Section 12 Vector Autoregression, Integration, and Cointegration, and ARCH models

Time-series econometrics was a frantically active field in the 1980s and 1990s, eventually leading to a shared Nobel Prize in 2003 for co-authors Clive Granger and Robert Engle. We will study three kinds of models from this literature: VAR models, cointegrated and error-correction models, and ARCH models.

Vector autoregression

- 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.
- Two-variable VAR(*p*) for *X* and *Y*:
    - $$\begin{aligned} Y_t &= \beta_{Y0} + \beta_{Y1} Y_{t-1} + \ldots + \beta_{Yp} Y_{t-p} + \gamma_{Y1} X_{t-1} + \ldots + \gamma_{Yp} X_{t-p} + u_{Yt} \\ X_t &= \beta_{X0} + \beta_{X1} Y_{t-1} + \ldots + \beta_{Xp} Y_{t-p} + \gamma_{X1} X_{t-1} + \ldots + \gamma_{Xp} X_{t-p} + u_{Xt} \end{aligned}$$
    - Note absence of current values of the "other" variable 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 means that we can't think of u_{Yt} as a "pure shock to *Y*" **and** u_{Xt} as a pure shock to *X*: one of them will have to be responding to the other in order for them to be correlated.
- 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 BIC on the system as a whole to evaluate:

      $BIC(p) = \ln\left[\det\left(\hat{\Sigma}_u\right)\right] + k(kp+1)\frac{\ln T}{T},$ 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.

- What can we use VARs for?
  - Granger causality tests
    - The setup is natural for bidirectional (or multidirectional) Granger causality tests.
  - 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.
- 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**.
    - Is u_X a pure shock to X with no effect on Y?
    - Is u_Y a pure shock to Y with no effect on X?
    - Both cannot generally be true because they are correlated.
    - Two possible interpretations (identifying restrictions)
      - u_X is a pure X shock and some part of u_Y is a response to u_X and the remainder is a pure Y shock.
        - X is "first" and Y responds to X in the current period. Y_t does not affect X_t.
      - u_Y is a pure Y shock and some part of u_X is a response to u_Y and the remainder is a pure X shock.
        - Opposite assumption about contemporaneous causality.
    - 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, 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.
    - Shock of one unit (we often use one standard deviation instead) to u_X causes a one-unit increase in X_t and a change in Y_t that depends on the covariance between u_X and u_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 u 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 $\Delta X_{t+s}/\Delta u_{Xt}$ and $\Delta Y_{t+s}/\Delta u_{Xt}$ 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 u_{Yt} has no effect on u_{Xt} 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 series 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 can 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.

Integrated variables

We have encountered the idea of unit roots and nonstationary series before when we discussed stochastic trends. We now formalized the idea of **integrated time series** and then extend the concept of integration to multiple variables that are jointly nonstationary but that tend to move together on their nonstationary paths (cointegration).

- **Unit roots and integration**

- Consider the general AR(p) process $Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + u_t$, which we write in lag-operator notation as $\beta(L) Y_t = \beta_0 + u_t$.
- We noted above that the stationarity properties of Y are determined by whether the roots of $\beta(L) = 0$ are outside the unit circle (stationary) or on it (nonstationary).
  - $\beta(L)$ is an order-p polynomial in the lag operator $1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_p L^p$.
  - We can factor $\beta(L)$ as $1 - \beta_1 L - \beta_2 L^2 - \ldots - \beta_p L^p = (1 - \phi_1 L)(1 - \phi_2 L)\cdots(1 - \phi_p L)$, where $\frac{1}{\phi_1}, \frac{1}{\phi_2}, \ldots, \frac{1}{\phi_p}$ are the roots of $\beta(L)$.
  - 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 $|\phi_j| \leq 1$.
  - 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 $\beta(L) = (1 - \phi_1 L)\cdots(1 - \phi_{p-k} L)(1 - L)^k$, where we number the roots so that the first $p - k$ are greater than one.
  - Let $\gamma(L) = \frac{\beta(L)}{(1 - L)^k} = (1 - \phi_1 L)\cdots(1 - \phi_{p-k} L)$. Then $\beta(L) Y_t = \gamma(L)(1 - L)^k Y_t = \gamma(L)(\Delta^k Y_t) = \beta_0 + u_t$.
  - Because $\gamma(L)$ has 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**.
    - 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 different is stationary, then Y is $I(2)$, etc.
    - In practice, most economic time series are $I(0)$, $I(1)$, or occasionally $I(2)$.

- Testing for unit roots
  - We discussed the Dickey-Fuller, augmented Dickey-Fuller, and Phillips-Peron tests before.
    - All of these tests have a null hypothesis that the series is $I(1)$ against the alternative that it is $I(0)$.

        - These tests tend to have low power: our graphs showed how difficult it can be to distinguish a series with a unit root from one with a barely stationary root.
        - Low power means we will often accept false null hypotheses, in this case concluding that the series is *I*(1) when in fact it may be *I*(0).
- 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:

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

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

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

$$V_t = \delta_0 X_{1t} + e_t.$$

        - 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* series with corrected critical values (S&W Table 16.1).
    - DF-GLS test for H_0: *Y* is *I*(1) vs. H_1: *Y* is stationary around deterministic trend
        - Quasi-difference series:

$$V_t = \begin{cases} Y_t, \text{ for } t = 1, \\ Y_t - \left(1 - \frac{13.5}{T}\right) Y_{t-1}, \text{ for } t = 2, 3, \ldots, T \end{cases}$$

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

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

- Run "trend" regression
  $V_t = \delta_0 X_{1t} + \delta_1 X_{2t} + e_t$
- 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 note that this test has considerably more power to distinguish borderline stationary series from non-stationary series.

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. 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 can't 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 $\Delta Y_t = \Delta X_t + \Delta u_t$.
    - 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 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 about 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 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 in differences 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, we can't rely on our test statistics because the variables (though not the error term) are nonstationary.
- The appropriate model for the cointegrated case is the **error-correction model** of Hendry and Sargan.
    - ECM consists of two equations:
        - Long-run (cointegrating) equation: $Y_t = \theta_0 + \theta_1 X_t + u_t$, where (for the true values of β_0 and β_1 u is $I(0)$
        - Short-run (ECM) adjustment equation:
          $$\Delta Y_t = \beta_0 + \beta_1 \Delta Y_{t-1} + \ldots + \beta_p \Delta Y_{t-p} + \gamma_1 \Delta X_{t-1} + \ldots + \gamma_p \Delta X_{t-p} + \alpha\left(Y_{t-1} - \theta_0 - \theta_1 X_{t-1}\right) + v_t$$
    - Note the presence of the error-correction term with coefficient α 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.
    - Because both Y and X are $I(1)$, their differences are $I(0)$. Because they are cointegrated with **cointegrating vector** θ_0, θ_1, the difference in the error-correction term is also $I(0)$.
        - It would not be if they weren't cointegrated and the ECM regression would be invalid.

- 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).
- Stock and Watson recommend an alternative "dynamic OLS" estimator for the cointegration equation:
  - $Y_t = \theta_0 + \theta_1 X_t + \sum_{j=-p}^{p} \delta_j \Delta X_{t-j} + u_t$
  - This can be estimated by OLS and the HAC-robust standard errors are valid for θ.
  - Don't include the δ terms in the error-correction term in the ECM regression, which remains $Y_{t-1} - \hat{\theta}_0 - \hat{\theta}_1 X_{t-1}$.
- 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 θ rather than known with certainty.
  - However, because the estimators of θ are "super-consistent" in the cointegration case, they converge asymptotically faster to the true θ than the β and γ 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.
    - 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:
    - Regress 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.

Autoregressive conditional heteroskedasticity (ARCH)

- 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.**
- Engle modeled this by making the variance of the error term depend on the square of recent error terms:
    - $Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t,$
      $u_t \sim N\left(0, \sigma_t^2\right),$
      $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_p u_{t-p}^2.$
    - 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.
- 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.