Econ 423 Lecture Notes: Additional Topics in Time Series[1]

John C. Chao

April 25, 2017

[1] These notes are based in large part on Chapter 16 of Stock and Watson (2011). They are for instructional purposes only and are not to be distributed outside of the classroom.

Vector Autoregression (VAR)

- **Motivation:** One may be interested in forecasting two or more variables; such as rate of inflation, rate of unemployment, growth rate of GDP, and interest rates. In this case, it is beneficial to develop a single model that allows you to forecast all these variables in a systemic approach.
- **Definition:** A $VAR\,(p)$, i.e., a vector autoregression of order p, a set of m time series regressions, in which the regressors are the p lagged values of the m time series variables.
- **Example:** $m = 2$ case

$$\begin{aligned} Y_{1t} &= \beta_{10} + \beta_{11} Y_{1t-1} + \cdots + \beta_{1p} Y_{1t-p} \\ &\quad + \gamma_{11} Y_{2t-1} + \cdots + \gamma_{1p} Y_{2t-p} + u_{1t}, \\ Y_{2t} &= \beta_{20} + \beta_{21} Y_{1t-1} + \cdots + \beta_{2p} Y_{1t-p} \\ &\quad + \gamma_{21} Y_{2t-1} + \cdots + \gamma_{2p} Y_{2t-p} + u_{2t}. \end{aligned}$$

Estimation and Inference

- Algebraically, the VAR model is simply a system of m **linear** regressions; or, to put it another way, it is a multivariate linear regression model.
- The coefficients of the VAR can be estimated by estimating each equation by OLS.
- Under appropriate conditions, the OLS estimators are consistent and have a joint normal distribution in large samples in the stationary case.
- In consequence, in the stationary case, inference can proceed in the usual way; for example, 95% confidence interval on coefficients can be constructed based on the usual rule:

$$\text{estimated coefficients} \pm 1.96 \times \text{standard errors.}$$

Estimation and Inference (con't)

- **An Advantage of the VAR:** By modeling the dynamics of m variables as a system, one can test joint hypotheses that involve restrictions **across** multiple equations.
- **Example:** In a two-varlable $VAR(1)$, one might be interested in testing the null hypothesis

$$H_0 : \beta_{11} - \beta_{21} = 0$$

on the unrestricted model

$$\begin{aligned} Y_{1t} &= \beta_{10} + \beta_{11} Y_{1t-1} + \gamma_{11} Y_{2t-1} + u_{1t}, \\ Y_{2t} &= \beta_{20} + \beta_{21} Y_{1t-1} + \gamma_{21} Y_{2t-1} + u_{2t}. \end{aligned}$$

- Since the estimated coefficients have a jointly normal large sample distribution, the restrictions on these coefficients can be tested by computing the t- or the F-statistic.
- Importantly, many hypotheses of interest to economists can be formulated as cross-equation restrictions.

Modeling Issues

- How many variables should be included in a VAR?
  - (i) Having more variables leads to having more coefficients to estimate which, in turn, increases estimation error and can result in a deterioration of forecast accuracy.
  - (ii) Concrete example: A VAR with 5 variables and 4 lags will have 21 coefficients to estimate in each equation, leading to a total of 105 coefficients that must be estimated.
  - (iii) A preferred strategy is to keep m relatively small and to make sure that the variables included are plausibly related to each other, so that they will be useful in forecasting one another.
  - (iv) For example, economic theory suggests that the inflation rate, the unemployment rate, and the short-term interest rate are related to one another, suggesting that it would be useful to model these variables together in a VAR system.

Modeling Issues (con't)

- Determining the lag order in VAR's
  (i) We can estimate the lag length of a VAR using either the F-test or an information criterion, but the latter is preferred as it trades off between goodness of fit and the dimension of the model, whereas the F-test does not.
  (ii) **BIC in the vector case:**

$$BIC\left(p\right)=\ln\left[\det\left(\widehat{\Sigma}_u\right)\right]+m\left(mp+1\right)\frac{\ln T}{T},$$

where $\widehat{\Sigma}_u$ is an estimate of Σ_u, the $m\times m$ covariance matrix of the VAR errors. Let $\widehat{u}_{it}$ and $\widehat{u}_{jt}$ be, respectively, the OLS residual for the i^{th} and j^{th} equations, respectively, and note that the $(i,j)^{th}$ element of $\widehat{\Sigma}_u$ is given by

$$\widehat{\Sigma}_u\left(i,j\right)=\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}\widehat{u}_{jt},$$

i.e., an estimate of $Cov\left(u_{it},u_{jt}\right)$. Moreover, $\det\left(\widehat{\Sigma}_u\right)$ denotes the determinant of the matrix $\widehat{\Sigma}_u$.

Modeling Issues (con't)

- Determining the lag order in VAR's (con't)
  - (iii) **AIC in the vector case:** Analogous to BIC,

$$AIC\left(p\right) = \ln\left[\det\left(\widehat{\Sigma}_u\right)\right] + m\left(mp+1\right)\frac{2}{T}.$$

  - (iv) Note that "penalty term" for both AIC and BIC involves the factor

$$m\left(mp+1\right),$$

whih is the total number of coefficients in an m-variable $VAR(p)$ model, as there are m equations each having an intercept as well as p lags of the m variables.

  - (v) Note also that increasing the lag order by one leads to an additional m^2 coefficients to estimate.
  - (vi) Let $\mathcal{P} = \{0, 1, 2, ..., \overline{p}\}$. As in the univariate case, we can select the lag order based on BIC or AIC using the following estimation rule

$$\begin{aligned} \widehat{p}_{BIC} &= \arg\min_{p\in\mathcal{P}} BIC\left(p\right), \\ \widehat{p}_{AIC} &= \arg\min_{p\in\mathcal{P}} AIC\left(p\right). \end{aligned}$$

Empirical Example: A VAR Model of the Rates of Inflation and Unemployment

- Estimating a $VAR(4)$ model for ΔInf_t and $Unemp_t$ using data from 1982:I to 2004:IV gives the following result:

$$\begin{aligned}
\widehat{\Delta Inf}_t &= \underset{(0.55)}{1.47} - \underset{(0.12)}{0.64}\Delta Inf_{t-1} - \underset{(0.10)}{0.64}\Delta Inf_{t-2} - \underset{(0.11)}{0.13}\Delta Inf_{t-3} \\
&\quad - \underset{(0.09)}{0.13}\Delta Inf_{t-4} - \underset{(0.58)}{3.49}\, Unemp_{t-1} + \underset{(0.94)}{2.80}\, Unemp_{t-2} \\
&\quad + \underset{(1.07)}{2.44}\, Unemp_{t-3} - \underset{(0.55)}{2.03}\, Unemp_{t-4}, \quad \left(\overline{R}^2 = 0.44\right);
\end{aligned}$$

$$\begin{aligned}
\widehat{Unemp}_t &= \underset{(0.12)}{0.22} + \underset{(0.017)}{0.005}\Delta Inf_{t-1} - \underset{(0.018)}{0.004}\Delta Inf_{t-2} - \underset{(0.018)}{0.007}\Delta Inf_{t-3} \\
&\quad - \underset{(0.014)}{0.003}\Delta Inf_{t-4} + \underset{(0.11)}{1.52}\, Unemp_{t-1} - \underset{(0.18)}{0.29}\, Unemp_{t-2} \\
&\quad - \underset{(0.21)}{0.43}\, Unemp_{t-3} + \underset{(0.11)}{0.16}\, Unemp_{t-4}, \quad \left(\overline{R}^2 = 0.982\right).
\end{aligned}$$

Empirical Example (con't)

- **Granger Causality Tests**

  (a) Write the $VAR(4)$ model for ΔInf_t and $Unemp_t$ as

$$\begin{aligned}\Delta Inf_t &= \beta_{10} + \beta_{11}\Delta Inf_{t-1} + \beta_{12}\Delta Inf_{t-2} + \beta_{13}\Delta Inf_{t-3} \\ &\quad + \beta_{14}\Delta Inf_{t-4} + \gamma_{11} Unemp_{t-1} + \gamma_{12} Unemp_{t-2} \\ &\quad + \gamma_{13} Unemp_{t-3} + \gamma_{14} Unemp_{t-4} + u_{1t}\end{aligned}$$

$$\begin{aligned}Unemp_t &= \beta_{20} + \beta_{21}\Delta Inf_{t-1} + \beta_{22}\Delta Inf_{t-2} + \beta_{23}\Delta Inf_{t-3} \\ &\quad + \beta_{24}\Delta Inf_{t-4} + \gamma_{21} Unemp_{t-1} + \gamma_{22} Unemp_{t-2} \\ &\quad + \gamma_{23} Unemp_{t-3} + \gamma_{24} Unemp_{t-4} + u_{2t}.\end{aligned}$$

  (b) Test Granger non-causality of lagged unemployment rates on changes in inflation, i.e.,

$$H_0 : \gamma_{11} = \gamma_{12} = \gamma_{13} = \gamma_{14} = 0.$$

  In this case, $F = 11.04$ with p-value $= 0.001$, so that H_0 is rejected.

Empirical Example (con't)

- **Granger Causality Tests (con't)**

(c) In this case, we conclude that unemployment rate is a useful predictor for changes in inflation, given lags in inflation.

(d) Test Granger non-causality of lagged changes in inflation rates on the unemployment rate, i.e.,

$$H_0 : \beta_{21} = \beta_{22} = \beta_{23} = \beta_{24} = 0.$$

Here, $F = 0.16$ with p-value $= 0.96$, so that H_0 is not rejected in this case.

- **Intuitive Notion of Common Stochastic Trend:**
  It is possible that two or more time series with stochastic trends can move together so closely over the long run that they appear to have the same trend component. In this case, they are said to share a **common stochastic trend**.
- **Orders of Integration, Differencing, and Stationarity**
  1. If Y_t is integrated of order **one** (denoted $Y_t \equiv I(1)$); then, its first difference ΔY_t is stationary, i.e., $\Delta Y_t \equiv I(0)$. In this case, Y_t has a unit autoregressive root.
  2. If Y_t is integrated of order **two** (denoted $Y_t \equiv I(2)$); then, its second difference $\Delta^2 Y_t$ is stationary. In this case, $\Delta Y_t \equiv I(1)$.
  3. If Y_t is integrated of order d (denoted $Y_t \equiv I(d)$); then, $\Delta^d Y_t$ is stationary, i.e., Y_t must be differenced d times in order to produce a series that is stationary.

- **Definition of Cointegration:**
  Suppose that X_t and Y_t are integrated of order one. If, for some coefficient θ, $Z_t = Y_t - \theta X_t$ is integrated of order zero; then, X_t and Y_t are said to be **cointegrated**. The coefficient θ is called the cointegrating coefficient.
- **Remark:**
  If X_t and Y_t are cointegrated, then they have the same, or common, stochastic trend. Computing the difference $Y_t - \theta X_t$ then eliminates this common stochastic trend.

Deciding If Variables Are Cointegrated

Three ways to decide whether two variables is cointegrated:

1. Use expert knowledge and economic theory.
2. Graph the series and see whether they appear to have a common stochastic trend.
3. Perform statistical test for cointegration.

Testing for Cointegration

- **Some Observations:** Let Y_t and X_t be two time series such that $Y_t \equiv I(1)$ and $X_t \equiv I(1)$.
  1. If Y_t and X_t are cointegrated with cointegrating coefficient θ, then $Y_t - \theta X_t \equiv I(0)$.
  2. On the other hand, if Y_t and X_t are not cointegrated, then $Y_t - \theta X_t \equiv I(1)$.
  3. 1. and 2. suggest that we can test for the presence of cointegration by testing

$$H_0 : Y_t - \theta X_t \equiv I(1) \text{ versus } H_1 : Y_t - \theta X_t \equiv I(0)$$

- **Two Cases**
  1. θ is known, i.e., a value for θ is suggested by expert knowledge or by economic theory. In this case, one can simply construct the time series

$$Z_t = Y_t - \theta X_t$$

and test the null hypothesis $H_0 : Y_t - \theta X_t \equiv I(1)$ using the augmented Dickey-Fuller test.

Testing for Cointegration (con't)

2. θ is unknown: In this case, perhaps the easiest approach is to adopt a two-step procedure

   1. **Step 1:** Estimate the cointegrating coefficient θ by OLS estimation of the regression
   $$Y_t = \alpha + \theta X_t + Z_t$$
   and obtain the residual series $\widehat{Z}_t = Y_t - \widehat{\alpha} - \widehat{\theta} X_t$.
   2. **Step 2:** Apply a unit root test, such as the augmented Dickey-Fuller test, to test whether the residual series $\widehat{Z}_t$ is an $I(1)$ process. (Engle and Granger, 1987, and Phillips and Ouliaris, 1990).

3. **Remark:** A complication which arises when θ is unknown is that, under H_0, $\widehat{Z}_t \equiv I(1)$, so that the regression of Y_t on X_t is a spurious regression, which implies, in particular, that $\widehat{\theta}$ is not a consistent estimator. As a result, we cannot use the same critical values which apply in Case 1 discussed earlier.
4. The two-step procedure can be extended in a straightforward manner to cases with more than one regressor (e.g., the case with k regressors $X_{1t}, ..., X_{kt}$) by running the multiple regression

$$Y_t = \alpha + \theta_1 X_{1t} + \cdots + \theta_k X_{kt} + Z_t$$

and testing the residual process $\widehat{Z}_t = Y_t - \widehat{\alpha} - \widehat{\theta}_1 X_{1t} - \cdots - \widehat{\theta}_k X_{kt}$ for the presence of a unit root. Critical values for the residual-based cointegration test do depend on the number of regressors, however.

Testing for Cointegration (con't)

Table: Critical Values for Residual-Based Tests for Cointegration

# of X's in the regression	10%	5%	1%
1	−3.12	−3.41	−3.96
2	−3.52	−3.80	−4.36
3	−3.84	−4.16	−4.73
4	−4.20	−4.49	−5.07

Vector Error Correction Model

- Suppose that $X_t \equiv I(1)$ and $Y_t \equiv I(1)$, and suppose that X_t and Y_t are cointegrated. Then, it turns out that a bivariate VAR model in terms of the first differences ΔX_t and ΔY_t is misspecified.
- The correct model will include the term $Y_{t-1} - \theta X_{t-1}$ in addition to the lagged values of ΔX_t and ΔY_t.
- More specifically, the correct model is of the form

$$\begin{aligned}\Delta Y_t &= \beta_{10} + \beta_{11}\Delta Y_{t-1} + \cdots + \beta_{1p}\Delta Y_{t-p} \\ &\quad + \gamma_{11}\Delta X_{t-1} + \cdots + \gamma_{1p}\Delta X_{t-p} \\ &\quad + \alpha_1 \left(Y_{t-1} - \theta X_{t-1}\right) + u_{1t}, \\ \\ \Delta X_t &= \beta_{20} + \beta_{21}\Delta Y_{t-1} + \cdots + \beta_{2p}\Delta Y_{t-p} \\ &\quad + \gamma_{21}\Delta X_{t-1} + \cdots + \gamma_{2p}\Delta X_{t-p} \\ &\quad + \alpha_2 \left(Y_{t-1} - \theta X_{t-1}\right) + u_{2t}.\end{aligned}$$

- This model is known as the **vector error correction model** (VECM), and the term $Y_{t-1} - \theta X_{t-1}$ is called the **error correction term**.

Vector Error Correction Model (con't)

- **Remarks:**
  1. In a VECM, past values of the error correction term $Y_t - \theta X_t$ help to predict future values of ΔY_t and/or ΔX_t.
  2. Note also that a VAR model in first differences is misspecified in this case precisely because it omits the error correction term.
- In the case where θ is known; set $Z_{t-1} = Y_{t-1} - \theta X_{t-1}$, and we have

$$\begin{aligned}
\Delta Y_t &= \beta_{10} + \beta_{11} \Delta Y_{t-1} + \cdots + \beta_{1p} \Delta Y_{t-p} \\
&\quad + \gamma_{11} \Delta X_{t-1} + \cdots + \gamma_{1p} \Delta X_{t-p} \\
&\quad + \alpha_1 Z_{t-1} + u_{1t},
\end{aligned}$$

$$\begin{aligned}
\Delta X_t &= \beta_{20} + \beta_{21} \Delta Y_{t-1} + \cdots + \beta_{2p} \Delta Y_{t-p} \\
&\quad + \gamma_{21} \Delta X_{t-1} + \cdots + \gamma_{2p} \Delta X_{t-p} \\
&\quad + \alpha_2 Z_{t-1} + u_{2t},
\end{aligned}$$

so that the parameters of the VECM can be estimated by linear least squares in this case.

- In the case where θ is unknown; then, the VECM is nonlinear in parameters, so that one cannot directly apply linear least squares.
- In this case, there are a few different approaches to estimating the parameters of a VECM.

1. **Approach 1: Two-step procedure.**
   (i) **Step 1:** Estimate θ by a preliminary OLS regression

$$Y_t = \alpha + \theta X_t + Z_t$$

   and obtain the residual $\widehat{Z}_{t-1} = Y_{t-1} - \widehat{\theta} X_{t-1}$.

(ii) **Step 2:** Plug $\widehat{Z}_{t-1}$ into the VECM specification to obtain

$$
\begin{aligned}
\Delta Y_t &= \beta_{10} + \beta_{11}\Delta Y_{t-1} + \cdots + \beta_{1p}\Delta Y_{t-p} \\
&\quad + \gamma_{11}\Delta X_{t-1} + \cdots + \gamma_{1p}\Delta X_{t-p} \\
&\quad + \alpha_1 \widehat{Z}_{t-1} + \widehat{u}_{1t},
\end{aligned}
$$

$$
\begin{aligned}
\Delta X_t &= \beta_{20} + \beta_{21}\Delta Y_{t-1} + \cdots + \beta_{2p}\Delta Y_{t-p} \\
&\quad + \gamma_{21}\Delta X_{t-1} + \cdots + \gamma_{2p}\Delta X_{t-p} \\
&\quad + \alpha_2 \widehat{Z}_{t-1} + \widehat{u}_{2t},
\end{aligned}
$$

The remaining parameters of the VECM can then be estimated by linear least squares. Note that this approach exploits the fact $\widehat{\theta}$ is a consistent estimator of θ if the assumption of cointegration is correct. Moreover, rate of convergence for this estimator is T which is faster than the usual $\sqrt{T}$ convergence rate.

Vector Error Correction Model (con't)

2. **Approach 2:** A more efficient approach is to estimate all the parameters θ, $\left(\beta_{10}, ..., \beta_{1p}, \beta_{20}, ..\beta_{2p}\right)$, $(\gamma_{11}, ..., \gamma_{1p}, \gamma_{21}, ..., \gamma_{2p})$, and (α_1, α_2) in the model

$$\begin{aligned}\Delta Y_t &= \beta_{10} + \beta_{11}\Delta Y_{t-1} + \cdots + \beta_{1p}\Delta Y_{t-p} \\ &\quad +\gamma_{11}\Delta X_{t-1} + \cdots + \gamma_{1p}\Delta X_{t-p} \\ &\quad +\alpha_1 \left(Y_{t-1} - \theta X_{t-1}\right) + u_{1t},\end{aligned}$$

$$\begin{aligned}\Delta X_t &= \beta_{20} + \beta_{21}\Delta Y_{t-1} + \cdots + \beta_{2p}\Delta Y_{t-p} \\ &\quad +\gamma_{21}\Delta X_{t-1} + \cdots + \gamma_{2p}\Delta X_{t-p} \\ &\quad +\alpha_2 \left(Y_{t-1} - \theta X_{t-1}\right) + u_{2t}.\end{aligned}$$

jointly by full system maximum likelihood. This is the approach that has been developed by Soren Johansen (see Johansen 1988, 1991).

Models of Conditional Heteroskedasticity - Motivation

- Consider again the $AR(1)$ model

$$Y_t = \beta Y_{t-1} + u_t,$$

where $|\beta| < 1$ and $\{u_t\} \equiv i.i.d.\left(0, \sigma^2\right)$.

- Note that for this model

$$E\left[Y_{t+1}\right] = 0$$

but

$$E\left[Y_{t+1} | Y_t, Y_{t-1}, ...\right] = E\left[Y_{t+1} | Y_t\right] = \beta Y_t,$$

so that by using information about current and past values of Y_t, this model allows one to improve on ones forecast of the mean-level of Y_{t+1} over that which can be obtained when this information is not used.

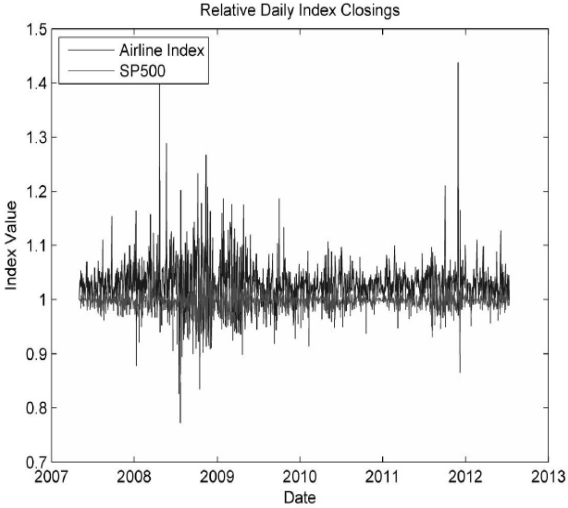
Models of Conditional Heteroskedasticity - Motivation

- **Shortcoming of this model:** The same improvement is not achieved when forecasting the error variance with this model since

$$E\left[u_{t+1}^2 | Y_t, Y_{t-1}, ...\right] = E\left[u_{t+1}^2\right] = \sigma^2$$

- **Observation:** This model is not rich enough to allow for better prediction of the error variance based on past information. In particular, the independence assumption on the errors precludes any forecast improvement.
- On the other hand, many financial and macroeconomic time series exhibit "volatility clustering." Volatility clustering suggests the possible presence of time dependent variance or time-varying heteroskedasticty that may be forecastable. Interestingly, this can occur even if the time series itself is close to being serially uncorrelated so that the mean-level is difficult to forecast.

Models of Conditional Heteroskedasticity - Empirical Motivation



Why would there be interest in forecasting variance?

- First, in finance, the variance of the return to an asset is a measure of the risk of owning that asset. Hence, investors, particularly those who are risk averse, would naturally be interested in predicting return variances.
- Secondly, the value of some financial derivatives, such as options, depends on the variance of the underlying assets. Thus, an options trader would want to obtain good forecasts of future volatility to help her or him decide on the price at which to buy or sell options.
- Thirdly, being able to forecast variance could allow one to have more accurate forecast intervals that adapt to changing economic conditions.

AutoRegressive Conditional Heteroskedasticity (ARCH) Models

- Here, we will discuss two frequently used models of time-varying heteroskedasticity: the **autoregressive conditional heteroskedasticity (ARCH)** model and its extension, the **generalized ARCH (or GARCH)** model.
- **ARCH(**1) process**:** Consider the ADL(1,1) regression

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t.$$

Instead of modeling $\{u_t\}$ as an independent sequence of random variables, as we have before, the ARCH**(**1) process takes

$$u_t = \varepsilon_t \left[\alpha_0 + \alpha_1 u_{t-1}^2\right]^{1/2},$$

where $\alpha_0 > 0$, $0 < \alpha_1 < 1$, and $\{\varepsilon_t\} \equiv i.i.d.N(0,1)$.

- **Remark:** We have described here a ADL(1,1) model with ARCH errors; but, in principle, an ARCH process can be applied to model the error variance for any time series regression.

- Some Moment Calculations
  - (i) **Conditional Mean:**

$$\begin{aligned} E\left[u_t | u_{t-1}, u_{t-2}, ...\right] &= \left[\alpha_0 + \alpha_1 u_{t-1}^2\right]^{1/2} E\left[\varepsilon_t | u_{t-1}, u_{t-2}, ...\right] \\ &= \left[\alpha_0 + \alpha_1 u_{t-1}^2\right]^{1/2} E\left[\varepsilon_t\right] \\ &= 0 \end{aligned}$$

  - (ii) **Unconditional Mean:**

$$\begin{aligned} E\left[u_t\right] &= E\left(E\left[u_t | u_{t-1}, u_{t-2}, ...\right]\right) \\ & \quad \text{(by law of iterated expectations)} \\ &= E\left[0\right] \\ &= 0. \end{aligned}$$

ARCH Models

- (iii) **Conditional Variance:**

$$
\begin{aligned}
E\left[u_t^2 | u_{t-1}, u_{t-2}, ...\right] &= \left[\alpha_0 + \alpha_1 u_{t-1}^2\right] E\left[\varepsilon_t^2 | u_{t-1}, u_{t-2}, ...\right] \\
&= \left[\alpha_0 + \alpha_1 u_{t-1}^2\right] E\left[\varepsilon_t^2\right] \\
&= \left[\alpha_0 + \alpha_1 u_{t-1}^2\right]
\end{aligned}
$$

(iv) **Autocovariances:** Let j be any positive integer, and note that

$$
\begin{aligned}
E\left[u_t u_{t-j}\right] &= E\left(u_{t-j} E\left[u_t | u_{t-1}, u_{t-2}, ...\right]\right) \\
&\quad \text{(by law of iterated expectations)} \\
&= E\left[u_{t-j} \times 0\right] \\
&= 0.
\end{aligned}
$$

- **Remark:** Interestingly, an ARCH process is serially uncorrelated but not independent. These features are important for the modeling of asset returns.

- **More Moments:** It can also be shown that

$$
\begin{aligned}
Var\left(u_t\right) &= E\left[u_t^2\right] = \frac{\alpha_0}{1-\alpha_1} , \\
E\left[u_t^4\right] &= \left[\frac{3\alpha_0^2}{\left(1-\alpha_1\right)^2}\right]\left[\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right].
\end{aligned}
$$

(**Note:** we assume that $\alpha_0 > 0$ and $0 < \alpha_1 < 1$).

- **Remark:** Note that since

$$\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 1,$$

we have that

$$E\left[u_t^4\right] = \left[\frac{3\alpha_0^2}{(1-\alpha_1)^2}\right]\left[\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right] > \frac{3\alpha_0^2}{(1-\alpha_1)^2} = 3\left(E\left[u_t^2\right]\right)^2.$$

On the other hand, if u_t had been normally distributed, say $\{u_t\} \equiv i.i.d.N\left(0,\sigma^2\right)$; then, we would have $E\left[u_t^4\right] = 3\left(E\left[u_t^2\right]\right)^2 = 3\sigma^4$. Hence, the ARCH error process has "fatter-tails" than that implied by the normal distribution.

ARCH Models

- **ARCH(p) process:** A straightforward extension of the ARCH(1) model is the p-th order ARCH process given by

$$u_t = \varepsilon_t \left[\alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2\right]^{1/2},$$

where $\{\varepsilon_t\} \equiv i.i.d.N(0,1)$; $\alpha_i > 0$ for $i = 0, 1, ..., p$; and

$$\alpha_1 + \cdots + \alpha_p < 1.$$

GARCH Models

- **GARCH(p,q) process:** A useful generalization of the ARCH model is the following GARCH model due to Bollerslev (1986).

$$u_t = h_t^{1/2} \varepsilon_t,$$

where

$$h_t = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + \delta_1 h_{t-1} + \cdots + \delta_q h_{t-q}.$$

- Assumptions:
  (i) $\{\varepsilon_t\} \equiv i.i.d.N(0,1)$;
  (ii) $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, ..., p$;
  (iii) $\delta_j \geq 0$ for $j = 1, ..., q$.
- **Remark:** Note that even a GARCH(1,1) model will allow h_t to depend on u_t^2 from the distant past. Thus, GARCH provides a clever way of capturing slowly changing variances without having to specify a model that has a lot of parameters to estimate.
- **Remark:** Both ARCH and GARCH can be estimated using the method of maximum likelihood.

Empirical Illustration

- A simple model of stock return with time-varying volatility is the following

$$R_t = \mu + u_t$$

where $\{u_t\}$ follows a GARCH(1,1) process, i.e.,

$$\begin{aligned} u_t &= h_t^{1/2}\varepsilon_t, \\ h_t &= \alpha_0 + \alpha_1 u_{t-1}^2 + \delta_1 h_{t-1}. \end{aligned}$$

- The textbook provides empirical results of fitting this model to daily percentage changes in the NYSE index using data on all trading days from January 2, 1990 to November 11, 2005. The results are

$$\begin{aligned} \widehat{R}_t &= \widehat{\mu} = \underset{(0.012)}{0.049} \\ \widehat{h}_t &= \underset{(0.0014)}{0.0079} + \underset{(0.005)}{0.072} u_{t-1}^2 + \underset{(0.006)}{0.919} h_{t-1} \end{aligned}$$