Section 9 Time Series Regression with Non-Stationary Variables

The TSMR assumptions include, critically, the assumption that the variables in a regression are stationary. But many (most?) time-series variables are nonstationary. We now turn to techniques—all quite recent—for estimating relationships among nonstationary variables.

Dickey-Fuller tests 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.
- The first and simplest test for unit-root nonstationarity is the **Dickey-Fuller test**. It comes in several variants depending on whether we allow a non-zero constant and/or a deterministic trend
- **Testing the null that y is random walk without drift: DF test with no constant or trend**
  - Consider the AR(1) process $y_t = \rho y_{t-1} + u_t$
    - The null hypothesis is that y is $I(1)$, so H_0: $\rho = 1$.
    - Under the null hypothesis, y follows a random walk without drift.
    - Alternative hypothesis is one-sided: H_1: $\rho < 1$ and y is stationary AR(1) process
  - We can't just run an OLS regression of this equation and test $\rho = 1$ with a conventional *t* test because the distribution of the *t* statistic is not asymptotically normal under the null hypothesis that y is $I(1)$.
  - If we subtract y_{t-1} from both sides, we get $\Delta y_t = (\rho - 1) y_{t-1} + u_t = \gamma y_{t-1} + u_t$, with $\gamma \equiv \rho - 1$.
    - If the null hypothesis is true ($\rho = 1$ or $\gamma = 0$) then y is non-stationary, Δy is stationary, and the coefficient on the right is zero.
    - 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 stricter than those of the normal.
      - See Table 18.2 and 18.3 on pp. 611 and 613 of Wooldridge for some critical values.
      - Stata will compute critical values for you based on your model and data.
    - If the DF statistic is less than the (negative) 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 (left-tailed) is appropriate here because $\gamma = \rho - 1$ should always be negative. Otherwise, it would imply $\rho > 1$, which is non-stationary in a way that cannot be rectified by differencing.
    - The intuition of the DF test relates to the mean-reversion property of stationary processes:
        - $\Delta y_t = \gamma y_{t-1} + u_t$.
        - If $\gamma < 0$, then when y is positive (above its zero mean) Δy will tend to be negative, pulling y back toward its (zero) mean.
        - If $\gamma = 0$, then there is no tendency for the change in y to be affected by whether y is currently above or below the mean: there is no mean reversion and y is nonstationary.
- **Testing the null that *y* is a random walk with drift: DF test with constant but no trend**
    - In this case, the null hypothesis is that y follows a random walk with drift
    - Alternative hypothesis is stationarity
    - $$y_t = \alpha + \rho y_{t-1} + u_t$$ $$\Delta y_t = \alpha + (\rho - 1) y_{t-1} + u_t = \alpha + \gamma y_{t-1} + u_t$$
    - $$H_0 : \rho = 1\ (\gamma = 0)$$ $$H_1 : \rho < 1\ (\gamma < 0)$$
    - Very similar to DF test without a constant but critical values are different (See Table 18.3)
- **Testing the null that *y* is "trend stationary": DF test with constant and trend**
    - In this case, the null is that the deviations of y from a deterministic trend are a random walk
    - Alternative is that these deviations are stationary
    - $$y_t = \alpha + \lambda t + \rho y_{t-1} + u_t$$ $$\Delta y_t = \alpha + \lambda t + (\rho - 1) y_{t-1} + u_t = \alpha + \lambda t + \gamma y_{t-1} + u_t$$
    - Note that under the alternative hypothesis, y is nonstationary (due to the deterministic trend) unless $\lambda = 0$.
- Is u serially correlated?
    - Probably, and the properties of the DF test statistic assume that it is not.
    - By adding some lags of Δy on the RHS we can usually eliminate the serial correlation of the error.
    - $\Delta y_t = \alpha + (\rho - 1) y_{t-1} + a_1 \Delta y_{t-1} + \ldots + a_p \Delta y_{t-p} + u_t$ is the model for the **Augmented Dickey-Fuller (ADF) test**, which is similar but has a different distribution that depends on p.
    - Stata does DF and ADF tests with the dfuller command, using the lags(#) option to add lagged differences.

- An alternative to the ADF 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-Perron 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.

- **DF-GLS test**
  - Another useful test that can have more power is the DF-GLS test, which tests the null hypothesis that the series is $I(1)$ against the alternative of either $I(0)$ or that the series is stationary around a deterministic trend.
    - Available for download from Stata as dfgls command.
    - DF-GLS test for H_0: y is $I(1)$ vs. H_1: y is $I(0)$
      - Quasi difference series:

        $$z_t = \begin{cases} y_t, \text{ for } t = 1, \\ y_t - \left(1 - \frac{7}{T}\right) y_{t-1}, \text{ for } t = 2, 3, \ldots, T. \end{cases}$$

        $$x_{1t} = \begin{cases} 1, \text{ for } t = 1, \\ \frac{7}{T}, \text{ for } t = 2, 3, \ldots, T. \end{cases}$$

      - Regress z_t on x_{1t} with no constant (because x_{1t} is essentially a constant):

        $z_t = \delta_0 z_{1t} + u_t$.
      - Calculate a "detrended" (really demeaned here) y series as

        $y_t^d \equiv y_t - \hat{\delta}_0$.
      - Apply the DF test to the detrended y^d series with corrected critical values (S&W Table 16.1 provide critical values).
    - DF-GLS test for H_0: y is $I(1)$ vs. H_1: y is stationary around deterministic trend

- Quasi-difference series:

$$z_t = \begin{cases} y_t, \text{ for } t = 1, \\ y_t - \left(1 - \dfrac{13.5}{T}\right) y_{t-1}, \text{ for } t = 2, 3, \ldots, T \end{cases}$$

$$x_{1t} = \begin{cases} 1, \text{ for } t = 1, \\ \dfrac{13.5}{T}, \text{ for } t = 2, 3, \ldots, T \end{cases}$$

$$x_{2t} = \begin{cases} 1, \text{ for } t = 1, \\ t - \left(1 - \dfrac{13.5}{T}\right)(t-1), \text{ for } t = 2, 3, \ldots, T \end{cases}$$

- Run "trend" regression

  $z_t = \delta_0 x_{1t} + \delta_1 x_{2t} + u_t$

- Calculate detrended y as $y_t^d \equiv y_t - \left(\hat{\delta}_0 + \hat{\delta}_1 t\right)$
- Perform DF test on y_t^d using critical values from S&W's Table 16.1.

▪ Stock and Watson argue that this test has considerably more power to distinguish borderline stationary series from non-stationary series.

Integration, differencing, and cointegration

- It is possible for two integrated series to "move together" in a nonstationary way, for example, so that their difference (or any other linear combination) is stationary. Such series follow a **common stochastic trend**. These series are said to be **cointegrated**.
  - Stationarity is like a rubber band pulling a series back to the fixed mean.
  - Cointegration is like a rubber band pulling the two series back to (a fixed relationship with) each other, even though both series are not pulled back to a fixed mean.
- If y and x are both integrated, we cannot rely on OLS standard errors or t statistics. By differencing, we can avoid spurious regressions:
  - If $y_t = \beta_0 + \beta_1 x_t + u_t$ then $\Delta y_t = \beta_1 \Delta x_t + \Delta u_t$.
    - Note the absence of a constant term in the differenced equation: the constant cancels out.
    - If a constant were to be in the differenced equation, that would correspond to a linear trend in the levels equation.
    - Δu is stationary as long as u is $I(0)$ or $I(1)$
  - The differenced equation has no "history." Is u stationary or nonstationary?
    - Suppose that u is $I(1)$.

            - This means that the difference $u_t = y_t - \beta_0 - \beta_1 x_t$ is not mean-reverting and there is no long-run tendency for *y* to stay in the fixed relationship with *x*.
                - No cointegration between *y* and *x*.
            - Δu is $I(0)$.
            - "Bygones are bygones:" if y_t is high (relative to x_t) due to a large positive u_t, then there is no tendency for *y* to come back to a defined relationship with *x* after *t*.
            - Estimation of differenced equation is appropriate.
        - Now suppose that *u* is $I(0)$.
            - That means that the levels of *y* and *x* tend to stay close to the relationship given by the equation.
            - Suppose that there is a large positive u_t that puts y_t above its long-run equilibrium level in relation to x_t.
            - With stationary *u*, we expect the level of *y* to return to the long-run relationship with *x* over time: stationarity of *u* implies that $\text{corr}(u_t, u_{t+s}) \to 0$ as $s \to \infty$.
            - Thus, future values of Δy should tend to be smaller (less positive or more negative) than those predicted by Δx in order to close the gap. In terms of the error terms, a large positive u_t should be followed by negative Δu_t values to return *u* to zero *if u is stationary*.
                - This is the situation where *y* and *x* are cointegrated.
            - This is *not* reflected in the differenced equation, which says that "bygones are bygones" and future values of Δy are only related to the future Δx values—there is no tendency to eliminate the gap that opened up at *t*.
    - In the cointegrated case
        - If we estimate the regression in differenced form we are missing the "history" of knowing how *y* will be pulled back into its long-run relationship with *x*.
        - If we estimate in levels, our test statistics are unreliable because the variables (though not the error term) are nonstationary.
- The appropriate model for the cointegrated case is the **error-correction model (ECM)** of Hendry and Sargan.
    - ECM consists of two equations:
        - Long-run (cointegrating) equation: $y_t = \beta_0 + \beta_1 x_t + u_t$, where (for the true values of β_0 and β_1) *u* is $I(0)$

        - Short-run (ECM) adjustment equation:
          $\Delta y_t = -\alpha\left(y_{t-1} - \beta_0 - \beta_1 x_{t-1}\right) + \theta_1 \Delta y_{t-1} + \ldots + \theta_p \Delta y_{t-p} + \delta_0 \Delta x_t + \ldots + \delta_q \Delta x_{t-q} + v_t$
    - Note the presence of the error-correction term with coefficient $-\alpha$ in the ECM equation.
        - This term reflects the distance that y_{t-1} is from its long-run relationship to x_{t-1}.
        - If $-\alpha < 0$, then y_{t-1} above its long-run level will cause Δy_t to be negative (other factors held constant), pulling y back toward its long-run relationship with x.
    - There is no constant term in the differenced regression (though many people include one) because the constant term in the x, y relationship cancels out in the differencing process.
    - Because both y and x are $I(1)$, their first differences are $I(0)$. Because they are cointegrated with **cointegrating vector** β_0, β_1, the difference in the error-correction term is also $I(0)$.
        - This term would not be stationary if they weren't cointegrated and the ECM regression would be invalid.
- Estimation of cointegrated models
    - The ECM equation can be estimated by OLS without undue difficulty because all the variables are stationary.
    - The cointegrating regression can be estimated "super-consistently" by OLS (although the estimates will be non-normal and the standard errors will be invalid).
    - HGL suggest estimating both equations together by nonlinear least squares
    - Stock and Watson recommend an alternative "dynamic OLS" estimator for the cointegration equation:
        - $y_t = \beta_0 + \beta_1 x_t + \sum_{j=-p}^{p} \phi_j \Delta x_{t-j} + v_t$
        - This can be estimated by OLS and the HAC-robust standard errors are valid for β.
        - Don't include the ϕ terms in the error-correction term in the ECM regression, which remains $y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1}$.
    - Normally, we would have to correct the standard errors of the ECM for the fact that the error-correction variable is calculated based on estimated β rather than known with certainty.
        - However, because the estimators of β are "super-consistent" in the cointegration case, they converge asymptotically faster to the true β than the δ and θ estimates and can be treated as if they were true parameter values instead of estimates.

- Multivariate cointegration
    - The concept of cointegration extends to multiple variables
    - With more than two variables, there can be more than one cointegrating relationship (vector)
        - Interest rates on bonds issued by Oregon, Washington, Idaho might be related by $r_O = r_W = r_I$. Two equal signs means two cointegrating relationships.
    - Vector error-correction models (VECM) allow for the estimation of error-correction regressions with multiple cointegrating vectors. We will study these soon.
        - Stata does this using the vec command.
- **Testing for cointegration**
    - The earliest test for cointegration is Engle and Granger's extension of the ADF test:
        - Estimate the cointegrating regression by OLS.
        - Test the residuals with an ADF test, using revised critical values as in S&W's Table 16.2.
    - Other, more popular tests include the Johansen-Juselius test, which generalizes easily to multiple variables and multiple cointegrating relationships.

VAR models (new, based on Daily Problem)

- VAR was developed in the macroeconomics literature as an attempt to characterize the joint time-series of a set (vector) of variables without making the restrictive (and perhaps false) assumptions that would allow the identification of structural dynamic models.
- VAR can be thought of as a reduced-form representation of the joint evolution of the set of variables.
    - However, in order to use the VAR for conditional forecasting, we have to make assumptions about the causal structure of the variables in the model.
    - The need for identifying restrictions gets pushed from the estimation to the interpretation phase of the model.
- Consider general relationship between two stationary variables y and x based on symmetric ARDL(1,1)
    - $$\begin{aligned} x_t &= \alpha_0 + \alpha_1 x_{t-1} + \theta_0 y \ + \theta_1 y_{t-1} + \varepsilon_t^x \\ y_t &= \phi_0 + \phi_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + \varepsilon_t^y \end{aligned}$$
    - $\text{var}\left(\varepsilon_t^x\right) = \sigma_x^2,\ \text{var}\left(\varepsilon_t^y\right) = \sigma_y^2,\ \text{cov}\left(\varepsilon_t^x, \varepsilon_t^y\right) = \sigma_{xy}.$
    - (We know this system will not be identified.)
- Solving yields

- $$x_t = \frac{\alpha_0 + \theta_0\phi_0}{1-\theta_0\delta_0} + \frac{\alpha_1 + \theta_0\delta_1}{1-\theta_0\delta_0}x_{t-1} + \frac{\theta_1 + \theta_0\phi_1}{1-\theta_0\delta_0}y_{t-1} + \frac{\theta_0\varepsilon_t^y + \varepsilon_t^x}{1-\theta_0\delta_0}$$
  $$y_t = \frac{\phi_0 + \theta_0\alpha_0}{1-\theta_0\delta_0} + \frac{\phi_1 + \delta_0\theta_1}{1-\theta_0\delta_0}x_{t-1} + \frac{\delta_1 + \delta_0\alpha_1}{1-\theta_0\delta_0}y_{t-1} + \frac{\delta_0\varepsilon_t^x + \varepsilon_t^y}{1-\theta_0\delta_0}.$$
    - Or
- $$x_t = \beta_{x,0} + \beta_{x,1}x_{t-1} + \gamma_{x,1}y_{t-1} + v_t^x$$
  $$y_t = \beta_{y,0} + \beta_{y,1}x_{t-1} + \gamma_{y,1}y_{t-1} + v_t^y.$$
    - $$\operatorname{var}\left(v_t^x\right) = \frac{1}{\left(1-\theta_0\delta_0\right)^2}\left(\sigma_x^2 + \theta_0^2\sigma_y^2 + 2\theta_0\sigma_{xy}\right)$$
      $$\operatorname{var}\left(v_t^y\right) = \frac{1}{\left(1-\theta_0\delta_0\right)^2}\left(\delta_0^2\sigma_x^2 + \sigma_y^2 + 2\delta_0\sigma_{xy}\right)$$
      $$\operatorname{cov}\left(v_t^x, v_t^y\right) = \frac{2}{\left(1-\theta_0\delta_0\right)^2}\left(\delta_0\sigma_x^2 + \theta_0\sigma_y^2 + \left(1+\delta_0\theta_0\right)\sigma_{xy}\right).$$

- Note absence of current values of the variables on the RHS of each equation.
    - This reflects uncertainty about whether the correlation between y_t and x_t is because x causes y or because y causes x.
    - Correlation between y_t and x_t will mean that the two error terms **are correlated** with one another, however. (This is assumed not to happen in the simplified HGL example, but in practice they are always correlated.) This means that we can't think of v_t^y as a "pure shock to y" **and** v_t^x as a pure shock to x: one of them will have to be responding to the other in order for them to be correlated.
- There are 11 parameters in the structural form (including variances and covariance), but only 9 that we can estimate in the VAR (again including variances and covariances).
    - The system lack identification: **we need two restrictions** in order to make the number of parameters match up.
- Note that each of the following identifying assumptions leaves the *form* of the VAR exactly the same:
    - $$\theta_0 = 0$$
      $$\delta_0 = 0$$
      $$\sigma_{xy} = 0.$$
    - Thus, we can't use the estimates of the β parameters of the VAR to discriminate between these assumptions. But making two of these three assumptions allow us to identify all of the structural parameters from the VAR coefficients.
    - In practice, we always assume that the underlying shocks are uncorrelated (third condition) and usually one of the other two.
    - There are alternative assumptions that can be made (Structural VAR) but the first two are most common.
- Estimate by OLS—SUR is identical because regressors are same in each equation

- How many variables?
  - More adds p coefficients to each equation
  - Generally keep the system small (6 variables is large)
- How many lags?
  - Can use the AIC or Schwartz criterion on the system as a whole to evaluate: $SC(p) = \ln\left[\det\left(\hat{\Sigma}_u\right)\right] + k(kp+1)\frac{\ln T}{T},$ where k is the number of variables/equations and p is the number of lags. The determinant is of the estimated covariance matrix of the errors, calculated as the sample variances and covariances of the residuals.

Identified vs. unidentified VARs

- What *can* we do with the unidentified VAR?
  - Granger causality tests
    - The setup is natural for bidirectional (or multidirectional) Granger causality tests.
    - What is null hypothesis?
      - Lagged x does not help predict y given the presence of lagged y values. (And vice versa)
    - Is this really causality?
      - No. It requires two critical assumptions
        - That no causality can be *only* instantaneous. There is always a dynamic fingerprint if one variable causes another.
        - That the present can't cause the past, so any correlation between y_t and x_{t-1} must be x causing y.
  - Forecasting
    - VAR is a simple generalization of predicting a single variable based on its own lags: we are predicting a vector of variables based on lags of all variables.
    - We can do forecasting without any assumptions about the underlying structural equations of the model. No identification issues for forecasting.
    - To do multi-period forecasts, we just plug in the predicted values for future years and generate longer-term forecasts recursively.
- What *can't* we do without identification assumptions?
  - Examine the effects of a shock to x or a shock to y (the ε terms) on the system
- Identification in VARs: **Impulse-response functions and variance decompositions**
  - In order to use VARs for simulation of shocks, we need to be able to **identify the shocks**.

- We nearly always assume that the structural shocks are uncorrelated: $\sigma_{xy} = 0$
  - Note that the VAR error terms will still be correlated $\text{cov}\left(v_t^x, v_t^y\right) \neq 0$.
- Identifying assumption in coefficients:
  - Is v^x a pure shock to x with no effect on y?
  - Is v^y a pure shock to y with no effect on x?
  - Both cannot generally be true if the two v terms are correlated.
  - Two possible interpretations (identifying restrictions)
    - v^x is a pure x shock; some part of v^y is a response to v^x and the remainder of v^y is a pure y shock.
      - x is "first" and y responds to x in the current period. y_t does not affect x_t.
      - This is equivalent to assuming that $\delta_0 \neq 0$ and $\theta_0 = 0$
    - v^y is a pure y shock; some part of v^x is a response to v^y and the remainder of v^x is a pure x shock.
      - Opposite assumption about contemporaneous causality: $\theta_0 \neq 0$ and $\delta_0 = 0$
  - If we don't make one of these assumptions, then shocks are not identified and we can't do simulations. (Can still forecast and do Granger causality, though.)
- If we make one or the other identifying restriction, (along with $\sigma_{xy} = 0$) then we can conduct simulations of the effects of shocks to x or y. Suppose that we assume that x affects y contemporaneously, but not the other way around ($\delta_0 \neq 0$ and $\theta_0 = 0$).
  - Shock of one unit (we often use one standard deviation instead) to v^x causes a one-unit increase in x_t and a change in y_t that depends on δ_0, which determines the covariance between v^x and v^y, which we can estimate.
    - In $t + 1$, the changes to x_t and y_t will affect x_{t+1} and y_{t+1} according to the coefficients β_{x1}, β_{y1}, γ_{x1}, and γ_{y1}. (We assume that all v terms are zero after t.)
    - Then in $t + 2$, the changes to x_t, y_t, x_{t+1}, and y_{t+1} will affect the values in $t + 2$. This process feeds forward indefinitely.
    - The sequence $\frac{\Delta x_{t+s}}{\Delta v_t^x}$ and $\frac{\Delta y_{t+s}}{\Delta v_t^x}$ for $s = 0, 1, 2, \ldots$ is called the

      impulse-response function with respect to a shock to x.
  - We can analyze a one-unit shock to y in the same basic way, except that by assumption v_t^y has no effect on v_t^x or x_t. This gives the IRF with respect to a shock to y.

    - Note that the **IRF will vary depending on our choice of identifying condition**. If we assume that y_t affects x_t but not vice versa (rather than the other way around), then we get a different IRF.
  - The identification and IRF calculation is similar for more than two variables. With $k > 2$, the most common identification scheme is identification by "ordering assumption." We pick one variable that can affect all the others contemporaneously, but is not immediately affected by any others. Then we pick the second variable that is affected only by the first in the immediate period but can affect all but the first.
    - This amounts to an ordering where variables can have a contemporaneous effect only on variable below them in the list.
    - (Of course, all variable in the model are assumed to affect all others with a one-period lag.)
  - The other common "output" from a VAR is the **variance decomposition**. This asks the same question about how the various shocks affect the various variables, but from the other direction.
    - "How much of the variance in y_{t+s} is due to shocks to x_t, shocks to y_t and shocks to other variables?"
    - The variance decomposition breaks down the variance of y_{t+s} into the shares attributed to each of the shocks.
    - We won't talk about the formulas used to calculate these.
  - The Enders text on the reading list provides more details.
- In Stata, there is a collection of var commands that begin with var
  - var varlist , lags(1/4) does estimation
    - Can also include exog(vars) to specify variables that are exogenous
  - After var command:
    - fcast compute (then fcast graph)
    - irf (create, graph, table)
      - oirf does ordered identification
      - irf alone assumes neither has immediate effect
    - vargranger
    - varlmar (test for autocorrelated residuals)
    - varsoc (criteria for lag length)
    - varstable (check stability of model)
  - Also var svar command for structural var

Vector error-correction models

- What if x and y are $I(1)$ variables that are cointegrated?
  - We can do a two-variable (vector) error-correction model

- With two variables, there can be only one cointegrating relationship linking the long-run paths of the two variables together.
  - With m variables, there can be $m - 1$ cointegrating relationships, but we won't worry about this.
- We can use OLS to estimate the cointegrating regression $y_t = \beta_0 + \beta_1 x_t + e_t$ and calculate the residuals $\hat{e}_t$, which are $I(0)$ if x and y are cointegrated.
- We can then estimate a VAR in the differences of x and y, using the residuals as cointegration terms on the right-hand side of each equation.
- For example:

$$\Delta y_t = \beta_{y0} + \alpha_y \hat{e}_{t-1} + \beta_{y1}\Delta y_{t-1} + \ldots + \beta_{yp}\Delta y_{t-p} + \gamma_{y1}\Delta x_{t-1} + \ldots + \gamma_{yp} x_{t-p} + v_t^y$$
$$\Delta x_t = \beta_{x0} + \alpha_x \hat{e}_{t-1} + \beta_{x1}\Delta y_{t-1} + \ldots + \beta_{xp}\Delta y_{t-p} + \gamma_{x1}\Delta x_{t-1} + \ldots + \gamma_{xp} x_{t-p} + v_t^x$$

- In Stata, vec command

Time-varying volatility: Autoregressive conditional heteroskedasticity (ARCH) models

- Financial economists have noted that volatility in asset prices seems to be autocorrelated: if there are highly volatile returns on one day, then it is likely that returns will have high volatility on subsequent days as well. This is called **volatility clustering.**
- Does "high volatility" (large error variance) tend to persist over time?
  - Heteroskedasticity in a time-series context means that the error variance depends on t: σ_t^2
  - One possibility would be to model σ_t^2 as a deterministic function of t (a trend?) or of other time-dependent variables
  - We can also model time-dependent error variance as a random variable
- ARCH models error variance as an AR or MA process: This is a particular pattern of heteroskedasticity where there are positive or negative shocks to the error variance each period and where shocks tend to persist
- Engle modeled this by making the variance of the error term depend on the square of recent error terms:
  - $y_t = \beta_1 + \beta_2 y_{t-1} + \gamma_1 x_{t-1} + u_t,$
    $u_t \sim N\left(0, \sigma_t^2\right),$
    $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_p u_{t-p}^2.$
  - This error structure is the ARCH(p) model.
    - Note that the conditional heteroskedasticity here is really moving average rather than autoregressive because there are no lagged σ^2 terms.
  - (HGL leave out the lagged y and x terms to look only at a single stationary variable)

- A now-more-common generalization is the GARCH(p, q) model:
    - $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_p u_{t-p}^2 + \phi_1 \sigma_{t-1}^2 + \ldots + \phi_q \sigma_{t-q}^2$
- ARCH and GARCH models (and a variety of other variants) are estimated by ML.
- In Stata, the arch command estimates both ARCH and GARCH models, as well as other variants.