}(u_t, u_s) = 0, \quad s \neq t,$$

replaces the random-sampling assumption. It is relatively weak and may occasionally be strengthened to independence.

Gauss-Markov for time-series regression

Theorem

*With assumptions TS.1–TS.5, OLS is the best linear unbiased estimator conditional on X. (**Gauss-Markov Theorem**)*

Theorem

With assumptions TS.1–TS.5, the variance of the OLS estimator conditional on X is $\mathrm{var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$.

Theorem

With assumptions TS.1–TS.5, the estimator

$$\hat{\sigma}^2 = \frac{SSR}{n-k-1}$$

is an unbiased estimator of σ^2. $n-k-1$ is also called the degrees of freedom df.

Trends

Many economic variables tend to grow over time. Such trends in y and x, say, can be shown to yield significant coefficients in regressions $y = \beta_0 + \beta_1 x + u$, even if y and x are unrelated: **spurious regression**.

Including a (linear, sometimes quadratic) time trend in the regression as an additional regressors often removes the spurious regression problem, though not always:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 t + u_t$$

Note: regressing both y and x separately on linear trends and then regressing the 'purged' or 'de-trended' variable $\tilde{y}$ on $\tilde{x}$ yields the same OLS coefficient estimate $\hat{\beta}_1$.

Seasonality

Economic variables observed at quarterly or monthly frequency often exhibit **seasonality** that may distort coefficient estimates in regression. Often, the inclusion of seasonal dummy variables (1 for a specific quarter or month, 0 otherwise) serves to isolate the effects properly:

$$y_t = \beta_0 + \gamma_1 spring_t + \gamma_2 summer_t + \gamma_3 fall_t + \beta_1 x_{t,1} + u_t$$

Note: you can only use 3 (or 11) seasonal dummies plus a constant, or alternatively 4 (or 12) but no intercept, otherwise there will be perfect multicollinearity (dummy variable trap).

Stationarity and weak dependence

The $(k+1)$–dimensional random sequence $(y_t, x_t')'$ can be seen as following a *stochastic process*. A stochastic process z_t is said to be (covariance) *stationary* iff

$$\begin{aligned} \mathrm{E} z_t &= \mu \quad \forall t; \\ \mathrm{var} z_t &= \sigma_z^2 \quad \forall t; \\ \mathrm{cov}(z_t, z_{t-h}) &= C(h) \quad \forall t, h; \end{aligned}$$

Technically, stationarity does not suffice for good asymptotic properties. It should also hold that $C(h) \to 0$ as h gets large: *weak dependence*.

Assumptions for asymptotic OLS properties in time-series regression

The new assumption $[\mathbf{TS.1}]'$ consists of the linear form

$$y_t = \beta_0 + \beta_1 x_{t,1} + \ldots + \beta_k x_{t,k} + u_t$$

plus stationarity and weak dependence of $(y_t, x_t')'$. It guarantees that central limit theorems work.

The assumption $[\mathbf{TS.2}]'$ on non-multicollinearity is as before.

The contemporaneous exogeneity assumption $[\mathbf{TS.3}]'$ states that $\mathrm{E}(u_t|x_t') = 0$ and is slightly weaker than **[TS.3]**. It admits well specified dynamic regressions.

Consistency of OLS in time-series regression

Theorem

With $TS.1'$, $TS.2'$, $TS.3'$, the OLS estimator is consistent:

$$\text{plim}\hat{\beta} = \beta, \quad \beta' = (\beta_0, \beta_1, \ldots, \beta_k)$$

Unbiasedness of OLS would require $\mathrm{E}(u_t|X) = 0$ and does not hold in dynamic regression. Consistency requires $\mathrm{E}(u_t|x_t) = 0$ only, which holds in all well-specified regression models, even in dynamic ones. If the model is incorrectly specified, e.g. correlated errors in dynamic regressions, consistency breaks down. Such models are really bad.

Dynamic regression models do have a small OLS bias that disappears in larger samples. OLS should still be used, though sometimes care should be exercised. Not so bad models.

Additional assumptions for asymptotic normality

The assumption **TS.4**′ on contemporaneous homoskedasticity

$$\mathrm{var}(u_t|x_t') = \sigma^2$$

is less restrictive than homoskedasticity **TS.4**.

The same holds for the assumption **TS.5**′ on no serial correlation

$$\mathrm{E}(u_t u_s|x_t', x_s') = 0.$$

Asymptotic normality of OLS in time-series regression

Theorem

With assumptions $\mathbf{TS.1}'$–$\mathbf{TS.5}'$*, the OLS estimator is asymptotically normally distributed.*

This is only an asymptotic result. The usual inference statistics (such as Wald, LR, LM) with sometimes exact t (or F) distributions in static regression are only guaranteed to have approximate normal (or χ^2) distributions in large samples. Some people use t or F distributions, nevertheless, and can cite some support by simulation evidence.