ECMT3150: The Econometrics of Financial Markets

1c. Linear Time Series Analysis

Simon Kwok
University of Sydney

Semester 1, 2021

Outline

1. Regression with Time Series Errors
   1.1 When would a regression break down?
   1.2 Spurious Regression
   1.3 Cointegration
2. Unit Root Models
   2.1 Stochastic vs Deterministic Trends
   2.2 Unit Root Tests
3. Seasonal Models
4. Long Memory Models

Regression Model with Time Series Errors

Let $\{x_t\}$ and $\{y_t\}$ be two time series. Let's say we run the regression

$$y_t = \underset{1\times k}{x_t'} \cdot \underset{k\times 1}{\beta} + \varepsilon_t. \tag{1}$$

The OLS estimate of β is

$$\hat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_t. \tag{2}$$

Suppose the conditional covariance of $\hat{\beta}$ is estimated by

$$\widehat{Cov}(\hat{\beta}|x) = \hat{\sigma}_\varepsilon^2 \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1},$$

where $\hat{\sigma}_\varepsilon^2$ is the OLS variance estimator of the residuals $\{\hat{\varepsilon}_t\}$, given by $\hat{\sigma}_\varepsilon^2 = \frac{1}{T-k}\sum_{t=1}^{T} \hat{\varepsilon}_t^2$.

Q: Upon diagnostic checks, we may find that $\{\hat{\varepsilon}_t\}$ are heteroskedastic and/or serially correlated. Is $\hat{\beta}$ consistent for β, and $\widehat{Cov}(\hat{\beta}|x)$ consistent for $Cov(\hat{\beta}|x)$?

A: It depends on the true data generating process (DGP).

Regression Model with Time Series Errors

Scenario 1: Let $x = [x_1, x_2, \ldots, x_T]$, a $k \times T$ matrix of full row rank. The true DGP is (1), where $\{x_t\}$ is a covariance stationary process with $E(\|x_t\|^2) < \infty$ and $E[x_t x_t']$ being positive definite, $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$, and the two processes $\{x_t\}$ and $\{\varepsilon_t\}$ are independent.
Consequence: $\hat{\beta}$ is consistent, and $\widehat{Cov}(\hat{\beta}|x)$ is consistent.

Sketch of proof: Substitute (1) into (2).

$$\hat{\beta} = \beta + \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t \varepsilon_t.$$

Rewrite the equation as

$$\hat{\beta} - \beta = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t\right).$$

Since $E(\|x_t\|^2) < \infty$ and $E(\varepsilon_t^2) < \infty$, which imply that $E(\|x_t \varepsilon_t\|) < \infty$ by Cauchy-Schwartz inequality, we apply the strong law of large numbers and obtain

$$\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \stackrel{a.s.}{\to} E(x_t x_t'), \qquad \frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t \stackrel{a.s.}{\to} E(x_t \varepsilon_t) = 0,$$

as $T \to \infty$, so that $\hat{\beta} - \beta \stackrel{a.s.}{\to} 0$ as $T \to \infty$.

Regression Model with Time Series Errors

Conditional on $x = [x_1, x_2, \ldots, x_T]$, the covariance matrix of $\hat{\beta}$ is

$$\begin{aligned} Cov(\hat{\beta}|x) &= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'] \\ &= \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} Cov\left(\sum_{t=1}^{T} x_t \varepsilon_t \middle| x\right) \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}. \end{aligned}$$

Now let us compute

$$\begin{aligned} Cov\left(\sum_{t=1}^{T} x_t \varepsilon_t \middle| x\right) &= E\left[\left(\sum_{t=1}^{T} x_t \varepsilon_t\right)\left(\sum_{t=1}^{T} x_t' \varepsilon_t\right) \middle| x\right] \\ &= E\left[\sum_{t=1}^{T} x_t x_t' \varepsilon_t^2 + \sum_{s=1}^{T}\sum_{t=s+1}^{T} (x_s x_t' + x_t x_s')\varepsilon_s \varepsilon_t \middle| x\right] \\ &= \sum_{t=1}^{T} x_t x_t' E(\varepsilon_t^2|x) + \sum_{s=1}^{T}\sum_{t=s+1}^{T} (x_s x_t' + x_t x_s') E(\varepsilon_s \varepsilon_t|x). \end{aligned}$$

Since $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$ and $E(x_t \varepsilon_t) = 0$, we have $E(\varepsilon_t^2|x) = E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_s \varepsilon_t|x) = E(\varepsilon_s \varepsilon_t) = 0$ for all $s \neq t$, and so $Cov\left(\sum_{t=1}^{T} x_t \varepsilon_t \middle| x\right) = \sigma_\varepsilon^2 \left(\sum_{t=1}^{T} x_t x_t'\right)$. It follows that $Cov(\hat{\beta}|x) = \sigma_\varepsilon^2 \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}$.

Since $E(\varepsilon_t^2) < \infty$, we have, by the strong law of large numbers, $\hat{\sigma}_\varepsilon^2 \xrightarrow{a.s.} \sigma_\varepsilon^2$, and hence,

$$\widehat{Cov}(\hat{\beta}|x) = \hat{\sigma}_\varepsilon^2 \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \xrightarrow{a.s.} \sigma_\varepsilon^2 \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} = Cov(\hat{\beta}|x) \text{ as } T \to \infty.$$

Regression Model with Time Series Errors

Scenario 2: The true DGP is (1) with serially uncorrelated but heteroskedastic ε_t.
Consequence: $\hat{\beta}$ is consistent, but $\widehat{Cov}(\hat{\beta}|x)$ is inconsistent.

Solution: Use the White (1980) heteroskedasticity consistent (HC) estimator for $Cov(\hat{\beta}|x)$:

$$\widehat{Cov}(\hat{\beta}|x)_{HC} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\sum_{t=1}^{T} \hat{\varepsilon}_t^2 x_t x_t'\right) \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}.$$

Regression Model with Time Series Errors

Scenario 3: The true DGP is (1) with serially correlated and heteroskedastic $\{\varepsilon_t\}$.
Consequence: $\hat{\beta}$ is consistent, but $\widehat{Cov}(\hat{\beta}|x)$ and $\widehat{Cov}(\hat{\beta}|x)_{HC}$ are inconsistent.

Solution: Use the Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) estimator:

$$\begin{aligned}
\widehat{Cov}(\hat{\beta}|x)_{HAC} &= \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \hat{C}_{HAC} \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}, \\
\hat{C}_{HAC} &= \sum_{t=1}^{T} \hat{\varepsilon}_t^2 x_t x_t' + \sum_{j=1}^{\ell} w_j \sum_{t=j+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j} (x_t x_{t-j}' + x_{t-j} x_t').
\end{aligned}$$

One needs to pick the truncation parameter ℓ and the weights $\{w_j\}_{j=1}^{\ell}$ (e.g., $\ell = \left\lfloor 4\left(\frac{T}{100}\right)^{2/9} \right\rfloor$ and $w_j = 1 - \frac{j}{\ell+1}$).

Regression Model with Time Series Errors

Scenario 4: The true DGP contains lagged values of y_t, and $\{\varepsilon_t\}$ is serially correlated. E.g.,

$$\begin{aligned} y_t &= \beta y_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \theta \varepsilon_{t-1} + u_t, \end{aligned}$$

where $\{u_t\} \sim wn(0, \sigma_u^2)$.
Consequence: $\hat{\beta}$ is inconsistent.
To see this, first note that this model is the same as regression (1) with one regressor $x_t = y_{t-1}$ and an ε_t with $\{\varepsilon_t\} \sim AR(1)$. Now, let us compute

$$\begin{aligned} E(x_t \varepsilon_t) &= E[y_{t-1}\varepsilon_t] \\ &= E[y_{t-1}(\theta \varepsilon_{t-1} + u_t)] \\ &= \theta E(y_{t-1}\varepsilon_{t-1}) + E[y_{t-1}u_t]. \end{aligned}$$

However, $E(y_{t-1}\varepsilon_{t-1}) = E[(\beta y_{t-2} + \varepsilon_{t-1})\varepsilon_{t-1}] \neq 0$, and $E[y_{t-1}u_t] = 0$, so that $E(x_t \varepsilon_t) \neq 0$. As a result, the proof of the consistency of $\hat{\beta}$ under scenario 1 breaks down.

Solution: Use MLE instead of OLS.

Unit Root Nonstationarity

A process is nonstationary if some of its unconditional moments vary with time. Some examples are:

- random walk (unit root / stochastic trend process): $y_t = y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$.
- random walk with drift: $y_t = c + y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$.
- trend-stationary time series: $y_t = a + bt + u_t$, where u_t is stationary.
- ARIMA(p, d, q): $\{\Delta^d u_t\} \sim ARMA(p, q)$, where $\Delta^d y_t$ is the d^{th} order difference of y_t.[1] i.e., $[1 - \phi(L)]\,(\Delta^d u_t) = [1 + \theta\,(L)]\,\varepsilon_t$, where $\phi(\cdot)$ and $\theta(\cdot)$ are polynomial function of orders p and q, and $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$.

[1] For any positive integer $d \geq 1$, the d^{th} order difference is defined as $\Delta^d y_t = (1 - L)^d y_t$, where L is the lag operator. E.g.,
$\Delta^1 y_t = \Delta y_t = (1 - L) y_t = y_t - y_{t-1}$, and
$\Delta^2 y_t = (1 - L)^2 y_t = y_t - 2y_{t-1} + y_{t-2}$.

Random Walk (RW)

For $t \geq 1$, $y_t = y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$. $y_0 =$ initial value

Write y_t in terms of the noises: $y_t = y_0 + \varepsilon_t + \varepsilon_{t-1} + \cdots \varepsilon_1$.

Interpretation: any previous shock ε_{t-j} has a permanent effect on y_t.

Conditional on $\mathcal{F}_0$, the mean is $E(y_t|\mathcal{F}_0) = y_0$ and the variance is $Var(y_t|\mathcal{F}_0) = t\sigma_\varepsilon^2$. The variance grows linearly with time.

Ex: Show that $\hat{y}_t(\ell) = E[y_{t+\ell}|\mathcal{F}_t] = y_t$ for all $\ell > 0$.

This shows that $\{y_t\}$ is a *martingale*. Interpretation: the best point forecast of a RW is given by its current value.

Ex: Show that the forecast error $\hat{e}_t(\ell) = y_{t+\ell} - \hat{y}_t(\ell)$ has variance $\ell\sigma_\varepsilon^2$, which diverges as $\ell \to \infty$ (hopeless to forecast RW in the distant future).

Ex: Show that the ACF is $\rho_j = 1$ for all integers j (long memory).

Random Walk with Drift

For $t \geq 1$, $y_t = c + y_{t-1} + \varepsilon_t$, where $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$. $y_0 =$ initial value
Write y_t in terms of the noises:

$$y_t = \underbrace{y_0}_{\text{initial point}} + \underbrace{ct}_{\text{deterministic trend}} + \underbrace{\varepsilon_t + \varepsilon_{t-1} + \cdots \varepsilon_1}_{\text{stochastic trend}}.$$

$E(y_t|\mathcal{F}_0) = y_0 + ct$ (so $c =$ average rate of change in y_t over time).
Conditional on $\mathcal{F}_0$, $y_0 + ct$ is deterministic, so $Var(y_t|\mathcal{F}_0) = t\sigma_\varepsilon^2$ (same as RW without drift).

Ex: Show that $\hat{y}_t(\ell) = E[y_{t+\ell}|\mathcal{F}_t] = c\ell + y_t$ for all $\ell > 0$.

Ex: Show that $Var[\hat{e}_t(\ell)] = \ell\sigma_\varepsilon^2$.

Trend-Stationary Time Series

For $t \geq 1$, $y_t = a + bt + u_t$, where $\{u_t\}$ is stationary (e.g., an ARMA model) with mean zero and variance σ_u^2. $\{y_t\}$ has a deterministic linear trend $a + bt$ but no stochastic trend. $E(y_t) = a + bt$ (so $b =$ average rate of change in y_t over time). As $a + bt$ is deterministic, $Var(y_t) = \sigma_u^2$, which is time-invariant if it exists.

Ex: Show that $\hat{y}_t(\ell) = E[y_{t+\ell}|\mathcal{F}_t] = a + b(t + \ell)$ for all $\ell > 0$.

Ex: Show that $Var[\hat{e}_t(\ell)] = \sigma_u^2$.

Dickey-Fuller (DF) Test

Consider the regression

$$y_t = \phi_1 y_{t-1} + \varepsilon_t.$$

We want to test: $H_0 : \phi_1 = 1$ vs $H_a : \phi_1 < 1$.

Run the regression, get the OLS estimate $\hat{\phi}_1 = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$, and obtain the residual variance $\hat{\sigma}_\varepsilon^2 = \frac{1}{T-1} \sum_{t=1}^{T} (y_t - \hat{\phi}_1 y_{t-1})^2$.

The standard error of $\hat{\phi}_1$ is $s.e.(\hat{\phi}_1) = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\sum_{t=1}^{T} y_{t-1}^2}}$.

The DF test statistic is the t ratio of $\hat{\phi}_1$ under H_0:

$$DF = \frac{\hat{\phi}_1 - 1}{s.e.(\hat{\phi}_1)} = \frac{\sum_{t=1}^{T} \varepsilon_t y_{t-1}}{\hat{\sigma}_\varepsilon \sqrt{\sum_{t=1}^{T} y_{t-1}^2}}.$$

Under H_0, DF converges to a nonstandard distribution (function of standard Brownian motion) as $T \to \infty$ (need to use simulation to get critical value).

Dickey-Fuller Test

$$DF = \frac{\sum_{t=1}^{T} \varepsilon_t y_{t-1}}{\hat{\sigma}_\varepsilon \sqrt{\sum_{t=1}^{T} y_{t-1}^2}}.$$

Q: What is the asymptotic distribution of DF under H_0?

Sketch of proof: Let $W(t)$ be the standard Brownian motion (in continouous time). As $T \to \infty$, by the strong law of large numbers,

- $\hat{\sigma}_\varepsilon^2 \xrightarrow{a.s.} \sigma_\varepsilon^2$.

Also, applying the *functional central limit theorem,*

- $\frac{1}{T} \sum_{t=1}^{T} y_{t-1} \varepsilon_t \xrightarrow{d} \frac{\sigma_\varepsilon^2}{2} [W(1)^2 - 1]$,
- $\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma_\varepsilon^2 \int_0^1 W(s)^2 ds$.

Combining the limits using Slutsky's theorem, we obtain

$$DF \xrightarrow{d} \frac{\frac{1}{2} [W(1)^2 - 1]}{\sqrt{\int_0^1 W(s)^2 ds}}.$$

Augmented Dickey-Fuller (ADF) Test

We augment the regression model with a deterministic intercept c_t and $p-1$ lagged differenced series $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$

$$y_t = c_t + \beta y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + \varepsilon_t.$$

We want to test: $H_0 : \beta = 1$ vs $H_a : \beta < 1$.
The ADF test statistic is the t ratio of $\hat{\beta}$ (OLS estimate of β):

$$ADF = \frac{\hat{\beta} - 1}{s.e.(\hat{\beta})}.$$

Under H_0, ADF converges to a different nonstandard distribution as $T \to \infty$ (need to use simulation to get critical value).
Equivalently, we may run the *error-correction* regression of Δy_t:

$$\Delta y_t = c_t + \beta_c y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + \varepsilon_t.$$

Note that $\beta_c = \beta - 1$.

Spurious Regression

Suppose $\{y_t\}$ and $\{x_t\}$ contain a unit root.

Q: After running the regression (1), we may detect a unit root in the residuals (e.g., as revealed by ADF test), and find a statistically significant $\hat{\beta}$. Is the inference on $\hat{\beta}$ reliable?

A: $\hat{\beta}$ can be spuriously significant, and R^2 spuriously high. This is known as *spurious regression*.

Solutions:

- Take the first-order difference of $\{y_t\}$ and $\{x_t\}$, and run the regression $\Delta y_t = \alpha + \beta \Delta x_t + \varepsilon_t$. Check for serial correlations of the residuals by looking at their ACF. Add lags of Δy_t, Δx_t and ε_t if necessary. The OLS estimates $\hat{\alpha}$ and $\hat{\beta}$ are $\sqrt{T}$-consistent and asymptotically normal as $T \to \infty$.
- Add lagged values of y_t and x_t to regression (1). However, the asymptotic distributions of $\hat{\alpha}$ and $\hat{\beta}$ are non-standard.
- Apply Cochrane-Orcutt adjustment.

Cointegration

Suppose $\{y_t\}$ and $\{x_t\}$ contain a unit root.

Q: After running regression (1), we find that the residuals are stationary. What does $\hat{\beta}$ represent?

A: In this case, $\{y_t\}$ and $\{x_t\}$ are *cointegrated*, We say that the $\{(y_t, x_t)\}$ pair displays a *cointegrating relationship* given by (1) with cointegrating vector $(1, -\beta)$. As for inference, the OLS estimate $\hat{\beta}$ is super-consistent (T-consistent), and the asymptotic distribution is non-standard.

Seasonal Models

Time series may exhibit cyclical patterns (e.g., weekly pattern for daily series, monthly pattern for weekly series).

Q: Suppose $\{y_t\}$ has a cyclical pattern of periodicity s. How to carry out analysis?

A: If $\{y_t\}$ is stationary, we may remove the cyclicity by applying the seasonal adjustment:

$$\begin{aligned} \Delta_s y_t &= (1 - L^s) y_t \\ &= y_t - y_{t-s}. \end{aligned}$$

If $\{y_t\}$ has a unit root, we may apply the seasonal adjustment to the first-differenced series:

$$\begin{aligned} \Delta_s(\Delta y_t) &= (1 - L^s)(1 - L) y_t = (y_t - y_{t-1}) - (y_{t-s} - y_{t-s-1}) \\ &= y_t - y_{t-1} - y_{t-s} + y_{t-s-1}. \end{aligned}$$

Seasonal Models

Multiplicative season model:

$$w_t := (1 - L^s)(1 - L)y_t = (1 - \theta L)(1 - \lambda L^s)u_t \qquad (3)$$

Ex: What is the ACF of $\{w_t\}$?

If $\lambda = 1$, then the seasonal factor $(1 - L^s)$ appears on both sides of (3). This suggests that the seasonal pattern is deterministic. Exact-likelihood estimation can reveal this and is recommended.

Long Memory / Fractionally differenced Model

Let $\{\varepsilon_t\} \sim wn(0, \sigma_\varepsilon^2)$, and suppose that $\phi(\cdot)$ and $\theta(\cdot)$ are polynomial functions of orders p and q. We say that y_t follows an *autoregressive fractionally integrated moving-average* model, $ARFIMA(p, d, q)$, if

$$[1 - \phi(L)]\Delta^d y_t := [1 + \theta(L)]\varepsilon_t.$$

- $d \in (-0.5, 0)$: long-range negative dependence, with ACF $\rho_j \sim j^{2d-1}$ (hyperbolic decay) as $j \to \infty$.[2]
- $d = 0$: e.g., for $AR(1)$, $\rho_j = \phi_1^{|j|}$ (exponential decay).
- $d \in (0, 0.5)$: long-range positive dependence, $\rho_j \sim j^{2d-1}$ (hyperbolic decay) as $j \to \infty$.
- $d \in [0.5, 1)$: mean-reverting, non-stationary process.
- $d = 1$: martingale, unit root process, $\rho_j \equiv 1$ for all integers j.

[2] "$\sim$" means "is proportional to".