ECON 2007: Quant Econ and Econometrics
Regression with Time Series

Dr. Áureo de Paula

Department of Economics
University College London

Introduction

Time-series data have a few subtle departures from cross-sectional datasets. Most importatly:

- There is a natural order: time flows in one direction... (A sequence of random variables is called a **stochastic process**.)
- Whereas in cross-sectional settings, random sampling is a typical paradigm, time-series observations will usually be display some dependence. Tomorrow is not completely independent from today!
- Example: Life Expectancy in UK.

Introduction

- To learn how today and tomorrow are related, we would like to get many realizations for a stochastic process of interest. The thought experiment of restarting the stochastic process to collect a different realization of the trajectory of random variables is nevertheless infeasible.

- If we observe many “todays” and “tomorrows” and the relation between them is stable we might still be able to use only one realization of the process nonetheless.

Static Models

A **static model** is one in which only contemporaneous variables appear. For example,

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

An example would be a (*static*) *Philips curve*:

$$inf_t = \beta_0 + \beta_1 unem_t + u_t$$

Another example would be the relation between life expectancy and income per capita:

$$life_t = \beta_0 + \beta_1 \log inc_t + u_t$$

Finite Distributed Lag Models

A **FDL model** is one in which lagged explanatory variables appear as regressors. For example,

$$y_t = \alpha_0 + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + u_t$$

which is a FDL model of order 2.

δ_j is the change in y j periods after a temporary change in x.

Finite Distributed Lag Models

δ_0 is the **impact propensity** or **impact multiplier**.

A graph of δ_j against j is known as the **lag distribution** and gives the dynamic effect of a temporary increase in x on y.

$\sum_{i=0}^{q} \delta_j$ is the **long-run propensity** or **long-run multiplier** in a FDL model of order q and gives the overall change in y from a permanent change in x.

Finite Sample Properties of OLS

Consider again the static linear model:

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_{1t} + \cdots + \beta_k x_{kt} + u_t \\
&= \beta^\top \mathbf{x}_t + u_t
\end{aligned}
$$

Under what conditions will OLS estimate β *without bias*?

As in the previous term, we need the above linear specification (TS.1) and no perfect collinearity (TS.2) (i.e. no explanatory variable is a perfect linear combination of the others).

Finite Sample Properties of OLS

We also need to avoid counfounding between x_t and u_t to separate the influence of either on y_t.

With a random sample where observations were independent from each other this was achieved by the assumption that

$$E(u_t|\mathbf{x}_t) = 0.$$

This **contemporaneous exogeneity** is *not* enough to guarantee unbiasedness since observations will not necessarily be independent any longer. (It is nonetheless sufficient for consistency.)

Finite Sample Properties of OLS

Since we do not have a random sample, for OLS to be unbiased we need to strengthen the condition above to **strict exogeneity** (TS.3):

$$E(u_t|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) = 0$$

for $t = 1, \ldots, T$.

Anything that breaks **contemporaneous exogeneity** (e.g., omitted variables or measurement error) would break TS.3. It would also break if there are lagged outcomes as explanatory variables or any type of feedback from past outcomes into explanatory variables: explanatory variables that are strictly exogenous cannot react to what has happened to y_t in the past.

Finite Sample Properties of OLS

Theorem (Unbiasedness of OLS)

Under TS.1, TS.2 and TS.3, OLS estimators are unbiased (conditional on $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T$) and unconditionally.

Finite Sample Properties of OLS

With a random sample, homoskedasticity would allow us to obtain an expression for the variance of the estimators. This is not enough with time series.

Here, homoskedasticity is stated in terms of all realizations of $\mathbf{x}_t$:

$$Var(u_t|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) = Var(u_t) = \sigma^2 \quad \text{(TS.4)}$$

This means that the residual variance does not depend on covariates and does not change through time! It would fail for example if policy changes affect the distribution of residuals, changing their variance.

Finite Sample Properties of OLS

Yet another assumption is that there be no serial correlation:

$$Corr(u_t, u_s | \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T) = 0 \quad \text{(TS.5)}$$

for all $t \neq s$.

If these hold, then:

Theorem (OLS Sampling Variances)

Under TS.1-TS.5,

$$Var(\hat{\beta}_j | \mathbf{x}_1, \ldots, \mathbf{x}_T) = \sigma^2 / [SST_j(1 - R_j^2)]$$

Finite Sample Properties of OLS

TS.1-TS.5 also guarantee unbiasedness of the usual estimator for σ^2 (i.e., $SSR/n-k-1$) and the Gauss-Markov Theorem (i.e., OLS is best linear unbiased estimator).

Finite Sample Properties of OLS

Why these additional conditions were not necessary with cross-sectional data?

With cross-sectional data, *random sampling* complements **contemporaneous exogeneity** to guarantee **strict exogeneity**. It also complements the usual homoskedasticity assumption to guarantee the stricter version we presented above. Finally, it also guarantees no serial correlation.

Dummies, Seasons and Trends

We can use transformations of the data (e.g., natural logs) much as was done in the Term 1.

It is also customary to employ *dummy variables* as explanatory variables in the analysis of time series. Because observations are indexed by time, a dummy variable can be used to single-out particular time-periods.

Those may be systematically different from remaining observations because of the implementation of a new policy or changes in government. (Of course, remember that these still need to be **strictly exogenous** for the OLS estimator to have good finite sample properties.)

Dummies, Seasons and Trends

Dummy variables are commonly used as a simple way to control for seasonal patterns. Some economic activities such as tourism might be influenced by weather patterns for example.

Although many datasets are pre-adjusted, one may still encounter unadjusted data. To account for seasonal patterns we can simply include dummy variables among the regressors. For example, for monthly data,

$$y_t = \beta^\top \mathbf{x}_t + \delta_1 feb_t + \delta_2 mar_t + \cdots + \delta_{11} dec_t + u_t$$

where $feb_t, mar_t, \ldots$ are dummies indicating whether t is February, March, ... If there is no seasonality in y_t (after accounting for $\mathbf{x}_t$), the δ parameters are zero.

Dummies, Seasons and Trends

Many time series display a trend which needs to be accounted for, lest one might erroneously infer association between two trending variables when indeed their change is unrelated to each other.

If a variables grows by the same amount on average every period, its behavior is typically captured well by a linear trend model:

$$y_t = \alpha_0 + \alpha_1 t + e_t$$

When the log of the variable has a linear trend, we say that the variable itself displays an exponential trend: its *growth rate* is on average the same over time:

$$\log y_t = \alpha_0 + \alpha_1 t + e_t$$

Dummies, Seasons and Trends

Trends may be important in regression analysis as well.

"[W]e must be careful to allow for the fact that unobserved, trending factors that affect y_t might also be correlated with the explanatory variables. If we ignore this possibility, we may find a spurious relationship between y_t and one or more explanatory variables. $(\ldots)$ Fortunately, adding a time trend eliminates this problem." (Wooldridge, p.363)

More concretely, one adds the time trend as an extra explanatory variable. For example,

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

Dummies, Seasons and Trends

Because we can always "partial out" the regression (remember last term?), we can interpret the results as a regression using detrended data.

In other words, running OLS on the above example would give us $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$. The same estimates for β_0 and β_1 could be obtained if we

1. Run a regression of y_t on t and save the residuals as $\ddot{y}_t$.
2. Run a regression of x_t on t and save the residuals as $\ddot{x}_t$.
3. Run a regression of $\ddot{y}_t$ on $\ddot{x}_t$.

It turns out that the intercept and slope coefficient estimates one obtains are numerically identical to $\hat{\beta}_0$ and $\hat{\beta}_1$ from the original regression!

Dummies, Seasons and Trends

If y_t and x_t are trending, one may then include a time trend in the regression to prevent from finding a relationship when there is none.

Even when y_t has no discernible trend, if x_t has a trend we might want to control for that. Otherwise, x_t may appear as though it has not relationship with y_t when it might have one.

Dummies, Seasons and Trends

For example, Castillo-Freeman and Freeman (1992) examine the effect of US minimum wage on employment in Puerto Rico. Their model is a more sophisticated version of:

$$\log(prepop_t) = \beta_0 + \beta_1 \log(mincov_t) + \beta_2 \log(usgnp_t) + u_t$$

where $prepop_t$ is the employment rate in Puerto Rico, $usgnp_t$ is the real US gross national product (in billions of dollars) and $mincov_t$ is an index measure the importance of minimum wage relative to average wages.

Dummies, Seasons and Trends

Initial results give

$$\log(\widehat{prepop_t}) = \underset{(0.77)}{-1.05} \ \underset{(.065)}{-.154} \log(mincov_t) \ \underset{(.089)}{-.012} \log(usgnp_t)$$

With a time trend nevertheless,

$$\log(\widehat{prepop_t}) = \underset{(1.30)}{-8.70} \ \underset{(.044)}{-.169} \log(mincov_t) \ \underset{(0.18)}{+1.06} \log(usgnp_t) \ \underset{(.005)}{-.032}\, t$$

Dummies, Seasons and Trends

Some caveats:

- Be careful when y_t and x_t have different trends (i.e. linear versus exponential)
- To account for the trend in y_t, use detrended y_t in constructing the R^2 for the regression with time trend. Otherwise, the R^2 will be overstated.

Stationarity and Weak Dependence

- The importance of stationarity and weak dependence for us is that these notions are used in some LLN and CLT results we will use in time series analysis.

- A stationary process is a series whose probability distribution remains stable over time: if we shift the sequence ahead h steps, the join probability distribution is unchanged:

  The stochastic process $\{x_t : t = 1, 2, \dots\}$ is stationary if for every collection of time indices $1 \leq t_1 \leq t_2 \leq \cdots \leq t_m$, the joint distribution of $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ is the same as the joint distribution of $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ for all $h \geq 1$.

Stationarity and Weak Dependence

- Sometimes we only require a weaker form of stationarity:

  The stochastic process $\{x_t : t = 1, 2, \dots\}$ with finite second moment $E(x_t^2) < \infty$ is covariance stationary if (i) $E(x_t)$ is constant; (ii) $Var(x_t)$ is constant; and (iii) for any $t, h \geq 1$, $Cov(x_t, x_{t+h})$ depends only on h and not on t.

Stationarity and Weak Dependence

- The notion of weak dependence relates to how related two rancom variables x_t and x_{t+h} are as h gets large.
- If a process is weakly dependent, x_t and x_{t+h} are almost independent as h gets large.
- Different ways of mathematically formalizing “almost independence” give rise to different forms of weak dependence.

Examples

- MA(1) process:

$$x_t = u_t + \rho u_{t-1}, \quad t = 1, 2, \ldots$$

where u_t is iid with mean zero and variance σ^2.

- AR(1) process:

$$y_t = \rho y_{t-1} + u_t, \qquad t = 1, 2, \ldots$$

with $|\rho| < 1$ and u_t is iid with mean zero and variance σ^2.

- "Trend-stationary" series. (Notice that these series are non-stationary).

Asymptotic Properties of OLS

- For unbiasedness, we used three assumptions: TS.1-3.
- For consistency, we will maintain the no perfect collinearity condition (TS.2)=(TS.2').
- We will also require the linear specification to hold (TS.1):

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t,$$

  but we will append to it the requirement that $\{(y_t, \mathbf{x}_t) : t = 1, 2, \ldots\}$ be stationary and weakly dependent. (TS.1')
- Stationarity and weak dependence are used to establish Laws of Large Numbers and Central Limit Theorems.

Asymptotic Properties of OLS

- For unbiasedness, we required *strict exogeneity*. To attain consistency, we can weaken that condition to *contemporaneous exogeneity* (TS.3'):

$$E(u_t|\mathbf{x}_t) = 0.$$

- This allows for feedback from past outcomes into explanatory variables and lagged values of y_t as regressors.

Asymptotic Properties of OLS

Under these conditions:

Theorem (Consistency of OLS)

Assume TS.1'-3', the OLS estimators are consistent:

$$\textit{plim}\hat{\beta} = \beta$$

Asymptotic Properties of OLS

- An important model satisfying TS.1'-3' is the AR(1) model:

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

where

$$E(u_t | y_{t-1}, y_{t-2}, \dots) = 0$$

- The model is clearly linear. For stationarity and weak dependence to hold, we need that $|\beta_1| < 1$. In this case, TS.1' holds.
- TS.3' also holds.
- OLS consistently estimates β_0 and β_1.
- β_1 is biased. The magnitude of the bias depends on sample size and how close β_1 is to 1.

Asymptotic Properties of OLS

- To obtain asymptotic distribution results, we need a version of TS.4 and TS.5:
- (TS.4') The errors are contemporaneously homoskedastic: $var(u_t|\mathbf{x}_t) = \sigma^2$.
- (TS.5') No serial correlation: $E(u_t u_s|\mathbf{x}_t \mathbf{x}_s) = 0$.
- This last condition may be delicate in static or finite distributed lag regression models.

Asymptotic Properties of OLS

Under these conditions:

Theorem (Asymptotic Normality of OLS)

Assume TS.1'-5', the OLS estimators are asymptotically normal. The usual OLS standard errors, t stats, F stats and LM stats are asymptotically valid.

Example: Efficient Markets Hypothesis

Let y_t be weekly percentage returns in the NYSE. A version of the EMH states that:

$$E(y_t | y_{t-1}, y_{t-2}, \dots) = E(y_t)$$

Otherwise, one could use past information to predict returns and once such investment opportunities are noticed, any predictability would disappear.

Using weekly data from 1976 to 1989, one gets:

$$\begin{array}{lccl} \widehat{return} = & 0.180 & +0.059 & return_{t-1} \\ & (0.081) & (0.038) & \end{array}$$

$(n = 689, R^2 = .0035)$. The autoregressive coefficient is not statistically different from zero at usual significance levels. An AR(2) model also supports the EMH.

Highly Persistent Time Series

- Many economic time series are not characterized by weak dependence (e.g. T-bill).
- To analyse those, we need to transform the data for it to be used in statistical analysis.

Random Walk

- An example of a highly persistent time series is the *random walk*:

$$y_t = y_{t-1} + u_t, \qquad t = 1, 2, \ldots$$

  where u_t is independent and identically distributed with mean zero and constant variance σ^2. It is also assumed that the initial value y_0 is independent of u_t (and typically set to zero).
- It can be shown that

$$E(y_t) = E(y_0), \qquad \text{for all } t$$

  and

$$Var(y_t) = \sigma^2 t$$

Random Walk

- The persistent behavior of the random walk is reflected in the fact that its present value is important even in the very distant future:

$$E(y_{t+h}|y_t) = y_t, \qquad \text{for all } h \geq 1$$

- No matter how far in the future, our best forecast is the current value of the series.
- In contrast, for the AR(1) process we studied earlier (with $|\rho| < 1$):

$$E(y_{t+h}|y_t) = \rho^h y_t$$

Random Walk

- It can be deduced that

$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}.$$

- The correlation depends on the starting point is goes to zero (as $h \to \infty$), but slowly as $t \to \infty$.

Random Walk

- One generalization of the RW is a *unit root process*, which allows u_t to be dependent.
- Another generalization is the RW with drift:

$$y_t = \alpha + y_{t-1} + u_t, \qquad t = 1, 2, \ldots$$

- This introduces a trend in the expected value of y_t:

$$E(y_t) = \alpha t + E(y_0), \qquad \text{for all } t$$

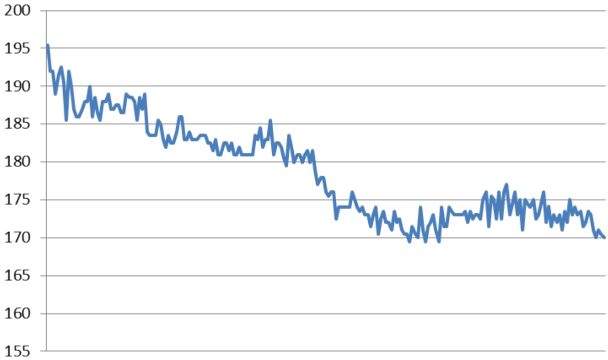
- It is important to note that trending and highly persistent behaviors are different. Many highly persistent series (i.e., interest rates, inflation rates and unemployment rates) do not present an obvious trend.

Random Walk

- One of the problems of a unit root is that OLS estimators of the autoregressive coefficient ($= 1$) are consistent, but their asymptotic distribution is not normal!
- In fact, the estimator of the autoregressive coefficient tends to be severely biased towards zero: the expected value of $\hat{\beta}_1$ is approximately $1 - 5.3/T$. With 20 years of quarterly data (=80 observations), the expectation is approximately 0.934.
- Because the distribution has a long left tail, t-tests will also be affected.

An Example

- Consider for example the following “misterious” time series:



- This series is statistically indistinguishable from a unit root (even at 10%).

Transformations of Highly Persistent Time Series

- Unit root processes can lead to errenous inference if the assumptions guaranteeing the asymptotic properties fail.
- Simple transformations may nonetheless be enough to address some of the issues.
- For example, if we first-difference a random walk process we end up with a stationary process:

$$y_t - y_{t-1} = u_t$$

- It would also remove the time trend from a trending variable $y_t = \alpha_0 + \alpha_1 t + u_t$:

$$y_t - y_{t-1} = \alpha_1 + u_t - u_{t-1}$$

Transformations of Highly Persistent Time Series

To estimate the elasticity of hourly wages with respect to output per hour, one can estimate:

$$\log(hour_t) = \beta_0 + \beta_1 \log(outphr_t) + \beta_2 t + u_t$$

where the time trend is included to accommodate the fact that both series are clearly upward trending.

Results are given below:

$$\begin{array}{lcccccc} \log(\widehat{hrwage_t}) = & -5.33 & +1.64 & \log(outputhr_t) & -.018 & t \\ & (0.37) & (0.09) & & (0.002) & \end{array}$$

Transformations of Highly Persistent Time Series

But even after linearly detrending, the first order autocorrelation is still high for both variables (0.967 and 0.945).

To accommodate the possibility of a unit root, the model is reestimated in differences:

$$\begin{array}{rccl} \Delta\log(\widehat{hrwage_t}) = & -.0036 & +.809 & \Delta\log(outputhr_t) \\ & (.0042) & (.173) & \end{array}$$

which substantially reduces the estimated elasticity.

These slides covered:

Wooldridge 10, 11.1-3, Stock and Watson 14 and 15.