Section 9  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
    \[
    y_t = f(x_t, x_{t-1}, x_{t-2}, \ldots) .
    \]
- 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?

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 47</th>
<th>F(  1,    45) = 619.89</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3.4342e+15</td>
<td>1</td>
<td>3.4342e+15</td>
<td>Prob &gt; F = 0.0000</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>2.4930e+14</td>
<td>45</td>
<td>5.5400e+12</td>
<td>R-squared = 0.9323</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3.6835e+15</td>
<td>46</td>
<td>8.0077e+13</td>
<td>Adj R-squared = 0.9308</td>
<td>Root MSE = 2.4e+06</td>
</tr>
</tbody>
</table>

| 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 \( L(x) \) or \( Lx_t \equiv x_{t-1} \) to represent lags.
  - We often use higher-order lags: \( L^sx_t \).
- 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:
    \( \Delta^2 x_t = \Delta(\Delta x_t) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = x_t - 2x_{t-1} + x_{t-2} \)
    \( = (1 - L)^2x_t = (1 - 2L + L^2)x_t \)
  - \( \Delta^s x_t = (1 - L)^sx_t \)
- 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
  o 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
  o 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 \( \text{cov}(x_t, x_{t-s}) \)
  o We generally assume that the autocovariance depends only on s, not on t.
  o This is analogous to our Assumption #0: that all observations follow the same model (or were generated by the same data-generating process)
  o This is one element of a time series being stationary
• Autocorrelation of order s (\( \rho_s \)) is the correlation coefficient between \( x_t \) and \( x_{t-s} \).
  o \( \rho_s = \frac{\text{cov}(x_t, x_{t-k})}{\text{var}(x_t)} \)
    \[ \rho_k = \frac{1}{T-k} \sum_{t=k+1}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x}) \]
    • We estimate with \( r_k = \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x}) \)
  ▪ 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 \( \infty \).
• \( \rho_k \) 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 $\rho_0$, 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 $y$.
    - $y_t = E[y_t | y_{t-1}, y_{t-2}, \ldots] + v_t$
    - 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 $v_t$ 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**: $y_t = \beta_0 + \beta_1 y_{t-1} + v_t$, where $e$ is white noise.
  - In this case, our one-period-ahead forecast is $E[y_t | y_{t-1}] = \beta_0 + \beta_1 y_{t-1}$ and the forecast error is $v_t$.
  - For simplicity, suppose that we have removed the mean from $y$ so that $\beta_0 = 0$.
    - Consider the effect of a one-time shock $v_1$ on the series $y$ from time one on, assuming (for simplicity) that $y_0 = 0$ and all subsequent $v$ values are also zero.
      - $y_1 = \beta_1 (0) + v_1 = v_1$
      - $y_2 = \beta_1 y_1 + v_2 = \beta_1 v_1$
      - $y_3 = \beta_1 y_2 + v_3 = \beta_1^2 v_1$
      - $y_s = \beta_1^{s-1} v_1$.
    - 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 $y_t = y_{t-1} + v_t$ or $\Delta y_t = v_t$.
        - The first difference of a random walk is stationary and is white noise.
If \( y \) follows a stationary AR(1) process, then \( \rho_1 = \beta_1, \rho_2 = \beta_1^2, \ldots, \rho_s = \beta_1^s \).

- 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 \( \beta_1 \).
- 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.

**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**: \( E(e_t | x_t, x_{t-1}, \ldots) = 0 \).
  - **Strict exogeneity**: \( E(e_t | \ldots, x_{t+2}, x_{t+1}, x_t, x_{t-1}, x_{t-2}, \ldots) = 0 \).

- Assumptions of time-series regression:
  - **TSMR2**: \( y \) and \( x \) are stationary and \( x \) is strictly exogenous
  - **TSMR3**: \( E(e_t) = 0 \)
  - **TSMR4**: \( \text{var}(e_t) = \sigma^2 \)
  - **TSMR5**: \( \text{cov}(e_t, e_s) = 0, t \neq s \)
  - **TSMR6**: \( e_t \sim N(0, \sigma^2) \)

- However, nearly all time-series regressions are prone to having serially correlated error terms, which violates TSMR5.
  - 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, \( \sqrt{T} r_k \sim N(\rho_k, 1) \), 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 \( \chi^2 \)
  - Box-Ljung \( Q \) test for null hypothesis that the first \( k \) autocorrelations are zero:
    \[
    Q_k = T (T + 2) \sum_{j=1}^{k} \frac{f_j^2}{T - j}
    \]
    is asymptotically \( \chi^2_k \).
  - 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, \( v \) if error term is \( e_t = pe_{t-1} + v_t \)), 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

\[ b_2 = \beta_2 + \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x}) e_t \]

In this formula, \( \text{plim } \bar{x} = \mu_x \), \( \text{plim } \left( \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x})^2 \right) = \sigma_x^2 \).

So \( \text{plim } (b_2 - \beta_2) = \frac{\text{plim } (\frac{1}{T} \sum_{t=1}^{T} (x_t - \mu_x) e_t)}{\sigma_x^2} = \frac{\text{plim } (\bar{u})}{\sigma_x^2} \), 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, \( \text{var } (b_2) = \text{var } \left( \frac{\bar{u}}{\sigma_x^2} \right) = \frac{\text{var } (\bar{u})}{\sigma_x^4}. \)

- Under IID assumption, \( \text{var } (\bar{u}) = \frac{1}{T} \text{var } (u_t) = \frac{\sigma^2}{T} \), 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:

\[ \text{var } (\bar{u}) = \text{var } \left( \frac{u_1 + u_2 + \ldots + u_T}{T} \right) \]

\[ = \frac{1}{T^2} \left[ \sum_{t=1}^{T} \sum_{j=1}^{T} E(u_t u_j) \right] \]

\[ = \frac{1}{T^2} \sum_{t=1}^{T} \left( \text{var } (u_t) + \sum_{j=t}^{T} \text{cov } (u_t, u_j) \right) \]

\[ = \frac{1}{T^2} \left[ T \text{var } (u_t) + 2(T-1) \text{cov } (u_t, u_{t-1}) + 2(T-2) \text{cov } (u_t, u_{t-2}) + \ldots + 2 \text{cov } (u_t, u_{(T-1)}) \right] \]

\[ = \frac{\sigma_x^2}{T} f_T, \]

where

\[ f_T \equiv 1 + 2 \sum_{j=1}^{T-1} \left( \frac{T-j}{T} \right) \text{corr } (u_t, u_{t-j}) \]

\[ = 1 + 2 \sum_{j=1}^{T-1} \left( \frac{T-j}{T} \right) \rho_j. \]
Thus, \( \text{var}(b_2) = \left[ \frac{1}{T} \frac{\sigma_u^2}{\sigma_x^2} \right] f_T \), 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 \( \rho_1 \), we have lots of information because there are \( T-1 \) pairs of values for \((u_t, u_{t-1})\) in the sample.
- For \( \rho_{T-1} \), 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 \( m = 0.75T^{3/2} \) 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:
    \[
    y_t = \beta_0 + \beta_1 x_t + e_t,
    \]
    \[
    e_t = \rho e_{t-1} + \nu_t,
    \]
  where \( \nu_t \) 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
  \[
  \tilde{y}_t = \begin{cases} y_t \sqrt{1-\rho^2}, & t = 1, \\ y_t - \rho y_{t-1}, & t = 2, 3, \ldots, T, \end{cases}
  \]
  \[
  \tilde{x}_t = \begin{cases} x_t \sqrt{1-\rho^2}, & t = 1, \\ x_t - \rho x_{t-1}, & t = 2, 3, \ldots, T, \end{cases}
  \]
  \[
  \tilde{e}_t = \begin{cases} e_t \sqrt{1-\rho^2}, & t = 1, \\ e_t - \rho e_{t-1}, & t = 2, 3, \ldots, T. \end{cases}
  \]

Then
  \[
  \tilde{y}_t = (1-\rho)\hat{\beta}_1 + \beta_2 \tilde{x}_t + \tilde{e}_t.
  \]
The error term in this regression is equal to \( v_t \) for observations 2 through \( T \) and is a multiple of \( e_1 \) for the first observation.

By assumption, \( v \) is white noise and values of \( v \) in periods after 1 are uncorrelated with \( e_1 \), 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 \( \rho \)?

Need to estimate \( \rho \) to calculate feasible GLS estimator.

Traditional estimator for \( \rho \) is \( \hat{\rho} = \text{corr}(\hat{e}_t, \hat{e}_{t-1}) \) using OLS residuals.

This estimation can be iterated to get a new estimate of \( \rho \) 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: \( \rho \) is not estimated consistently if \( e_t \) 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 \( \rho \) and \( \beta \) 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 \( \rho \) 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 \( y_t = \alpha + \sum_{i=0}^{\infty} \beta_i x_{t-i} + \epsilon_t \). But of course, we cannot estimate an infinite number of lag coefficients \( \beta_i \), 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
  - \( \beta_0 \) is the “impact multiplier”
• \( \sum_{s=0}^{t} \beta_s \) is the cumulative or “interim” multiplier after \( s \) periods. It measures the cumulative effect of a permanent change in \( x \) on \( y \) 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.

• **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 TSMR1–6, 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 must be used.

  ○ 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**

  ○ \( y_t = \delta + \theta_1 y_{t-1} + \delta_0 x_t + v_t \)

  ○ \( \frac{\partial y_t}{\partial x_t} = \delta_0 \)

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

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

    \( \frac{\partial y_{t+k}}{\partial x_t} = \theta_1^k \delta_0 \)

  Thus, dynamic multipliers start at \( \delta_0 \) 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 \( \sum_{k=0}^{t} \frac{\partial y_{t+k}}{\partial x_t} = \delta_0 \sum_{k=0}^{t} \theta_1^k \).

  ○ Long-run effect of a permanent change is \( \delta_0 \sum_{k=0}^{t} \theta_1^k = \frac{\delta_0}{1 - \theta_1} \).

  ○ Estimation has the potential problem of **inconsistency** if \( v_t \) 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:
  \[ y_t = \delta + \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p} + \delta_0 x_t + \nu_t. \]
- This can be written \( \theta(L) y_t = \delta + \delta_0 x_t + \nu_t. \)
- Same general principles apply:
  - Worry about stationarity of lag structure: roots of \( \theta(L) \)
  - If \( \nu \) 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 \( \theta_p \) is statistically significant
  - 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)
- \( y_t = \delta_0 + \theta_1 y_{t-1} + \ldots + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \ldots + \delta_q x_{t-q} + \nu_t \)
  - Can add more \( x \) variables with varying lag lengths
  \( \theta(L) y_t = \delta_0 + \delta(L) x_t + \nu_t, \)
  - \( y_t = \frac{\delta_0 + \delta(L) x_t + \nu_t}{\theta(L)} \).
• 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:
  - \( E(y_t | y_{t-1}, y_{t-2}, \ldots, y_{t-p}, x_t, x_{t-1}, \ldots, x_{t-q}) = 0 \)
  - \((y_t, x_t)\) 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 TSMR assumptions that apply to most time-series models.

### Forecasting with time-series models

- **With AR(\( p \)) model**
  - \( y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_p y_{t-p} + \nu_t \), with \( \nu \) assumed to be serially uncorrelated.
  - We have observations for \( t = 1, 2, \ldots, T \)
  - Forecast for \( T+1 \): \( \hat{y}_{T+1} = \delta + \hat{\theta}_1 y_T + \hat{\theta}_2 y_{T-1} + \ldots + \hat{\theta}_p y_{T-p+1} \) with expected value of \( \nu \) at zero because of no serial correlation
    - If error term were autoregressive, then conditional expectation of \( \nu_{T+1} \) would be \( \rho \nu_T \), so would include residual of observation \( T \)

○ Forecast error is \( u_t = y_{T+1} - \hat{y}_{T+1} = (\delta - \hat{\delta}) + \sum_{s=1}^{p} \left( \theta_s - \hat{\theta}_s \right) y_{T+s} + \nu_{T+1} \)
  - HGL assume that \( \text{var}(\nu_{T+1}) >> \text{var} \left[ (\delta - \hat{\delta}) + \sum_{s=1}^{p} \left( \theta_s - \hat{\theta}_s \right) y_{T+s} \right] \), so they ignore the latter.
  - I’m not willing to ignore this (although it is often true that the variance of the error is larger, as in Problem 6.17.)
  - I will write \( e_{h,k} = (\delta - \hat{\delta}) + \sum_{s=1}^{p} \left( \theta_s - \hat{\theta}_s \right) y_{T+k-s} \) and \( \text{var}(e_{h,k}) \) to be that component of the \( k \)-period-ahead forecast and keep it in the equation
    - \( u_t = e_{h,1} + \nu_{T+1} \)
    - \( \text{var}(u_t) = \text{var}(e_{h,1}) + \sigma^2 \)

○ 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:
    \[
    \hat{y}_{T+2} = \delta + \hat{\theta}_1 \hat{y}_{T+1} + \hat{\theta}_2 y_T + \ldots + \hat{\theta}_p y_{T+p-1}
    \]
- Forecast error is
  \[ u_2 = y_{T+2} - \hat{y}_{T+2} = \left( \delta - \hat{\delta} \right) + \theta_1 (x_{T+1} - y_{T+1}) + \sum_{s=1}^{p} (\theta_s - \hat{\theta}_s) y_{T+2-s} + \nu_{T+2} \]
  \[ = e_{b,2} + \theta_1 e_{b,1} + \nu_{T+2} + \theta_1 \nu_{T+1}. \]

- Variance of forecast error is
  \[ \text{var}(u_2) = \text{var}(e_{b,2}) + \theta_1^2 \text{var}(e_{b,1}) + (1 + \theta_1^2) \sigma_y^2. \]

- Similarly,
  \[ \text{var}(u_3) = \text{var}(e_{b,3}) + \theta_1^2 \text{var}(e_{b,2}) + (\theta_1^2 + \theta_2^2) \text{var}(e_{b,1}) + \left(1 + \theta_1^2 + (\theta_2 + \theta_1^3) \right) \sigma_y^2 \]