1 Today’s Agenda

1. Multivariate Systems (VAR): Reduced forms
2. A Basic Structural model: Consumption-based model
3. Risk Return Trade-Off (ICAPM)
4. The Importance of a Model in Risk-Return relation
5. Volatility Modelling: ARCH/GARCH, Realized Volatility, Stochastic Volatility, MIDAS
6. Next: GMM

Aside: Monetary Policy and the Taylor Rule
2 Vector Autoregressions (VAR)

• First, we should not confuse VAR (vector autoregression) with VaR (Value at Risk).

• The VAR is a natural generalization of the autoregressive process, for multivariate series.

• Suppose we are interested in the joint, dynamic interaction between a few series, say returns and volatility.
  – We suppose that higher volatility must lead to higher returns
  – We also think that there might be some feedback effect from returns to volatility Campbell and Hentschel (1992).
  – Ultimately, we are not sure about the dynamic relationship between those two series
Then, we write

\[ Y_t = \begin{bmatrix} r_t \\ \sigma_t \end{bmatrix} \]

where \( r_t \) is the return and \( \sigma_t \) is its volatility (standard deviation) at time \( t \). We want to write the following dynamic relationship

\[ r_t = \alpha_1 + \beta_{11} r_{t-1} + \beta_{12} \sigma_{t-1} + \varepsilon_{1,t} \]
\[ \sigma_t = \alpha_2 + \beta_{21} r_{t-1} + \beta_{22} \sigma_{t-1} + \varepsilon_{2,t} \]

- Both series depend on their own lagged values (as in the AR process)
- Both series depend on the lagged realization of the other process
- The residuals \([\varepsilon_{1t} \varepsilon_{2t}]\) have a covariance matrix \( \Sigma \).
- This system can be written in a more elegant form as:

\[ Y_t = \alpha + \Phi Y_{t-1} + \varepsilon_t \]

where \( \Phi = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \) and \( \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \).
The beauty of VARs

• Very easy to estimate: Run regressions of variables on their lagged values and on the lagged values of all other variables

• Run the regressions, equation by equation, and then stack them together.

• We don’t need a priori theory to know what variable causes what variable, etc.

• The VARs help us make the notion of statistical “causality” very precise.
Statistical (Granger) causality:
- Example: Martian observes that when people bring umbrellas, it starts raining. She concludes that people bringing umbrellas causes rain.
- But it is the other way around...The anticipation of rain makes people bring their umbrellas.
- In this example, the statistical causality runs from “umbrellas” to “rain”
- The structural causality runs from “rain” to “umbrellas”.
- Weather and orange juice futures example. – argument for market efficiency.
- Problem: In empirical work, if we don’t have a model (or more information about the system), we cannot distinguish between the two alternatives.
- The OJ futures example: OJ futures forecast the weather, the weather does not forecast OJ futures prices
- Is this evidence for or against EMH?
Causality in a bivariate VAR:
\[
    r_t = \alpha_1 + \beta_{11} r_{t-1} + \beta_{12} \sigma_{t-1} + \varepsilon_{1,t}
\]
\[
    \sigma_t = \alpha_2 + \beta_{21} r_{t-1} + \beta_{22} \sigma_{t-1} + \varepsilon_{2,t}
\]

- Suppose that $\beta_{12} = 0$. Then $r_t = \alpha_1 + \beta_{11} r_{t-1} + \varepsilon_{1,t}$, or $r_t$ is not affected by $\sigma_{t-1}$. In such a case, we say that $\sigma_{t-1}$ does not Granger-cause (or just cause) $r_t$.

- However, since $\beta_{21} \neq 0$, $r_{t-1}$ does Granger-cause $\sigma_t$.

- It is important to understand that Granger-causality gives us the timing (umbrella, then rain) but not the economic story.

- This is a very common mistake in academia, in practice, and in everyday life.
Test for Granger-causality:

- We run the two regressions.
- For the hypothesis: \( \sigma_{t-1} \) Granger-causes \( r_t \), we test \( \beta_{12} = 0 \)
- For the hypothesis: \( r_{t-1} \) Granger-causes \( \sigma_t \), we test \( \beta_{21} = 0 \)
- It is always a good idea to start an empirical work with some background Granger-causality tests, be it only to get a feel for the data.
- But it is not a good idea to start weaving economic stories based on that evidence alone.
- We will come back to VARs.
Here is a real research problem:

- **Fact:** The effect of the FED actions on the stock market is negative. In other words, a contractionary monetary policy (aimed at curbing inflation) will result in lower returns for some time in the future.

- **Question:** Why is this so? There are two possibilities
  - Fed has a better forecast of the state of the economy, but its policy has no real effect on stock fundamentals (the umbrella). This is only Granger-causation without structural effect.
  - Fed has an impact on the economy and by contracting the economy, cash flows go down, returns decrease (the cloud). This is Granger-causation and a structural effect.

- **Q:** Is there something about the stock market that would help us distinguish between the two alternatives?
3 A Basic Structural model: Consumption-based model (CCAPM)

- So far, we have estimated the APT and the CAPM
- The CAPM and the APT capture risk and return, but are they related to our more fundamental needs: consumption of goods.
- Some of you have asked me to clarify: What do you mean by “The equity premium puzzle is too high”?
- We will work out a “simple” model where assets are priced explicitly relative to our utility from consumption.
- This explicit model will generate a familiar stochastic discount factor pricing relation.
• First, we have to model the behavior of a representative investor
• Think of this investor as the average person in the economy.
• The investor invests primarily so he will consume goods (bread, cheese, or Ferraris) tomorrow.
• The utility function of this investor, today and tomorrow is:
  \[ U(c_t, c_{t+1}) = u(c_t) + \beta u(c_{t+1}) \]
  where \( c_t \) denotes consumption at date \( t \)
• \( \beta \) is a subjective discount factor, usually \( \beta < 1 \).
• Note: Some people have argued that \( \beta \geq 1 \) (Behavioral finance—fitness club experiments).
• Ultimately, we won’t be estimating \( \beta \). Set \( \beta = 0.995 \).
• At the beginning, we will work with $u(c_t)$, where
  – $u(.)$ is concave, reflecting a decreasing marginal value of consumption
  – $u(.)$ is increasing, reflecting a insatiable desire for more consumption
  – the curvature generates aversion to risk and to inter-temporal substitution: The investor prefers a consumption stream that is steady over time and across states of nature.
• But to make the model operational (estimable), we have to give a functional form to $u(.)$.
• We will assume that
  $$u(c_t) = \frac{1}{1 - \gamma} c_t^{1-\gamma}$$
  where $\gamma$ captures the curvature.
Here is the game:
- We want to value an asset with an uncertain payoff $x_{t+1}$
- It is a two-period problem (today/tomorrow or young/old, etc.)
- The problem is:
  \[
  \max_{\zeta} u(c_t) + E_t [\beta u(c_{t+1})]
  \]
  such that:
  \[
  c_t = e_t - p_t \zeta \\
  c_{t+1} = e_{t+1} + x_{t+1} \zeta
  \]
  - $e_t$ is the original endowment of the individual (cash he inherited from his parents)
  - At time $t$, he has an endowment $e_t$.
  - He decided to purchase $\zeta$ shares of the asset at price $p_t$
  - Whatever is left over after the purchase of the asset $e_t - p_t \zeta$, is used for consumption.
  - At time $t + 1$, we has an endowment $e_{t+1}$ but also the payoff from the asset $x_t$ (think $p_{t+1} + d_{t+1}$) time the number of shares.
  - Since the individual “dies” at $t + 1$, he must consume everything, so $c_{t+1} = e_{t+1} + x_{t+1} \zeta$.
  - No bequest–Can’t analyze inheritance here.
To recapitulate: We want to find the number of shares that would be bought at price $p_t$, given that the payoff from this investment is $x_{t+1}$.

Implicitly, we will find the price of the asset as a function of the payoff and everything else.

This is called an equilibrium model.

This is also a structural model.

Solving the maximization problem yields the FOC:

$$p_t u'(c_t) = E_t [\beta u'(c_{t+1}) x_{t+1}]$$

or

$$p_t = E_t \left[ \beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right]$$
The equation
\[ p_t = E_t \left[ \beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right] \]

is quite interesting.

1. Note that if we write \( m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)} \) and interpret \( m_{t+1} \) as the stochastic discount factor, then we get the familiar pricing equation, where \( m_{t+1} \) is a function of consumption. Hence, this is called a consumption-based pricing model.
\[ p_t = E_t [m_{t+1} x_{t+1}] \]

2. Note that we cannot go beyond this point without specifying a functional form for \( u() \), and hence, for \( u'(c) = c^{-\gamma} \) and the FOC is:
\[ p_t = E_t \left[ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} x_{t+1} \right] \]

3. Now, we can rewrite the model as:
\[ 1 = E_t \left[ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} \frac{x_{t+1}}{p_t} \right] = E_t \left[ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + \frac{1}{1 + R_{t+1}}) \right] \]

4. If we have data on \( R_{t+1} \) and on \( c_t \), we can think of a way to estimate the parameters \( \gamma \) and \( \beta \).

5. Problem: The above relationship is nonlinear...we only know how to run linear regressions....
Here is a good insight: The equation

\[ 1 = E_t \left[ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + R_{t+1}) \right] \]

must hold for any asset. Indeed, we did not specify a particular asset when deriving the above equations.

- Let’s consider a particular asset, the risk free asset with return \( R^f \) (and forget uncertainty and expectations for a moment)

- The pricing equation can be rewritten as

\[ (1 + R^f) = \frac{1}{\beta} \left( \frac{c_{t+1}}{c_t} \right)^\gamma \]

- Take logs:

\[ \ln(1 + R^f) = r^f = -\ln \beta + \gamma (\ln c_{t+1} - \ln c_t) \]

- Note that \( \ln c_{t+1} - \ln c_t \) is nothing but the growth in consumption between \( t + 1 \) and \( t \).

- We can estimate the above equation if we have data on \( r^f \) and \( c_t \).

- What is the interpretation of \( \gamma \)?
Now, we have to take care of the uncertainly. For that we will use log-normality.

If \( X \) is conditionally lognormally distributed, it has the convenient property

\[
\ln E_t (X) = E_t (\ln(X)) + \frac{1}{2} Var_t (\ln(X))
\]

Recall our discussion of: \( g(E(X)) \neq E(g(X)) \). In the above example, \( g() = \ln() \).

In addition, we will assume that \( Var_t (\ln(X)) = Var (\ln(X)) = \sigma^2_x \). (What is this assumption called?)
Recall

\[ 1 = E_t \left[ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + R_{t+1}) \right] \]

Taking logs

\[ 0 = \ln E_t \left[ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + R_{t+1}) \right] \]

\[ = E_t \left[ \ln \left( \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + R_{t+1}) \right) \right] + \frac{1}{2} \text{Var} \left( \ln \left( \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + R_{t+1}) \right) \right) \]

\[ = E_t \left[ \ln (1 + R_{t+1}) + \ln \beta - \gamma (\ln c_{t+1} - \ln c_t) \right] + \frac{1}{2} \text{Var} \left( \ln (1 + R_{t+1}) + \ln \beta - \gamma (\ln c_{t+1} - \ln c_t) \right) \]

\[ = E_t r_{t+1} + \ln \beta - \gamma E_t \left[ (\ln c_{t+1} - \ln c_t) \right] + \frac{1}{2} \left[ \sigma_r^2 + \gamma^2 \sigma_{\Delta c}^2 - 2\gamma \sigma_{r, \Delta c} \right] \]
The equation
\[ 0 = E_t r_{t+1} + \ln \beta - \gamma E_t [(\ln c_{t+1} - \ln c_t)] + \frac{1}{2} \left[ \sigma_r^2 + \gamma^2 \sigma_{\Delta c}^2 - 2\gamma \sigma_{r,\Delta c} \right] \]
is valid for any asset. It holds for the risk free asset, where \( \sigma_{rf}^2 = \sigma_{rf,\Delta c} = 0 \). Therefore, we can write
\[ r^f = - \ln \beta + \gamma E_t [(\ln c_{t+1} - \ln c_t)] - \frac{1}{2} \gamma^2 \sigma_{\Delta c}^2 \]
Comparing this formula to what we had before, the term \(-\frac{1}{2} \gamma^2 \sigma_{\Delta c}^2\) adjusts for the variance, or uncertainly in consumption, provided that all processes are lognormal.
• The advantage of this formula is that we can handle uncertainty and any other asset.

• From

\[ E_t r_{t+1} = - \ln \beta + \gamma E_t [\ln c_{t+1} - \ln c_t] - \frac{1}{2} \left[ \sigma_r^2 + \gamma^2 \sigma_{\Delta c}^2 - 2 \gamma \sigma_r \Delta c \right] \]

• And

\[ r^f = - \ln \beta + \gamma E_t [\ln c_{t+1} - \ln c_t] - \frac{1}{2} \gamma^2 \sigma_{\Delta c}^2 \]

we obtain the cool expression

\[ E_r r_{t+1} - r^f = - \gamma \sigma_r \Delta c - \frac{\sigma_r^2}{2} \]

\[ = - \gamma Cov (r_{t+1}, (\ln c_{t+1} - \ln c_t)) - \frac{Var (r_{t+1})}{2} \]

• In words, the excess return is equal to the covariance of the asset return with consumption growth.

• What is this result reminiscent of?

• This is called the Consumption CAPM (or CCAPM)?

• But then:
  – The CAPM does not hold. Does CCAPM hold?
  – Why is this model better? (before \( \beta \), now \( \gamma \))
• The CCAPM is as successful as the CAPM, or even less, but
  – The coefficient $\gamma$ has a very nice interpretation: It measures our aversion to risk.
  – We have a consumption variable in the pricing kernel.
  – To test the CAPM we needed the market portfolio (Roll’s critique). Similarly, now we need consumption.
  – The CCAPM (with the added log-linearity restrictions) is easy to test using regressions.
• Note that we have two regressions that we can run in order to estimate $\gamma$

• First, using the riskless rate

$$r^f = -\ln \beta + \gamma E_t [(\ln c_{t+1} - \ln c_t)] - \frac{1}{2} \gamma^2 \sigma^2_{\Delta c}$$

• Second, using the risky rate

$$E_r r_{t+1} - r^f = -\gamma \sigma_{r,\Delta c} - \frac{\sigma^2_r}{2}$$

• Note that both equations must give us the same result (statistically speaking, at least).

• The trouble is that the estimates of $\gamma$ in those regressions are in total disagreement.

• Either the risky rate has been “too high” or the riskless rate has been “too low” to reconcile the model with the data.

• What’s next: The assumption that the risk premium $E_r r_{t+1} - r^f$ does not vary with time has been seen, lately, as being particularly bothersome.

• A few models have tried to relax this assumption, while keeping the economic story of the model.
4 Risk Return Trade-Off (ICAPM)

  \[ E_{t-1} R_{M,t} = \gamma \sigma^2_{M,t} + \eta \sigma_{M,F,t} \]

- One of the main tenets of modern finance is that we have to be compensated with higher returns if we are exposed to higher risk.

- We measure risk with the variance.

- So far, we have assumed that \( \text{Var}(r_t) = \sigma^2 \).

- This is clearly an untenable assumption.

- There is no reason why the variance of return will stay constant.

- Why does the variance fluctuate (Schwert (1990))?
  - Macroeconomic factors
  - Microeconomic factors
  - Who knows why?

- Bottom line: \( \text{Var}(r_t) = \sigma^2_t \) (What does this imply for returns?)

- Modelling \( \sigma_t \) will be a separate discussion.
• Suppose we have an estimate of $\sigma_t$.

• We can run the regression

$$r_t = \alpha + \beta\sigma_{t-1} + \varepsilon_t$$

• Under the null hypothesis that there is a risk-return trade-off, we expect $\beta > 0$.

• If the relationship does not hold, there are several possibilities:
  – The relationship is not well specified (perhaps see VAR)
  – The volatility is not well estimated
  – There are other variables that must enter into a VAR (investment opportunity set is time-varying).
  – A combination of the above

• Surprisingly, thus far, the evidence for $\hat{\beta} > 0$ is non-existent.

• Several papers find $\hat{\beta} < 0$. 
A naive way to estimate $\sigma_t$ (it is unobservable)

- Note that before we can estimate the conditional second moment, we have to estimate the conditional first moment. A lot of people forget to specify the conditional first moment and get garbage. From the definition, \( \text{Var}_t (r_{t+1}) = E_t \left( (r_{t+1} - E_t r_{t+1})^2 \right) = E_t \left( r_{t+1}^2 \right) - (E_t (r_{t+1}))^2 \)

- Here are the steps.
  - Run daily returns on lagged daily returns and lagged dividend yields. The residuals from this regression will be
    \[ \hat{e}_t = r_t - \hat{E}_{t-1} (r_t) \]
    where $\hat{e}$ denotes the residual from the OLS regression.
  - Using the daily data, we can form an estimate for the monthly volatility as in:
    \[ \hat{\sigma}_t^2 = \left( \frac{1}{n} \sum_{i=1}^{n} \hat{e}_i^2 \right) \times 22 \]
    where $i = 1, \ldots, n$ are the daily observations within month $t$.
  - This is a particularly simple way to estimate the monthly $\sigma_t^2$. It is non-parametric (no parameters to estimate, just $\sigma_t$).
• If we plot $\hat{\sigma}_t^2$ over time, it becomes immediately clear that $\hat{\sigma}_t^2$ is not white noise, but follows a certain process.

• Recall, how we characterized the levels $Y_t = \phi Y_{t-1} + \varepsilon_t$.

• We want to find a similar model for the volatility, something like $\hat{\sigma}_t^2 = \phi \hat{\sigma}_{t-1}^2 + \varepsilon_t$.

• But a simple AR model would not quite work. Why?

• Alternatives:
  – Model $\sigma_t$ [GARCH/Realized Volatility/Stochastic Volatility]
  – Model $r_t$ and $\sigma_t$ jointly [GARCH-M and leverage literature]
  – Get implied $\sigma_t$ from options—careful as $\sigma_t^{IMPLIED} \neq \sigma_t^{HISTORICAL}$ [Why?]
5 The Importance of a Model: The Gordon-Growth Model

• Recall the definition of return:

\[ R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} - 1 \]

which can be re-written, if we assume that

\[ E_t R_{t+1} = R, \]

as:

\[ P_t = E_t \left[ \frac{P_{t+1} + D_{t+1}}{1 + R} \right] \]

\[ = E_t \left[ \sum_{i=1}^{K} \left( \frac{1}{1 + R} \right)^i D_{t+i} \right] + E_t \left[ \left( \frac{1}{1 + R} \right)^K P_{t+K} \right] \]

\[ \approx E_t \left[ \sum_{i=1}^{\infty} \left( \frac{1}{1 + R} \right)^i D_{t+i} \right] \]

Note:

\[ E_t P_{t+1} = (1 + R) P_t - E_t D_{t+1} \]

1. A constant expected return, as assumed, does not imply that \( P_t \) would follow a martingale.

2. To obtain a martingale, we must construct a portfolio for which all dividend payments are re-invested in the stock. The value of this portfolio is a martingale (Campbell et al. p. 257)!
Now, suppose that $D_{t+1} = (1 + G) D_t + \varepsilon_t$, where $\varepsilon_t$ is iid. Then $E_t D_{t+i} = (1 + G)^i D_t$ and

$$P_t = E_t \left[ \sum_{i=1}^{\infty} \left( \frac{1}{1+R} \right)^i D_{t+i} \right]$$

$$= \frac{(1 + G) D_t}{R - G} = \frac{E_t D_{t+1}}{R - G}$$

This is called the Gordon-Growth model.

Note the unrealistic assumptions: $E_t R_{t+1} = R$ and $G$ is constant.

This is how people think about prices.

But the definition seems circuitous: Prices are determined by the discount rate, what is the discount rate determined by?
• But it must be the case that (I am cheating a bit...changing assumptions):

\[ E_t R_{t+1} = a + \gamma \sigma_t^2 \]

• Hence:

\[ P_t \approx \frac{E_t D_{t+1}}{E_t R_{t+1} - G} \]

• Q: Why is this useful?
• A: Suppose \( \alpha = 0.08 \) and \( \gamma = 2.5 \), and \( \sigma_t^2 = 0.15^2 \). Then \( E_t R_{t+1} = 0.1362 \).

• Suppose that \( E_t D_{t+1} = 1 \), and \( G = 0 \). Then:

\[
P_t \approx \frac{1}{E_t R_{t+1}} = \frac{1}{0.1362} = 7.34
\]

• Now, suppose that volatility increases at \( t + 1 \), or \( \sigma_{t+1}^2 = 0.25^2 \). Hence, \( E_{t+1} R_{t+2} = 0.2363 \).

This implies that:

\[
P_{t+1} \approx \frac{1}{0.2363} = 4.23
\]

• What happened to \( R_{t+1} \) (no dividend got paid)?

\[
R_{t+1} = \frac{P_{t+1} - P_t}{P_t} = \frac{4.23 - 7.34}{4.23} < 0
\]

• MORAL: \( COV \left( R_{t+1}, \sigma_{t+1}^2 \right) < 0 \).

• BUT: \( COV(R_{t+1}, \sigma_t^2) > 0 \), on average.

• ALSO: \( COV \left( R_{t+1}, \Delta \sigma_{t+1}^2 \right) < 0 \).
- When you look for a positive risk-return trade-off, you must look at the correct lead-lag relationship.
- You should not look at the contemporaneous relationship.
- Forecasting is not the same as estimation
6 Autoregressive Conditional Heteroskedasticity (ARCH)

- Suppose we have the returns process $\{r_t\}_{t=1}^T$.
- First, we model the conditional mean, for example as
  \[
  r_t = \mu + \phi r_{t-1} + u_t
  \]
- We know that the unconditional first moments are $E(r_t) = \mu \frac{1}{1-\phi}$ and $Var(r_t) = \frac{\sigma_u^2}{1-\phi^2}$.
- We also know that the conditional first moment $E_{t-1}(r_t) = \mu + \phi r_t$ is time varying, even though the unconditional moment is not!
- Q: Can we also have the same situation for the second conditional moment, i.e. to have a time-varying conditional second moment, although the unconditional second moment constant over time?
- A: Yes.
- Note: The unconditional second moment of $u_t$ is $\sigma_u^2$. 
• Intuitively, we want to model $u_t^2$ to follow an AR process just as we had $r_t$ follow an AR process.

• We can write such a process as
  
  $$u_t^2 = \zeta + \alpha u_{t-1}^2 + w_t$$

  where $w_t$ is a white noise process with $E(w_t^2) = \sigma_w^2$.

• Note that $E_{t-1}(u_t^2) = \zeta + \alpha u_{t-1}^2$, so that the conditional moment is time-varying although $E(u_t^2) = \sigma_u^2$.

• The above process is called an autoregressive conditional heteroskedasticity model of order 1, or ARCH(1).

• We can generalize to an ARCH(p) model as:

  $$u_t^2 = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_p u_{t-p}^2 + w_t$$
• Note that a variance cannot be negative. We need to place certain restrictions on
\[ u_t^2 = \zeta + \alpha u_{t-1}^2 + w_t \]
in order to insure that \( u_t^2 \) is always positive.

• We need:
  – \( w_t \) to be bound from below by \(-\zeta\), where \( \zeta > 0 \).
  – \( \alpha \geq 0 \).
  – For covariance stationarity of \( u_t^2 \), we also need \( \alpha < 1 \) as in the other AR model.

• With all those conditions, we can see that
\[ \text{Var} (u_t) = E (u_t^2) = \sigma_u^2 = \frac{\zeta}{1 - \alpha} \]
This is the ARCH(1) model and its historical relation to what we have done before.

It is more convenient, but less intuitive, to present the ARCH(1) model as:

\[ u_t = \sqrt{h_t} v_t \]

where \( v_t \) is iid with mean 0, and \( E \left( v_t^2 \right) = 1 \).

Suppose that

\[ h_t = \zeta + au_{t-1}^2 \]

then combining the above equations, we obtain:

\[ u_t^2 = h_t v_t^2 \]

Now, since \( v_t \) is iid then

\[
E_{t-1} \left( u_t^2 \right) = E_{t-1} \left( h_t^2 v_t^2 \right) \\
= E_{t-1} \left( h_t^2 \right) E_{t-1} \left( v_t^2 \right) \\
= \zeta + au_{t-1}^2
\]

as before.
• Reconciling the two definitions.
• From one side, we have $u_t^2 = h_tv_t^2$
• From another side, we have $u_t^2 = h_t + w_t$. Therefore $h_tv_t^2 = h_t + w_t$
  or

\[ w_t = h_t \left( v_t^2 - 1 \right) \]

• From here, we can see that $E_{t-1} (w_t^2)$ is time varying, whereas $E (w_t^2) = \sigma_w^2$

• Note: The unconditional second moment of $w_t$ (the unconditional fourth moment of $u_t$) does not always exist for an ARCH model. Not a big deal, but might be annoying if we want to look at conditional kurtosis.
• The ARCH process gives us conditional heteroskedasticity, but it turns out that $\sigma^2_u$ is a very persistent process.

• We can capture such a process with an ARCH(p) process $u_t^2 = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + ... + \alpha_p u_{t-p}^2 + w_t$ where $p$ is very large.

• This solution is inefficient. There are too many parameters to estimate!

• What to do? GARCH

• The GARCH, Generalized ARCH allows us to capture the persistence of conditional volatility in a parsimonious way.
• Recall that we could write:

\[ u_t = \sqrt{h_t} v_t \]

where \( h_t = \zeta + a u_{t-1}^2 \) for an ARCH process.

• GARCH: Suppose, we specify \( h_t \) as

\[ h_t = \zeta + \delta h_{t-1} + a u_{t-1}^2 \]

• The direct link between \( h_t \) and \( h_{t-1} \) is exactly what is needed to capture the dependence between \( \sigma_t^2 \) and \( \sigma_{t-1}^2 \).

• A process with \( h_t = \zeta + \delta h_{t-1} + a u_{t-1}^2 \) is called a GARCH(1,1)

• Of course, we can generalize to a GARCH(p,q) as:

\[ u_t^2 = \zeta + \delta_1 h_{t-1} + \delta_2 h_{t-2} + \ldots + \delta_p h_{t-p} + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_q u_{t-q}^2 + \ldots \]
• For most practical purposes a GARCH(1,1) is GREAT.
  – There is a trade-off. You introduce more parameters to capture the accurate dynamics, but there are more parameters to estimate
  – Those parameters have restrictions. The estimation is tricky.
  – Bottom line, for 99% of the applications, GARCH(1,1) does a great job.

• GARCH is successful, because it can capture the persistence in $\sigma_t^2$, which is the most significant feature that needs to be captured.
Another useful model to estimate is the IGARCH model, or integrated GARCH.

The IGARCH(1,1) is a GARCH(1,1) where
\[ \delta + \alpha = 1 \]

If this condition is satisfied, it can be shown that the conditional variance of \( u_t \) is infinite.

The processes \( u_t \) and \( u_t^2 \) are not covariance stationary.

However, the process \( u_t \) is stationary (i.e. its conditional density does not depend on \( t \)).

The IGARCH is important because it captures the important case of a strong dependence that leads to non-stationarity.
Asymmetricities: We can go crazy with an autoregressive specification. For instance, Glosten, Jagannathan, and Runkle (1993) use a specification of the following kind:

\[ \sigma_t^2 = \kappa + \beta \sigma_{t-1}^2 + \alpha u_{t-1}^2 + \eta u_{t-1}^2 1_{\{u_{t-1}>0\}} \]

- This type of an effect is suitable to test the hypothesis that negative surprises increase volatility more than positive surprises. If this hypothesis is true, we expect

  \[ \eta < 0 \]
The GARCH literature has gone crazy chasing after the perfect conditional heteroskedasticity model. Some of the models we have are:

- ARCH in Means
- Exponential GARCH
- Nonlinear GARCH
- Asymmetric GARCH
- Fractionally Integrated GARCH (FIGARCH)
- ABS.ASYMM.FIGARCH ?????
• The interest in forecasting conditional volatility using past volatility has been greater in the applied field more so than in academia.

• The GARCH models are unsatisfactory, from an economic perspective, because
  – Explaining vol with past vol tells us nothing about the underlying economic factors that cause the volatility to move.
  – If a structural break occurs, a recursive model will fail miserably...a structural model might not.
  – In general, the GARCH models are difficult to generalize to a multivariate setting.
• Fact: Volatility is very persistent. It might be more persistent than what is allowed by an exponentially decaying GARCH-type model.

• There is another type of GARCH that plays an increasingly important role in finance. It is called Fractionally integrated GARCH.
6.1 Realized Volatility Estimators

- French, Schwert and Stambaugh (1987): The idea is to use higher frequency data to estimate the variance as:

\[ \sigma^2_t = \frac{1}{k} \sum_{d=1}^{k} \varepsilon_{t+d}^2 \]

- where \( \varepsilon_t \) are measured in days, and we estimate monthly variance.
- This produces a monthly sequence \( \{ \hat{\sigma}^2_t \} \) of estimated variances.
- There is nothing wrong with this scheme.

- Another method: AR model for volatility:

\[ |\varepsilon_t| = \eta + \gamma |\varepsilon_{t-1}| + \nu_t \]

- where the \( \varepsilon_t \) are estimated from a first step procedure.
- There is nothing wrong with this method. It provides another model for stochastic volatility.

- Since we don’t observe true volatility, we can’t really say which method is the best at capturing it.
• GARCH in Mean or GARCH-M models (Engle, Lilien, and Robins (1987)):

\[ r_t = a + bx_{t-1} + c\sigma_t^2 + \varepsilon_t \]

\[ \varepsilon_t = z_t\sigma_t \]

\[ \sigma_t^2 = \kappa + \beta (L) \sigma_{t-1}^2 + \alpha (L) \varepsilon_{t-1}^2 \]

• The difference from the previous models is that the volatility enters also in the mean of the return.

• This is exactly what Merton’s (1973, 1980) ICAPM produces–risk-return tradeoff.

• It must be the case that \( b > 0 \).

• The GARCH-M is estimated with ML or QML.

• The evidence on the risk-return tradeoff is not good.

• French, Schwert and Stambaugh conduct similar tests, but their method is a two-step procedure (inefficient, and potentially problematic.)
6.2 MIDAS estimators—Mixed Data Sampling estimators (Ghysels, Santa-Clara, Valkanov (2004, 2005))

• Idea: Use data at different frequencies to estimate the risk-return tradeoff

\[ R_{t+1} = \alpha + \beta \left( \sum_{d=1}^{D} w_d r_{t-d}^2 \right) + \varepsilon_t \]

• where \( R_t \) is at monthly frequency and \( r_t \) is at daily.

• The weights \( w_d \) sum up to one.

• Given that vol is persistent, there might be many weights to estimate, which would result in inefficient estimators.

• Hence, we parameterize \( w_d (\theta) \) and estimate the shape of the weights.

• There are several advantages:
  – Higher frequency data, i.e. better estimates of vol.
  – Joint estimation of \( \theta, \alpha, \beta \)
  – Flexibility of weights
  – Easy to implement other variables, asymmetries.
6.3 Stochastic volatility models:
\[ r_t = a + br_{t-1} + \varepsilon_t \]
\[ \varepsilon_t = z_t \sigma_t \]
\[ \sigma_t = \kappa + \beta \sigma_{t-1} + v_t \]

- The difference here is that the shocks that govern the volatility are not necessarily \( \varepsilon_t^2 \)’s.
- In fact, if \( \rho = \text{corr}(\varepsilon_t, v_t) \neq 0 \), then the (marginal) distribution of returns will be skewed
  - If \( \rho < 0 \), then \( \text{skew}(r_t) < 0 \).
- This is really a discretization of a continuous-time model, where the mean and the variance follow two OU processes.
- The shocks \( \varepsilon_t \) and \( v_t \) don’t have to be Gaussian (or Gaussian with jumps). For the consistent estimation of the parameters, we only need \( E(\varepsilon_t|I_{t-1}) = 0 \)
- Stochastic vol models can be estimated by NL, MLE (see later) or other methods.