Section 8 Regression with Stationary Time Series

How time-series regression differs from cross-section

- Natural ordering of observations contains information
  - Random reshuffling of observations would obscure dynamic economic relationship, but leave traditional regression unchanged
  - How can we incorporate this dynamic information into our regression model?
- We usually think of the data as being drawn from a potentially infinite **data-generating process** rather than from a finite population of observations.
- Variables are often call "time series" or just "series" rather than variables
  - Index observations by time period t
  - Number of observations = T
- Dynamic relationship means that not all of the effects of x_t occur in period t.
  - A change in x_t is likely to affect y_{t+1}, y_{t+2}, etc.
  - By the same logic, y_t depends not only on x_t but also on x_{t-1}, x_{t-2}, etc.
  - We model these dynamic relationships with **distributed lag** models, in which $y_t = f\left(x_t, x_{t-1}, x_{t-2}, \ldots\right)$.
- We will need to focus on the dynamic elements of both the deterministic relationship between the variables and the stochastic relationship (error term)
- The dynamic ordering of observations means that the error terms are usually **serially correlated** (or autocorrelated over time)
  - Shocks to the regression are unlikely to completely disappear before the following period
    - Exception: stock market returns, where investors should respond to any shock and make sure that next period's return is not predictable
  - Two observations are likely to be more highly correlated if they are close to the same time than if they are more widely separated.
  - Covariance matrix of error term will have non-zero off-diagonal elements, with elements lying closest to the diagonal likely being substantially positive and decreasing as one moves away from the diagonal.
- **Nonstationary** time series create problems for econometrics.
  - We will study implications of and methods for dealing with nonstationarity in Section 12.
  - Example will illustrate nature of problem ("spurious regressions")
    - Regression of AL attendance on Botswana real GDP
    - Correlation = 0.9656
    - R^2 = 0.9323

- Coefficient has t of 24.90.
- Good regression?

```
      Source |       SS       df       MS              Number of obs =      47
-------------+------------------------------           F(  1,    45) =  619.89
       Model |  3.4342e+15     1  3.4342e+15           Prob > F      =  0.0000
    Residual |  2.4930e+14    45  5.5400e+12           R-squared     =  0.9323
-------------+------------------------------           Adj R-squared =  0.9308
       Total |  3.6835e+15    46  8.0077e+13           Root MSE      =  2.4e+06

------------------------------------------------------------------------------
    ALAttend |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rgdpl2 |    3285.11   131.9447    24.90   0.000      3019.36     3550.86
       _cons |    8029710   640681.3    12.53   0.000      6739311     9320108
------------------------------------------------------------------------------
```

- Correlation is spurious because both series are trending upward, so most of each series' deviation from mean is due to separate trends.
- Much of the last 20 years in econometrics has been devoted to understanding how to deal with nonstationary time series.
- We will study this intensively in a few weeks.
- Nonstationarity forces us to remove the common trend (often by differencing) before interpreting the correlation or regression

Lag operators 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 autocorrelation

- Autocovariance of order *s* is $\text{cov}(x_t, x_{t-s})$
  - 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* (which we write as ρ_s) is the correlation coefficient between x_t and x_{t-s}.
  - $\rho_k = \dfrac{\text{cov}(x_t, x_{t-k})}{\text{var}(x_t)}$
  - We estimate with $r_k = \dfrac{\frac{1}{T-k}\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-s} - \bar{x})}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})}$.
    - We sometimes subtract one from both denominators, or sometimes ignore the different fractions in front of the summations since their ratio goes to 1 as *T* goes to ∞.
- ρ_k as a function of *k* is called the **autocorrelation function** of the series and its plot is often called a **correlogram**.

Some simple 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 $y_t = v_t$, where v_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.
    - $y_t = E\left[y_t \mid y_{t-1}, y_{t-2}, \ldots\right] + \varepsilon_t$
    - 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 v 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**: $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$, where ε is white noise.
  - In this case, our one-period-ahead forecast is $E\left[y_t \mid y_{t-1}\right] = \beta_0 + \beta_1 y_{t-1}$ and the forecast error is ε_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 ε_1 on the series y from time one on, assuming (for simplicity) that $y_0 = 0$ and all subsequent ε values are also zero.
    - $y_1 = \beta_1(0) + \varepsilon_1 = \varepsilon_1$
    - $y_2 = \beta_1 y_1 + \varepsilon_2 = \beta_1 \varepsilon_1$
    - $y_3 = \beta_1 y_2 + \varepsilon_3 = \beta_1^2 \varepsilon_1$
    - $y_s = \beta_1^{s-1} \varepsilon_1$.
    - 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 $y_t = y_{t-1} + \varepsilon_t$ or $\Delta y_t = \varepsilon_t$. 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.
- 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.

ARMA processes and lag polynomials

- Higher-order AR processes
  - AR(p) process:
$$\begin{aligned} y_t &= \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \\ &= \alpha + \sum_{i=1}^{p} \phi_i L^i y_t + \varepsilon_t \\ &= \alpha + \left(\sum_{i=1}^{p} \phi_i L^i \right) y_t + \varepsilon_t \\ &= \alpha + \phi(L) y_t + \varepsilon_t \end{aligned}$$
  - Is this stationary?
    - Depends on the ϕ parameters

Stationarity

- **Formal definition**
  - $$\begin{aligned} E(y_t) &= \mu \\ \text{var}(y_t) &= \sigma^2 \\ \text{cov}(y_t, y_{t-s}) &= \gamma_s \end{aligned}$$
- The key point of this definition is that all of the first and second moments of *y* are the same for all *t*
- Stationarity implies **mean reversion**: that the variable reverts toward a fixed mean after any shock

Kinds of nonstationarity

- Like most rules, nonstationarity can be violated in several ways
- **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.

- **Deterministic trends** are constant increases in the mean of the series over time, though the variable may fluctuate above or below its trend line randomly.
  - $y_t = \alpha + \lambda t + v_t$
  - v is stationary disturbance term
  - If the constant rate of change is in percentage terms, then we could model ln$y$ as being linearly related to time.
  - This violates the stationarity assumptions because $E(y_t) = \alpha + \lambda t,$ which is not independent of t
- **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: $y_t = y_{t-1} + \varepsilon_t$ where e is white noise.
  - Random walk is limiting case of stationary AR(1) process $y_t = \rho y_{t-1} + \varepsilon_t$ as $\rho \to 1$
  - Solving recursively (conditional on given initial value y_0),
    - $y_1 = y_0 + \varepsilon_1,$
    - $y_2 = y_1 + \varepsilon_2 = y_0 + \varepsilon_1 + \varepsilon_2,$
    - $y_t = y_0 + \varepsilon_1 + \ldots + \varepsilon_t = y_0 + \sum_{s=1}^{t} \varepsilon_s = y_0 + \sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}.$

- This violates stationarity assumptions because $\text{var}(y_t \mid y_0) = \text{var}\left(\sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}\right) = t\sigma_\varepsilon^2$, which depends on *t*, and unconditional variance is infinite: $\text{var}(y_t) = \text{var}\left(\sum_{\tau=0}^{\infty} \varepsilon_\tau\right) = \sum_{\tau=0}^{\infty} \sigma_\varepsilon^2 = \infty.$
- Note comparison with stationary AR(1):
  - $y_t = y_0 + \sum_{s=1}^{t} \rho^t \varepsilon_t,$
  - $\text{var}(y_t) = \text{var}\left(\sum_{\tau=0}^{\infty} \rho^\tau \varepsilon_{t-\tau}\right) = \sigma_\varepsilon^2 \sum_{\tau=0}^{\infty} (\rho^2)^\tau = \frac{\sigma_\varepsilon^2}{1-\rho^2} < \infty$

- Random walk with drift allows for non-zero average change: $y_t = \alpha + y_{t-1} + \varepsilon_t$
  - This also violates the constant-mean assumption:
  - $y_1 = \alpha + y_0 + \varepsilon_1,$
    $y_2 = \alpha + y_1 + \varepsilon_2 = y_0 + 2\alpha + \varepsilon_1 + \varepsilon_2,$
    $y_t = y_0 + t\alpha + \sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}.$
  - $E(y_t \mid y_0) = y_0 + t\alpha,$
    $\text{var}(y_t \mid y_0) = t\sigma_e^2.$
  - Both conditional mean and conditional variance depend on *t*
  - Both unconditional mean and unconditional variances are infinite
  - For AR(1) with non-zero mean:
    - $y_t = y_0 + \sum_{\tau=0}^{t-1} \rho^\tau (\alpha + \varepsilon_{t-\tau}) = \sum_{\tau=0}^{\infty} \rho^\tau (\alpha + \varepsilon_{t-\tau}).$
    - $E(y_t) = \alpha \sum_{\tau=0}^{\infty} \rho^\tau = \frac{\alpha}{1-\rho},$
    - $\text{var}(y_t) = \sigma_v^2 \sum_{\tau=0}^{\infty} (\rho^2)^\tau = \frac{\sigma_e^2}{1-\rho^2}.$
    - Both unconditional mean and variance are finite and independent of *t*.

- **Difference between deterministic and stochastic trend**
  - Consider large negative shock *e* in period *t*
    - In deterministic trend, the trend line remains unchanged.
      - Because *e* 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

Unit roots and integration in AR models

- Note that the random-walk model is just the AR(1) model with ρ = 1.
- In general, the stationarity of a variable depends on the parameters of its AR representation:
  - AR(*p*) is $y_t = \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t$, or $\phi(L) y_t = \varepsilon_t$.
    - (Can generalize to allow *e* to be any stationary process, not just white noise.)
  - The stationarity of *y* depends on the roots (solutions) to the equation $\theta(L) = 0$.
    - $\phi(L)$ is a *p*-order polynomial that has *p* roots, which may be real or imaginary-complex numbers.
    - AR(1) is first-order, so there is one root: $\phi(L) = 1 - \phi_1 L$,

      $\phi(L) = 0 \Leftrightarrow 1 - \phi_1 L = 0 \Leftrightarrow 1 = \phi_1 L \Leftrightarrow L = \frac{1}{\phi_1}$, so $1/\theta_1$ is the root of the AR(1) polynomial. (Or $1/\rho$ in the simpler AR(1) notation we used above.)
  - If the *p* roots of $\phi(L) = 0$ are all greater than one in absolute value (formally, because the roots of a polynomial can be complex, we have to say "outside the unit circle of the complex plane"), then *y* is stationary.
    - By our root criterion for stationarity, the AR(1) is stationary if $\left|\frac{1}{\phi_1}\right| > 1$, or $|\phi_1| < 1$.
    - This corresponds to the assumption we presented earlier that $|\rho| < 1$.
- If one or more roots of $\phi(L) = 0$ are equal to one and the others are greater than one, then we say that the variable has a **unit root**.
  - We call these variables **integrated** variables for reasons we will clarify soon.
  - Integrated variables are just barely nonstationary and have very interesting properties.
  - (Variables with roots less than one in absolute value simply explode.)
  - The random-walk is the simplest example of an integrated process:
    - $$y_t = y_{t-1} + \varepsilon_t$$
      $$y_t - y_{t-1} = \varepsilon_t$$
      $$(1 - L) y_t = \phi(L) y_t = \varepsilon_t$$
    - The root of 1 – *L* = 0 is *L* = 1, which is a unit root.
- **Integrated processes**

- Consider the general AR(p) process $y_t = \alpha + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t$, which we write in lag-operator notation as $\phi(L) y_t = \alpha + \varepsilon_t$.
- We noted above that the stationarity properties of y are determined by whether the roots of $\phi(L) = 0$ are outside the unit circle (stationary) or on it (nonstationary).
    - $\phi(L)$ is an order-p polynomial in the lag operator $1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$.
    - We can factor $\phi(L)$ as $1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p = (1 - f_1 L)(1 - f_2 L)\cdots(1 - f_p L)$, where $r_1 = \frac{1}{f_1}, r_2 = \frac{1}{f_2}, \ldots, r_p = \frac{1}{f_p}$ are the roots of $\phi(L)$.
    - We rule out allowing any of the roots to be inside the unit circle because that would imply explosive behavior of y, so we assume $|f_j| \leq 1$.
    - Suppose that there are $k \leq p$ roots that are equal to one (k unit roots) and $p - k$ roots that are greater than one (outside the unit circle in the complex plane). We can then write $\phi(L) = (1 - f_1 L)\cdots(1 - f_{p-k} L)(1 - L)^k$, where we number the roots so that the first $p - k$ are greater than one.
    - Let $\gamma(L) = \frac{\phi(L)}{(1 - L)^k} = (1 - f_1 L)\cdots(1 - f_{p-k} L)$. Then $\phi(L) y_t = \gamma(L)(1 - L)^k y_t = \gamma(L)(\Delta^k y_t) = \alpha + \varepsilon_t$.
    - Because $\gamma(L)$ has all of its roots outside the unit circle, the series $\Delta^k y_t$ is stationary.
    - We introduce the terminology "**integrated of order *k***" (or $I(k)$) to describe a series that has ***k* unit roots** and that is **stationary after being differenced *k* times**.
        - The term "integrated" should be thought of as the inverse of "differenced" in much that same way that integrals are the inverse of differentiation.
            - The "integration" operator $(1 - L)^{-1}$ accumulates a series in the same way that the difference operator $1 - L$ turns the series into changes.
            - Integrating the first differences of a series reconstructs the original series: $(1 - L)^{-1} \Delta y_t = (1 - L)^{-1}(1 - L) y_t = y_t$
        - If y is stationary, it is $I(0)$.

- If the first difference of y is stationary but y is not, then y is $I(1)$. Random walks are $I(1)$.
- If the first difference is nonstationary but the second difference is stationary, then y is $I(2)$, etc.
- In practice, most economic time series are $I(0)$, $I(1)$, or occasionally $I(2)$.

- **Impacts of integrated variables** in a regression
  - If y has a unit root (is integrated of order > 0), then the OLS estimates of coefficients of an autoregressive process will be biased downward in small samples.
  - Can't test $\phi_1 = 0$ in an autoregression such as $y_t = \alpha + \phi_1 y_{t-1} + \varepsilon_t$ with usual tests
  - 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.
    - This is exactly the kind of problem illustrated by the baseball attendance/Botswana GDP example
    - Show the Granger-Newbold results/tables

Assumptions of time-series regression

- 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. (GHL's Section 9.3)
- Can estimate time-series regressions by OLS as long as y and x are stationary and x is exogenous.
  - **Exogeneity**: $E\left(u_t \mid x_t, x_{t-1}, \ldots\right) = 0.$
  - **Strict exogeneity**: $E\left(u_t \mid \ldots, x_{t+2}, x_{t+1}, x_t, x_{t-1}, x_{t-2}, \ldots\right) = 0.$
- Assumptions of time-series regression:
  - **TS1**: linear model
  - **TS2:** no perfect collinearity
  - **TS3:** $E\left(u_t \mid x\right) = 0$
  - **TS4:** $\operatorname{var}\left(u_t \mid x\right) = \sigma^2$
  - **TS5:** $\operatorname{cov}\left(u_t, u_s\right) = 0,\ t \neq s$
  - **TS6:** $u_t \mid x \sim N\left(0, \sigma^2\right)$
- However, nearly all time-series regressions are prone to having serially correlated error terms, which violates TS5.
  - 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

Detecting autocorrelation

- We can test the autocorrelations of a series to see if they are zero.
    - Asymptotically, $\sqrt{T} r_k \sim N\left(\rho_k, 1\right)$, so we can compute this as a test statistic and test against the null hypothesis $\rho_k = 0$.
- **Breusch-Godfrey Lagrange multiplier test** for autocorrelation:
    - Regress *y* (or residuals) on *x* and lagged residuals (first-order, or more)
    - Use *F* test of residual coefficient(s) in *y* regression or TR^2 in residual regression as χ^2
- **Box-Ljung *Q* test** for null hypothesis that the first *k* autocorrelations are zero:

  $Q_k = T\left(T+2\right)\sum_{j=1}^{k}\frac{r_j^2}{T-j}$ is asymptotically χ_k^2.
- **Durbin-Watson test** used to be the standard test for first-order autocorrelation, but was difficult because critical values depend on *x*. Not used much anymore.

Estimation with autocorrelated errors

- OLS with autocorrelated errors
    - Assumption TSMR4 is violated, which leads to inefficient estimators and biased standard errors just like in case of heteroskedasticity
    - **Important special case:** We will see that a common distributed lag model puts y_{t-1} on the right-hand side as a regressor. This causes special problems when there is serial correlation because
        - e_{t-1} is part of y_{t-1}
        - e_{t-1} is correlated with e_t

        - Therefore e_t is correlated with one of the regressors, which leads to bias and inconsistency in the coefficient estimators.
        - If we can transform the model into one that has no autocorrelation (for example, v_t if error term is $e_t = \rho e_{t-1} + v_t$), then we can get consistent OLS estimators as long as all the x variables are exogenous (but not necessarily strictly exogenous) with respect to v.

- **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

$$b_2 = \beta_2 + \frac{\frac{1}{T}\sum_{i=1}^{T}(x_t - \bar{x})e_t}{\frac{1}{T}\sum_{i=1}^{T}(x_t - \bar{x})^2}.$$

    - In this formula, $\operatorname{plim}\bar{x} = \mu_X$, $\operatorname{plim}\left(\frac{1}{T}\sum_{i=1}^{T}(x_t - \bar{x})^2\right) = \sigma_X^2$.
    - So $\operatorname{plim}(b_2 - \beta_2) = \frac{\operatorname{plim}\left(\frac{1}{T}\sum_{i=1}^{T}(x_t - \mu_X)e_t\right)}{\sigma_X^2} = \frac{\operatorname{plim}(\bar{u})}{\sigma_X^2}$, where $\bar{u} = \frac{1}{T}\sum_{t=1}^{T}u_t$ and $u_t \equiv (X_t - \mu_X)e_t$.
    - And in large samples, $\operatorname{var}(b_2) = \operatorname{var}\left(\frac{\bar{u}}{\sigma_X^2}\right) = \frac{\operatorname{var}(\bar{u})}{\sigma_X^4}$.
        - Under IID assumption, $\operatorname{var}(\bar{u}) = \frac{1}{T}\operatorname{var}(u_t) = \frac{\sigma_u^2}{T}$, 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 u_t terms:

$$
\begin{aligned}
\operatorname{var}(\bar{u}) &= \operatorname{var}\left(\frac{u_1 + u_2 + \ldots + u_T}{T}\right) \\
&= \frac{1}{T^2}\left[\sum_{i=1}^{T}\sum_{j=1}^{T} E(u_i u_j)\right] \\
&= \frac{1}{T^2}\sum_{i=1}^{T}\left(\operatorname{var}(u_i) + \sum_{j \neq i}\operatorname{cov}(u_i, u_j)\right) \\
&= \frac{1}{T^2}\left[T\operatorname{var}(u_t) + 2(T-1)\operatorname{cov}(u_t, u_{t-1}) + 2(T-2)\operatorname{cov}(u_t, u_{t-2}) + \ldots + 2\operatorname{cov}\left(u_t, u_{t-(T-1)}\right)\right] \\
&= \frac{\sigma_u^2}{T} f_T,
\end{aligned}
$$

where

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

- Thus, $\operatorname{var}(b_2) = \left[\frac{1}{T}\frac{\sigma_u^2}{\sigma_X^4}\right] f_T$, 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 u 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 (u_t, u_{t-1}) in the sample.
  - For ρ_{T-1}, there is only one pair $(u_t, u_{t-(T-1)})$—namely (u_T, u_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 $\hat{f}_T = 1 + 2\sum_{j=1}^{m-1}\left(\frac{m-j}{m}\right)r_j$.
  - m must be large enough to provide a reasonable correction but small enough relative to T to allow the r 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)

- **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 e_t follows and AR(1) process:

$$y_t = \beta_0 + \beta_1 x_t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t,$$

where ε is a white-noise error term and $-1 < \rho < 1$.

  - In practice, $\rho > 0$ nearly always

- GLS transforms the model into one with an error term that is not serially correlated.
- Let

$$\tilde{y}_t = \begin{cases} y_t\sqrt{(1-\rho^2)}, & t=1, \\ y_t - \rho y_{t-1}, & t = 2, 3, \ldots, T, \end{cases} \qquad \tilde{x}_t = \begin{cases} x_t\sqrt{(1-\rho^2)}, & t=1, \\ x_t - \rho x_{t-1}, & t = 2, 3, \ldots, T, \end{cases}$$

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

- Then $\tilde{y}_t = (1-\rho)\beta_1 + \beta_2 \tilde{x}_t + \tilde{u}_t$.
  - 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 ρ?
  - Need to estimate ρ to calculate feasible GLS estimator.
  - Traditional estimator for ρ is $\hat{\rho} = \text{corr}(\hat{u}_t, \hat{u}_{t-1})$ using OLS residuals.
  - This estimation can be iterated to get a new estimate of ρ based on the GLS estimator and then re-do the transformation: repeat until converged.
- Two-step estimator using FGLS based on $\hat{\rho}$ is called the **Prais-Winsten** estimator (or **Cochrane-Orcutt** when first observation is dropped).
- Problems: ρ 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 ρ 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 ρ as the coefficient on y_{t-1} in an OLS model, with or without the restriction implied in HGL's equation (9.44).

Distributed-lag models

- Modeling the **deterministic part** of a dynamic relationship between two variables
- In general, the distributed-lag model has the form $y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + u_t$. 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.
    - **Multipliers**
        - $\frac{\partial E(y_t)}{\partial x_{t-s}} = \beta_s$ is the *s*-period delay multiplier, telling how much a one-time, temporary shock to *x* would be affecting *y s* periods later
        - β_0 is the "impact multiplier"
        - $\sum_{r=0}^{s} \beta_r$ is the cumulative or "interim" multiplier after *s* periods. It measures the cumulative effect of a permanent change in *x* on *y s* periods later.
        - $\sum_{s=1}^{\infty} \beta_s$ is the "total multiplier" measuring the final cumulative effect of a permanent change in *x* on *y*.
    - We can easily have additional regressors with either the same or different lag structures.
- **Dynamically complete models**
    - The dynamics of the error term and the dynamics of the lag structure are intimately related
        - In model with no lags and AR(1) error,

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

          $u_t = \rho u_{t-1} + e_t,$

          but $u_{t-1} = y_{t-1} - \beta_0 - \beta_1 x_{t-1},$

          so $y_t = \beta_0 + \beta_1 x_t + \rho(y_{t-1} - \beta_0 - \beta_1 x_{t-1}) + e_t$ or

          $y_t = (1-\rho)\beta_0 + \rho y_{t-1} + \beta_1 (x_t - \rho x_{t-1}) + e_t.$
        - This is, or course, just the GLS transformation with y_{t-1} on the right-hand side, but we could estimate it as $y_t = \gamma_0 + \gamma_1 y_{t-1} + \gamma_2 x_t + \gamma_3 x_{t-1} + e_t$ with a white-noise error.
            - Note that there is one lost degree of freedom because there is an unimposed linear restriction among the γ coefficients, but this is not a big problem unless there are lots of *x* variables
        - This shows that putting lags of *y* and/or *x* on the RHS can filter the model in a way that eliminates serial correlation of the error

    - This is the most common was of dealing with serial correlation now rather than using GLS
    - If a model's dynamic structure (in the deterministic part) is sufficient to reduce the error to white noise, we say that the model is **dynamically complete**.

- **Finite distributed lags**
  - $y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_q x_{t-q} + e_t = \alpha + \beta(L) x_t + e_t$
  - This is finite distributed-lag model of order *q*
  - Under assumptions TS.1–6, the model can be estimated by OLS. If there is serial correlation, then the appropriate correction must be made to standard errors or a GLS model should be used, but if *q* is large enough we should be able to make the model dynamically complete
  - Problems with finite DL model
    - If *x* is strongly autocorrelated, then collinearity will be a problem and it will be difficult to estimate individual lag weights accurately
    - Difficult to know appropriate lag length (can use AIC or SC)
    - If lags are long and sample is short, will lose lots of observations.
- **Koyck lag: *y* is AR(1) with regressors**
  - $y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + u_t$
  - $\frac{\partial y_t}{\partial x_t} = \delta_0$

    $\frac{\partial y_{t+1}}{\partial x_t} = \theta_1 \delta_0$

    $\frac{\partial y_{t+2}}{\partial x_t} = \theta_1^2 \delta_0$

    $\frac{\partial y_{t+s}}{\partial x_t} = \theta_1^s \delta_0$

    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 $\sum_{k=0}^{s} \frac{\partial y_{t+k}}{\partial x_t} = \delta_0 \sum_{k=0}^{s} \theta_1^k$.
  - Long-run effect of a permanent change is $\delta_0 \sum_{k=0}^{\infty} \theta_1^k = \frac{\delta_0}{1-\theta_1}$.
  - 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: $y_t = \delta + \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p} + \delta_0 x_t + u_t$.
  - This can be written $\theta(L) y_t = \delta + \delta_0 x_t + u_t$.
  - Same general principles apply:
    - Worry about stationarity of lag structure: roots of $\theta(L)$
    - If u is serially correlated, OLS will be biased and inconsistent
    - Dynamic multipliers are determined by coefficients of infinite lag polynomial $[\theta(L)]^{-1}$
    - If more than on x, all have same lag structure
  - How to determine length of lag p?
    - Can keep adding lags as long as θ_p is statistically significant
    - Add lags until serial correlation of u is eliminated (dynamic completeness)
    - Can choose to max the Akaike information criterion (AIC) or Bayesian (Schwartz) information criterion (SC).
    - 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 SC.
      - AIC came first, so is still used a lot
      - SC is asymptotically unbiased
    - Stata calculates info criteria by estat ic (after regression)

- **ADL(*p*, *q*) Model: "Rational" lag**
  - We can also add lags to the x variable(s)
  - $y_t = \delta_0 + \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \ldots + \delta_q x_{t-q} + u_t$
    - Can add more x variables with varying lag lengths

  $\theta(L) y_t = \delta_0 + \delta(L) x_t + u_t$,
  - $$y_t = \frac{\delta_0}{\theta(L)} + \frac{\delta(L)}{\theta(L)} x_t + \frac{u_t}{\theta(L)}.$$
    - Multipliers are the (infinite) coefficients on the lag polynomial $\frac{\delta(L)}{\theta(L)}$

- Stationarity depends only on $\theta(L)$, 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 $s \to \infty$
  - No perfect multicollinearity
- These are general TS assumptions that apply to most time-series models.

Forecasting with time-series models

- **With AR(p) model**
  - $y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_p y_{t-p} + u_t$, with v assumed to be serially uncorrelated.
  - We have observations for $t = 1, 2, \ldots, T$
  - Forecast for $T + 1$: $\hat{y}_{T+1} = \hat{\delta} + \hat{\theta}_1 y_T + \hat{\theta}_2 y_{T-1} + \ldots + \hat{\theta}_p y_{T-p+1}$ with expected value of u at zero because of no serial correlation
    - If error term were autoregressive, then conditional expectation of u_{T+1} would be ρu_T, so would include residual of observation T
  - Forecast error is $e_1 = y_{T+1} - \hat{y}_{T+1} = \left(\delta - \hat{\delta}\right) + \sum_{s=1}^{p}\left(\theta_s - \hat{\theta}_s\right) y_{T+1-s} + u_{T+1}$
    - Some textbooks assume that $\operatorname{var}\left(u_{T+1}\right) >> \operatorname{var}\left[\left(\delta - \hat{\delta}\right) + \sum_{s=1}^{p}\left(\theta_s - \hat{\theta}_s\right) y_{T+1-s}\right]$, so they ignore the latter.
    - I'm not willing to ignore this.
    - I will write $e_{b,k} = \left(\delta - \hat{\delta}\right) + \sum_{s=1}^{p}\left(\theta_s - \hat{\theta}_s\right) y_{T+k-s}$ and $\operatorname{var}\left(e_{b,k}\right)$ to be that component of the k-period-ahead forecast and keep it in the equation
    - $u_1 = e_{b,1} + v_{T+1},$
    - $\operatorname{var}\left(u_1\right) = \operatorname{var}\left(e_{b,1}\right) + \sigma_v^2$
  - What about forecast for $T + 2$?
    - y_{T+1} appears on the right-hand side of the T+ 2 equation, so we substitute our one-period-ahead forecast of it: $\hat{y}_{T+2} = \hat{\delta} + \hat{\theta}_1 \hat{y}_{T+1} + \hat{\theta}_2 y_T + \ldots + \hat{\theta}_p y_{T+2-p}$

- Forecast error is

$$u_2 = y_{T+2} - \hat{y}_{T+2} = \left(\delta - \hat{\delta}\right) + \theta_1\left(\hat{y}_{T+1} - y_{T+1}\right) + \sum_{s=1}^{p}\left(\theta_s - \hat{\theta}_s\right) y_{T+2-s} + v_{T+2}$$
$$= e_{b,2} + \theta_1 e_{b,1} + v_{T+2} + \theta_1 v_{T+1}.$$

- Variance of forecast error is

$$\operatorname{var}\left(u_2\right) = \operatorname{var}\left(e_{b,2}\right) + \theta_1^2 \operatorname{var}\left(e_{b,1}\right) + \left(1 + \theta_1^2\right)\sigma_v^2.$$

- Similarly,

$$\operatorname{var}\left(u_3\right) = \operatorname{var}\left(e_{b,3}\right) + \theta_1^2 \operatorname{var}\left(e_{b,2}\right) + \left(\theta_1^2 + \theta_2\right)^2 \operatorname{var}\left(e_{b,1}\right) + \left(1 + \theta_1^2 + \left(\theta_2 + \theta_1^2\right)^2\right)\sigma_v^2$$