Section 11 Basics of Time-Series Regression

- What's different about regression using time-series data?
  - Dynamic effects of X on Y
  - Ubiquitous autocorrelation of error term
  - Difficulties of nonstationary Y and X
    - Correlation is common just because both variables follow trends
    - This can lead to "spurious regression" if we interpret the common trend movement as true correlation
  - Focus is on estimating the properties of the data-generating process rather than population parameters
  - Variables are often called "time series" or just "series"
- **Lags and differences**
  - With time-series data we are often interested in the relationship among variables at different points in time.
  - Let X_t be the observation corresponding to time period t.
    - The first lag of X is the preceding observation: X_{t-1}.
    - We sometimes use the **lag operator** $L(X_t)$ or $LX_t \equiv X_{t-1}$ to represent lags.
    - We often use higher-order lags: $L^s X \equiv X_{t-s}$.
  - The first difference of X is the difference between X and its lag:
    - $\Delta X_t \equiv X_t - X_{t-1} = (1 - L)X_t$
    - Higher-order differences are also used:
      $\Delta^2 X_t = \Delta(\Delta X_t) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$
      $= (1 - L)^2 X_t = (1 - 2L + L^2)X_t$
    - $\Delta^s X_t = (1 - L)^s X_t$
  - Difference of the log of a variable is approximately equal to the variable's growth rate: $\Delta(\ln X_t) = \ln X_t - \ln X_{t-1} = \ln(X_t / X_{t-1}) \approx X_t / X_{t-1} - 1 = \Delta X_t / X_t$
    - Log difference is exactly the continuously-compounded growth rate
    - The discrete growth-rate formula $\Delta X_t / X_t$ is the formula for once-per-period compounded growth
  - Lags and differences in Stata
    - First you must define the data to be time series: tsset year
      - This will correctly deal with missing years in the year variable.
      - Can define a variable for quarterly or monthly data and set format to print out appropriately.
      - For example, suppose your data have a variable called month and one called year. You want to combine into a single time variable called time.
        - gen time = ym(year, month)

        - This variable will have a %tm format and will print out like 2010m4 for April 2010.
        - You can then do tsset time
    - Once you have the time variable set, you can create lags with the lag operator l. and differences with d.
        - For example, last period's value of x is l.x
        - The change in x between now and last period is d.x
        - Higher-order lags and differences can be obtained with l3.x for third lag or d2.x for second difference.
- **Autocovariance and autocorrelations**
    - Autocovariance of order *s* is cov(X_t, X_{t-s})
        - We generally assume that the autocovariance depends only on *s*, not on *t*.
        - This is analogous to our Assumption #0: that all observations follow the same model (or were generated by the same data-generating process)
        - This is *one element* of a time series being stationary
    - Autocorrelation of order *s* (ρ_s) is the correlation coefficient between X_t and X_{t-s}.
        - $\rho_s = \dfrac{\text{cov}(X_t, X_{t-s})}{\text{var}(X_t)}$
        - We estimate with $\hat{\rho}_s = \dfrac{\frac{1}{T-s-1}\sum_{t=s+1}^{T}\left(X_t - \bar{X}_{s+1,T}\right)\left(X_{t-s} - \bar{X}_{1,T-s}\right)}{\frac{1}{T-1}\sum_{t=1}^{T}\left(X_t - \bar{X}_{1,t}\right)}$, where

          $\bar{X}_{t_1,t_2} \equiv \dfrac{1}{t_2 - t_1 + 1}\sum_{t=t_1}^{t_2} X_t$ is the mean of *X* over the range of observations designated by the pair of subscripts.
            - We sometimes ignore the different fractions in front of the summations since their ratio goes to 1 as *T* goes to ∞.

Univariate time-series models

- We sometimes represent a variable's time-series behavior with a univariate model.
- **White noise:** The simplest univariate time-series process is called white noise $Y_t = u_t$, where u_t is a mean-zero IID error (usually normal).
    - The key point here is the autocorrelations of white noise are all zero (except, of course, for ρ_0, which is always 1).
    - Very few economic time series are white noise.
        - Changes in stock prices are probably one.
    - We use white noise as a basic building block for more useful time series:
        - Consider problem of forecasting Y_t conditional on all past values of *Y*.

- $Y_t = E[Y_t \mid Y_{t-1}, Y_{t-2}, \ldots] + u_t$
- Since any part of the past behavior of Y that would help to predict the current Y should be accounted for in the expectation part, the error term u should be white noise.
- The one-period-ahead forecast error of Y should be white noise.
- We sometimes call this forecast-error series the "fundamental underlying white noise series for Y" or the "innovations" in Y.

- The simplest autocorrelated series is the **first-order autoregressive (AR(1)) process**: $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$, where u is white noise.
  - In this case, our one-period-ahead forecast is $E[Y_t \mid Y_{t-1}] = \beta_0 + \beta_1 Y_{t-1}$ and the forecast error is u_t.
  - For simplicity, suppose that we have removed the mean from Y so that $\beta_0 = 0$.
    - Consider the effect of a one-time shock u_1 on the series Y from time one on, assuming (for simplicity) that $Y_0 = 0$ and all subsequent u values are also zero.
    - $Y_1 = \beta_1(0) + u_1 = u_1$
    - $Y_2 = \beta_1 Y_1 + u_2 = \beta_1 u_1$
    - $Y_3 = \beta_1 Y_2 + u_3 = \beta_1^2 u_1$
    - $Y_s = \beta_1^{s-1} u_1$.
    - This shows that the effect of the shock on Y "goes away" over time only if $|\beta_1| < 1$.
      - The condition $|\beta_1| < 1$ is necessary for the AR(1) process to be **stationary**.
    - If $\beta_1 = 1$, then shocks to Y are permanent. This series is called a **random walk**.
      - The random walk process can be written $Y_t = Y_{t-1} + u_t$ or $\Delta Y_t = u_t$. The first difference of a random walk is stationary and is white noise.
  - If Y follows a stationary AR(1) process, then $\rho_1 = \beta_1$, $\rho_2 = \beta_1^2$, …, $\rho_s = \beta_1^s$.
    - One way to attempt to identify the appropriate specification for a time-series variable is to examine the **autocorrelation function** of the series, which is defined as ρ_s considered as a function of s.
    - If the autocorrelation function declines exponentially toward zero, then the series might follow an AR(1) process with positive β_1.
    - A series with $\beta_1 < 0$ would oscillate back and forth between positive and negative responses to a shock.

        - The autocorrelations would also oscillate between positive and negative while converging to zero.

- The **AR(*p*)** process
    - We can generalize the AR(1) to allow higher-order lags:
      $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + u_t$
    - We can write this compactly using the lag operator notation:
      $Y_t = \beta_0 + \beta_1 L Y_t + \beta_2 L^2 Y_t + \ldots + \beta_p L^p Y_t + u_t$
      $Y_t - \beta_1 L Y_t - \beta_2 L^2 Y_t - \ldots - \beta_p L^p Y_t = \beta_0 + u_t$
      $\left(1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_p L^p\right) Y_t = \beta_0 + u_t$
      $\beta(L) Y_t = \beta_0 + u_t,$
      where $\beta(L)$ is a degree-p polynomial in the lag operator with zero-order coefficient of 1.
    - We can again analyze the dynamic effect on Y of a shock u_0 most easily by assuming a zero mean ($\beta_0 = 0$), no previous shocks ($Y_0 = Y_{-1} = \ldots = Y_{-p} = 0$), and no subsequent shocks ($u_2 = u_3 = \ldots = 0$).
        - It gets complicated very quickly:
          $Y_1 = u_1$
          $Y_2 = \beta_1 u_1$
          $Y_3 = \beta_1^2 u_1 + \beta_2 u_1$
          $Y_4 = \beta_1\left(\beta_1^2 + \beta_2\right) u_1 + \beta_2 \beta_1 u_1 + \beta_3 u_1$
        - We can most easily see the behavior of this by utilizing the polynomial in the lag operator:
          $\beta(L) Y_t = u_t$
          $Y_t = \frac{1}{\beta(L)} u_t$
          where the division is the equation is polynomial division.
            - You can easily verify that for the AR(1) process,
              $\beta(L) = 1 - \beta_1 L$
              $\frac{1}{\beta(L)} = 1 + \beta_1 L + \beta_1^2 L^2 + \ldots = \sum_{i=0}^{\infty} \beta_1^i L^i$
              $\frac{1}{\beta(L)} u_t = \sum_{i=0}^{\infty} \beta_1^i u_{t-i}$
              so $Y_{t+s+1} = \beta_1^s u_1$ is the effect of the shock in period 1 on the value of Y s periods later.
            - Note that this process is stationary only if $|\beta_1| < 1$.

- Consider solving for the roots of the equation $\beta(L) = 0$. In the AR(1) case, this is

  $$1 - \beta_1 L = 0$$
  $$\beta_1 L = 1$$
  $$L = \frac{1}{\beta_1}.$$

- Thus the roots of $\beta(L) = 0$ must be *greater than 1 in absolute value* if the process is to be stationary.
- A **unit root** means that $\beta_1 = 1$, which is the random walk process.

  - This pattern of association between the stationarity of the AR process and the roots of the equation $\beta(L) = 0$ carries over into higher-order processes.
    - With a higher-order process, the roots of the equation can be complex rather than real.
    - The corresponding stationarity condition is that the roots of $\beta(L) = 0$ must lie *outside the unit circle in the complex plane*.
    - As we will discuss at more length later, for each unit root—lying on the unit circle in the complex plane—we must difference Y once to achieve stationarity.
    - With roots that lie inside the unit circle, the model is irretrievably nonstationary.

- **Moving-average processes**
  - We won't use them or study them much, but it's worth briefly discussing the possibility of putting lags on u_t in the univariate model.
  - $$\begin{aligned} Y_t &= \alpha_0 + u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \ldots + \alpha_q u_{t-q} \\ &= \alpha_0 + \alpha(L) u_t \end{aligned}$$

    is called a moving-average process of order q or MA(q)
    - Note that the coefficient of u_t is normalized to one.
  - Finite MA processes are always stationary because the effect of a shock to u dies out after q periods.
    - The autocorrelations of an MA(q) process are zero for lags greater than q.
  - When we divided by $\beta(L)$ to solve the AR process, we created an infinite MA process on the right side.
- We can combine AR and MA processes to get the ARMA(p, q) process:

  $$Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + u_t + \alpha_1 u_{t-1} + \ldots + \alpha_q u_{t-q}$$

$\beta(L)Y_t = \beta_0 + \alpha(L)u_t$, which solves to

$$Y_t = \frac{\beta_0}{\beta(L)} + \frac{\alpha(L)}{\beta(L)}.$$

  - Stationarity of the ARMA process depends entirely on the roots of $\beta(L) = 0$ lying outside the unit circle because all finite MA processes are stationary.
- Considerations in estimating ARMA models
  - Autoregressive models can in principle be estimated by OLS, but there are some pitfalls.
    - By including enough lagged Y terms we should be able to eliminate any autocorrelation in the error term: make u white noise.
      - If we have autocorrelation in u, then u_t will be correlated with u_{t-1}, which is part of Y_{t-1}. So Y_{t-1} is correlated with the error term and OLS is biased and inconsistent.
    - If Y is highly autocorrelated, then its lags are highly correlated with each other, which makes multicollinearity a concern in trying to identify the coefficients.
  - MA models cannot be estimated by OLS because we have no values for lagged u terms through which to estimate the α coefficients.
    - Box and Jenkins developed a complicated iterative process for estimating models with MA errors.
      - First we estimate a very long AR process (because an MA can be expressed as an infinite AR just as an AR can be expressed as an infinite MA) and use it to calculate residuals.
      - We then plug in these residuals ("back-casting" to move generate the pre-sample observations) and estimate the α coefficients.
      - Then we recalculate the residuals and iterate under the estimates of the α coefficients converge.
    - Stata will estimate ARMA (or ARIMA) models with the arima command.

Regression with serially correlated error terms

Before we deal with issues of specifications of Y and X, we will think about the problems that serially correlated error terms cause for OLS regression.

- Can estimate time-series regressions by OLS as long as Y and X are stationary and X is exogenous.
  - **Exogeneity**: $E(u_t \mid X_t, X_{t-1}, \ldots) = 0.$
  - **Strict exogeneity**: $E(u_t \mid \ldots, X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \ldots) = 0.$

- However, nearly all time-series regressions are prone to having serially correlated error terms.
    - Omitted variables are probably serially correlated
- This is a particular form of violation of the IID assumption.
    - Observations are correlated with those of nearby periods
- As long as the other OLS assumptions are satisfied, this causes a problem not unlike heteroskedasticity
    - OLS is still unbiased and consistent
    - OLS is not efficient
    - OLS estimators of standard errors are biased, so cannot use ordinary t statistics for inference
- To some extent, adding more lags of Y and X to the specification can reduce the severity of serial correlation.
- Two methods of dealing with serial correlation of the error term:
    - GLS regression in which we transform the model to one whose error term is not serially correlated
        - This is analogous to weighted least squares (also a GLS procedure)
    - Estimate by OLS but use standard error estimates that are robust to serial correlation
- **GLS with an AR(1) error term**
    - One of the oldest time-series models (and not used so much anymore) is the model in which u_t follows and AR(1) process:

      $Y_t = \beta_0 + \beta_1 X_t + u_t$

      $u_t = \phi u_{t-1} + \varepsilon_t,$

      where ε is a white-noise error term and $-1 < \phi < 1$.
        - In practice, $\phi > 0$ nearly always
    - GLS transforms the model into one with an error term that is not serially correlated.
    - Let

      $$\tilde{Y}_t = \begin{cases} Y_t\sqrt{\left(1-\phi^2\right)}, & t = 1, \\ Y_t - \phi Y_{t-1}, & t = 2, 3, \ldots, T, \end{cases} \quad \tilde{X}_t = \begin{cases} X_t\sqrt{\left(1-\phi^2\right)}, & t = 1, \\ X_t - \phi X_{t-1}, & t = 2, 3, \ldots, T, \end{cases}$$

      $$\tilde{u}_t = \begin{cases} u_t\sqrt{\left(1-\phi^2\right)}, & t = 1, \\ u_t - \phi u_{t-1}, & t = 2, 3, \ldots, T. \end{cases}$$

    - Then $\tilde{Y}_t = (1-\phi)\beta_0 + \beta_1 \tilde{X}_t + \tilde{u}_t.$
        - The error term in this regression is equal to ε_t for observations 2 through T and is a multiple of u_1 for the first observation.

- By assumption, ε is white noise and values of ε in periods after 1 are uncorrelated with u_1, so there is no serial correlation in this transformed model.
- If the other assumptions are satisfied, it can be estimated efficiently by OLS.

- But what is ϕ?
  - Need to estimate ϕ to calculate feasible GLS estimator.
  - Traditional estimator for ϕ is $\hat{\phi} = \text{corr}(\hat{u}_t, \hat{u}_{t-1})$ using OLS residuals.
  - This estimation can be iterated to get a new estimate of ϕ based on the GLS estimator and then re-do the transformation: repeat until converged.
- Two-step estimator using FGLS based on $\hat{\phi}$ is called the **Prais-Winsten** estimator (or **Cochrane-Orcutt** when first observation is dropped).
- Problems: ϕ is not estimated consistently if u_t is correlated with X_t, which will always be the case if there is a lagged dependent variable present and may be the case if X is not strongly exogenous.
  - In this case, we can use nonlinear methods to estimate ϕ and β jointly by search.
  - This is called the **Hildreth-Lu** method.
- In Stata, the prais command implements all of these methods (depending on option). Option corc does Cochrane-Orcutt; ssesearch does Hildreth-Lu; and the default is Prais-Winsten.
- You can also estimate this model with ϕ as the coefficient on Y_{t-1} in an OLS model, as in S&W's Section 15.5.

HAC consistent standard errors (Newey-West)

- As with White's heteroskedasticity consistent standard errors, we can correct the OLS standard errors for autocorrelation as well.
- We know that

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{T}\sum_{i=1}^{T}(X_t - \bar{X})u_t}{\frac{1}{T}\sum_{i=1}^{T}(X_t - \bar{X})^2}.$$

- In this formula, $\text{plim}\,\bar{X} = \mu_X$, $\text{plim}\left(\frac{1}{T}\sum_{i=1}^{T}(X_t - \bar{X})^2\right) = \sigma_X^2$.
- So $\text{plim}(\hat{\beta}_1 - \beta_1) = \frac{\text{plim}\left(\frac{1}{T}\sum_{i=1}^{T}(X_t - \mu_X)u_t\right)}{\sigma_X^2} = \frac{\text{plim}(\bar{v})}{\sigma_X^2}$, where $\bar{v} = \frac{1}{T}\sum_{i=1}^{T} v_t$ and $v_t \equiv (X_t - \mu_X)u_t$.

- And in large samples, $\text{var}\left(\hat{\beta}_1\right) = \text{var}\left(\frac{\bar{v}}{\sigma_X^2}\right) = \frac{\text{var}\left(\bar{v}\right)}{\sigma_X^4}$.
  - Under IID assumption, $\text{var}\left(\bar{v}\right) = \frac{1}{T}\text{var}\left(v_t\right) = \frac{\sigma_v^2}{T}$, and the formula reduces to one we know from before.
  - However, serial correlation means that the error terms are not IID (and X is usually not either), so this doesn't apply.
- In the case where there is serial correlation we have to take into account the covariance of the v_t terms:

$$\begin{aligned}
\text{var}\left(\bar{v}\right) &= \text{var}\left(\frac{v_1 + v_2 + \ldots + v_T}{T}\right) \\
&= \frac{1}{T^2}\left[\sum_{i=1}^{T}\sum_{j=1}^{T} E\left(v_i v_j\right)\right] \\
&= \frac{1}{T^2}\sum_{i=1}^{T}\left(\text{var}\left(v_i\right) + \sum_{j\neq i}\text{cov}\left(v_i, v_j\right)\right) \\
&= \frac{1}{T^2}\left[T\,\text{var}\left(v_t\right) + 2\left(T-1\right)\text{cov}\left(v_t, v_{t-1}\right) + 2\left(T-2\right)\text{cov}\left(v_t, v_{t-2}\right) + \ldots + 2\,\text{cov}\left(v_t, v_{t-(T-1)}\right)\right] \\
&= \frac{\sigma_v^2}{T} f_T,
\end{aligned}$$

  where

$$\begin{aligned}
f_T &\equiv 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\text{corr}\left(v_t, v_{t-j}\right) \\
&= 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j.
\end{aligned}$$

- Thus, $\text{var}\left(\hat{\beta}_1\right) = \left[\frac{1}{T}\frac{\sigma_v^2}{\sigma_X^4}\right] f_T$, which expresses the variance as the product of the no-autocorrelation variance and the f_T factor that corrects for autocorrelation.
- In order to implement this, we need to know f_T, which depends on the autocorrelations of v for orders 1 through $T-1$.
  - These are not known and must be estimated.
  - For ρ_1 we have lots of information because there are $T-1$ pairs of values for (v_t, v_{t-1}) in the sample.
  - For ρ_{T-1}, there is only one pair $(v_t, v_{t-(T-1)})$—namely (v_T, v_1)—on which to base an estimate.
  - The **Newey-West** procedure truncates the summation in f_T at some value $m-1$, so we estimate the first $m-1$ autocorrelations of v using the OLS residuals and compute $\hat{f}_T = 1 + 2\sum_{j=1}^{m-1}\left(\frac{m-j}{m}\right)\hat{\rho}_j$.

        - *m* must be large enough to provide a reasonable correction but small enough relative to *T* to allow the ρ values to be estimated well.
            - Stock and Watson suggest choosing $m = 0.75T^{\frac{1}{3}}$ as a reasonable rule of thumb.
    - To implement in Stata, use hac option in xtreg (with panel data) or post-estimation command newey , lags(m)

Distributed-lag models

- Univariate time-series models are an interesting and useful building block, but we are almost always interested not just in *Y*'s behavior by itself but also in how *X* affects *Y*.
- In time-series models, this effect is often dynamic: spread out over time.
    - This means that answering the question "How does *X* affect *Y*?" involves not just $\partial Y/\partial X$, but a more complex set of **dynamic multipliers** $\partial Y_t/\partial X_t$, $\partial Y_{t+1}/\partial X_t$, $\partial Y_{t+2}/\partial X_t$, etc.
- We estimate the dynamic effects of *X* on *Y* with **distributed-lag models**.
- In general, the distributed-lag model has the form $Y_t = \beta_0 + \sum_{i=1}^{\infty} \beta_i X_{t-i} + u_t$. But of course, we cannot estimate an infinite number of lag coefficients β_i, so we must either truncate or find another way to approximate an infinite lag structure.
    - We can easily have additional regressors with either the same or different lag structures.
- **Finite distributed lag**
    - Simplest lag distribution is to simply truncate the infinite distribution above and estimate $Y_t = \beta_0 + \sum_{i=1}^{r+1} \beta_i X_{t-i-1} + u_t$ by OLS.
        - Dynamic multipliers in this case are $\frac{\partial Y_t}{\partial X_{t-s}} = \beta_{s+1}$, for *s* = 0, 1, …, *r* and 0 otherwise.
        - Cumulative dynamic multipliers (effect of permanent change in *X*) are $\sum_{v=0}^{s} \beta_{s+1}$.
- **Koyck lag: AR(1) with regressors**
    - $Y_t = \beta_0 + \beta_1 Y_{t-1} + \delta_0 X_t + u_t$
    - $\frac{\partial Y_t}{\partial X_t} = \delta_0$

      $\frac{\partial Y_{t+1}}{\partial X_t} = \beta_1 \delta_0$

$$\frac{\partial Y_{t+2}}{\partial X_t} = \beta_1^2 \delta_0$$

$$\frac{\partial Y_{t+s}}{\partial X_t} = \beta_1^s \delta_0$$

Thus, dynamic multipliers start at δ_0 and decay exponentially to zero over infinite time. Thus, this is effectively a distributed lag of infinite length, but with only 2 parameters (plus intercept) to estimate.

- Cumulative multipliers are $\sum_{v=0}^{s} \frac{\partial Y_{t+v}}{\partial X_t} = \delta_0 \sum_{v=0}^{s} \beta_1^v$.
- Long-run effect of a permanent change is $\delta_0 \sum_{v=0}^{\infty} \beta_1^v = \frac{\delta_0}{1-\beta_1}$.
- Estimation has the potential problem of inconsistency if u_t is serially correlated.
  - This is a serious problem, especially as some of the test statistics for serial correlation of the error are biased when the lagged dependent variable is present.
- Koyck lag is parsimonious and fits lots of lagged relationships well.
- With multiple regressors, the Koyck lag applies the same lag structure (rate of decay) to all regressors.
  - Is this reasonable for your application?
  - Example: delayed adjustment of factor inputs: can't stop using expensive factor more quickly than you start using cheaper factor.

- **ARX(*p*) Model**
  - We can generalize the Koyck lag model to longer lags:
    $Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \delta_0 X_t + u_t$.
  - Same general principles apply:
    - Worry about stationarity of lag structure: roots of $\beta(L)$
    - If u is serially correlated, OLS will be biased and inconsistent
    - Dynamic multipliers are determined by coefficients of infinite lag polynomial $[\beta(L)]^{-1}$
    - If more than on X, all have same lag structure
  - How to determine length of lag p?
    - Can keep adding lags as long as β_p is statistically significant
    - Can choose to max the **Akaike information criterion (AIC)**:
      $$\text{AIC}(p) = \ln\left(\frac{SSR(p)}{T}\right) + \frac{2}{T}(p+1).$$
    - Can choose to max the **Bayesian (Schwartz) information criterion (BIC)**: $\text{BIC}(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\ln T}{T}$.

        - Note that regression can use as many as $T-p$ observations, but should use the same number for all regressions with different p values in assessing information criteria.
        - AIC will choose longer lag than BIC.
            - AIC came first, so is still used a lot
            - BIC is asymptotically unbiased
        - Stata calculates info criteria by estat ic (after regression)
    - Can also include moving-average terms to get ARMAX(p, q)
- **ADL(p, q) Model: "Rational" lag**
    - We can also add lags to the X variable(s)
    - $Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + \delta_0 X_t + \delta_1 X_{t-1} + \ldots + \delta_q X_{t-q} + u_t$
        - Note that S&W omit current effect here.
        - Can add more X variables with varying lag lengths

      $\beta(L) Y_t = \beta_0 + \delta(L) X_t + u_t,$
    - $Y_t = \frac{\beta_0}{\beta(L)} + \frac{\delta(L)}{\beta(L)} X_t + \frac{u_t}{\beta(L)}.$
    - Stationarity depends only on $\beta(L)$, not on $\delta(L)$.
    - Can easily estimate this by OLS assuming:
        - $E\left(u_t \mid Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}, X_t, X_{t-1}, \ldots, X_{t-q}\right) = 0$
        - (Y_t, X_t) has same mean, variance, and autocorrelations for all t
        - (Y_t, X_t) and (Y_{t-s}, X_{t-s}) become independent as $s \to \infty$
        - All variables have finite, non-zero fourth moments
        - No perfect multicollinearity
    - These are general assumptions that apply to most time-series models.
- **Granger causality**
    - Granger attempted to test causality by testing whether a variable's $\delta(L)$ polynomial was zero. (Need to leave X_t out of regression here.)
        - F test of set of coefficients on all lags of X, given effects of lagged Y and any other regressors.
        - Rejection means X causes Y.
    - Should not be called causality: the real question being answered is whether X helps to predict Y given the path that Y would follow based on its own lags (and perhaps on other regressors)

Nonstationarity due to trends

- Many macroeconomic variables have trends: output, prices, wages, money supply, productivity, etc. all tend to increase over time.

- **Deterministic trends** are constant increases over time, though the variable may fluctuate above or below its trend line randomly.
  - $Y_t = \beta_0 + \beta_1 t + u_t$
  - u is stationary error term
  - If the constant rate of change is in percentage terms, then we could model $\ln Y$ as being linearly related to time.
- **Stochastic trends** allow the trend change from period to period to be random, with given mean and variance.
  - Random walk is simplest version of stochastic trend: $Y_t = Y_{t-1} + u_t$ where u is white noise.
  - Random walk with drift allows for non-zero average change: $Y_t = \beta_0 + Y_{t-1} + u_t$
- Difference between deterministic and stochastic trend
  - Consider large negative shock u in period t
    - In deterministic trend, the trend line remains unchanged.
      - Because u is assumed stationary, its effect eventually disappears and the effect of the shock is temporary
    - In stochastic trend, the lower Y is the basis for all future changes in Y, so the effect of the shock is permanent.
  - Which is more appropriate?
    - No clear rule that always applies
    - Stochastic trends are popular right now, but they are controversial
- Note that the random-walk model is just the AR(1) model with $\beta_1 = 1$.
  - This is a **unit root** to the $\beta(L) = 0$ equation.
- Impacts of stochastic trends
  - Estimates of coefficients of an autoregressive process will be biased downward in small samples.
    - Can't test $\beta_1 = 0$ with usual test in OLS autoregression
  - Distributions of t statistics are not t or close to normal
  - Spurious regression
    - Non-stationary time series can appear to be related with they are not.
    - Show the Granger-Newbold results/tables
- **Testing for unit roots**
  - Since the desirable properties of OLS (and other) estimators depend on the stationarity of Y and X, it would be useful to have a test for a unit root.
  - **Dickey-Fuller test** is such a test.
  - We can't just run $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$ and test $\beta_1 = 1$, because the distribution is not reliable under the null hypothesis of nonstationarity.
  - If we subtract Y_{t-1} from both sides, we get $\Delta Y_t = \beta_0 + (\beta_1 - 1) Y_{t-1} + u_t$.

        - If the null hypothesis is true ($\beta_1 = 1$) then the dependent variable is non-stationary and the coefficient on the right is zero.
        - Under the null hypothesis, Y follows a random walk with drift (β_0).
        - We can test this hypothesis with an OLS regression, but because the regressor is nonstationary (under the null), the t statistic will not follow the t or asymptotically normal distribution. Instead, it follows the Dickey-Fuller distribution, with critical values much higher than those of the normal.
        - If the DF statistic exceeds the critical value at our desired level of significance, then we *reject the null hypothesis of non-stationarity* and conclude that the variable is stationary.
            - Note that a one-tailed test is appropriate here because $\beta_1 - 1$ should always be negative. Otherwise, it would imply $\beta_1 > 1$, which is non-stationary in a way that cannot be rectified by differencing.
    - Can we be sure that u is not serially correlated?
        - Probably not, but by adding some lags of ΔY on the RHS we can usually eliminate the serial correlation of the error.
        - $\Delta Y_t = \beta_0 + (\beta_1 - 1)Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \ldots + \gamma_p \Delta Y_{t-p} + u_t$ is the model for the Augmented Dickey-Fuller (ADF) test, which is similar but has a different distribution that depends on p.
    - We can also use an ADF test to test the null hypothesis that the series is stationary around a deterministic trend ("trend stationary") by adding a linear trend term δt to the equation.
    - Intuition of DF and ADF tests: Stationary series are "mean-reverting." They tend to come back to their means after a disturbance, so the future *changes* depend on the current *level* of the series.
        - If the level Y_{t-1} affects the change in time t (negatively), then the series will tend to revert to a fixed mean and is stationary.
        - If the change in time t is independent of the level, then the series is floating away from any fixed mean and is nonstationary.
    - Stata does DF and ADF tests with the dfuller command, using the lags(#) option to add lagged differences.
    - An alternative test is to use Newey-West HAC robust standard errors in the original DF equation rather than adding lagged differences to eliminate serial correlation of u. This is the Phillips-Peron test: pperron in Stata.
- **Nonstationary vs. borderline stationary series**
    - $Y_t = Y_{t-1} + u_t$ is a nonstationary random walk
    - $Y_t = 0.999Y_{t-1} + u_t$ is a stationary AR(1) process

- They are not very different when $T < \infty$.
- Show graphs of three series
- Can we hope that our ADF test will discriminate between nonstationary and borderline stationary series? Probably not without longer samples than we have.
- Since the null hypothesis is nonstationarity, a low-power test will usually fail to reject nonstationarity and we will tend to conclude that some highly persistent but stationary series are nonstationary.
- Note: The ADF test does not prove nonstationarity; it fails to prove stationarity.

Nonstationarity due to breaks

- Breaks in a series/model are the time-series equivalent of a violation of Assumption #0.
  - The relationship between the variables (including lags) changes either abruptly or gradually over time.
- With a known potential break point (such as a change in policy regime or a large shock that could change the structure of the model):
  - Can use Chow test based on dummy variables to test for stability across the break point.
  - Interact all variables of the model with a sample dummy that is zero before the break and one after. Test all interaction terms (including the dummy itself) = 0 with Chow *F* statistic.
- If breakpoint is unknown:
  - Quandt likelihood ratio test finds the largest Chow-test *F* statistic, excluding (trimming) the first and last 15% (or more or less) of the sample as potential breakpoints to make sure that each sub-sample is large enough to provide reliable estimates.
  - QLR test statistic does not have an *F* distribution because it is the max of many *F* statistics.