FE670 Algorithmic Trading Strategies

Basic Models and Empirics

Steve Yang

Stevens Institute of Technology

steve.yang@stevens.edu

09/14/2023

Overview

Financial Modeling Fundamentals

Model Return, Volatility and Jumps

Stylized Facts for Asset Returns

Financial Modeling Fundamentals

- Financial engineering is a process of synthesis in the sense that the objective of the engineering process is to construct purposeful artifacts, such as portfolios, investment strategies or derivative products.
- Constructive methodoloties are available only when we arrive at the point where we can optimize, that is, codify our design in terms of variables and express the quality of our design in terms of a goal function defined on the design variables.
  - *Science is analytic:* We have the models to analyze a given system.
  - *Design is a constructive process:* We need to synthesize a design starting from general high-level specification.
  - *Constructive design is performed iteratively:* We make an approximate design and analyze it.
  - *Design automation:* The process of design can be automated only when we arrive at the stage of expressing the design quantitatively in terms of a goal function.

Learning, Theoretical, and Hybrid Approaches to Modeling

There are three basic approaches to finanical modeling: **the learning approach, the theoretical approach, and the learning-theoretical approach**.

- **The learning approach** to financial modeling is in principle a consequence of the diffusion of low-cost high-performance computers. It is based on using a family of models that
  (1) include an unlimited number of parameters and
  (2) can approximiate sample data with high precision.

- Neural networks are a classical example. With an unrestricted number of layers and nodes, a neural network can approximate any function with arbitrary precision.

- However, practice has shown that if we represent sample data with very high precision, we typically obtain poor forecasting performance.

- **The Theoretical Approach** to financial modeling is based on human creativity. In this approach, models are the result of new scientific insights that have been embodied in theories.
  - Laws such as the Maxwell equations of electromagnetism were discovered not through a process of learning but by a stroke of genius.
  - The Capital Asset Pricing Model (CAPM) is the most well-known example of a theoretical model in financial economics.
- **The Hybrid Approach** to financial modeling retains characteristics of both the theoretical and learning approaches. It uses a theoretical foundation to identify families of models but uses a learning approaches to choose the correct model within the family.
  - For example the ARCH/GARCH family of models is suggested by theoretical considerations while the right model is selected through a learning approach that identifies the model parameters.

Biases

- **Survivorship Bias** is exhibited by samples selected on the basis of criteria valid at the last point in the sample population.

  - In the presence of survivorship biases in our data, return processes relative to firms that ceased to exist prior to that date are ignored.
  - For example, while poorly performing mutual funds often close down (and therefore drop out of the sample), better performing mutual funds continue to exist (and therefore remain in the sample).
  - In this situation, estimating past returns from the full sample would result in overestimation due to survivorship bias.

Biases

- **Selection Bias** is an error in choosing the individuals or groups to take part in a scientific study.

- Intrinsic in common indexes such the Russell 1000 universe (large-cap stocks). In order to understand the selection bias, we can apply a selection rule similar to that of the Russell 1000 to artificially generated random walks.
- Assume we have 10,0000 independent random walk price processes, each representing the price of a company's stock, over 1,000 periods using the recursive formula:

$$\begin{aligned} P_i(2) &= (1 + R_i(2)) \times P_i(1) = 1 + 0.007 \times \epsilon_i(2) \\ P_i(3) &= (1 + R_i(3)) \times P_i(2) = (1 + 0.007 \times \epsilon_i(3)) \times (1 + 0.007 \times \epsilon_i(2)) \\ &\dots \\ P_i(n) &= (1 + 0.007 \times \epsilon_i(n)) \times \dots \times (1 + 0.007 \times \epsilon_i(3)) \times (1 + 0.007 \times \epsilon_i(2)) \end{aligned} \tag{1}$$

Pitfalls in Choosing from Large Datasets

STATEMENT: Any statistical test, regardless of its complexity and power, will fail in a certain number of cases simply by chance.

- *For example, pairs trading is based on selecting pairs of stocks that stay close together. Suppose we know that the price paths of two stocks will stay close together. When they are at their maximum distance, we can go long in the stock with the highest value and short in the other stock. When their distance is reduced or changes sign a profit is realized.*
    - Given a large universe of stocks, a pairs trading strategy will look for cointegrate pairs. A typical approach will consist in running a cointegration test on each pair. Actually test can consist of multiple tests that each pair has to pass in order to be accepted as cointegrated.
    - However, a pair can appear cointegrated in a sample period purely by chance. Or a truly cointegrated pair may fail the test.

Pitfalls in Choosing from Large Datasets

To illustrate this phenomenon, let's consider a set of 1,000 artificial arithmetic random walk paths that are 1,000 steps long. Consider that is the sample set there are $(1,000 \times 1,000 - 1,000)/2 = 1,000 \times 999 \times 0.5 = 499,500$ different pairs of processes.

- The random walk is defined by the following recursive equation:

$$P_i(t) = P_i(t-1) + 0.007 \times \epsilon_i(t)$$

where the $\epsilon_i(t)$ are independent draws from a standard normal distribution $N(0,1)$.

  - Using the ADF test at 1% significance level, in run 1, 1.1% pass the cointegration test, in run 2, 0.8%.
  - Using the Johansen test at 99% significance level, in run 1, 2.7% pass the cointegration test, in run 2, 1.9%.
  - Using the Johansen maximum eigenvalue test, in run 1, 1.7% pass the cointegration test, in run 2, 1.1%.

Pitfalls in Selection of Data Frequency

- In financial theory, we have both discrete-time and continuous-time models. For example, the Black-Scholes option pricing equation, under certain assumptions, can be solved in a closed-form format. In other cases, we have to look for numeric solutions.
- Let's look at a discrete-time models. For example a vector autoregressive model of order 1:

$$X_t = AX_{t-1} + E_t$$

  Such a model is characterized by a time step. If the X are returns, the time steps could be days, weeks, or months.
- Given a process that we believe is described by a given model, can we select the time step arbitrarily? Or are different time steps characterized by different models?
- There is no general answer to these questions. Most models currently used are not invariant after time aggregation.

Model Return, Volatility and Jumps

- Time Series Models.
  (1) ARIMA Linear Models
  (2) ARCH-GARCH Variance Models
  (3) Point Process Models
- Advanced Topics
  (1) State-Space Modeling
  (2) Regime Switching and Change-Point Models
  (3) *Neural Network Models - to be covered later*
  (4) *Reinforcement Learning Models - to be covered later*

Stationarity and Ergodicity

- Much statistical inference relies on the law of large numbers (LLN) and central limit theorem (CLT).

  LLN: says that the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed.

  CLT: states that, given certain conditions, the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed

  Financial time series data are by nature dependent, therefore we rely on alternative: *stationarity* and *ergodicity*.

- **Ergodicity:**
  - A time series is *ergodic*, if its local stochastic behavior is (possibly in the limit) independent of the starting point, that is, initial conditions.
  - An ergodic process eventually "forgets" where it started.

- **Stationarity:**
  - A time series r_t is said to be strictly stationary if the joint distribution of $(r_{t_1}, \ldots, r_{t_k})$ is identical to that of $(r_{t_1+t}, \ldots, r_{t_k+t})$ for all t, where k is an arbitrary positive integer and $(t_1, \ldots, t_k)$ is a collection of k positive integers.
  - A time series r_t is said to be weakly stationary if both mean of r_t and the covariance between r_t and $r_{t-\ell}$ are time invariant, where ℓ is an arbitrary integer.

  Strict: distributions are time-invariant
  Weak: first 2 moments are time-invariant
    - $E(r_t) = \mu$, which is a constant.
    - $Cov(r_t, r_{t-\ell}) = \gamma_\ell$, which only depends on ℓ.

What does weak stationarity mean in practice?

Ans: it enables one to make inferences concerning future observations (e.g. prediction)

Weak Stationarity:

- The covariance $\gamma_\ell = Cov(r_t, r_{t-\ell})$ is called lag-ℓ autocovarance of r_t. It has two important properties:
  - $\gamma_0 = Var(r_t)$, which is a constant.
  - $\gamma_{-\ell} = \gamma_\ell$.
- *The second property holds because:*

$$\begin{aligned} Cov(r_t, r_{t-(-\ell)}) &= Cov(r_{t-(-\ell)}, r_t) \\ &= Cov(r_{t+\ell}, r_t) = Cov(r_{t_1}, r_{t_1-\ell}) \end{aligned}$$

,where $t_1 = t + \ell$.

- In practice:
  - It is common to assume that an asset return series is weakly stationary.
  - It can be empirically checked, provided that a sufficient number of historical returns available.

Estimation for a given sample data $\{r_1, ..., r_T\}$:

- Past: time plot of $\{r_t\}$ varies around a fixed level within a finite range.
- Future: the first 2 moments of future $\{r_t\}$ are the same as those of the data so that meaningful inferences can be made.
- Mean (or expectation) of returns:

$$\mu = E(r_t)$$

- Variance (variability) of returns:

$$Var(r_t) = E[(r_t - \mu)^2]$$

- Estimate using sample mean and sample variance of returns.

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t \text{ and } Var(r_t) = \frac{1}{T-1}\sum_{t=1} T(r_t - \bar{r})^2$$

Correlation:

- **The correlation coefficient between two random variables X and Y is defined as:**

$$\rho_{x,y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$
$$= \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{E(X-\mu_x)^2 E(Y-\mu_y)^2}}$$

  where μ_x and μ_y are the mean of X and Y, respectively, and it is assumed that the variances exist. This coefficient measures the strength of linear dependence between X and Y, and it can be shown that $-1 \leq \rho_{x,y} \leq 1$ and $\rho_{x,y} = \rho_{y,x}$.

- **Properties:**
  - The two random variables are uncorrelated if $\rho_{x,y} = 0$.
  - If both X and Y are normal random variables, then $\rho_{x,y} = 0$ if and only if X and Y are independent.

Sample Correlation:

- When the sample $\{(x_t, y_t)\}_{t=1}^{T}$ is available, the sample correlation can be estimated as:

$$\hat{\rho_{x,y}} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2 \sum_{t=1}^{T}(y_t - \bar{y})^2}}$$

  where $\bar{x} = (\sum_{t=1} Tx_t)/T$ and $\bar{y} = (\sum_{t=1} Ty_t)/T$ are the sample mean of X and Y, respectively
- Consider a weakly stationary return series r_t, when the linear dependence between r_t and its past values r_{t-i} is of interest, the concept of correlation is generalized to autocorrelation. The correlation coefficient between r_t and $r_{t-\ell}$ is called the lag-ℓ autocorrelation of r_t and is commonly denoted by ρ_ℓ, which under the weak stationarity assumption is a function of ℓ only. Let $\bar{r}$ be the sample mean, and $\bar{r} = (\sum_{t=1}^{T} r_t)/T$.
- **Lag-ℓ autocovariance:**

$$\gamma_\ell = Cov(r_t, r_{t-\ell}) = E[(r_t - \bar{r})(r_{t-\ell} - \bar{r})].$$

Autocorrelation Function (ACF):

- **Serial (or auto-) correlations:**

$$\rho_\ell = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{Var(r_t)Var(r_{t-\ell})}} = \frac{Cov(r_t, r_{t-\ell})}{Var(r_t)} = \frac{\gamma_\ell}{\gamma_0}$$

  where the property $Var(r_t) = Var(r_{t-\ell})$ for a weakly stationary series is used.

  Note: $\rho_0 = 1$ and $\rho_\ell = \rho_{-\ell}$ and $-1 \leq \rho_\ell \leq 1$ for $\ell \neq 0$. Why?

- A weakly stationary series r_t is not serially correlated if and only if $\rho_\ell = 0$ for all $\ell > 0$.

- *Sample autocorrelation function (ACF)*

$$\hat{\rho}_\ell = \frac{\sum_{t=1}^{T-\ell}(r_t - \bar{r})(r_{t+\ell} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}$$

  where $\bar{r}$ is the sample mean and T is the sample size.

Test zero serial correlations (market efficiency):

- Individual test: for example, asymptotically normal $N(0,1)$.

$$H_0 : \rho_\ell = 0 vs. H_a : \rho_\ell \neq 0$$
$$t = \frac{\hat{\rho}_\ell}{\sqrt{1 + 2\sum_{i=1}^{\ell-1} \hat{\rho}_\ell^2}}$$

Decision rule: Reject H_0 if $|t| > Z_{\alpha/2}$ or p-value less than α.

- Joint test (Ljung-Box statistics):

$$H_0 : \rho_1 = ... = \rho_m = 0 vs. H_a : \rho_i \neq 0$$
$$Q(m) = T(T+2)\sum_{\ell=1}^{m} \frac{\hat{\rho_\ell^2}}{T-\ell}$$

Decision rule: Reject H_0 if $Q(m) > \chi^2_m(\alpha)$ or p-value less than α.

A proper perspective: at a time point t

- Available data $\{r_1, r_2, ..., r_{t-1}\} \equiv F_{t-1}$
- The return can be decomposed into two parts as

$$\begin{aligned} r_t &= \text{predictable part} + \text{not predictable part} \\ &= \text{function of elements of } F_{t-1} + a_t \end{aligned}$$

In other words, given information F_{t-1}

$$r_t = \mu_t + a_t = E(r_t | F_{t-1}) + \sigma_t \epsilon_t$$

- μ_t: conditional mean of r_t
- a_t: shock or innovation at time t
- ϵ_t: an iid sequence with mean zero and variance 1
- σ_t: conditional standard deviation (commonly called volatility in finance)

- Model for μ_t: **mean equation**
- Model for σ_t^2: **volatility equation**

- **White Noise:** a time series r_t is called a white noise if $\{r_t\}$ is a sequence of independent and identically distributed random variables with finite mean and variance.
    - *Gaussian White Noise:* if r_t is normally distributed with mean zero and variance σ^2:
    - For a white noise series, all the ACFs are zero.
    - In practice, if all sample ACFs are close to zero, then the series is a white noise series.

- **Linear Time Series:**
    - A time series r_t is said to be linear if it can be written as:

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

where μ is the mean of r_t, $\psi_0 = 1$ and $\{a_t\}$ is a sequence of independent and identically distributed random variables with mean zero and a well-defined distribution.

General ARMA Models:

- A general ARMA(p,q) model is in the form

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}, \tag{2}$$

where $\{a_t\}$ is a white noise series and p and q are no-negative integers. The AR and MA models are special cases of ARMA(p,q) model. Using the back-shift operator, the model can be written as

$$(1 - \phi_1 B - ... - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - ... - \theta_q B^q) a_t.$$

The polynomial $1 - \phi_1 B - ... - \phi_p B^p$ is the AR polynomial of the model. Similarly, $1 - \theta_1 B - ... - \theta_q B^q$ is the MA polynomial. We require that there are no common factors between the AR and MA polynomials.

Identifying ARMA Models:

- The ACF and PACF are not informative in determining the order of an ARMA model. There are two ways to determine the order of ARMA models:
  - *The Extended Autocorrelation Function (EACF)* - If we can obtain a consistent estimate of the AR component of an ARMA model, then we can derive the MA component. From the derived MA series, we can use the ACF to identify the order of the MA component.
  - *The Information Criteria (AIC or BIC)* - Typically, for some prespecified positive integers P and Q, one computes AIC (or BIC) for ARMA(p,q) models, where $0 \leq p \leq P$ and $0 \leq q \leq Q$, and selects the model that gives the minimum AIC (or BIC).
- Once an ARMA(p,q) model is specified, its parameters can be estimated by either the conditional or exact likelihood method.
- The Ljung-Box statistics of the residuals can be used to check the adequacy of a fitted model.

Forecasting Using ARMA Models:

- Denote the forecast origin by h and the available information by F_h. The 1-step ahead forecast of r_{h+1} becomes:

$$\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i} - \sum_{i=1}^{q} \theta_i a_{h+1-i}.$$

  and the forecast error is $e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$. The variance of 1-step ahead forecast error is $Var[e_h(1)] = \sigma_a^2$.

- For the ℓ-step ahead forecast, we have:

$$\hat{r}_h(\ell) = E(r_{h+\ell}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i) - \sum_{i=1}^{q} \theta_i a_h(\ell - i).$$

  where it is understood that $\hat{r}_h(\ell - i) = r_{h+\ell-i}$ if $\ell - i \leq 0$ and $a_h(\ell - i) = 0$ if $\ell - i > 0$ and $a_h(\ell - i) = a_{h+\ell-i}$ if $\ell - i \leq 0$.

Forecasting Using ARMA Models:

- Thus, the multi-step ahead forecasts of an ARMA model can be computed recursively. The associated forecast error is

$$e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell - i).$$

Three Model Representations for an ARMA Model

- The three representations of a stationary ARMA(p,q) models serve different purposes. Knowing these representations can lead to a better understanding of the model.
- The first representation is the ARMA(p,q) model in Eq. 3. This representation is compact and useful in parameter estimation. It is also useful in computing recursively multi-step ahead forecasts of r_t.
- For the other two representations, we use long division of two polynomials.

Three Model Representations for an ARMA Model

- Given the following two polynomials:

$$\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$$

$$\theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$$

we can obtain, by long divisions, that

$$\frac{\theta(B)}{\phi(B)} = 1 + \psi_1 B + \psi_2 B^2 + ... \equiv \psi(B) \tag{3}$$

$$\frac{\phi(B)}{\theta(B)} = 1 + \pi_1 B + \pi_2 B^2 + ... \equiv \pi(B) \tag{4}$$

from the definition $\psi(B)\pi(B) = 1$.

Three Model Representations for an ARMA Model

- For instance, if $\theta(B) = 1 - \theta_1 B$ and $\phi(B) = 1 - \phi_1 B$, then

$$\frac{\phi_0}{\theta_1} = \frac{\phi_0}{1 - \theta_1 - \ldots - \theta_q} \text{ and } \frac{\phi_0}{\phi_1} = \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p}$$

- **AR Representation** (shows the dependence of the current return r_t on the past return r_{t-i} where $i > 0$)

$$r_t = \frac{\phi_0}{1 - \theta_1 - \ldots - \theta_q} + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \pi_3 r_{t-3} + \ldots \tag{5}$$

- **MA Representation** (shows explicitly the impact of the past shock $a_{t-i} (i > 0)$ on the current return r_t.)

$$r_t = \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p} + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \ldots \tag{6}$$

Unit-Root Nonstationary Time Series

- Consider an ARMA model, if one extends the model by allowing the AR polynomial to have 1 as a characteristic root, then the model becomes the well-known autoregressive integrated moving-average (ARIMA) model. A conventional approach for handling unit-root nonstationarity is to use *differencing*.
- A time series is said to be an ARMA(p, 1, q) process if the change series

$$c_t = y_t - y_{t-1} = (1 - B)y_t,$$

follows a stationary and invertable ARMA(p, q) model. In finance, price series are commonly believed to be nonstationary, but the log return series is stationary.

$$r_t = ln(p_t) - ln(p_{t-1})$$

Unit-Root Nonstationary Time Series

- In some scientific field, a time series y_t may contain multiple unit roots and need to be differenced multiple times to become stationary. More complex model ARIMA(p, d, q) can be applied.
- *Unit-Root Test*

  whether the log price p_t of an asset follows a random walk or a random walk with drift.

$$p_t = \phi_1 p_t + e_t,$$
$$p_t = \phi_0 + \phi_1 p_{t-1} + e_t,$$

  the null hypotheses $H_0 : \phi_1 = 1$ versus the alternative hypothesis $H_a : \phi_1 < 1$. This is well-known *Dickey-Fuller* test.
- ARIMA also removes the trend and provides a trend-stationary time series.

Models for Variance of Asset Returns

- To put the volatility models in a proper perspective, it is informative to consider the conditional mean and variance of r_t given F_{t-1}; that is

$$\mu_t = E(r_t|F_{t-1}), \sigma_t^2 = Var(r_t|F_{t-1}) = E[(r_t - \mu_t)^2|F_{t-1}], \tag{7}$$

  where F_{t-1} denote the information set, and typically F_{t-1} consists of all linear functions of the past returns.
- We can model r_t as a stationary $ARMA(p, q)$ model with some explanatory variables.

$$r_t = \mu_t + a_t, \mu_t = \phi_0 + \sum_{i=1}^{k} \beta_i x_{it} + \sum_{i=1}^{p} \phi_i r_{t-i} - \sum_{i=1}^{q} \theta_i a_{t-i}, \tag{8}$$

  for r_t, where k, p and q are non-negative integers, and x_{it} are explanatory variables.

Models for Variance of Asset Returns

- Combine the last two equations, we have

$$\sigma_t^2 = Var(r_t|F_{t-1}) = E[(a_t|F_{t-1})],$$

- The conditional heteroscedastic models are concerned with the evolution of σ_t^2. The manner under which σ_t^2 evolves over time distinguishes one volatility model from another. They can be classified into two general categories:

  1. Those use an exact function to govern the evolution of σ_t^2;
  2. Those use a stochastic equation to describe σ_t^2.

- Autoregressive Conditional Heteroscedastic (ARCH) model of Engle (1982), and the Generalized Autoregressive Conditional Heteroscedastic (GARCH) model of Bollerslev (1986) belong to the first category whereas the Heston (1993) stochastic volatility model is in the second category.

Heston Stochastic Volatility Model

- The basic Heston model assumes that S_t, the price of the asset, is determined by a stochastic process:

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dW_t^S$$
$$d\nu_t = \kappa(\theta - \nu_t)d_t + \xi\sqrt{\nu_t} dW_t^\nu$$

  where where ν_t follows a Cox-Ingersoll-Ross process, and W_t^S, W_t^ν are Wiener processes (i.e., random walks) with correlation ρ, or equivalently, with covariance ρdt. The parameters in the above equations represent the following:

  - μ is the rate of return of the asset.
  - θ is the long variance, or long run average price variance; as t tends to infinity, the expected value of ν_t tends to θ.
  - κ is the rate at which ν_t reverts to θ.
  - ξ is **the volatility of the volatility**, or 'vol of vol', and determines the variance of ν_t.

* If the parameters obey $2\kappa\theta > \xi^2$ (known as the Feller condition) then the process ν_t is strictly positive.

GARCH Volatility Model

1. Specify a mean equation by testing for serial dependence in the data and, if necessary, building an econometric model (e.g. an ARMA model) for the return series to remove any linear dependence.

2. Use the residuals of the mean equation to test for ARCH effects.

$$a_t = \sigma_t \epsilon_t, \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + ... + \alpha_m a_{t-m}^2.$$

   where $\{\epsilon\}$ is a sequence of independent and identically distributed (iid) random variables with mean zero and variance 1, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$.

3. Specify a volatility model if ARCH effects are statistically significant and perform a joint estimation of the mean and volatility equations.

4. Check the fitted model carefully and refine it if necessary.

Volatility GARCH Model

Volatility GARCH(m,n) process combines the ARCH(m) process with the AR(n) process for lagged variance:

$$\sigma_t^2 = \omega + a_1\epsilon_{t-1}^2 + a_2\epsilon_{t-2}^2 + \ldots + a_m\epsilon_{t-m}^2 +$$
$$+b_1\sigma_{t-1}^2 + b_2\sigma_{t-2}^2 + \ldots + b_n\sigma_{t-n}^2 +$$

The simple GARCH(1,1) model is widely used in financial applications:

$$\sigma_t^2 = \omega + a\epsilon_{t-1}^2 + b\sigma_{t-1}^2 \tag{9}$$

It can be transformed into

$$\sigma_t^2 = \omega + (a+b)\sigma_{t-1}^2 + a[\epsilon_{t-1}^2 - \sigma_{t-1}^2] \tag{10}$$

Volatility GARCH Model

- The last term in (10) conditioned on information available at time $t-1$ has zero mean and can be treated as a shock to volatility.

  Therefore, the unconditional expectation of volatility for the GARCH(1,1) model equals

$$E[\sigma_t^2] = \omega/(1-a-b)$$

- This implies that the GARCH(1,1) process is weakly stationary when $a+b<1$. The advantage of the stationary GARCH(1,1) model is that it can be easily used for forecasting. Namely, the conditional expectation of volatility at time $(t+k)$ equals:

$$E[\sigma_{t+k}^2] = (a+b)^k[\sigma_t^2 - \omega/(1-a-b)] + \omega/(1-a-b) \tag{11}$$

Volatility GARCH Model

- The GARCH(1, 1) model (10) can be rewritten as

$$\sigma_t^2 = \omega/(1-b) + a(\epsilon_{t-1}^2 + b\epsilon_{t-2}^2 + b^2\epsilon_{t-3}^2 + \ldots) \tag{12}$$

- This implies that the GARCH(1,1) model is equivalent to the infinite ARCH model with exponentially weighted coefficients. This explains why the GARCH models are more efficient than the ARCH models.
- Several other GARCH models have been derived fro addressing specifics of various economic and financial time series. One popular GARCH(1,1) model, in which $a + b = 1$, is called *integrated* GARCH (IGARCH).

$$\begin{aligned} \epsilon_t &= \sqrt{h_t} \cdot a_t, \\ h_t &= \omega_0 + \omega_1 h_{t-1} + (1-\omega_1)\epsilon_{t-1}^2 \end{aligned} \tag{13}$$

- ARCH model can be estimated by both OLS and ML method, whereas GARCH model has to be estimated by ML method.

Point Process Jump Models

- To deal with such time-varying variances scenario, modeling techniques such as EGARCH (Nelson, 1991), TGARCH (Zakoian, 1994), APARCH (Ding et al.,1993), GJR-GARCH (Glosten et al., 1993), AVGARCH (Taylor, 2008),NGARCH (Higgins and Bera, 1992), stochastic volatility (SV) (Barndorff-Nielsen, 2002; Barndorff-Nielsen and Shepherd, 2004) are widely applied.
- The classic models mentioned above usually assume continuous prices,which may be unwarranted as jumps are detected in real market data (Boller-slev et al., 2008; Andersen et al., 2007, 2010; Barndorff-Nielsen and Shep-herd, 2006).
- Cox and Lewis (1996) pointed out that there exist jumps in market prices, and these jumps can considered as discrete events. Point process can be introduced to model its behaviors.
- McNeil and Rüdiger (2000) used the extreme value theory (EVT) and incorporated Poisson processes into the design of the innovation distribution which provides the best model fitting amongst all tested leptokurtic models.
- Duan et al. (2006), and Ornthanalai (2014) introduced Poisson and Lévy style jumps to formulate GARCH-Jump models.
- Following a similar idea, Chan et al. (2007) developed a non-parametric approach to deal with the heavy-tail issue after fitting to GARCH.

Point Process Jump Models

- Poisson process is a natural choice for modeling jumps. Duan et al. (2006) modeled the asset price $S_{i\Delta t}$ to follow the process:

$$\frac{S_{i\Delta t}}{S_{(i-1)\Delta t}} = e^{f_{i\Delta t}(\Delta t)+\sqrt{h_{i\Delta t}}J_i(\Delta t)\sqrt{\Delta t}}$$

$$J_i(\Delta t) = X_i^{(0)} + \sum_{j=1}^{N_i(\Delta t)} X_i^{(j)}(\Delta t) \tag{14}$$

where $X_i^{(0)} \sim N(0,1)$ and $X_i^{(j)}(\Delta t) \sim N(\mu(\Delta t), \gamma^2(\Delta t))$ for $j = 1, 2, ...,$ and $N_i(\Delta t), i = 1, 2, ...$ are a sequence of independent Poisson random variables with parameter $\lambda\Delta t$.

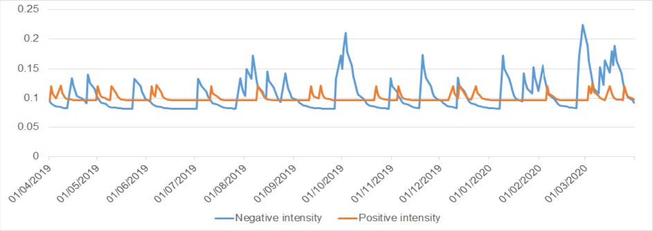
- Furthermore, for $i = 1, 2, ..., n$,

$$Corr(X_i^{(j)}, X_i'^{(j')}) = \left\{ \begin{array}{ll} \rho & \text{if } i = i', \text{ and } j = j' \\ 0 & |z| > 5 \end{array} \right\} \tag{15}$$

- When λ is released from 0, the innovations are a random mixture of normals. This model, called the GARCH-Jump model, was empirically tested by DRS (2004), who showed that the inclusion of "jumps"" significantly improved the fit of historical time series of the S&P 500, as well as helping to explain a significant portion of the volatility smile in option prices.

Point Process Jump Models

- Hawkes process is a non-homogeneous Poisson process - a class of multivariate point process that models the relationship of event arrivals where the occurrence of an event increases or decreases the probability of the occurrences of future events.
- Since its introduction in 1971, the Hawkes process and its variants have been successfully applied to model seismic events, community crimes, social media, biological neuron, finance, etc.



- Market events, such as changes in prices, index values or market crashes, can be described as realizations of multivariate point process (PP) $\{T_i, Z_i\}_{i\in\{1,2,\dots\}}$, where T_i is the occurrence time of the ith event, and Z_i indicates the type of the ith event.

Hawkes nonhomogeneous point process

- Conditional intensity:

$$\lambda(t) = \lim_{\delta \to 0} \frac{E[N(t+\delta t) - N(t)|\mathcal{H}_t]}{\delta t} \tag{16}$$

- Simple self-exciting process:

$$\lambda(t) = \mu(t) + \int_0^t \gamma(t-s)dN(s) = \mu(t) + \sum_{t_i<t} \gamma(t-t_i). \tag{17}$$

  where the $\mu(t)$ represents the exogenous component, and $\int_0^t \gamma(t-s)dN(s)$ represents the self-exciting or endogenous component.

- For more recent work, please see Chavez-Demoulin et al. (2005), in which extreme returns are modeled through self-exciting marked point processes. Follow-ing this research, Chavez-Demoulin et al. (2014) proposed a non-parametric extension to produce an evolution of the marked point process parameters.

State-Space Modeling

- State-Space models were initially developed by control systems engineers to measure a signal contaminated by noise.
- The signal at time t is taken to be a linear combination of variables, called state variables that form the so-called state vector at time t.
- The key property of the state vector is that it contains information from past and present data but the future behavior of the system is independent of the past values and depends only on the present values.
- Thus, the latent state vector evolves according to the Markov property:

$$Z_t = \Phi_t Z_{t-1} + a_t \tag{18}$$

the observation equation as

$$Y_t = H_t Z_t + N_t \tag{19}$$

where a_t and N_t are independent white-noise processes; a_t is a vector white-noise with covariance matrix Σ_a and N_t has variance of σ_N^2.

- The state space form of an $ARIMA(p, d, q)$ process $\Phi(B)Y_i = \theta(B)\epsilon_t$ has been studied by Reinsel (2002).

State-Space Modeling

- For the state-space model, define the finite sample estimates of the state vector Z_{t+1} based on observations $Y_t, ..., Y_l$ as $\hat{Z}_{t+l|t} = E[Z_{t+l}|Y_t, ..., Y_l]$, with $V_{t+l} = E[(Z_{t+l} - \hat{Z}_{t+l|t})(Z_{t+l} - \hat{Z}_{t+l|t})']$
- A convenient computational procedure, known as *Kalman filter equations*, is used to obtain the current estimate $\hat{Z}_{t|t}$ (see Reinsel (2002)).
- Note for ARIMA models with $Y_t = HZ_t$ and $H = [1, 0, ..., ..]$, this Kalman filtering procedure provides an alternative way to obtain finite sample forecasts.
- Another alternative setup is to assume that the deviation in the observation and transition equations are related. The transition equation (19) is modified as

$$Z_t = \Phi Z_{t-1} + \alpha N_t \tag{20}$$

the observation equation as

$$Y_t = H_t Z_t + N_t \tag{21}$$

This model studied by Ord, Koehler, and Snyder (1997) with a single source of error N_t is shown to be closely related to various exponential smoothing procedures.

Regime Switch and Change-Point Models

- Build upon the elegant model introduced by Hamilton (1989), the regime switching model can be written as:

$$y_t = \mu_{s_t} + \epsilon_t \tag{22}$$

where y_t is the observed variable, such as asset return and s_t represents two distinct regimes and ϵ_t are i.i.d. $\sim N(0, \sigma^2)$.

- Let $\{F_{t-1}\}$ be the information set available as of $t-1$. The transition between the regimes is taken to be Markovian,

$$Prob(s_t = j | s_{t-1} = i, F_{t-1}) = p_{ij}, i, j = 1, 2; \tag{23}$$

- Thus leading to an $AR(1)$ model for μ_{s_t} as

$$y_t = \phi_0 + \phi_1 \mu_{s_{t-1}} + a_t \tag{24}$$

where a_t by definition can take four possible values depending on s_t and s_{t-1}.

Regime Switch and Change-Point Models

- Note $\phi_0 = p_{21}\mu_1 + p_{12}\mu_2$ and $\phi_1 = p_{11} - p_{21}$. With aggregation $y_t \sim ARMA(1,1)$, but because of the discrete nature of a_t (23) is a non-linear process.
- The unknown parameters are $(\mu_1, \mu_2, \sigma, p_{11}, p_{22})' = \lambda$ and they can be estimated via maximum likelihood:
  - Note $y_t|(s_t = j, F_{t-1}) \sim N(\mu, \sigma^2)$.
- The predictive density of y_t is given as a finite mixture model,

$$f(y_t|F_{t-1}) = \sum_{i=1}^{2} \text{Prob}(s_t = i|F_{t-1}) f(y_t|s_i = i, F_{t-1}) \tag{25}$$

  and estimate of λ is obtained by maximizing the likelihood function $L(\lambda) = \sum_{t=1}^{T} logf(y_t|F_{t-1}, \lambda)$.
- The forecast for $'k'$ period ahead can be stated as:

$$E(y_{t+k}|F_{t-1}) = \mu + \theta_l^k \sum_{i=1}^{2} (\mu_i - \mu) \text{Prob}(s_t = i|F_{t-1}), \tag{26}$$

  where $\mu = \theta_0/(1 - \theta_1)$. Interested readers should refer to Hamilton (2016) and references therein.

Stylized Facts for Asset Returns

- **Stylized Fact 1:** Absence of Autocorrelation in Returns: Linear autocorrelations are often insignificant for lower frequency data.
- **Stylized Fact 2:** Heavy Tails: The returns are likely to display a heavy tailed distribution such as power-law or Pareto.
- **Stylized Fact 3:** Asymmetry: When prices fall, they tend to fall faster than when they tend to rise, thus drawdowns tend to be larger than upward rises.
- **Stylized Fact 4:** Positive Excess Kurtosis: Returns generally exhibit high values for the fourth moment due to different phases of high and low trading activity.
- **Stylized Fact 5:** Volatility Clustering: High volatility events tend to cluster in time and this can be measured through positive correlations that exist over several lags.
- **Stylized Fact 6:** Volume Volatility Relationship: There is a positive correlation between conditional volatility and volume.
- **Stylized Fact 7:** Negative Correlation in Returns: Non-synchronous trading and bid-ask bounce both can introduce negative lag-1 autocorrelation in returns.