Mathematical Models for Quantitative Finance Notes (2022)

Manuel Luci[1]

v.1 September 2023

[1] manuel.luci@live.it

Notes for readers

These notes are derived from the ”Mathematical Models for Quantitative Finance: Market Microstructure, Networks, and Systemic Risk” course of the year 2022 by Professors Fabrizio Lillo and Piero Mazzarisi. There are likely to be errors, so I recommend using these notes only as a starting point and delving deeper through books and articles. If you notice any errors, please report them to my email, or create a branch in the course repository:GitHub . I would like to thank Professors Lillo and Mazzarisi for conducting this interesting course. Enjoy reading and studying!

Contents

I

ARMA

1.1 Some Definitions

Let's give some definitions:

- Stocastic process (univariate): sequence of random variables Y_t
- Second order process: $\mathbf{E}[Y_t^2] < +\infty \ \forall t$
- Mean $\mu_t = \mathbf{E}[Y_t]$ and variance $\sigma^2 = \mathbf{E}(Y_t - \mu_t)^2$
- Autocovariance $\gamma_t(k) \equiv Cov(Y_t, Y_{t-k}) = \mathbf{E}[Y_t - \mu_t][Y_{t-k} - \mu_{t-k}]$, from this we can define $\sigma_t^2 = \gamma_t(0)$
- Autocorrelation:
$$\rho_t(k) \equiv Corr(Y_t, Y_{t-k}) = \frac{\gamma_t(k)}{\sqrt{\gamma_k(0)\gamma_{t-k}(0)}}$$

We give some definitions of stationarity:

Strict stationarity

$$(Y_1, \dots, Y_n) \overset{d}{=} (Y_{1+k}, \dots, Y_{n+k}) \quad \forall n > 1, k \tag{I.1}$$

we use $\overset{d}{=}$ as converges in distribution.

Weak/second-order/covariance stationarity

Under this conditions:

- $\mathbb{E}\left[Y_t\right] = \mu$
- $\mathbb{E}\left[Y_t - \mu\right]^2 = \sigma^2 < +\infty$
- $\mathbf{E}[Y_t - \mu_t][Y_{t-k} - \mu_{t-k}] = \gamma(|k|)$ (independent of t for each k) than:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} \tag{I.2}$$

This is the most common definition of stationarity

Ergodicity

We consider three different definition:

- In mean: $\bar{y} \equiv \frac{1}{T}\sum_{t+1}^{T} Y_t \xrightarrow{p} \gamma(k)$
- In second moments: $\frac{1}{T}\sum_{t+1}^{T}(Y_t - \mu)(Y_{t-k} - \mu) \xrightarrow{p} \gamma(k)$
- if a stationary process Y_t satisfy $\sum_{k=0}^{+\infty} |\gamma(k) < \infty$ then Y_t is ergodic in mean.

This one is less used

We can give this intepretations:

- uncodintional mean and variance are constant.
- mean reversion
- shocks are transient
- covariance between Y_t and Y_{t-k} tends to 0 if $k \to \infty$

We define white noise as:

Weak (uncorrelated)

- $\mathbb{E}\left[\epsilon_t\right] = 0 \ \forall t$
- $Var(\epsilon_t) = \sigma^2 \ \forall t$
- $Corr(\epsilon_t, \epsilon_s) = 0 \ \forall s \neq t$

Strong (independece)

$\epsilon_t \sim i.i.d(0, \sigma^2)$

Gaussian (weak = strong)

$\epsilon_t \sim n.i.d(0, \sigma^2)$ (if uncorrelated, than are independent)

Introducing the lag operator proves advantageous to furthers discussions:

Lag operator

The lag operator is defined as:

$$LY_t \equiv Y_{t-1} \tag{I.3}$$

Some properties of lag operator:

- it is a linear operator:

$$L(\alpha X_t + \beta Y_t) = \alpha L(X_t) + \beta L(Y_t) = \alpha X_{t-1} + \beta Y_{t-1}$$

- it admits power exponent:

$$\begin{aligned} L^k Y_t &= Y_{t-k} \\ L^{-k} Y_t &= Y_{t+k} \end{aligned}$$

Using lag operator we can define:

$$\begin{aligned} &\text{First diffent :} && \Delta Y_t = Y_t - Y_{t-1} = Y_t - LY_t = (1-L)Y_t \\ &AR(2) : && Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} = (\theta_1 L + \theta_2 L^2) Y_t \end{aligned}$$

A lag polynomial $\theta(L)$ is:

$$L(\theta) = \sum_{k=0}^{+\infty} \theta_k L^k$$

1.2 Moving Avarage (MA) process

Lag operator is useful to costruct a stationary process. For example we cas costruct a sort of weighted moving average of white noise ϵ_t:

MA(q)

$$\text{MA(q)} : Y_t = \theta(L)\epsilon_t = \epsilon_t + \theta_1\epsilon_{t-1} + \ldots + \theta_q\epsilon_{t-q} \tag{I.4}$$

It is useful studying the first order understand MA(q) structure:

$$\text{MA(1)} : Y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t$$

If we study the expected value and covariance:

$$\begin{aligned}
\mathbb{E}[Y_t] &= 0 \\
\gamma(0) &= \mathbb{E}[Y_tY_t] = \ldots = \sigma^2(1+\theta^2) \\
\gamma(1) &= \mathbb{E}[Y_tY_{t-1}] = \ldots = \sigma^2\theta \\
\gamma(k) &= \mathbb{E}[Y_tY_{t-k}] = \ldots = 0 \quad \forall k > 1
\end{aligned}$$

so that:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta}{1+\theta^2} \tag{I.5}$$

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = 0 \qquad \forall k > 1 \tag{I.6}$$

While white noise is 0-correlated, MA(1) is 1-correlated, this means that only the first correlation is different from zero. The first correlation is easy to evaluate and measuring. So that if we noticed that some samples have only the first correlation different from zero, it could come from a MA(1).
We can generalize it for MA(q):

$$\gamma(0) = \sigma^2(1 + \theta_1^2 + \ldots + \theta_q^2) \tag{I.7}$$

$$\gamma(k) = \sigma^2\sum_{j=0}^{q-k}\theta_j\theta_{j+k} \qquad \forall k \leq q \tag{I.8}$$

$$= 0 \qquad \forall k > q \tag{I.9}$$

MA(q) is q-correlated and any stationary q-correlated process can be represented as an MA(q).
At same time MA(q) process is not unique, but there is only one MA(q) that is invertible.
As early, we can study MA(1) and than generalize: given $Y_t = (1+\theta L)\epsilon_t$ we can expand it in a Taylor series:

$$(1+\theta L)^{-1} = (1 - \theta L + \theta^2L^2 + \ldots) = \sum_{i=0}^{+\infty}(-\theta L)^i$$

inverting the $\theta(L)$ lag polynomial:

$$(1 - \theta L + \theta^2 L^2 + \ldots)Y_t = \epsilon_t$$

which is AR(∞).
If an MA process can be written as AR(∞), than MA process is invertible.
For MA(1) process, the invertibility is given by $|\theta| < 1$. Generalizing it, a MA(q) process is invertible if the roots of the lag polynomial:

$$1 + \theta_1 z + \ldots + \theta_q z^q = 0$$

lie outside the unit circle.
Invertible has an important practical in application: given ϵ_t a non observable, it could be reconstructed from Y through a AR(∞) representation. Let's make an example: given the generic lag polynomial:

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$$

we have to find the series $\theta(L)^{-1} = \varphi_0 + \varphi_1 L + \ldots$ such that:

$$(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)(\varphi_0 + \varphi_1 L + \varphi_1 L^2 \ldots) = 1$$

coefficients φ_i can be computed recursively matching L^k (easy to compute)

1.3 Auto-Regressive Process (AR)

A general AR process is defined as:

$$\phi(L)Y_t = \epsilon_t \tag{I.10}$$

Analysing AR(1):

$$(1 - \phi L)Y_t = \epsilon \quad \text{or} \quad Y_t = \phi Y_{t-1} + \epsilon_t$$

inverting the lag polynomial $(1 - \phi L)$ the AR(1) can be written as:

$$Y_t = (1 - \phi L)^{-1}\epsilon_t = \sum_{i=0}^{\infty}(\phi L)^i = \sum_{i=0}^{\infty} \phi \epsilon_{t-i} = \text{MA}(\infty)$$

stationary condition is $|\phi| < 1$
. From this represention, we can compute $\rho(\cdot)$:

$$\rho(k) = \phi^{|k|} \quad \forall k$$

In general AR(p) process is define as:

AR(p)

$$\text{AR(p)} : Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t \tag{I.11}$$

AR(p) is stationarity if all roots of the characteristic equation of lag polynomial:

$$1 - \phi_1 Z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$$

are outside of unite circle.
Exposing AR(p) in its state space from:

$$\begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

It can be expresses as:

$$X_t = F X_{t-1} + v_t$$

The expected value of X_t satisfy:

$$\mathbb{E}\left[X_t\right] = F X_{t-1} \quad \text{and} \quad \mathbb{E}\left[X_{t+j}\right] = F^{j+1} X_{t-1}$$

whose dynamic is given by the eigenvalues of the matrix F

1.4 Partial Autocorrelation Function (PACF)

For an AR(p) process, the k-lag ACF ρ_k can be interpreted as simple regression:

$$Y_t = \rho_k Y_{t-k} + \text{error}$$

and the k-lag PACF:

$$a_t(k) \equiv Corr(Y_t, Y_{t-k} | Y_{t-1}, \ldots, Y_{t-k+1})$$

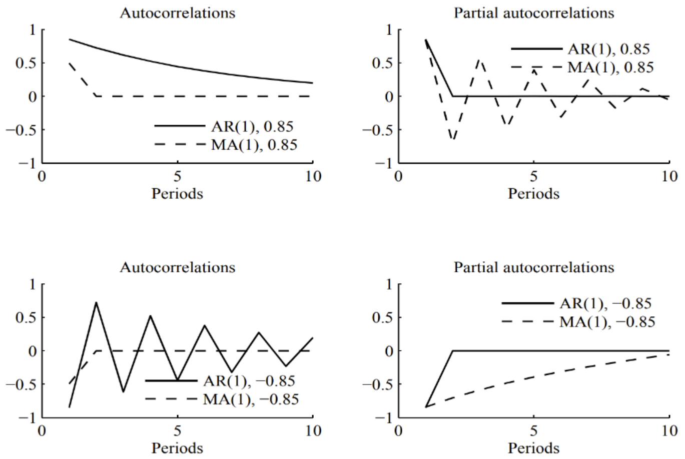
can be build as multiple regression:

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_k Y_{t-k} + \textbf{error}$$

that can be computed by solvig the Yule-Walker system:

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(k-1) \\ \gamma(1) & \gamma(2) & \cdots & \gamma(k-2) \\ \vdots & \vdots & \cdots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \cdots & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(k) \end{bmatrix}$$

We can conclude that AR(p) process are p-partially correlated. That is a useful tool to identify AR order



1.5 ARMA(p,q)

An ARMA(p,q) process is defined as

ARMA

$$\phi(L)Y_t = \theta(L)\epsilon_t \tag{I.12}$$

where $\phi(L)$ and $\theta(L)$ are p^{th} and q^{th} lag polynomials.
Process is stationary if all the roots of polynomial:

$$\phi(z) \equiv 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_{p-1} z^{p-1} - \phi_p z^p = 0$$

lie outside the unit circle → admits MA(∞) representation:

$$Y_t = \phi(L)^{-1}\theta(L)\epsilon_t$$

The process is invertible if all roots of

$$\theta(z) \equiv 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_{q-1} z^{1-1} + \theta_q z^q = 0$$

lie outside the unit circle → admits AR(∞) representation:

$$\epsilon_t = \theta(L)^{-1}\phi(L)Y_t$$

1.6 Estimate AR models

In time series the data are usually not i.i.d, so that it is better use prediction-error decomposition of likelihood:

$$L(y_T, y_{T-1}, \ldots, y_1; \theta) = f(y_T|\Omega_{T-1};\theta) \cdot f(y_{T-1}|\Omega_{T-2};\theta) \cdot \ldots \cdot f(y_1|\Omega_0;\theta)$$

Let's make an example for AR(1) (gaussian noise):

$$y_t = \phi_1 y_{t-1} + \epsilon_t$$

the full log-Likelihood is:

$$I(\phi) = \underbrace{f_{Y_1}}_{\text{marginal } 1^{\text{st}} \text{ obs}}(y_1;\phi) + \underbrace{\sum_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\phi)}_{\substack{\text{conditional likelihhod} \\ \text{under normality OLS = MLE} \\ \text{process AR(1), it is Markovian}}} = f_{Y_1}(y_1;\phi) - \frac{T}{2}\log(2\pi) - \sum_{t=1}^{T}\log\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \phi y_{t-1})^2}{\sigma^2}$$

Maximing conditional likelihood for ϕ → minimize:

$$\sum_{t=2}^{T}(y_t - \phi y_{t-1})^2$$

In general AR(p) process under gaussianity are asymptotically equivalent to MLE.

1.7 Estimation MA models

Let's consider MA(1):

$$y_t = \theta\epsilon_{t-1} + \epsilon_t$$

The full log-Likelihood is:

$$I(\phi) = \underbrace{f_{Y_1}}_{\text{marginal } 1^{\text{st}} \text{ obs}}(y_1;\phi) + \underbrace{\sum_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\phi)}_{\text{conditional likelihhod}} = f_{Y_1}(y_1;\phi) - \frac{T}{2}\log(2\pi) - \sum_{t=1}^{T}\log\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \epsilon y_{t-1})^2}{\sigma^2}$$

ϵ are not observed, I observe y, ϵ can be recover from y:

$$\epsilon_t = y_t - \theta\epsilon_{t-1} = (-\theta)^t\epsilon_0 + \sum_{i=1}^{t}(-\theta)^i y_{t-i}$$

if MA is invertible; minimaztion of RSS is non-linear in θ, we need MLE or Non-Linear Least Square

1.8 Estimate ARMA models

For general ARMA(p,q):

$$T_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \ldots \epsilon_q\epsilon_{t-q}$$

Y_{t-1} is correlated with $\epsilon_{t-1}, \ldots, \epsilon_{t-q}$, so $\mathbb{E}\left[\epsilon|X\right] \neq 0$, OLS is not onsistent. We need MLE with numerical optimization procedures.

1.9 Optimal Prediction

Optimal Prediction

If the Loss function of a prediction is a quadratic function of the prediction error i.e, the MSE:

$$MSE(\hat{Y}_t) \equiv \mathbb{E}\left[Y_t - \hat{Y}_t\right]^2$$

Than the optimal prediction of Y in terms of past values X is given by conditional expectation $\mathbb{E}\left[Y_{t+1}|X_t\right]$

If process Y is linear or normally distributed, linear projection: $\hat{Y}_t \equiv P(Y_{t+1}|X_t) = \alpha' X$ is the optimal prediction, $\alpha' = [\mathbb{E}\left[X_t X_t'\right]]^{-1}\mathbb{E}\left[X_t' Y_{t+1}\right] \simeq OLS$.
In general:

$$Y_{t+k} = \underbrace{\text{function of past values}}_{\text{determine } \mathbb{E}[Y_{t+k}]} + \underbrace{\text{fucntion of future values}}_{\text{determine} Var_t(Y_{t+k})}$$

1.10 Wold Theorem

Wold Theorem

Any mean zero covariance stationary process can be represented in the form:

$$Y_t = \underbrace{\sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j}}_{\text{random}} + \underbrace{k_j}_{\text{deterministic}} \tag{I.13}$$

where:

- $\sum_{j=0}^{\infty} \varphi_j^2 < \infty, \varphi_0 = 1$
- $\epsilon_t = Y_t - P(Y_t|Y_{t-1}, Y_{t-2}, \dots)$ are linear prediction errors
- $\{\varphi_j\}, \{\epsilon_j\}$ are unique
- k_j is linearly deterministic
- ϵ_{t-j}, k_j are uncorrelated.

1.11 Box-Jenkins Approach

Box-Jenkins Approach is a standard procedure models to find the best fit of a time-series model to past values of a time series.

- First of all check for stationarity (exists some statistical test), if it fails, try transformation.
- Identification: check autocorrelation function (MA(q) model are q-correlated process) and partial autocorrelation function (AR(p) are p-partially correlated process)
- Validation: checl appropriateness of the model by some measure of fit (AIC/Akaike, Bic/Schwarz, residua,...)

1.12 ARIMA

ARIMA model is an integrated ARMA model. ARIMA(p,1,q) denote a nonstationary process Y_t for which the first difference is an ARMA(p,q) process:

ARIMA(p,1,q)

$$Y_t - Y_{t-1} = (1-L)Y_t \tag{I.14}$$

Y_t is siad to be intefrated of order 1 ($I(1)$)
$I(2) \rightarrow Y_t$ is integrated of order 2 if : $Y_t - Y_{t-1} = (1-L)^2 Y_t$

1.13 ARFIMA

We can try to think on a difference operator and ask what is the meaning of a Lag-operator with exponent fractional.
The $k-$ difference operator $(1-L)^n$ with integer n can be generalized to a fractional operator $(1-L)^d$ with $0 < d < 1$ defined by the binomial expansion:

$$(1-L)^d = 1 - dL + d(d-1)L^2/2! + \dots$$

If $d < 0.05$, the process is cov stationary and admits an AR(∞) representation.
Fractional filter $(1-L)^d$ produces hyperbolic decaying autocorrelation, the so colled long memory. for ARFIMA(pd,q) processes:

$$\phi(L)(1-L)^d Y_t = \theta(L)\epsilon_t$$

its autocrrelation functions is proportional to:

$$\rho(k) \simeq ck^{2d-1}$$

II

Market microstructure models

2.1 Introduction

The dominant framework utilized in financial market modeling is based on the condition of no arbitrage opportunities.

Arbitrage

An arbitrage opportunity is present in a market when an economic actor can devise a trading strategy which is able to provide her a financial gain continuously, without initial investment, and without risk.

In an efficient market, the exploitation of an arbitrage opportunity typically leads to its rapid elimination. Considering fundamental analysis, Williams (1938), Graham and Dodd (1934) proposed thid definition of fundamental value:

Fundamental values

Intrinsic or Fundamental value of any security equals the discounted cash flow which that security gives title to, and actual price fluctuate around fundamental values, in formula:

$$P_t = \sum_{k=0}^{\infty} \frac{1}{(1+r)^{k+1}} \mathbb{E}\left[D_{t+k}\right] \tag{II.1}$$

where r is the constant discount rate and D_t is the dividend payed at time t

Working (1934) proposed that random walks generate patterns resembling stock prices, while Kendall (1953), Granger and Morgenstern (1963) conducted statistical analyses that revealed stock prices exhibit a random walk behavior.

> *"If stock prices were patternless, was there any point to fundamental analysis? (LeRoy 1989)"*

When we say that two information sets I and J satisfy the condition $I \subset J$, it means that set J contains more information or is more comprehensive than set I.
" In other words, J is superior or more extensive in terms of information compared to I".

Law of Iterated Expectations

Given a random variable X:

$$\mathbb{E}\left[X|I\right] = \mathbb{E}\left[\mathbb{E}\left[X|J\right]|I\right] \tag{II.2}$$

In simpler terms, if you have limited information denoted as I,the most accurate forecast you can make for a random variable X is essentially the forecast you would make for X if you had access to more comprehensive and superior information denoted as J.
In essence, having better information helps you make a more accurate prediction.
The law can be rewritten a:

$$\mathbb{E}\left[X - \mathbb{E}\left[X|J\right]|I\right] = 0$$

This implies that you cannot use limited information represented by I to predict the forecast error that you would make if you had superior information denoted as J. Limited information doesn't allow you to account for the additional insights and knowledge that superior information would provide, making it impossible to accurately predict the forecast errors that would arise with better information.
Suppose that at a given time t, the price of a security, denoted as P_t, can be expressed as the rational expectation of a fundamental value V, taking into account the information available at that time, denoted as I_t:

$$P_t = \mathbb{E}\left[V|I_t\right] \equiv \mathbb{E}\left[V\right]_t$$

the expectation of the change over next period is:

$$\mathbb{E}\left[P_{t+1} - P_t\right]_t = \mathbb{E}\left[\mathbb{E}\left[V\right]_{t+1} - \mathbb{E}\left[V\right]_t\right] = 0$$

due to Law of Iterated Expectations. The actual changes in prices, when they occur, cannot be predicted using the information contained in the set I_t.

"A capital market is said to be efficient if it fully and correctly reects all relevant information in determining security prices. Formally, the market is said to be efficient with respect to some information set if prices would be unaffected by revealing that information to all participants. Moreover, efficency with respect to an information set implies that it is impossible to make economic profits by trading on the basis of that information set.(Malkiel, 1992) "

Informations doesn't change the price, surprise does it!

Weak-form efficiency

The information set includes only the history of prices.

Semistrong-form efficiency

The information set includes all the publicly available information.

Strong-form efficiency:

The information set includes all information, i.e. also private information known to any market participant.

In the context of the Efficient Market Hypothesis, it follows that price fluctuations must be inherently unpredictable. In an increasingly efficient market, the sequence of price changes becomes more stochastic. The Random Walk Hypothesis is a commonly adopted model for prices, emphasizing market efficiency. It's important to note that the Random Walk Hypothesis imposes stricter limitations than the Martingale Hypothesis. As we have seen in last chapter, the existence of autocorrelated returns, implies predictability using linear time series methods.

Roll Model

- Time is discrete and advances one unit any time a trade occurs
- All trades are conducted through a monopolistic dealer
- The efficient price follows a random walk

  $$m_t = m_{t-1} + u_t \tag{II.3}$$

  where u_t is an independent identically distributed (i.i.d) noise with variance σ_u^2.
- At each time the dealer posts bid and ask prices.
- Dealer incurs a noninformational cost c per trade (due to

  fixed costs: computers,telephone).No adverse selection and asymmetric information.
- The bid and the ask prices are:

  $$b_t - m_t - c \quad a_t = m_t + c$$

  and thus the spread is $2c$.
- The transaction price p_t:

  $$p_t = m_t + q_t c \tag{II.4}$$

  where $q_t = +1(-1)$ if the customer is buyer (seller).

- Moreover q_t are assumed serially independent and independent from prices (m_t and p_t). As we will see, this assumption is strongly violated in real data.

Variance of transaction price increments is (remember: q_{t-1} and q_t are no correlated, q_t, u_t are independent):

$$\gamma_0 \equiv Var(\Delta p_t) = \mathbb{E}\left[(p_t - p_{t-1})^2)\right] = \ldots = 2c^2 + \sigma_u^2$$

That means: observed volatility is larger than the volatility of the efficient price!
The autocovariance of transaction price increments is:

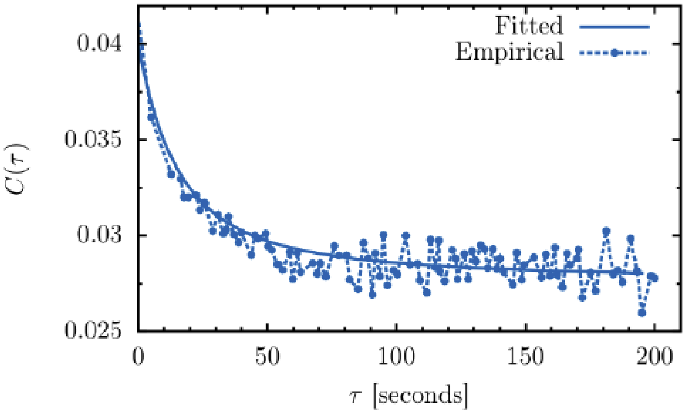
$$\gamma_1 \equiv Cov[\Delta p_{t-1} \Delta p_t] = -c^2$$

and it is zero for lags larger than one. Thus observed returns are autocorrelated even if the efficient price is a martingale.
Empirical observations have revealed that the one-lag autocorrelation function of trade-by-trade return, and occasionally even in the context of high-frequency data like one-minute returns, exhibits a negative value. This phenomenon is commonly referred to as "bid-ask bounce".
Roll model was used to determine the spread from transaction data.

According to the Roll model, the volatility over τ consecutive trades is given by

$$C(\tau) = \frac{\mathbb{E}\left[(p_{t+\tau} - p_t]\right)^2}{\tau} = \underbrace{\frac{\mathbb{E}\left[(\sum_{t=1}^{\tau} \Delta p_t)^2\right]}{\tau}}_{\textbf{diffusive process (random walk)}} = \sigma_u^2 + \frac{2c^2}{\tau}$$

Plot of $C(\tau)$ is called signature plot. More flat is plot, more diffusive is the process, more is random-walk.



In the Roll model, the autocovariance pattern of transaction price increments is identical to that of a Moving Average (MA) process with a lag of 1 (MA(1)). It's important to note that the Roll model is essentially a structural model, but its time series representation can be seen as a statistical model.

Assuming covariance stationarity allows us to apply the Wold theorem, which states that any zero-mean covariance stationary process x_t can be represented as:

$$x_t = \sum_{j=1}^{\infty} \theta_j \epsilon_{t-j} + \kappa_t$$

where ϵ_t is a zero mean white noise process and κ_t is a linearly deterministic process (can be predicted arbitrarily well by a linear projection on past observation of x_t).
Structural models typically incorporate unobserved variables. For instance, in the original Roll model, neither u_t nor q_t are directly observable. Consequently, the efficient price mt remains unobserved as well, although estimating it would be of significant interest.
On the other hand, statistical models are well-suited for forecasting purposes. In the Roll model, one can directly observe p_t and estimate γ_0 and γ_1 with relative ease. In an MA(1) process, $x_t = \epsilon_t + \theta\epsilon_{t-1}$, they are related to the model's parameters as:

$$\gamma_0 = (1+\theta^2)\sigma_\epsilon^2 \qquad \gamma_1 = \theta\sigma_\epsilon^2 \qquad \sigma_\epsilon^2 = Var[\epsilon_t]$$

If $|\theta| < 1$, exixsts a convergent AR(∞) which is useful for forecasting.

2.2 Asymmetric information models

Asymmetric information models are designed to depict how informed traders interact with uninformed intermediaries, often in the presence of noise traders (these are necessary, they make market less efficient). In these models, the security's payoff typically exhibits a common value nature, with the primary benefit of owning the security being its resale value or terminal liquidating dividend, which is consistent for all holders.
However, for trading to occur, there must also be private value components. These are individual agents' needs for diversification or exposure to specific risks that are unique to each agent.
Initially, public information consists of shared knowledge about the probability structure of the economy. This includes information about the unconditional distribution of terminal security values and the distribution of agent types.
As the trading process unfolds, the most significant updates to the public information set come from market data, such as bids, asks, and the prices and volumes of trades. Consequently, the trading process involves transitioning from one well-defined information set to another, and it lacks properties like stationarity, ergodicity, and time-homogeneity. So that the market 'learns' from others, it expands its information set.
We hace 2 different models:

- **Sequential trade models**: it involve traders arriving at the market one by one, randomly, independently, and individually. In a situation where an individual trader only engages in the market once, there is no necessity for them to consider the impact their actions might have on the subsequent decisions of other participants.

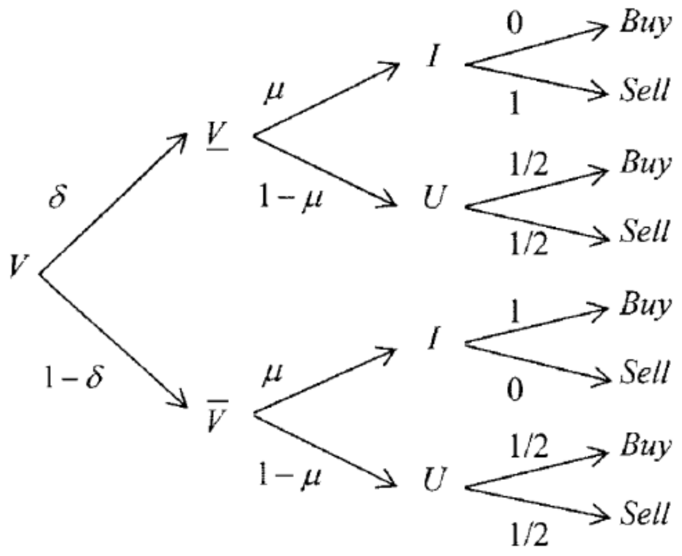
- **Strategic trader models:** a single informed agent is capable of trading multiple times within the market (Kyle 1985). However, when a trader revisits the market for subsequent trades, they must engage in strategic calculations that take into account various factors. In both models, whether it's a single or repeat trade, each transaction discloses certain aspects of the trader's private information.

2.2.1 Sequential model: Glosten and Milgrom (simplified)

Sequential model: Glosten and Milgrom

- One security with value (payoff) V that is either low ($\underline{V}$) or high ($\overline{V}$). The probability of the low outcome is δ (indepentent from trading).
- The value is revealed after the market closes and it is not affected by trading. It is determined by a random draw of nature before the market opens.
- Traders can be either informed (i.e. they know the value (V) outcome) or uninformed. The fraction of informed traders is μ.
- A dealer posts bid and ask quotes, B and A.
- He knows parameters ($\underline{V}$, $\overline{V}$, δ, μ)
- At each time step a trader is drawn at random from the population:
  - If she is informed, she buys if $V = \overline{V}$ and sells if $V =\underline{V}$
  - If she is informed, she buys and sells with equal probability.
- Transaction price is set by the dealer:
  - Buys(from traders) occur at dealer's ask price A
  - Sells (from traders) occur at dealer's bid price B.
- The dealer does not know whether the trader is informed.

Let's rapresent probability tree:



Total probabilities are obtained by multiplication along the path (U=Uninformed, I=informed):

$$P(V = \underline{V}, U, \text{Buy}) = \delta(1-\mu)\frac{1}{2}$$
$$P(\text{Buy}) = \frac{1+\mu(1-2\delta)}{\delta}$$

Let us consider the dealer:

- The purchases and sales are not sensitive to quotes.
- If she is monopolist, she sets $A = \infty$ and $B = 0$
- Competition and regulation: "We assume that competition among dealers drive expected profit to zero".

The dealer uses Bayes rule to update her beliefs on V. Assume that the first trade is a buy, the updated value of her belief on δ is:

$$\delta_1(\text{Buy}) = P(\underline{V}|\text{Buy}) = \frac{P(\underline{V}, \text{Buy})}{P(\text{Buy})} = \frac{\delta(1-\mu)}{1+\mu(1-2\delta)}$$

If the first is a sell, than her updated value of δ is:

$$\delta_1(\text{Sell}) = P(\underline{V}|\text{Sell}) = \frac{P(\underline{V}, \text{Sell})}{P(\text{Buy})} = \frac{\delta(1+\mu)}{1-\mu(1-2\delta)}$$

Assume that the first trade is a buy. At the end of the day, the dealer's realized profit on the first transaction is $\Pi = A - V$(because it is a buy). Her exception of the profit is:

$$\mathbb{E}\left[\Pi|\text{Buy}\right] = A - \mathbb{E}\left[V|\text{Buy}\right] = A - (\delta_1(\text{Buy})\underline{V} + (1 - \delta_1(\text{Buy})\overline{V}))$$

Due to competon, we impose $\mathbb{E}\left[\Pi|\text{Buy}\right] = 0$, than:

$$A = \mathbb{E}\left[V|\text{Buy}\right] = \frac{\underline{V}(1-\mu)\delta + \overline{V}(1-\delta)(1+\mu)}{1+\mu(1-2\delta)}$$

Analogues:

$$B = \mathbb{E}\left[V|\text{Sell}\right] = \frac{\underline{V}(1+\mu)\delta + \overline{V}(1-\delta)(1-\mu)}{1-\mu(1-2\delta)}$$

Therefore the spread (before the first transaction) is set by the dealer at:

$$A - B = \frac{4(1-\delta)\delta\mu(\overline{V}-\underline{V})}{1-(1-2\delta)^2\mu^2}$$

Simple case: $\delta = 1/2$:

$$A - B = (\overline{V}-\underline{V})\mu$$

The (initial) spread is proportional to the fraction of informed traders, the spread is a protection of the dealer to adverse selection from trading with informed traders.

Let us analysing the dynamics and let us see who loses money.
Assume that the first trade is a buy. From conditional expectation:

$$\mathbb{E}\left[V|\text{Buy}\right] = \mathbb{E}\left[V|U,\text{Buy}\right]P(U|\text{Buy}) + \mathbb{E}\left[V|I,\text{Buy}\right]P(I|\text{Buy})$$

Using $A = \mathbb{E}\left[V|\text{Buy}\right]$ and rearranging the terms:

$$(A - \mathbb{E}\left[V|U,\text{Buy}\right])P(U|\text{Buy}) = -(A - \mathbb{E}\left[V|I,\text{Buy}\right])P(I|\text{Buy})$$

The left is dealer's gain from trading with the uninformed, the right hand is dealer's gain from trading with the informed. Wealth is transferred from uninformed to informed traders.
After the initial trade, the dealer updates her beliefs and update the quotes accordingly by using the expressions for the map from prior to posterior probabilities:

$$\delta_k(\text{Buy}_k;\delta_{k-1}) = \frac{\delta_{k-1}(1-\mu)}{1+\mu(1-2\delta_{k-1})} \qquad \delta_k(\text{Sell}_k;\delta_{k-1}) = \frac{\delta_{k-1}(1+\mu)}{1-\mu(1-2\delta_{k-1})}$$

We noticed thath:

- Trade price is a martingale
- The spread declines over time and prices converges to the true value of V
- There is a price impact of prices, i.e., given a past history, a buy moves the price up and a sell moves the price down

General comments about this model:

- The Glosten Milgrom model makes the assumption that informed traders exclusively use market orders, which is not reflective of how trading typically occurs in limit order book markets.
- Nevertheless, this unrealistic assumption is employed for empirical purposes in estimating the probability of informed trading (known as PIN).
- The Glosten Milgrom model demonstrates that a significant portion of the spread is attributable to adverse selection, where informed traders exploit their information advantage.
- In classical microstructure analysis, the spread is typically decomposed into three components:
  - Fixed noninformational cost (similar to the Roll model).
  - Adverse selection cost (similar to the Glosten Milgrom model).
  - Inventory cost (representing the spread that dealers apply to manage the risk associated with holding a large inventory).

2.2.2 Strategic Model: Kyle

The model addresses a scenario characterized by information asymmetry, the mechanism by which information influences prices, and the strategic decision-making of both the dealer and the informed trader. This model operates within an equilibrium framework.
It can take on several variations, including single-period, multi-period, and continuous-time settings. In this model, there are typically three key agents involved:

- **Market Maker (MM) or Dealer**: This agent facilitates trading in the market.
- **Informed Trader**: This agent possesses private information that can impact trading decisions.
- **Noise Traders**: These are numerous traders who engage in the market without possessing any significant private information, and their trading behavior is often driven by random or non-informative factors.

Strategic model:Kyle (one period)

- The terminal (liquidation) value v of an asset is normally distributed with mean $p0$ and variance Σ_0.
- The informed trader knows v and enters a demand x (volume).
- Noise traders submit a net order flow u, which is Gaussian distributed with mean zero and variance σ_u^2

- x and u are signed order flows, i.e. they are the purchased volume if positive, and the sold volume if negative.
- The MM observes the total demand $y = x + u$ and then sets a price p. All the trades are cleared at p, any imbalance is exchanged by the MM.
- However the MM knows that there is an informed trader and if the total demand is large (in absolute value) she is likely to incur in a loss. Thus the MM protects herself by setting a price that is increasing in the net order flow.
- The solution to the model is an expression of this trade-off

Let us analsing informed trader setting. She conjectures that the MM uses a linear price adjustment rule $p = \lambda y + \mu$, where λ is inversely related to liquidity. Her profit is:

$$\pi = (v - p)x = x[v - \lambda(u + x) - \mu])$$

with expected profit:

$$\mathbb{E}\left[\pi\right] = x(v - \lambda x - \mu)$$

It is a parabola, the traders maximizes it if:

$$x = \frac{v - \mu}{2\lambda}$$

In Kyle's model the informed trader can loose money, but on average she makes profit.

Let us analsing informed MM. Under the hypotesis that trader's demand is linear in v: $x = \alpha + \beta v$. So that MM knows optimizazion problem, MM solves:

$$\frac{v - \mu}{2\lambda} = \alpha + \beta v$$

Solving it (similarity method) gives:

$$\alpha = -\frac{\mu}{2\lambda} \qquad \beta = \frac{1}{2\lambda}$$

As liquidity drops, the informed agent trade less, so MM observes y and sets:

$$p = \mathbb{E}\left[v|y\right]$$

In order to solve it, the following property is useful:

Expected value normal variables

If X and Y are bivariate normal variables:

$$\mathbb{E}\left[Y|X=x\right] = \mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x-\mu_x)$$

μ_i is the mean, σ_{XY} is covariance, σ_X variance.

we find:

$$\mathbb{E}\left[v|y\right] = \mathbb{E}\left[v|u+\alpha+\beta v\right]$$

solution is:

$$\alpha = -p_0\sqrt{\frac{\sigma_u^2}{\Sigma_0}} \qquad \mu = p_0 \qquad \lambda = \frac{1}{2}\sqrt{\frac{\Sigma_0}{\sigma_u^2}} \qquad \beta = \sqrt{\frac{\sigma_u^2}{\Sigma_0}}$$

substitute this coefficient in price formula, we obtain:

$$p = p_0 + \frac{1}{2}\sqrt{\frac{\Sigma_0}{\sigma_u^2}}y \tag{II.5}$$

we notice:

- more noise traders imply more liquidity market.
- $p - p_0 \propto y$, smaller λ means a more liquidity market.
- Informed trader trade more if her hide between noise
- Informed trader take profit from noise traders.
- noise traders loose money and MM breaks even

Informed agent volume is:

$$x = (v-p_0)\sqrt{\frac{\sigma_u^2}{\Sigma_0}} \Rightarrow \mathbb{E}\left[\pi\right] = \frac{(v-p_0)^2}{2}\sqrt{\frac{\sigma_u^2}{\Sigma_0}}$$

Strategic model:Kyle (multiple period)

- There are N auctions, each taking place at time $0 = t_0 < \ldots < t_N = 1$
- The liquidation value of the asset is v, normally distributed with mean p_0 and variance Σ_0
- The quantity traded by noise traders in auction n is $\Delta u_n = u_n - u_{n-1}$, where u_n is a Brownian motion with zero mean and instantaneous variance σ_u^2

- x_n is the aggregate position of the informed after the nth auction and $\Delta x_n = x_n - x_{n-1}$ is the quantity traded in this auction

- Each auction is divided in two steps:
  - The informed and the noise traders place the aggregate demand $y = \Delta x_n + \Delta u_n$
  - The market maker sets the liquidation price p_n

The informed trader's trading strategy is a vector of functions $X = \langle X_1, \dots, X_N \rangle$, where:

$$x_n = X_n(p_1, \dots, p_{n-1}, v)$$

while market maker pricing rule is a vector of function $P = \langle P_1, \dots, P_N \rangle$, where:

$$p_n = P_n(x_1 + u_1, \dots, x_n + u_n)$$

The profit of the informed on position acquired at auctions $n, \dots, N$ is:

$$\pi_n = \sum_{k=n}^{N} (v - p_k) x_k$$

A sequential auction equilibrium is a pair X, P such that:

- Profit maximization: $\forall n = 1, \dots, N$ and $\forall X'$ s.t. $X'_1 = X_1, \dots, X'_{n-1} = X_{n-1}$ it is:

$$\mathbb{E}\left[\pi_n(X, P) | p_1, \dots, p_{n-1}, v\right] \geq \mathbb{E}\left[\pi_n(X', P) | p_1, \dots, p_{n-1}, v\right]$$

- Market efficiency: $\forall n = 1, \dots, N$ it is:

$$p_n = \mathbb{E}\left[v | x_1 + u_1, \dots, x_n + u_n\right]$$

A linear equilibrium is a sequential auction equilibrium in which the functions X and P are linear, recursive linear equilibrium is a linear equilibrium such that, given $\exists \lambda_1, \dots, \lambda_N$, $\forall n = 1, \dots, N$:

$$p_n = p_{n-1} + \lambda_n (\Delta x_n + \Delta u_n) \tag{II.6}$$

Equilibrium Linear equilibrium Kyle Model multiple periods

Exist a unique linear equilibrium and this equilibrium is a recursive linear equilibrium. This equilibrium is given by:

$$\begin{aligned}
&\Delta x_n = \beta_n(v - p_{n-1})\Delta t_n \\
&\Delta p_n = \lambda_n(\Delta x_n + \Delta u_n) \\
&\Sigma_n \equiv var[v|\Delta x_1 + \Delta u_1, \dots, \Delta x_n + \Delta u_n] = (1 - \beta_n \lambda_n \Delta t_n)\Sigma_{n-1} \\
&\mathbb{E}\left[\pi_n | p_1, \dots, p_{n-1}, v\right] = \alpha_{n-1}(v - p_{n-1})^2 + \delta_{n-1}
\end{aligned}$$

coefficient are:

$$\begin{aligned}
\alpha_{n-1} &= \frac{1}{4\lambda_n(1 - \alpha_n \lambda_n)} \\
\delta_{n-1} &= \delta_n + \alpha_n \lambda_n^2 \sigma_u^2 \Delta t_n \\
\beta_n \Delta t_n &= \frac{1 - 2\alpha_n \lambda_n}{2\lambda_n(1 - \alpha_n \lambda_n)} \\
\lambda_n &= \beta_n \Sigma_n / \sigma_u^2 \\
\Sigma_n &= (1 - \beta_n \lambda_n \Delta t_n)\Sigma_{n-1}
\end{aligned}$$

Given final condition: $\alpha_n = 0, \delta_n = 0, \lambda_n(1 - \alpha_n \lambda_n) > 0$

Equilibrium Kyle model in continuous time ($N \to \infty$)

In continuous time, the equation for profit, informed order flow and price are:

$$\begin{aligned}
d\pi(t) &= [v - p(t)]dx(t) \\
dx(t) &= \beta(t)[v - p(t)]dt \\
dp(t) &= \lambda(t)[dx(t) + du(t)]
\end{aligned}$$

In the recursive continuous auction equilibrium, if the trading takes place in $t \in [0, 1]$, that is:

$$\begin{aligned}
\lambda(t) &= \sqrt{\Sigma_0 / \sigma_u^2} \\
\Sigma(t) &= (1 - t)\Sigma_0 \\
\beta(t) &= \sigma_u \Sigma_0^{-1/2} / (1 - t) \\
\alpha(t) &= \frac{1}{2}\sqrt{\sigma_u^2 / \Sigma_0} \\
\delta(t) &= \frac{1}{2}\sqrt{\sigma_u^2 \Sigma_0}(1 - t)
\end{aligned}$$

General comments on Kyle model:

- The informed agent splits her order flow in order to hide in the noise trader order flow
- Linear price impact
- Uncorrelated total order flow
- Permanent and fixed impact
- Cariance of fundamental value v declines

2.2.3 Market impact

There are different definition of price impact:

- Impact of an individual market order of size v
- The correlation of the average price change in a given time interval T with the total market order imbalance in the same interval (i.e. the sum of the signed volume $\pm v$ of all individual trades.)
- Cross impact, i.e. how do trades on asset A impact the price of asset B (important for portfolio).
- The impact of a given order of size Q, executed with many trades in a given direction, originating from the same agent.

During the course of the discussion, we will call Large order executed as: Metaorders.
For Kyle model, the average price variation due to a signed volume ϵv is:

$$\Delta p = \lambda \epsilon v$$

λ is the inverse of liquidity and ϵ is ± 1 if it is buyer(seller).
We can show that:

- the impact of single order is linear in volume and permanent:

$$R_{\text{so}}(T) = \mathbb{E}\left[(p_T - p_0) \cdot \epsilon_0\right] = \lambda \mathbb{E}\left[v\right]$$

- the impact of aggregated order flow is linear in the volume imbalance:

$$p_T = p_0 + \lambda \sum_{n=0}^{N-1} \epsilon_n v_n + \sum_{n=0}^{N-1} \eta_n$$

- The price impact of metaorder of total volume Q is linear:

$$R_{\text{mo}}(T|Q) = \mathbb{E}\left[(p_T - p_0) \cdot \epsilon_{\text{mo}} | \sum_{n \in \text{mo} = Q}\right] = \lambda Q$$

- time correlation properties of returns are 'inherited' by order flow

Empirical data shows a sublinear (concave) volume dependence of impact of individual orders:

$$\mathbb{E}\left[\Delta p|v\right] \equiv R_{\text{so}}(T=1|v) \propto v^{\psi}; \qquad \psi \in [0.1, 0.3]$$

or even a logarithmic dependence:

$$R_{\text{so}}(T=1|v) \propto \ln v$$

that is in contrast with Kyle prediction.

Market impact

The expected average price change between beginning and end of a metaorder of size Q, empirically fit by:

$$\Delta \ln p \equiv \mathcal{I}(Q) = \pm Y \sigma_d \left(\frac{Q}{V_D}\right)^{\delta} \tag{II.7}$$

where σ_D is daily volatily, V_D daily volume traded, ± if buy(sell), $\delta \sim 1/2$

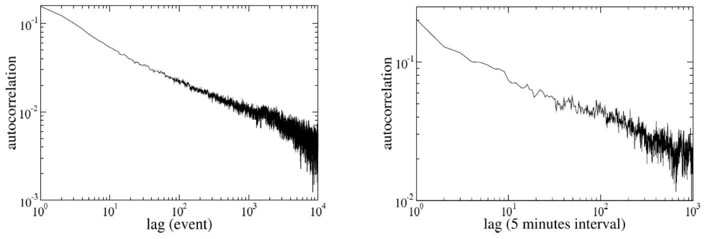
There is a weakness in this formula, it is temporally independent.

2.2.4 Market order flow and Autocorrelation function

In this context, our attention is directed towards orders that initiate transactions, specifically market orders. A buy market order tends to increase the price, while a sell market order tends to decrease it (on average). The flow of market orders is a reflection of the supply and demand for shares in the market. A market order is defined by two key attributes: a volume denoted as v and a sign, represented as $\epsilon = 1$ if buy, -1 if sell.
We are examining the time series in terms of market order time, which means that time progresses by one unit each time a new market order is placed.
Sample autocorrelation function of sign is:

$$C(\tau) = \frac{1}{N}\sum_t \epsilon_t \epsilon_{t+\tau} - \left(\frac{1}{N}\sum_t \epsilon_t\right)^2$$

where N is the length of the time series.



Asymptotically autcorrelation decays as:

$$C(\tau) \sim \tau^{-\gamma} = \tau^{2H-2}$$

H is called Hurst exponent and it is $H \simeq 0.75$

Long memory process

Let $\gamma(k)$ be the autocovariance function of a time series X_t, a process is long memory if in the time $k \to \infty$ it is:

$$\gamma(k) \sim k^{-\gamma} L(k) \qquad \gamma \in (0,1)$$

where $L(k)$ is slowly varying function

Hurst exponent is defined as $H = 1 - \gamma/2$
We can also provide a definition in the frequency domain.

Long memory process (frequency)

In long memory process, the spectral density diverges for low frequencies $\omega \to 0$ as:

$$g(\omega) \simeq \omega^{1-2H} L(\omega)$$

The integrated process is super diffusive: $Var(\sum_{s=0}^{t} X_s) \sim t^{2H}$.
So that $H > 1/2$, process grows really fast (remember, in random walk $Var(X) \sim t$) .
An example is fractional ARIMA (fARIMA) or fractional Brownian motion:

$$(1-L)^d X_t = \epsilon_t \qquad d = H - 1/2$$

Two proposed explanations have emerged for the origin of long-memory in order flow:

- Herding Among Market Participants (LeBaron and Yamamoto, 2007): Agents tend to exhibit herding behavior either because they are responding to the same signals or because they imitate each other's trading strategies. This interaction can be both direct and indirect.

- Order Splitting (Lillo, Mike, and Farmer, 2005): To prevent the disclosure of their true intentions, large investors fragment their trades into smaller pieces and execute them incrementally, a concept originally introduced by Kyle in 1985. This practice converts the heavy tail of large order volume distributions into correlated order flow.

Is it feasible to empirically measure the contributions of herding and order splitting to the autocorrelation observed in order flow? It's worth noting that this inquiry is part of the broader question regarding the origin of the diagonal effect as raised in the work of Biais, Hillion, and Spatt in 1995.
Assuming that we know the identity of the investor placing any market order, for each investor i we define a time series of market order sign ϵ_t^i which is equal to zero if the market order at time t was not placed by investor i and equal to the market order sign otherwise.
Under this condition, autocorrelation can be rewritten as:

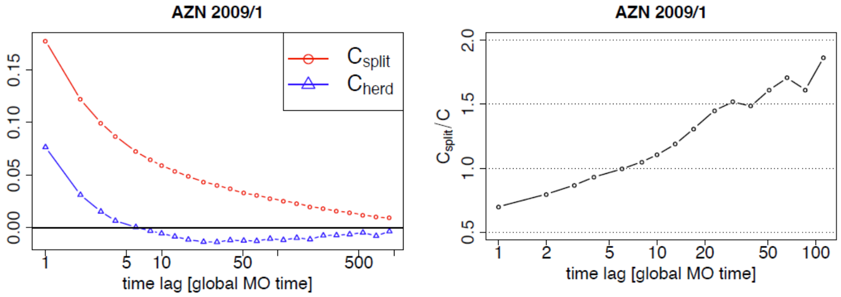
$$C(\tau) = \frac{1}{N} \sum_t \sum_{i,j} \epsilon_t^i \epsilon_{t+\tau}^j - \left(\frac{1}{N} \sum_t \sum_i \epsilon_t^i \right)^2$$

We can rewrrite it as $C(\tau) = C_{\text{split}}(\tau) + C_{\text{herd}}(\tau)$, where:

$$C_{\text{split}}(\tau) = \sum_i \left(\sum_i P^{ii}(\tau) \left[\frac{1}{N^{ii}(\tau)} \sum_t \epsilon_t^i \epsilon_{t+\tau}^i \right] - \left[P^i \frac{1}{N^i} \sum_t \epsilon_t^i \right]^2 \right)$$

$$C_{\text{heard}}(\tau) = \sum_{i \neq j} \left(\sum_i P^{ij}(\tau) \left[\frac{1}{N^{ij}(\tau)} \sum_t \epsilon_t^i \epsilon_{t+\tau}^j \right] - P^i P^j \left[\frac{1}{N^i} \sum_t \epsilon_t^i \right] \left[\frac{1}{N^j} \sum_t \epsilon_t^j \right] \right)$$

N^i is the number of market orders placed by agent i, $P^i = N^i/N$ (probability agent i take ordet at time t), $N^{ij}(\tau)$ is the number of time that an order from investor i at time t is followed by an order from investor j at time $t + \tau$ and $P^{ij}(\tau) = N^{ij}(\tau)/N$.
Empirically, hear decomposition is bigger.



III

Market impact models

A market impact model explains how prices react to trades, or more broadly, to orders placed in the market. If we consider m_n the midprice at time n:

$$\Delta m_n \equiv m_{n+1} - m_n = \\ F(\epsilon_n, \epsilon_{n-1}, \ldots; v_n, v_{n-1}, \ldots; \Omega_n, \Omega_{n-1}, \ldots; \Delta m_{n-1}, \Delta m_{n-2}, \ldots; x_n \ldots) + \eta_n$$

where:

- $\epsilon_n = \pm 1$ is the sign of traders
- v_n is the volume of the trade(s)
- Ω_n is the state of the order book
- x_n are other covariates
- η_n is noise term

The dynamics of the (mid)-price is deterministically given by the flow of all the orders (e.g. limit orders, market orders, and cancellations).

3.1 Response function and lagged return variance

We introduce two diagnostics to test (and calibrate) empirical market impact models:

- The **response function** is defined as:

  $$\mathcal{R}(l) \equiv \mathbb{E}\left[\epsilon_n \cdot (m_{n+l} - m_n)\right]$$

  and measures the expected price shift between n and $n + l$ conditioned to the sign of the trade at time n

- The **lagged return variance** is defined as:

  $$\mathcal{V}(l) \equiv \mathbb{E}\left[(m_{n+l} - m_n)^2\right]$$

  it is the variance of the price on a time scale l.

Plotting $\mathcal{V}/l$ versus l, we obtain signature plot, which is expected to tend to a constant if $l \to \infty$ (long term volatility)

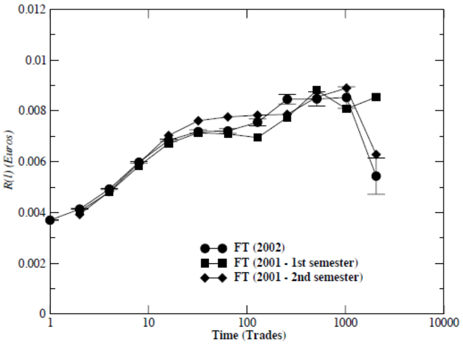


Figure: Average empirical response function $\mathcal{R}(\ell)$ for FT, during three different periods (first and second semester of 2001 and 2002). We have given error bars for the 2002 data. For the 2001 data, the $y-$axis has been rescaled such that $\mathcal{R}(1)$ coincides with the 2002 result. $\mathcal{R}(\ell)$ is seen to increase by a factor ~ 2 between $\ell = 1$ and $\ell = 100$.

3.1.1 A firs fixed permanent impact model

Fixed permanent impact model

- r_n is the midquote price change between just before the nth trade and just before the $n+1$th trade.
- Immediate impact, $\mathbb{E}\left[r_n|\epsilon_n v_n\right]$, is non zero and can be written as $\mathbb{E}\left[r_n|\epsilon_n v_n\right] = \epsilon f(v)$, where f is a function that grows with v
- Impact of a transaction is permanent, like in usual random walks, and the equation for the midquote price m_n at time n is:

  $$r_n = m_{n+1} - m_n = \epsilon_n f(v_n; \Omega_n) + \eta_n$$

  where η_n is random term describing price changes not directlu attrivuted to trading itself (for example news). We assume η_n is indepndent on the order glow and we set $\mathbb{E}\left[\eta\right] = 0$ and $\mathbb{E}\left[\eta^2\right] = \Sigma^2$

- We have incorporated the idea that the impact of a trade can be influenced by the current state of the order book Ω_n. This inclusion is justified on general principles: when a market order with a volume v_n interacts with a substantial queue of limit orders, it typically exerts only a minimal impact on the price. Conversely, one can anticipate a strong correlation between the state of the order book, Ω_n, and the magnitude of incoming market orders: larger volumes of limit orders tend to attract larger market orders.

We can express the midquote price as:

$$m_n = \sum_{k<n} \epsilon_k f(v_k; \Omega_k) + \sum_{k<n} \eta_k$$

This equation show a non-decaying nature of the impact, $\epsilon_k \partial m_n / \partial v_k$ $(k < n)$ does not decay as $n - k$ grows.

Lagged impact function $\mathcal{R}(l)$ and lagged return variance $\mathcal{V}(l)$ is:

$$\mathcal{R}(l) = \mathbb{E}\left[f\right] \qquad \mathcal{V}(l) = \left(\mathbb{E}\left[f^2\right] + \Sigma^2\right) l$$

constant price impact and pure price diffusion, close to what is indeed observed empirically on small tick, liquid contracts.

At the same time, the autocovariance of price returns:

$$\mathbb{E}\left[r_n r_{n+\tau}\right] \propto \mathbb{E}\left[\epsilon_n \epsilon_{n+\tau]}\right] \sim \tau^{-\gamma}$$

with γ small. That means that price returns are strongly autocorrelated in time. But this would violate market efficiency hypothesis, in fact price returns would be easy predictable.

So empirically observed long memory of order flow is incompatible with the random walk model if price are efficient.

"How can the market be statistically efficient (i.e. unpredictable) in the presence of an autocorrelated order flow? "

3.2 Madhavan, Richardson and Roomans (MRR) model

MRR model

- all trades have the same volume $v_n = v$ and the ϵ_n's are generated by a Markov process with correlation ρ, thus $\mathbb{E}\left[\epsilon_n | \epsilon_{n-1} = \rho\epsilon_{n-1}\right]$
- correlations decay exponentially fast, i.e. $C(l) = \mathbb{E}\left[\epsilon_i \epsilon_{i+l}\right] = \rho^l$ which does not conform to reality.

- The MRR model postulates that the mid-point m_n evolves only because of unpredictable external shocks (or news) and because of the surprise component in the order flow. This postulate of course automatically removes any predictability in the price returns and ensures efficiency:

  $$m_{n+1} - m_n = \theta[\epsilon_n - \rho\epsilon_{n-1}] + \eta_n$$

  where η iss shock component, θ measure the size of trade impact

We can rewrite price return as:

$$m_{n+l} - m_n = \sum_{j=n}^{n+l-1} \eta_j + \theta \sum_{j=n}^{n+l-1} [\epsilon_j - \rho\epsilon_{j-1}]$$

The full impact function is found to be constant, equal to:

$$\mathcal{R}(l) = \theta(1-\rho^2), \qquad \forall l$$

We can introduce the function $G(l)$ thet measure the influence of a single trade at time $n-l$ on the mid-point at time n:

$$m_n = \sum_{j=-\infty}^{n-1} \eta_j + \sum_{j=-\infty}^{n-1} G(n-j-1)\eta_j \tag{III.1}$$

in this way we can impose $G(l=0) = \theta$ and $G(l \geq 1) = \theta(1-\rho)$: a part of $\theta\rho$ of impact instantaneously decays to zero after first trade, whereas the rest of the impact is permanent.
The volatility in this model can be written as:

$$\mathcal{V}(l) = [\theta^2(1-\rho^2) + \Sigma^2]l$$

The MRR model with bid-ask spread

- The original MRR model assumes that it is the ‘true’ fundamental price p_n, rather than the midpoint m_n, which is impacted by the surprise in order flow, and hence:

  $$p_{n+1} - p_n = \eta_n + \theta[\epsilon_n - \rho\epsilon_{n-1}]$$

- Market Makers cannt guess the surprise of the next trade, so bid and ask price are given by:

  $$a_n = p_n + \theta[1 - \rho\epsilon_{n-1}] + \phi \qquad b_n = p_n + \theta[-1 - \rho\epsilon_{n-1}] - \phi$$

  where ϕ is extra compension due to shock component risk.

- The midpoint $m \equiv= (a+b)/2$ immediately before the nth trade is:

  $$m_n = p_n - \theta\rho\epsilon_{n-1}$$

  whereas the spread is given by $S = a - b = 2(\theta + \phi)$

If we neglecting ϕ for arbitrary correlations between signs:

$$m_{n+l} - m_n = \sum_{j=n}^{n+l-1} \eta_n + \theta \sum_{j=n}^{n+l-1} \left\{ \epsilon_j - \mathbb{E}\left[\epsilon_{j+1}\right]_j \right\}$$

Than, the impact function is:

$$\mathcal{R}(l) = \theta[1 - C(l)]$$

So that $\mathcal{R}(\infty) = \theta = S/2$, the long term profit of market makers is zero (Spread and impact are two sides of the same coin).
Long time impact is enhanced compared to the short term impact by factor:

$$\lambda \equiv \frac{\mathcal{R}(\infty)}{\mathcal{R}(1)} = \frac{1}{1 - C(1)} > 1$$

If $\phi \neq 0$:

$$S = 2(\theta + \phi) = 2(\mathcal{R}(\infty) + \phi) = 2\lambda\mathcal{R}(1) + 2\phi$$

where $\lambda = (1-\rho)^{-1}$.
We can write mid-point volatility on scale l:

$$\sigma_l^2 = \frac{1}{l}\mathbb{E}\left[(m_{l+i} - m_i)^2\right]$$

and it is the sum of trade induced volatility $\theta^2(1-\rho)^2$ and a news induced volatility Σ^2:

$$\sigma_1^2 = \mathbb{E}\left[(m_{n+1} - m + n)^2\right] = \Sigma^2 + \theta^2(1-\rho)^2$$
$$\sigma_\infty^2 = \Sigma^2 + \theta^2(1-\rho)^2\left(1 + 2\frac{\rho}{1-\rho}\right) = \Sigma^2 + \theta^2(1-\rho^2) \geq \sigma_1^2$$

The MRR model leads to two simple relations between spread, impact and volatility per trade:

$$S = 2\lambda\mathcal{R}(1) + 2\phi \qquad \sigma_1^2 = (1)^2 + \Sigma^2$$

With empirical data, it fits quite well.
Some general comments:

- The linear relationship between the spread and volatility per trade is not anticipated to remain valid when considering the volatility per unit of time σ. This is because the latter incorporates an additional factor, which is both stock-dependent and time-dependent, namely the trading frequency denoted as f which can be expressed as:

  $$\sigma = \sigma_1\sqrt{f}$$

- There are two complementary economic interpretations of the relation $\sigma_1 \sim S$ in small tick markets
  - Given that the typical liquidity available in the order book is relatively limited, market orders often capture a significant portion of the volume at the best available price. Additionally, it's observed that the size of the "gap" above the ask or below the bid is roughly comparable in magnitude to the bid-ask spread itself. Consequently, both the market impact and the volatility per trade are expected to be of the same order of magnitude as the stock price, denoted as "S," as has been empirically observed.
  - This relationship can also be interpreted inversely, where when the volatility per trade is high, the risk associated with placing limit orders increases. Consequently, the spread tends to widen until it becomes advantageous to use limit orders.

Let us continue the discussion on long-term memory. We have said that long term memory of trades is a priori paradoxical and hints towards a non trivial property of financial markets, which can be called long-term resilience.
Using equation III.1 and assuming that single trade impact is lag independent $G(l) = G$ and that volume fluctuations can be neglected, mid price variance can be computed as:

$$\mathcal{V}(l) \equiv \mathbb{E}\left[(m_{n+l} - m_n)^2\right] = \underbrace{\left[\Sigma^2 + G^2\right]}_{\text{diffusive process}} l + 2G\sum_{j=1}^{l}(l-j)C(j)$$

When $\gamma < 1$, the second therm of RHS can be approximated $(l\ell 1)$ by $2c_0 G l^{2-\gamma}/(1-\gamma)(2-\gamma)$, which grows faster than the first term.
The price would super diffuse, at long times, trend with a volatility diverging with the lag l. This phenomena does not occur, in fact the market reacts to trade correlations so as to prevent the occurrence of such trends.

3.3 Transient impact model (TIM)

Propagator model (Bouchaud 2004)

$$m_t = \sum_{t'<t}[G(t-t')\epsilon_{t'} + \eta_{t'}] + m_{-\infty} \tag{III.2}$$

The term $G(t - t')$ avoid super-diffusivity.
In differential form, setting $r_t = m_{t+1} - m_t$:

$$r_t = G(1)\epsilon_t + \sum_{t'<t}\mathcal{G}(t-t)\epsilon_{t'} + \eta_t \qquad \mathcal{G}(l) \equiv G(l+1) - G(l) \tag{III.3}$$

where $G(l \leq 0) \equiv 0$

In this setting, past order flow affects future returns, and we do not require efficiency (martingale assumption).
The lagged price variance can be computed as:

$$\mathcal{V}(l) = \sum_{0 \leq j < l} G^2(l-j) + \sum_{j>0} [G(l+j) - G(j)]^2 + 2\Delta(l) + \Sigma^2 l$$

where $\Delta(l)$ is the correlation induced contribution:

$$\begin{aligned}\Delta(l) = &\sum_{0 \leq j < k < l} G(l-j)G(l-k)C(k-j) \\ &+ \sum_{0 < j < k} [G(l+j) - G(j)][G(l+k) - G(k)]C(k-j) \\ &+ \sum_{0 \leq j < l} \sum_{k>0} G(l-j)[G(l+k) - G(k)]C(k+j)\end{aligned}$$

Assuming $G(l)$ decays at large l as $G(l) \sim \Gamma_0 l^{-\beta}$, if $\beta, \gamma < 1$, than:

$$\Delta(l) \sim \Gamma_0^2 c_o I(\gamma, \beta) l^{2-2\beta-\gamma}$$

where $I > 0$ is a numerical integral. General comments on this model:

- if single trade impact does not decay ($\beta = 0$), we have superdiffusive result.
- As impact decays faster, superdiffusion is reduced.
- At a critical value $\beta_c = (1-\gamma)/2$, grows linearly with l and contributes to the long term value of the volatility
- If $\beta > \beta_c$, $\Delta(l)$ grows sublinear with l, impact enhances high frequency value compared to long term Σ^2
- Long range correlation in order flow grows not induce long term correlations nor anticorrelations in price returns iff the impact of single trades is transiet ($\beta > 0$) but itself non-summable ($\beta < 1$)

Continuous time version of the TIM (Gatheral 2010)

Let us consider an unperturbed martingale dynamics of the price:

$$dS_t^0 = \sigma_t dW_t$$

where σ_t is volatility and W_t is Wiener process.
When trading $dX_t = \dot{X}_t dt$ shares in the interval $[]t, t+dt]$, the price follows this eqution:

$$S_t^X = S_0 + \int_0^T f(\dot{X}_s)G(t-s)ds + \int_0^t \sigma_s dW_s \tag{III.4}$$

$f(\dot{X}_s)$ is velocity function and $G(t-s)$ evaluate how the pass influence the future

If $f(x) = kx$, the settinf is the same as TIM.
The avarage impact function $\mathcal{R}(l)$ of model is:

$$\mathcal{R}(l) = G(l) + \sum_{0<j<l} G(l-j)C(j) + \sum_{j>0}[G(l+j) - G(j)]C(j)$$

So that $C(n)$ and $\mathcal{R}(l)$ are measurable quantity, from them we can evaluate $G(l)$.
An alternative method less sensitive to boundary effects is using response function $\mathcal{S}(l) = \mathbb{E}\left[r_{r+l} \cdots \epsilon_t\right]$ and $C(l)$:

$$\mathcal{S}(l) = \sum_{n\geq 0} \mathcal{G}(n)C(n-l)$$

whose solution represents the values of the kernel $\mathcal{G}(l)$.
Exist a relationship beween response and impact function:

$$\mathcal{R}(l) = \sum_{0\leq i<l} \mathcal{S}(i)$$

Asymptotically, when $G(l)$ decays as $\Gamma_0 l^{-\beta}$, if $\beta + \gamma < 1$:

$$\mathcal{R}(l) \sim l^{1-\beta-\gamma}$$

- $\beta < \beta_c$: $\mathcal{R}(l)$ diverges to $+\infty$ for large l
- $\beta > \beta_c$: $\mathcal{R}(l)$ diverges to $-\infty$ This means that when the decay of single trade impact is too fast, the accumulation of meatn reverting effects leads to a negative long term average impact.
- $\beta = \beta_c$ $\mathcal{R}(l) \to \mathcal{R}(\infty)$, the decay of single trade impact precisely offsets the positive correlation of the trades

IV

Optimal execution

4.1 Introduction

The investor's objective is to trade a specific number of shares while minimizing costs through incremental trading.
This process can be decomposed into three scales:

- **Portfolio Manager's Decision**: Initially, the portfolio manager determines how to split the order across different trading days.
- **Trader's Decision for Each Day**: For each trading day, the trader further divides the day into "macroscopic" intervals, such as 5 or 15 minutes, and decides how much to trade during each of these intervals.
- **Trading Strategy at Each Interval**: Finally, within each interval, the trader must decide on the type of orders to use (e.g., limit versus market orders) and the specific strategy to employ. This includes considerations like when to cross the spread if the price moves adversely.

The current focus is on the second level of optimization, which involves determining the trading quantities within each trading interval. This optimization can occur in both discrete time (with data calibration) and continuous time.
An investor wants to buy q_0 shares. dq_t is the number of shares traded in $[t; t+dt]$ and the price at time t is S_t.
Brokers offer their clients a wide range of services for buying or selling stocks. In addition to providing direct market access (DMA), they typically offer various trading strategies, which can be grouped into five main categories:

- Implementation Shortfall (IS) orders (aka Arrival Price orders)
- Target Close (TC) orders
- Percentage Of Volume (POV) orders

- Volume-Weighted Average Price (VWAP) orders
- Time-Weighted Average Price (TWAP) orders

Problem optimal execution setting

- $t \in [0, T]$: time interval of execution (that is not so banal)
- q_t: asset position at time t. We want to find the optimal position.
- $q_0 > 0$ shares to buy
- $q_T = 0$
- $dq_t = v_t dt$, with v_t trading velocity
- $S^0 = (S^0_t)_{t \geq 0}$ exogenously given asset price dynamics, here assumed to be a martingale on a probability space
- $S = (S_t)_{t \geq 0}$ asset price dynamics when the strategy $(q_t)_{t \geq 0}$ is used.

Consider the objective function at different setting:

Objective function Implementation Shortfall (IS) order

Fix the execution time interval $[0, T]$, the objective function to minimize is the expectation of:

$$\int_0^T \underbrace{S_t dq_t}_{\text{how much I spend to buy}} - \underbrace{q_0 S_0}_{\text{price at time 0}}$$

Objective function Target Close (TC) order

Fix the execution time interval $[0; T]$,the benchmark price is the closing price (typically unknown at the start of the execution), thus minimize the expectation of

$$\int_0^T S_t dq_t - q_0 S_{\text{close}}$$

Objective function Percentage of Volume (POV) orders

Traded volume as close as possible to a fixed percentage of the volume traded in the market. Fix the participation rate, the objective is to minimize the expected cost (traded volume is aleatory). Used to execute large blocks while following market flow.

Objective function Volume-Weighted Average Price (VWAP) orders

Obtain a price as close as possible to the average price over a given period of time. Benchmark for traders who buy or sell shares in line with their global investment strategies or to hedge a risky position.
Fix the execution time interval $[0;T]$, Let dVt the volume traded by the market in $[t;t+dt]$. The VWAP price is:

$$VWAP_0^T = \frac{\int_0^T S_t dV_t}{\int_0^T dV_t}$$

the objective is to minimize the expectation of:

$$\int_0^T S_T dq_t - q_0 VWAP_0^T$$

Time-Weighted Average Price (TWAP) orders. Similar to VWAP, with the assumption $V_t = V =$ const.

4.2 Almgren and Chriss discrete time

Almgren and Chriss discrete time

- An investor has q_0 shares to buy in N time periods. Let $v_k (k = 1, \dots, N)$ be the (signed) number of shares to be traded in interval k. Let $\tilde{p}_k$ be the price at which the investor trades at interval k (in general different from the average price in the interval p_k) and p_0 the price before the start of the execution

- A very used objective function is the Implementation Shortfall (IS) defined as:

$$C(\mathbf{v}) \equiv \sum_{k=1}^{N} v_k \tilde{p}_k - q_0 p_0$$

  in words: the difference between the cost and the cost in an infinitely liquid market.

- The implementation shortfall is in general a stochastic variable, therefore one often wants to minimize $\mathbb{E}[C(v)]$. This assumes a risk neutral profile.

- The mid-price of the stock at step k is equal to the previous price, plus a linear market impact and a random shock:

  $$p_k = p_{k-1} + \theta v_k + \eta_k \qquad \eta \sim i.i.d(0,\sigma)$$

  We consider effective price:

  $$\tilde{p}_k = p_k + \underbrace{\rho v_k}_{\text{temporary impact}} + \underbrace{\text{sign}(v_k) \cdot S/2}_{\text{spread}}$$

The equation for the executive costs is:

$$C(\mathbf{v}) = \sum_{k=1}^{N} v_k \tilde{p}_k - q_0 p_0 = \sum_{k=1}^{N} (\eta_k + \theta v_k) \sum_{k=1}^{k} v_j + \sum_{k=1}^{N} (\text{sign}(v_k) S/2 + \rho v_k) v_k$$

and the expecred value of the cost to be minimized is:

$$\mathbb{E}\left[C(\mathbf{v})\right] = \frac{\theta}{2} q_0^2 + (\rho + \theta) \sum_{k=1}^{N} v_k^2 + \theta \sum_{i \neq j} v_i v_j + S/2 \sum_{k=1}^{N} |v_k|$$

Under constraint:

$$\sum_{k=1}^{N} v_k = q_0$$

(I assume no mixed strategy: buy or sell, not a mix)

Optimal execution (simple) AC solution

Given the symmetry of the problem, the solution that minimizes the expected impact costs is:

$$\mathbf{v}^* \equiv \arg\min_{v} \mathbb{E}\left[C(\mathbf{v})\right] = \left(\frac{q_0}{N}, \frac{q_0}{N}, \ldots, \frac{q_0}{N}\right)^T$$

The solution is just trading at constant rate over the periods

If I add drift price:

$$p_k = p_{k-1} + \mu + \theta v_k + \eta_k \qquad \eta \sim i.i.d(0,\sigma)$$

Optimal execution (with drift) AC solution

Minimization of implementation shortfall gives:

$$v_k^* \propto q_0 \left(\frac{1}{N} + \frac{(N+1) - 2k}{2(2\rho + \theta)\mu} \right)$$

If drift is positive, we will accelerate a buy order. The amount of the acceleration depends positively on μ, of course, but it also depends inversely on ρ and θ.
A criticism of this model is that is static, market condition can change.

AM model consider also risk aversion for optimal execution, they minimize the sum of expected cost and costs'risk.
Using Markowitx portfolio optimization theory, optimal trading solution is:

$$\arg\min_v (\mathbb{E}\left[C(\mathbf{v})\right] + \lambda Var[C(\mathbf{v})])$$

The set of solutions to this problem for different values of λ is called "optimal frontier".
The variance of execution cost is:

$$Var[C(\mathbf{v})] = \mathbb{E}\left[(C(\mathbf{v} - \mathbb{E}\left[C(\mathbf{v})\right])^2\right] = \mathbb{E}\left[\left(\sum_{k=0}^{N-1} \eta_k \sum_{j=k+1}^{N-1} v_j\right)^2\right]$$

Assuming η_k independent, we obtain:

$$Var[C(\mathbf{v})] = \sigma^2 \sum_{k=0}^{N-1} \left(\sum_{j=k+1}^{N-1} v_j\right)^2$$

Using impact model, we can evaluate the midprice:

$$v_k = A\cosh(\beta(N-k))$$

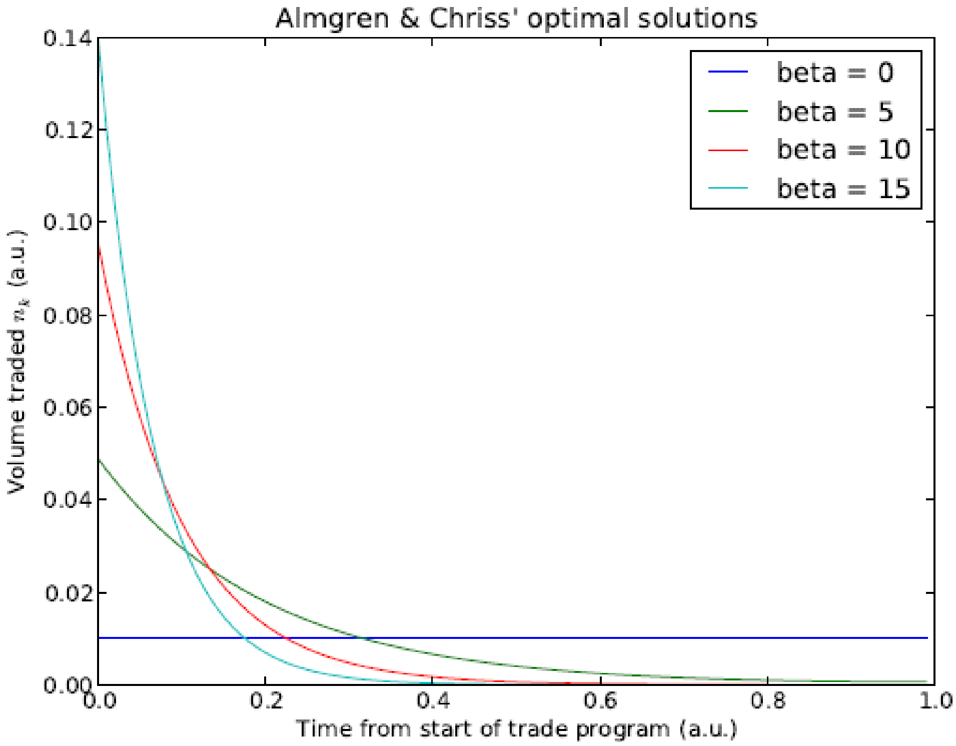
where A is a normalized constant and β solve the equation:

$$2[\cosh(\beta) - 1] = \frac{\lambda\sigma^2}{\rho + \theta}$$

Inverting and taking continuous time limit ($\sigma \to 0$):

$$\beta \simeq \sqrt{\frac{\lambda\sigma^2}{\rho + \theta}}$$

In word it could be translated as risk (volatility) / impact(liquidity).



4.3 Almgren and Chriss continuous time

The limitation of discrete model is:

- Non-linear temporary impact
- Non-linear permanent impact
- Benchmarks different from IS
- Utility function instead of mean-variance optimization
- Static vs dynamic strategies
- Multiple assets

to handle these, we use continuous setting

AC in continuous time

- Sell q_0 shares, $dq_t = v_t dt$
- Set of admissible strategies:
$$\mathcal{A} = \left\{ (v_t)_{t\in[0,T]} : \int_0^T v_t d_t = -q_0; \int_0^T |v_t| dt \text{ a.s. bounded } rv \right\}$$
- $dS_t = \sigma dW_t + kv_t dt$ usually price is given by geometric brownian motion, in this setting we have choosen arithmetic one because is easier and in intra-day arithmetic $\sim$ geometric
- $V_t d_T$ volume traded by other agents is $[t, t+dt]$: we consider it deterministic, continuous, positive and bounded
- X_t cash account:
$$dX_t = -v_t \left(S_t + g\left(\frac{v_t}{V_t}\right)\right) dt = -v_t S_t dt + V_t L\left(\frac{v_t}{V_t}\right) dt$$
here we consider $L(\rho) \equiv \rho g(\rho)$ s.t. $L(0) = 0$, it is convex and $\lim_{\rho\to+\infty} = \frac{L(\rho)}{\rho} = +\infty$. For AC we consider:
$$L(\rho) = \eta\rho^2$$

Permanent impat is linear in AC like model

Given the model (no temporary impact):
$$\begin{cases} dq_t = v_t dt \\ S_t = \sigma dW_t + f(v_t) dt \\ dX_t = -v_t S_T dt \end{cases}$$
there is a dynamic arbitrage if there exists $(v_t)_t$ and $t_1 < t_2$ s.t:
$$\int_{t_1}^{t_2} (|v_t| + |f(v_t)|) dt \in L^\infty(\Omega) \qquad \text{(boundedness)} \tag{IV.1}$$
$$\int_{t_1}^{t_2} v_t dt = 0 \quad \text{exist a round-trip: buy and sell same quantity]} \tag{IV.2}$$
$$\mathbb{E}\left[X_{t_2}|\mathcal{F}_{t_1}\right] > X_{t_1} \quad \text{(dynamic arbitrage)} \tag{IV.3}$$
In words: exist a round trip strategy that is profitable on average

Let us introduce the utility function.

Utility function

There exists an increasing and concave function $u(x)$ (utility function) such that a lottery X (real-valued random variable modeling the outcome of the gamble) is prefered to a lottery $Y \iff \mathbb{E}\left[u(X)\right] \geq \mathbb{E}\left[u(Y)\right]$

Jensen inequality

If u is concave, than:

$$\mathbb{E}\left[u(X)\right] \leq u(\mathbb{E}\left[X\right])$$

Absolute risk aversion function

Let u be a twice differentiable utility function s.t. $u' > 0$. We define the absolute risk aversion $\gamma(x)$ as:

$$\gamma(x) = -u''(x)/u(x)$$

which is invariant by increasing affine transformation of u

Constant absolute risk aversion (CARA)

CARA utility function is $u(x)$ s.t. $\gamma(x)$ is constant, so:

$$u(x) = -\exp(-\gamma x)$$

for $\gamma > 0$ and $u(x) = x$ for $\gamma = 0$

CRRA utility function

CRR utility function is:

$$u(x) = x^{1-\rho}/(1-\rho)$$

if $\rho \to 1$:

$$u(x) = \log x$$

Expected value CARA function

Let X be a real-valued Gaussian random variable $N(\mu, \sigma^2)$. Let $\gamma > 0$, then:

$$\mathbb{E}\left[-\exp(-\gamma X)\right] = -\exp\left(-\gamma\mu + \frac{1}{2}\gamma^2\sigma^2\right)$$

Cetainity equivalent

Let u be a continuous and increasing utility function. Let X be a real-valued random variable s.t. $X, u(X) \in L^1(\Omega)$. The certainty equivalent of X for utility function u is the unique $e \in \mathbf{R}$ s.t.:

$$\mathbb{E}\left[u(X)\right] = u(e)$$

The certainty equivalent of a lottery is the risk-free payoff that has the same expected utility as the lottery. If utility function is CARA with absolute risk aversion γ, the certainity equivalent is:

$$e = \mu - \frac{1}{2}\gamma\sigma^2$$

Deterministic strategies

Let us assume CARA risk averse investor, we have to maximize:

$$\mathbb{E}\left[-\exp(-\gamma X_T)\right]$$

If we restrict to deterministic strategies:

$$\mathcal{A}_{det} = \left\{(v_t)_{t\in[0,T]} \in \mathcal{A}, \forall t \quad v_t \text{ is } \mathcal{F}_0 - \text{measurable}\right\}$$

Terminal cash is:

$$X_t = X_0 + q_0 S_0 - \frac{k}{2}q_0^2 + \sigma\int_0^T q_t dW_t - \int_0^T V_t L\left(\frac{v_t}{V_t}\right)dt$$

assuming $c_t \in \mathcal{A}_{det}, X_t \sim \mathcal{N}(\mu_x, \sigma_X^2)$, than:

$$\mu_x = X_0 + q_0 S_0 - \frac{k}{2}q_0^2 - \int_0^T V_t L\left(\frac{v_t}{V_t}\right)dt$$

$$\sigma_X^2 = \sigma^2\int_0^T q_t^2 dt$$

Let consider two cases:

- $\gamma = 0$(risk neutral investor), $\mathbb{E}\left[X_T\right] = \mu_X$, can be minimized by $v_t \propto V_t$ due to convexity of L (due to Jensen inequality)

- $\gamma > 0$(risk averse investor) we can use Laplace trasform of the Gaussian and the problem can be rewritten as the minimization of the function $q(t)$:

$$J(q) = \int_0^T \left(V_t L\left(\frac{q'(t)}{V_t}\right) + \frac{\gamma}{2}\sigma^2 \int_0^T q^2(t)dt \right) dt$$

  in the set $\mathcal{C}$ of absolutely continuous functions $q(t)$ in $[0,T]$ with $q(0) = q_0$ and $q(T) = 0$. This problem is known as Bolza problem

Existing unique minimizer Bolza problem

There exists a unique minimizer q^* of the function J over the set $\mathcal{C}$, q^* is a monotone funciton:

- if $q_0 \geq 0$, q^* is a nonincreasing function of time
- if $q_0 \leq 0$, q^* is a nondecreasing function of time

In word: I cannot have a mixing buy-selling strategies

4.3.1 Lagrangian setting

We can translate this optimization problem in a Hamiltonian setting; Hamiltonian equations are:

$$\begin{cases} p'(t) = \gamma\sigma^2 q(t) \\ q'(t) = V_t H'(p(t)) \\ q(0) = q_0 \\ q(T) = 0 \end{cases}$$

Where H is the Legendre-Frenchel transform:

$$H(p) = \sup_\rho \rho p - L(\rho)$$

If L has continuous first two partial derivatives wrt all its arguments, one can introduce the Lagrangian:

$$\mathcal{L}(t, q(t), q'(t)) = V_t L\left(\frac{q'(t)}{V_t}\right) + \frac{\gamma}{2}\sigma^2 q^2(t)$$

and using Eular-Lagrande equation:

$$\frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial q'} = 0$$

with bounday conditions: $q(0) = q_0$ and $q(T) = 0$

Lagrange solution AC setting

In AC setting, $L(\rho) = \eta\rho^2$, than $H(p) = \frac{p^2}{4\eta}$, using Hamiltonian system and Euler-Langrange equation, we obtain:

$$q''(t) = \frac{\gamma\sigma^2 V_t}{2\eta} q(t)$$

Special case $V_t = V$, solution is:

$$q^*(t) = q_0 \frac{\sinh\left(\sqrt{\frac{\gamma\sigma^2 V}{2\eta}}(T-t)\right)}{\sinh\left(\sqrt{\frac{\gamma\sigma^2 V}{2\eta}}T\right)}$$

solution doesn not depend on impact coefficient k

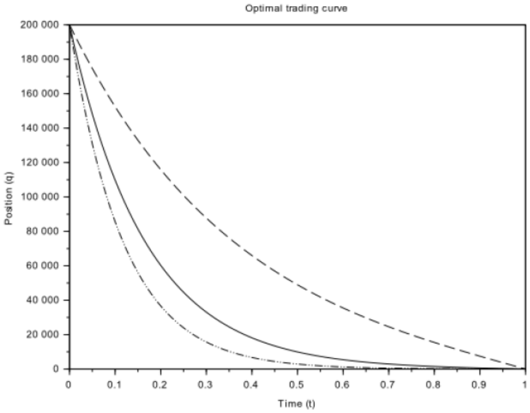


FIGURE 3.1: Optimal trading curve for $q_0 = 200{,}000$ shares over one day ($T = 1$), for different values of γ. Dash-dotted line: $\gamma = 10^{-5}$ €$^{-1}$. Solid line: $\gamma = 5.10^{-6}$ €$^{-1}$. Dashed line: $\gamma = 10^{-6}$ €$^{-1}$.

Let us analyse Stochastic strategies:

Stochastic Strategies

$$\sup_{v\in\mathcal{A}} \mathbb{E}\left[-\exp(-\gamma X_T)\right] = \sup_{v\in\mathcal{A}_{det}} \mathbb{E}\left[-\exp(-\gamma X_T)\right]$$

In CARA case, stochastic case does not have advantage on deterministic

Adding a participation rate constraint:

$$\mathcal{A}_{\rho_{\max}} = \left\{(v_t)_{t\in[0,T]}\text{prog. meas: } \int_0^T v_t dt = -q_0; |v_t| \leq \rho_{\max} V_t\right\}$$

We can use a trick in order to solve this problem: constraintthe objective function

$$J(q) = \int_0^T \left(V_t L_{\rho_{\max}}\left(\frac{q'(t)}{V_t}\right) + \frac{\gamma}{2}\sigma^2 \int_0^T q^2(t)dt\right) dt$$

where:

$$L_{\rho_{\max}} = \begin{cases} L(\rho) & |\rho| \leq \rho_{\max} \\ +\infty & |\rho| > \rho_{\max} \end{cases}$$

Solution AC adding participation rate constraint

There exists a unique minimizer q^* of the functional J over the set $\mathcal{C}$. If $q_0 \geq 0$, then q^* is a nonincreasing function of time. If $q_0 \leq 0$, then q^* is a nondecreasing function of time. Furthermore, q^* is uniquely characterized by:

$$\begin{cases} p'(t) = \gamma\sigma^2 q(t) \\ q'(t) = V_t H'_{\rho_{\max}}(p(t)) \\ q(0) = q_0 \\ q(T) = 0 \end{cases}$$

where $H_{\rho_{\max}}$ is the Legendre-Frenchel transform:

$$H_{\rho_{\max}}(p) = \sup_{|\rho| \leq \rho_{\max}} \rho p - L(\rho)$$

Let us discretization the Hamiltonian system: if $t_0 = 0 < \ldots < t_n = n\tau < \ldots < t_N = N\tau = T$, the discretized system is:

$$\begin{cases} p_{n+1} = p_n + \tau\gamma\sigma^2 q_{n+1}, & 0 \leq n < N-1 \\ q_{n+1} = q_n + \tau V_{n+1} H'(p_n), & 0 \leq n < N \\ q_0 = q_0 \\ q_N = 0 \end{cases}$$

We can solve this problem through fixed-point approach:

$$\begin{cases} p^\lambda_{n+1} = p^\lambda_n + \tau\gamma\sigma^2 q^\lambda_{n+1}, & 0 \leq n < N-1 \\ q^\lambda_{n+1} = q^\lambda_n + \tau V_{n+1} H'(p^\lambda_n), & 0 \leq n < N \\ q^\lambda_0 = q_0 \\ q^\lambda_N = 0 \end{cases}$$

can be solved through bisection.

4.3.2 AC multi-asset portfolio

AC model for a multi-asset portfolio

If a trader wants to trade simultaneously a set of $d > 1$ assets:

- **Number of shares** $q_t^i = q_0^i - \int_0^t v_s^i ds$
- **Price** $dS_t^i = \sigma^i dW_t^i - k^i v_t^i dt$. Σ is the covariance matrix of $\{\sigma^i W_t^i\}$
- **Cash** $dX_t = \sum_{i=1}^d v_t^i S_t^i dt - V_t^i L^i \left(\frac{v_t^i}{V_t^i}\right) dt$

There is no cross-impact term, but the "interaction" between stocks is only through the covariance of prices.

The value of terminal cash is:

$$X_T = X_0 + \sum_{i=1}^d q_0^i S_0^i - \sum_{i=1}^d \frac{k_i}{2}(q_0^i)^2 + \sum_{i=1}^d \int_0^T q_t^i \sigma^i dW_t^i - \sum_{i=1}^d \int_0^T V_t^i L^i \left(\frac{v_t^i}{V_t^i}\right) dt$$

The optimization problem is:

$$\sup_{(v_t)_t \in \mathcal{A}} \mathbb{E}\left[-\exp(-\gamma X_T)\right]$$

As in the single-asset case, deterministic strategies are optimal.
Using Hamiltionian setting, minimization problem is:

$$J(q) = \int_0^T \left(\sum_{i=1}^d V_t^i L^i \left(\frac{(q^i(t))'}{V_t^i} \right) + \frac{\gamma}{2} q(t) \cdot \Sigma q(t) \right)$$

Hamiltonm equation:

$$\begin{cases} p'(t) = \gamma \Sigma q(t) \\ (q^i(t))' = V_t^i (H(p^i(t)))' \\ q(0) = q_0 \\ q(T) = 0 \end{cases}$$

with

$$H^i(p) = \sup_{\rho} \rho p - L^i(\rho)$$

4.3.3 Hedging a liquidation with optimal execution of another asset

A trader could want to mitigate the risk linked to the execution process by introducing a second asset correlated with the first one. This approach involves solving a multi-asset liquidation problem, starting

with an initial position of $(q_0, 0)$. For instance, we can assume the presence of another asset denoted by the subscript "h" that has no associated execution costs, such as a futures contract.
The problem now is to minimize the functional:

$$J(q(t), q_h(t)) = \int_0^T \left(V_t L\left(\frac{q'(t)}{V_t}\right) + \frac{\gamma}{2}\left(\sigma^2 q^2(t) + 2\rho\sigma\sigma_h q(t)q_h(t) + \sigma^2 q_h^2(t)\right)\right) dt$$

The optimal hedge consist in setting $q_h(t) = -\rho\frac{\sigma}{\sigma_h}q(t)$, and minimize the functional:

$$\tilde{J}(q(t)) = J\left(q(t), -\rho\frac{\sigma}{\sigma_h}q(t)\right) = \int_0^T \left(V_t L\left(\left(\frac{q'(t)}{V_t}\right)\right) + \frac{\gamma}{2}\sigma^2(1-\rho^2)q^2(t)\right) dt$$

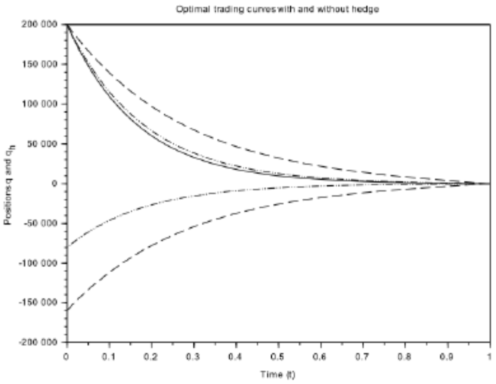


FIGURE 5.1: Optimal trading curve for $q_0 = 200{,}000$ shares over one day ($T = 1$), for $\gamma = 5.10^{-6}$ €$^{-1}$, with and without hedge. Solid line: without hedge. Dash-dotted lines: optimal strategy when $\sigma_h = \sigma$ and $\rho = 0.4$. Dashed lines: optimal strategy when $\sigma_h = \sigma$ and $\rho = 0.8$. Decreasing curves correspond to q, whereas increasing ones correspond to q_h.

Second asset perform a round trip, by building an initial position instantaneously and liquidating it progressively with an optimal trading curve given by: $q_h(t) = -\rho\frac{\sigma}{\sigma_h}q(t)$

4.3.4 Case Percentage of Volume (POV)

Changing set of admissible controls as:

$$\mathcal{A}_{POV} = \left\{(v_t)_{t\in\mathbb{R}^+}, \exists\rho > 0, \forall t \geq 0, v_t = -\rho V_t \mathbf{1}_{\{\int_0^t \rho V_s ds \leq q_0\}}\right\}$$

The terminal cash is:

$$X_T = q_0 S_0 - \frac{k}{2}q_0^2 - \frac{L(\rho)}{\rho}q_0 + \sigma\rho\int_0^T\int_t^T V_s ds dW_t$$

which is a Gaussian random variable.
The expected utility can be obtained using a Laplaca transformation, the function to minimize is:

$$J_{POV}(\rho) = \frac{L(\rho)}{\rho}q_0 + \frac{\gamma}{2}\sigma^2\rho^2\int_0^T\left(\int_t^T V_s ds\right)^2 dt$$

Existence global minimum case POV

There exists a $\rho^* > 0$ such that J_{POV} has a glbal minimium

If $V_t = V$, than:

$$J_{POV} = \frac{L(\rho)}{\rho} q_0 + \frac{\gamma}{6}\sigma^2 \frac{q_0^3}{\rho V}$$

If $L(\rho) = \eta|\rho|^{1+\phi} + \psi|\rho|$, optimal solution is:

$$\rho^* = \left(\frac{\gamma\sigma^2}{6\eta\phi} \frac{q_0^2}{V} \right)^{\frac{1}{1+\phi}}$$

The execution time will be fixed by $T^* = \frac{q_0}{V\rho^*}$. Some considerations:

- ρ^* does not depend on permanent impact and spread parameter ψ
- ρ^* increases with risk aversion, volatility, inventory
- ρ^* decreases with illiquidity (η) and ϕ

Through ρ^*, we can estimate the risk aversion γ of a trader from its executions:

$$\gamma = \frac{6\eta\phi V(\rho^*)^{1+\phi}}{\sigma^2 q_0^2}$$

4.4 Optimal execution with transient market impact

Almgren and Chriss make an assumption of a market impact that is linear, fixed, and permanent. However, as discussed in previous lectures, due to the correlation of order flow, it is evident that market impact is not solely fixed and permanent; instead, it exhibits a transient nature. This means that the past order flow influences future price impacts. To account for this phenomenon, one approach is to use a transient impact model (TIM).
TIM assumes that:

$$p_n = p_{-\infty} + \sum_{k=1}^{\infty} f(v_{n-k})G(k) + \sum_{k} \eta_k$$

where v_n is the signed order flow. So:

$$p_{n+1} - p_n = G(1)f(v_n) + \sum_{k=1}^{\infty}[G(k+1) - G(k)]f(v_{n-k}) + \eta_n$$

The decay of G make price diffusive (or approximately efficient), given the correlated order flow.
The efficiency leads to:

$$p_{n+1} - p_n = K(v_n - \hat{v}_n) + \eta_n$$

where $\hat{v}_n$ is the best predictor of v_n given $\mathcal{F}_n$
Let consider the propagator model in real time. We consider 5 minute intervals, p_n is the log mid price right before time t_n. We define the series of aggregated volume v_n in terms of the volume v_i^{tt} of single transaction:

$$v_n = \sum_{[t_n, t_{n+1}]} v_i^{tt}$$

with $v_i = \sum \text{buy} - \sum \text{sell}$.
Consider the normalized volume imbalance:

$$v_n^{nor} = \frac{\sum_{[t_n, t_{n+1}]} v_i^{tt}}{\sum_{[t_n, t_{n+1}]} |v_i^{tt}|}$$

The impact function $f(v^{nor})$ of the normalized volume imbalance is:

$$f(v^{nor}) = \mathbb{E}\left[r_n | v_n^{nor}\right]$$

and the propagator model in real time is:

$$r_j \equiv p_{k+1} - p_k = \sum_{k=0}^{j-1} \mathcal{G}(k) f(v_{j-k}^{nor}) + \eta_j \qquad \mathcal{G}(k) \equiv G(k+1) - G(k), G(0) = 0$$

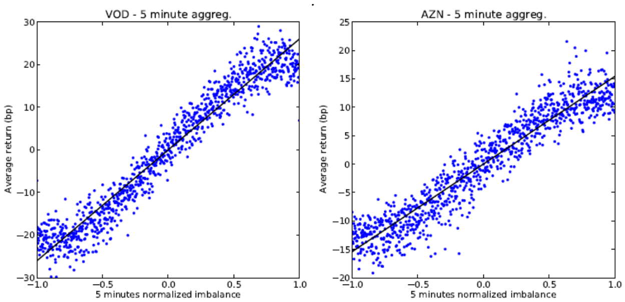
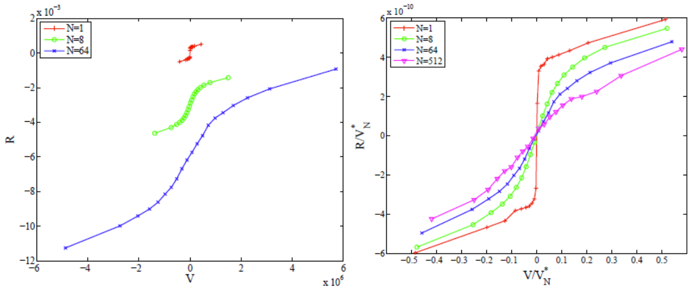
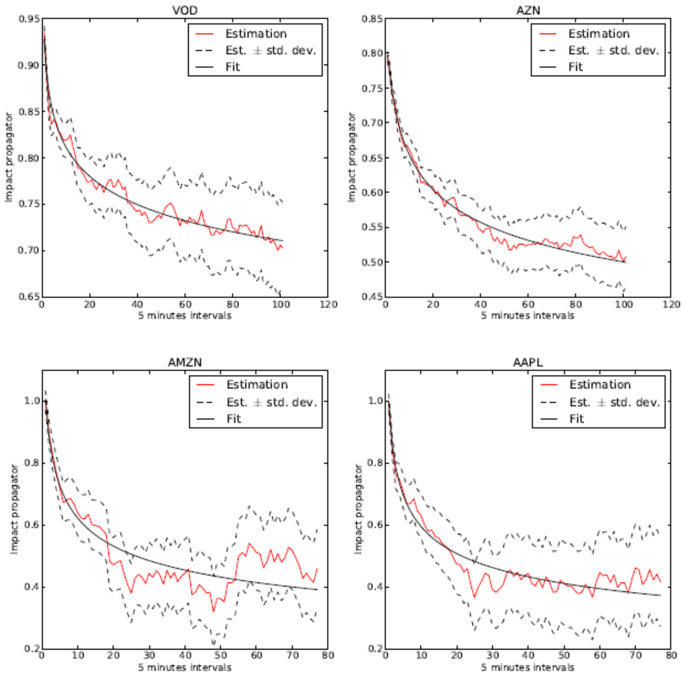


Figure: Impact (top) and propagator (bottom) from 5 min imbalance data

We noticed that at this level of aggragation, impact is roughly linear.
Considering a different aggregations (number of trades):



Market impact is a strongly concave function of volume at short scales, but it becomes progressively more linear on longer scale.



The TIM fits data quite well also on aggregated (time or trades) data.
Considering optimal execution, effective log-midprice $\tilde{p}_k$ is the logarithm of the average mid-price at which we trade the shares v_k between time t_k and time t_{k+1}, we assume:

$$\tilde{p}_k = \frac{p_k + p_{k+1}}{2}$$

the equation that describes the dynamics of effective price is:

$$\tilde{p}_n = p_0 + \sum_{k=0}^{n} [\eta_k + f(v_k)\tilde{G}(n-k)]$$

we degine effective propagator $\tilde{G}$ as:

$$\tilde{G}(0) = \frac{G(1)}{2}, \quad \tilde{G}(1) = \frac{G(1)+G(2)}{2}, \quad \tilde{G}(2) = \frac{G(3)+G(2)}{2}, \ldots$$

we define the logarithmic transiction cost $c(\mathbf{v})$ as:

$$c(\mathbf{v}) \equiv \sum_{k=0}^{N-1} v_k(\tilde{p}_k - p_0) \simeq \sum_{k=0}^{N-1} v_k \log\left(\frac{\tilde{P}_k}{P_0}\right) \simeq \sum_{k=0}^{N-1} v_k \left(\frac{\tilde{P}_k - P_0}{P_0}\right) = \frac{C(\mathbf{v})}{P_0}$$

The expected implementation shortfall is:

$$\mathbb{E}\left[C(\mathbf{v})\right] = \sum_{n=1}^{N} v_n \left[\sum_{k=1}^{n} f(v_k)\tilde{G}(n-k)\right]$$

if we assyme instantaneous impact linear: $f(v_k) = \theta_k v_k$, we obtain:

$$\mathbb{E}\left[C(\mathbf{v})\right] = 2\sum_{k,j} \theta_k \tilde{G}(|k-j|) v_k v_j = \mathbf{v}^T \mathcal{I} \mathbf{v}$$

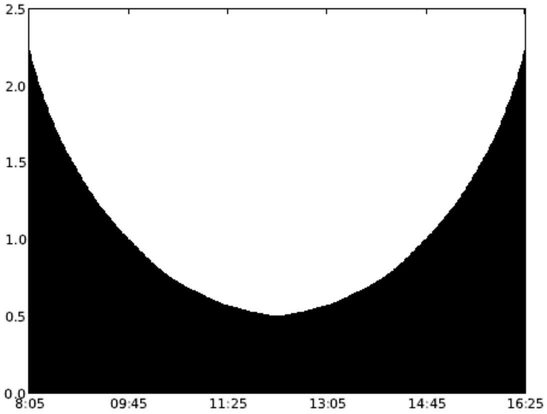
where $\mathcal{I}$ is Toeplitz matrix.
In this way we obtain a quadratic optimization problem:

$$\mathbf{v}^* = \arg\min_{v} \mathbf{v}^T \mathcal{I} \mathbf{v}, \quad \text{s.t.} \quad \sum_k v_k = \mathbf{1}^T \mathbf{v} = q_0$$

solving through Lagranfe multiplier:

$$\mathbf{v}^* = \frac{q_0}{\mathbf{1}\mathcal{I}^{-1}\mathbf{1}} \mathcal{I}^{-1}\mathbf{1}$$

If $G(k) = a(c+k)^{-\beta} \sim k^{-\beta}$



The solution is symmetric around $N/2$. U shape does not depend on the intraday profile of volume.

4.4.1 Adding risk term

The variance of the execution cost under the propagator model is:

$$Var[c(\mathbf{v})] = \mathbb{E}\left[(c(\mathbf{v}) - \mathbb{E}\left[c(\mathbf{v})\right])^2\right] = \mathbb{E}\left[\left(\sum_{k=1}^{N} v_k \sum_{j=0}^{k-1} \eta_j\right)^2\right] =$$

$$= \mathbb{E}\left[\left(\sum_{k=1}^{N} \eta_k \sum_{j=k}^{N} v_j\right)^2\right] = \sigma^2 \sum_{k=1}^{N} \left(\sum_{j=k}^{N} v_j\right)^2 = \sum_{k,j} \mathcal{V}_{k,j} v_k v_j$$

where σ^2 is variance of the residuals. Defining $\mathcal{F} \equiv \mathcal{I} + \lambda\mathcal{V}$, using Lagrange myltipliers we define:

$$\mathbf{v}^* = z\mathcal{F}^{-1}\mathbf{1} = \frac{q_0}{\mathbf{1}^T \mathcal{F}^{-1}\mathbf{1}} \mathcal{F}^{-1}\mathbf{1}$$

Plotting it, more is risk adversion, more trading at begging:

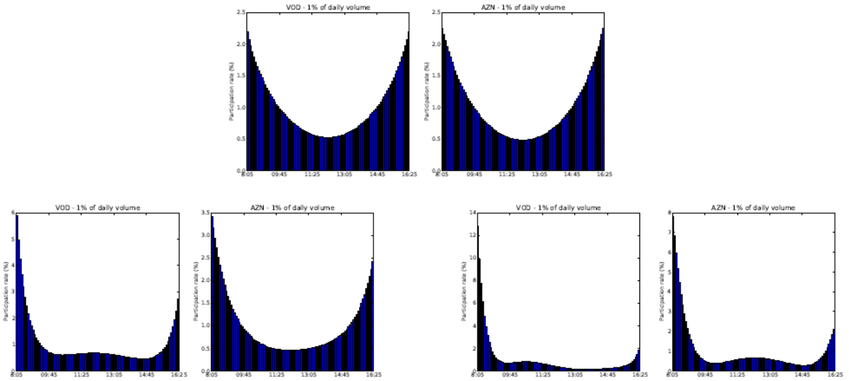


Figure: $\lambda = 0$ (top), $\lambda = 0.2$ (bottom left), $\lambda = 0.9$ (bottom right)

4.4.2 Including spread costs (no risk aversion)

In this derivation, we have not included transiction cost (fees, spread), spread costs change the qualitative results more.
Model price is:

$$\tilde{p}_n = p_0 + \sum_{k=0}^{n} [\eta_k + f(v_k)\tilde{G}(n-k)] + \operatorname{sign}(v_k)\delta_k$$

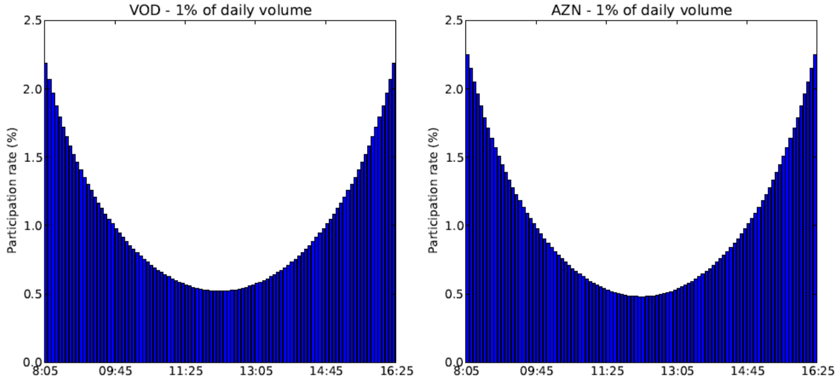
so that we payy half the bid-ask spread ib execution, we have defined:

$$\delta_k \equiv \frac{s_k/2}{P} = \frac{A_k - B_k}{A_k + B_k}$$

with A, B ask-bid spread. The objective function to optimize is:

$$F[\mathbf{v}] = \mathbb{E}\left[C(\mathbf{v})\right] + Spread(\mathbf{v}) = \mathbf{v}^T \mathcal{I} \mathbf{v} + \mathbf{D}^T |\mathbf{v}|$$

with $\mathbf{D} = (\{\delta_k\})^T$ is a vectore describing the spread cost during execution, for simplicity we assume $\mathbf{D} = \delta \mathbf{1}$. Solving it numerically, the shape of the solution does not qualitatively change.



General comments:

- The alternating (buy-sell) solution and the regularization achieved by the bid ask term is similar to what happens in portfolio optimization.
- By choosing a δ parameter much smaller than the fractional spread, one still recovers the U-shaped solution.

4.5 Transient Impact Model (TIM) in continuous time

Considering time interval $[0, T]$, the price S_t at time t is:

$$S_t = S_0 + \int_0^t f(\dot{q}_s) G(t-s) ds + \int_0^t \sigma_s dW_s$$

where $(\dot{q}_t)$ is the amount of shares sold by the execution time $[t, t+dt]$, W_s is a Wiener process, σ_s is deterministic function. The function f describes the instantaneous impact of the executed trades on price, we consider it linear:

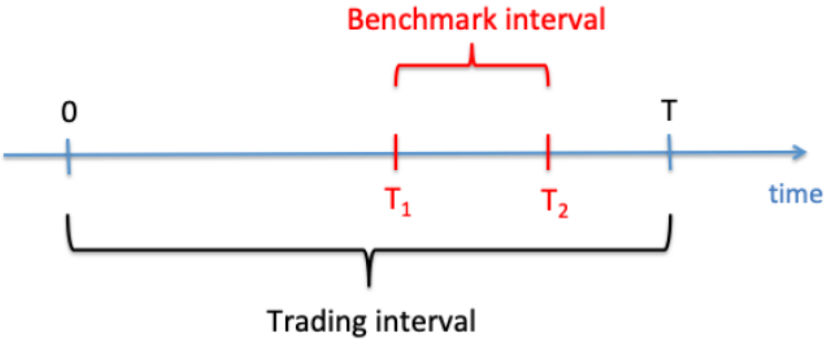
$$f(\dot{q}_t) = -k\dot{q}_t$$

The function $G(t)$ describes the delayed effect of trading on price and $G(t-s)$ characterizes how a trade at time s affects the price at time t.
Empirical evidences, shows power law kernel:

$$G(t) = t^{-\kappa}, \quad \kappa < 1$$

TIM continuous generalized VWAP

- borker receives from a fund a request of a VWAP sell execution of $q_0 > 0$ shares in a time windows $[T_1, T_2]$ termed the benchmark interval.
- The broker charges the fund the Volume Weighted Average Price in $[T_1, T_2]$.
- The broker is allowed to trade in a time window $[0, T] \supseteq [T_1, T_2]$



We will consider only deterministic strategies.
Let $V_t dt$ be the deterministic market volume traded in $[t, t+dt]$. VWAP benchmark is given by:

$$VWAP_{T_1}^{T_2} = \frac{\int_{T_1}^{T_2} S_t V_t dt}{\int_{T_1}^{T_2} V_t dt} = \int_0^T \eta_t S_t dt$$

where:

$$\eta_t = \frac{V_t}{\int_{T_1}^{T_2} V_s ds} \mathbb{1}_{t \in [T_1, T_2]}$$

The objective function of the broker is the difference between the cash she is able to obtain from the proceeds in the trading interval and the cash she will give back to the client, equal to the random variable $q_0 VWAP_{T_1}^{T_2}$, let use define the cash process (no temporary impact):

$$dX_t = \dot{q}_t S_t dt = v_t S_t dt \qquad X_0 = 0$$

Assuming CARA risk averse agent, objective function is:

$$U[\mathbf{q}] = \begin{cases} \mathbb{E}\left[X_T - q_0 q_0 VWAP_{T_1}^{T_2}\right] & \gamma = 0 \\ \mathbb{E}\left[-\exp(-2\gamma(X_T - q_0 VWAP_{T_1}^{T_2}))\right] & \gamma > 0 \end{cases}$$

where 2γ is risk aversion parameter.

Optimization problem TIM generalized VWAP

Under linear impact, $f(z) = -kz, k > 0$, the maximization of the utility function over the deterministic strategies is equivalent to the minimization of the functional:

$$C[\mathbf{x}] \equiv \frac{1}{2}\int_0^T \int_0^T \dot{q}_t \dot{q}_s G(|t-s|) ds dt - q_0 \int_0^T \eta_t dt \int_0^t G(t-s)\dot{q}_s ds \tag{IV.4}$$

$$+ \frac{\gamma}{k}\int_0^T \int_0^T dt dt' (\dot{q}_t - q_0\eta_t)(\dot{q}_{t'} - q_0 \eta_{t'}) \int_0^{t \wedge t'} \sigma_s^2 ds \tag{IV.5}$$

This is not a standard Lagrangian calculus of variation, we need a different approach form Euler-Lagrange.
If $\gamma = 0$, minimization is:

$$C[\mathbf{q}] = \frac{1}{2}\int_0^T \int_0^T dq_t dq_s G(|t-s|) - \frac{q_0}{T}\int_0^T dt \int_0^t G(t-s) dq_s \equiv Q[\mathbf{q}] + K[\mathbf{q}]$$

Considering a strategy:

$$dy_s = \delta_{t_2}(ds) - \delta_{t_1}(ds) \qquad 0 \le t_1 \le t_2 \le T$$

Let call $\mathbf{q}^*$ the optimal strategy, and setting $\mathbf{z} = \mathbf{q}^* + \alpha \mathbf{y}$, the integral equation satisfied the optimal strategy:

$$\left.\frac{\partial \mathbb{E}\left[C[\mathbf{z}]\right]_0}{\partial \alpha}\right|_{\alpha=0} = 0$$

Integral equation associated to Optimization problem TIM generalized VWAP

The strategy $\{q^*_t\}_0^T$ thath minimize the functional (with $\gamma = 0$), satisfies the integral equation:

$$\int_0^T G(|t-s|)dq^*_s - q_0 \int_t^T \eta_s G(s-t)ds = \lambda$$

where λ is a constant set by the normalization the total volume traded:

$$\int_0^T dq^*_s = q_0$$

We write the solution of the integral equation as:

$$w_s = \dot{q}^*_s - q_0 \eta_s$$

with $\int_0^T w_s ds = 0$, replacing it in integral equation:

$$\int_0^T G(|t-s|)w_s ds = \lambda - q_0 \int_0^t \eta_s G(t-s)ds$$

writing the solutuon $w_s = w_s^{(1)} + w_s^{(2)}$, where the second tem solves:

$$\int_0^T G(|t-s|)w_s^{(2)} ds = -q_0 \int_0^t \eta_s G(t-s)ds$$

and setting $q_0' = \int_0^T w_s^{(2)} ds$, the first term solves:

$$\int_0^T G(|t-s|)w_s^{(1)} ds = \lambda \qquad \int_0^T w_s^{(1)} ds = -q_0'$$

which is the equation when the objective function is the IS and nu,ber of shares is $-q_0'$. If $T_1 = T_2 = 0$, $\eta_t = 2\delta(t)$ and integral equation becomes:

$$\int_0^T G(|t-s|)dq^*_s = \lambda$$

Existence minimization value implementation Shortfall

Suppose G is positive definite. Then q^* minimizes expected cos $\iff$ $\exists\lambda$ s.t. q^*_t solves $\forall t$:

$$\int_0^T G(|t-s|)dq^*_s = \lambda$$

thus $C(q^*) = \frac{1}{2}\lambda q_0$

4.5.1 Obizhaeva-Wang model

Obizhaeva-Wang model

- Impact is linear and decays exponentially in time
- The decay is interpreted as the relaxation of the limit order book when shocked by a trade
- so that $v_t = \dot{q}_t$, the price during the execution is modeled as:

$$S_t = S_0 - k\int_0^T v_s \exp[-\rho(t-s)]ds + \int_0^t \sigma dW_t$$

The expected cost is:

$$\mathbb{E}\left[C[\mathbf{q}]\right] = k\int_0^T v_t dt \int_0^t v_s \exp[-\rho(t-s)]ds$$

Solution Obizhaeva-Wang model

The Fredholm integral equation of the first kind from the theorem is:

$$\int_0^T v_s \exp[-\rho|t-s|]ds = \lambda$$

where λ is integration constant. The exact solution is:

$$v_t = (q_0 - \rho T) + \rho + (q_0 - \rho T)\delta(t-T)$$

In special case $T_1 = T_2 = T$, integral equation is:

$$\int_0^T G(|t-s|)dq^*_s = \lambda + q_0 G(T-t)$$

with solution $\dot{q}^*_s = w^{(1)}_s + q_0\delta(T-t)$ (the sum of $q_0/2$ shares traded as in the IS case and the remaining $q_0/2$ shares traded at $t = T$).

Using power law kernel $G(t) = e^{-\rho t}$, the optimal strategy is:

$$v_t = \frac{q_0}{\rho T(2+\rho T)}[2(1+\rho T)\delta(t) + \rho(1+\rho T) - 2\delta(t-T)]$$

in word: sell a finite amount at time $t = 0$, selling at a constant rate for whole interval $(0, T)$, buying a finite amount at time $t = T$.

4.6 Solution TIM in discrete time

When we add more constraints, it is convenient to frame the problem in discrete time, we can make it at three different levels:

- express the cost function in discrete time and solve the optimization;
- use discrete time to obtain a quadrature of the integral equation;
- write the Transient Impact Model in discrete time, derive the corresponding cost and then minimize it.

The three method does not give the same results, if time intervals for discretization is sufficiently small, differences become negligible. We will use the last setting.

TIM Discrete Time

- Let us divide the interval $[0; T]$ in N equal intervals and define $\tau = T/N$.
- The strategy is now a vector $\mathbf{v} = (v_1, \dots, v_N)'$, with v_i we consider amount of shares traded in interval i
- The price dybamics of sell execution is:

$$S_l = S_0 - k\sum_{i=1}^{l} G(l-i)v_i + \tau^{1/2}\sum_{i=1}^{l}\epsilon_l \qquad l = \{0, \dots, N\}$$

that can be rewritten as:

$$\mathbf{S} = S_0\mathbf{1} - kG\mathbf{v} + \tau^{1/2}L\epsilon$$

where $\mathbf{S} = (S_1, \dots, S_N)', \mathbf{1} = (1, \dots, 1), L$ is lower triangular matrix of ones, G is the lower triangular matrix s.t:

$$G_{ij} = G[\tau(i-j)] \quad \text{if } i \geq j$$

and $\epsilon \sim \mathcal{N}(\mu, \mathbf{\Sigma})$ is Gaussian random vector describing the price dynamics.

Consider a $VWAP$ benchmark between $t = T_1$ and $t = T_2$, corresponding to $l_1 = \lfloor NT_1/T \rfloor$, $l_2 = \lfloor NT_2/T \rfloor$ are the rounding to the nearest integer. Introducing $B = \{l \in \mathbb{N} : l_1 \leq l \leq l_2\}$ and a vector η with components:

$$\eta_l = \frac{V_l}{||\eta||_1}\mathbb{1}_{l\in B}$$

where V_l is the market colume traded in interval l.
The benchmark is $q_0\eta'\mathbf{S}$, and normalization ensures that $\mathbf{1}'\eta = 1$.
We use utility function $\mathcal{U}[(\mathbf{v} - q_0\eta)'\mathbf{S}]$ and using CARA utility function with risk aversion 2γ, the expected utility is:

$$U[\mathbf{x}] = \mathbb{E}\left[(\mathbf{v} - q_0\eta)'\mathbf{S}\right]_0 - \gamma\mathbb{V}_0[(\mathbf{v} - q_0\eta)'\mathbf{S}]$$

Solution discrete TIM VWAP execution

Under CARA utility function iwth risk aversion 2γ, the optimal VWAP execution which maximizes the expected utility, is the solution of the quadratic optimization

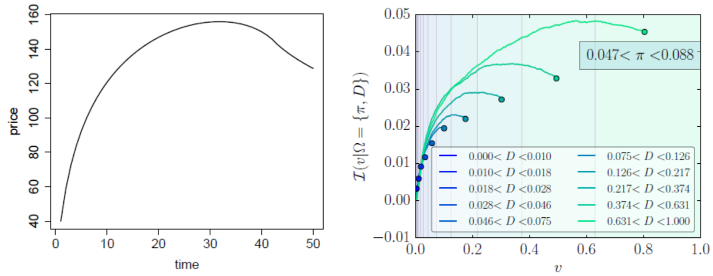
$$\min_{v}[\mathbf{v}'A\mathbf{v} - \mathbf{b}'\mathbf{v}] \qquad \text{s.t.}\mathbf{1}'\mathbf{v} = q_0$$

where:

$$A = kG + \gamma\tau L\Sigma L' \tag{IV.6}$$
$$\mathbf{b}' = kq_0\eta' G + 2\gamma\tau q_0\eta' L\Sigma L' + \tau^{1/2}\mu' L' \tag{IV.7}$$

The matrix A is positive and definite if Σ is positive definite. The solution of the quadratic optimization exists and is unique.

Due to quadratic form, several constraints can be added:



Left. Average price dynamics according tor the model with constraint on the sign of trades. Right. Average price dynamics from real executions (from Zarinelli et al 2015)

Note: with a fixed impact model, for example Almgren & Chriss (2000), the reversion of price during execution is not possible with constant trade signs.

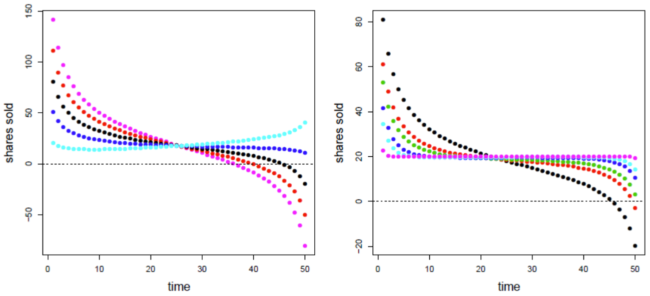


Figure: Left. Optimal VWAP schedule for a sell order by a risk neutral broker for different values of the price drift $\mu_i = 4$ (cyan), $\mu_i = 2$ (blue) $\mu_i = -2$ (red), and $\mu_i = -4$ (magenta). Black dots refer to the driftless benchmark case. Right. Optimal VWAP schedule for a risk averse broker under driftless price. The values of the risk aversion parameter γ are 0 (black), 0.5 (red), 1 (green), 3 (blue), 7 (cyan), 100 (magenta). In both panels the benchmark interval is coincident with the trading interval.

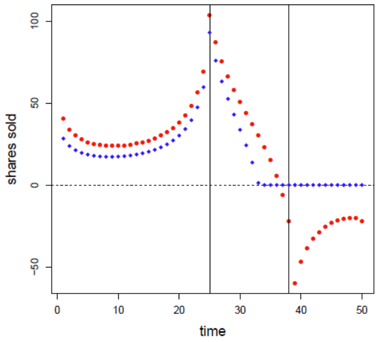


Figure: Optimal schedule without (red) and with (blue) constraint on trade sign for a VWAP with benchmark interval $T_1 = 25$ and $T_2 = 38$ (vertical lines).

From this picture we can notice the strategy is: selling a lot in the begging to buy in the end.

4.7 Price manipulation

Let us define:

- q_0 shares to be sold
- S_t^0= unperturbed price; S_t^q = price when strategy q is used
- Revenues:

$$\mathcal{R}_T(q) = -\int_0^T S_t^q dq_t$$

- Liquidation cost:

$$\mathcal{C}_T(q) = q_0 S_0^0 - \mathcal{R}_T(q)$$

  both stochastic.
- dq_t: number of shares traded in $[t, t+dt]$

Price manipulation

A round trip is an order execution strategy $(q_t)_{t\in[0,T]}$ with $q_0 = q_t = 0$. A price manipulation strategy is a round trip with:

$$\mathbb{E}\left[\mathcal{R}_T(q)\right] > 0$$

Transaction-triggered price manipulation (Alfonsi et al 2012)

A market impact model admits transaction-triggered price manipulation if the expected revenues of a sell (buy) program can be increased by intermediate buy (sell) trades. In other words, $\exists q_0, T > 0, \tilde{q}$, s.t.":

$$\mathbb{E}\left[\mathcal{R}_T(\tilde{q})\right] > \sup\{\mathbb{E}\left[\mathcal{R}_T(q)\right]\} \quad q \text{ is monotone}$$

Negative expected liquidation

A market impact model has negative expected liquidation costs if

$$\mathbb{E}\left[\mathcal{C}_T(q)\right] < 0$$

or $\mathbb{E}\left[\mathcal{R}_T(q)\right] > q_0 S_0$

Relation between manipulations (Klöck et all 2011)

- Any market impact model that does not admit negative expected liquidation costs does also not admit price manipulation.
- Suppose that asset prices are decreased by sell orders and increased by buy orders. Then the absence of transaction-triggered price manipulation implies that the model does not admit negative expected liquidation costs. In particular, the absence of transaction-triggered price manipulation implies the absence of price manipulation in the usual sense.

Let consider AC model: gicen two non-decreasing functions f and g, with $f(0) = g(0) = 0$, an absolutely contionuous strategy $(q_t)_{t\geq 0}$ leads to a price trajectory:

$$S_t^q = S_t^0 + \int_0^t \underbrace{f(\dot{q}_s)ds}_{\text{permanent price impact}} + \underbrace{g(\dot{q}_s)}_{\text{temporary-impact}}$$

assuming:

$$S_t^0 = S_0 + \sigma W_t$$

and W_t is a Wiener process

AC model price manipulation (Huberman and Stanzl 2004, Gatheral 2010)

If the AC model does not admit price manipulation, then $f(x) = \gamma x$ with $\gamma \geq 0$

This theorem show that non-linear permanent market impact is inconsistent with the principle of no price manipulation.

4.7.1 Transient impact models

Empirical evidences shows that the impact is transient (decay with time). We can generalize the TIM model as:

$$S_t^q = S_0 + \int_0^T f(\dot{q}_s)G(t-s)ds + \int_0^t \sigma dW_s$$

with expected cost:

$$\mathbb{E}\left[\mathcal{C}_T(q)\right] = \int_0^T \dot{q}_t dt + \int_0^t f(\dot{q}_s)G(t-s)ds$$

Non-linear Obizhaeva-Wang model manipulation

Considering:

$$S_t^q = S_0 + \eta \int_0^T f(v_s)\exp[-\rho(t-s)]ds + \int_0^t \sigma dW_t$$

If the temporary market impact decays exponentially, then price manipulation is possible unless $f(v) \propto v$

If we assume the instantaneous impact is linear $f(v) = \gamma v$, then:

$$S_t^q = S_t^0 + \int_{s<t} G(t-s)dq_s$$

$$C(q) \equiv \mathbb{E}\left[\mathcal{C}_q(q)\right] = \frac{1}{2}\int_0^T \int_0^T G(|t-s|)dq_s dq_t$$

Bochner Theorem

$C(q) \geq 0 \iff G(|x|)$ can be represented as the Fourier transform of a positive finite Borel measure μ on $\mathbb{R}$:

$$G(|x|) = \int \exp[ikx]\mu(dz)$$

Transaction-triggered price manipulation Linear case

Suppose G convex, satisfies $\int_0^t G(t)dt < \infty$ and there is an admissible strategy. Then there exists a unique admissible optimal strategy q_t^* which is monotone, i.e. there is no transaction-triggered price manipulation.

In words: with convexity, mixing strategy buy-sell is not admit.
Let consider now a non-linear transient impact model:

$$S_t^q = S_0 + \int_0^T f(\dot{q}_s)G(t-s)ds + \int_0^t \sigma dW_s$$

with

$$\int_0^T v_t dt = q_0$$

In these case, optimal solution is not known and if we consider VWAP strategy, $V_t = q_0/T$, the expected cost is:

$$C^{VWAP} = \frac{1}{T} f\left(\frac{q_0}{T}\right) \int_0^T dt \int_0^t G(t-s)ds$$

Price manipulation case non linear (Gatheral 2010)

If $G(t)$ is finite and continuous at $t = 0$ and f is nonlinear, then there is price manipulation

Considering the case:

$$f(v) = c\left(\frac{|v|}{V}\right)^\delta \text{sign}(v) \qquad G(t-s) = (t-s)^{-\gamma}$$

Special case (Gatheral 2010)

If $G(t) = t^{-\gamma}$ with $\gamma \in (0,1)$ and $f(x) \propto |x|^\delta \text{sign}(x)$ with $\delta > 0$, then price manipulation exists when one of the two following condition is verified:

$$\gamma + \delta \leq 1 \qquad \gamma \leq \gamma^* = 2 - \frac{\log 3}{\log 2} \simeq 0.415$$

The expected cost of a VWAP is:

$$C^{VWAP} = \frac{c}{(1-\gamma)(2-\gamma)} \frac{q_0^{\delta+1}}{V^\delta} T^{1-\gamma-\delta}$$

and the impact is:

$$\mathbb{E}\left[S_T^q - S_0\right] = \frac{c}{1-\gamma}\left(\frac{q_0}{V}\right)^\delta T^{1-\gamma-\delta}$$

we noticed that if $\gamma + \delta = 1$, the expected impact and cost do not depend on the execution time. If $\delta = \gamma = 0.5$ the impact is:

$$\mathbb{E}\left[S_T^q - S_0\right] = 2c\sqrt{\frac{q_0}{V}} = 2c\sqrt{T_d}\sqrt{\frac{q_0}{ADV}}$$

celebrate square root impact formula!

4.8 Integral equation for cost minimization and perturbative approach

Given $f \in C^1(\mathbb{R})$ and $G \in L^1[0,T]$, for class of functions x on $[0,T]$ satisfying:

- x is absolutely continuous on $(0,T)$.
- $f \circ v \in L^1[0,T]$

we have the following condition for stationarity of the cost functional:

$$\int_0^t f(v_s)G(t-s)ds + f'(v_t)\int_t^T v_s G(s-t)ds = \lambda$$

where λ is a constant set by the constraint equation.
Concave case ($\delta < 1$) there is no guarantee that the minimum is global.
Perform an expansion $f(v) = v^\delta = v^{1-\epsilon}$ with $0 < \epsilon \ll 1$ we solve exactly the perturbed equation, front loading for concave impact ($\delta < 1$), back loading for convex impact ($\delta > 1$) Solving the minimization numerically the cost on a discrete grid of N intervals in $[0,T]$:

$$\arg\min \sum_{i=1}^{N}\sum_{j=1}^{N} v_i^n f(v_j^n) A_{ij} \qquad \text{s.t.} \sum_{i=1}^{N} v_i = \frac{NX}{T}$$

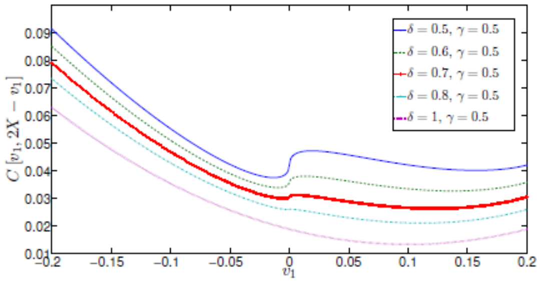
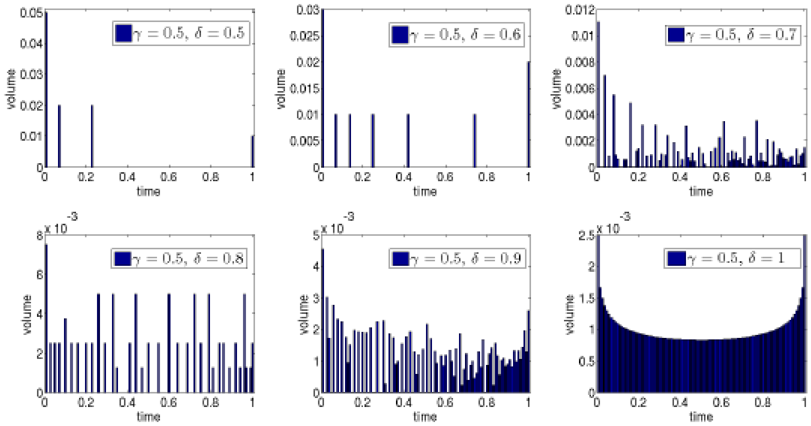


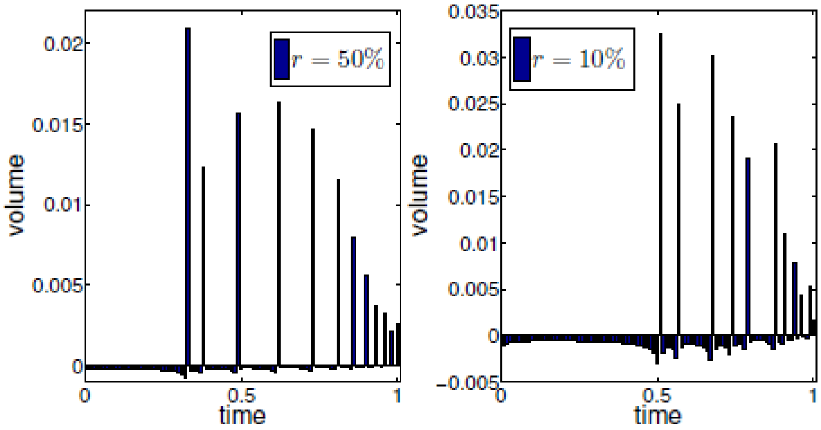
Figure: Cost function $C[v_1, 2X - v_1]$ for $X = 0.1, \gamma = 0.5$. For $\delta = 1$ the minimum is at $v_1 = X$. In the nonlinear case there are two local minima.

For buy program, strong nonlinearity is $\delta > 0.5$
Optimal strategies with the constraint $v \geq 0$:



The optimal strategy is to trade in bursts separated by no trading periods (buy, wait price go down, buy again).

Not linear optimal buy: small intense bursting periods + long period call



V

Market making

A Market Maker (MM) seeks to maximize their terminal wealth by engaging in trading activities through the use of limit orders. In a dealership market(we consider this situation) the MM has the freedom to set the quotes on their platforms. However, the challenge lies in the risk that customers might opt for another dealer who offers more convenient terms.
In a Limit Order Book (LOB) market, the strategy needs to account for the presence of other limit orders within the LOB. Failing to consider this can result in an extremely low probability of having orders executed.
Similar to any uninformed liquidity provider, the MM faces the risk of adverse selection by informed investors.
Even if the flow of investors is not driven by informed trading, there is an additional risk associated with accumulating a large inventory of securities. The key tool at the disposal of the MM is the relationship between the posted prices and the rate at which orders arrive:

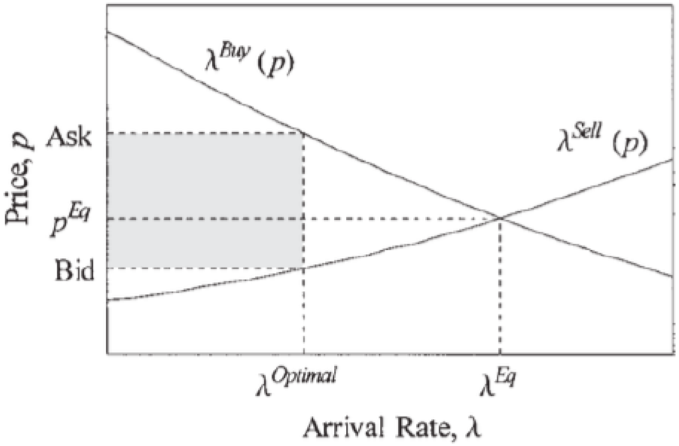
- By lowering the ask price, the MM can attract more buyers, while raising the ask price, the MM may attract fewer buyers.

- Lowering the bid price could result in fewer sellers, while raising the bid price may attract more sellers.

These adjustments in the posted prices allow the MM to manage and influence trading activity.

5.1 Garman model (1976)

Assume that $\lambda^{Buy}(Ask) = \lambda^{Sell}(Bid)$that is, supply and demand balance (on average).



The profits are defined by the shaded rectangle and the market maker sets the bid and ask to maximize this area.
If the market maker sets the Bid and Ask once for all, the holdings of stock follow a zero-drift random walk. Cash holdings follow a positive-drift random walk (due to the turn). Garman points out that in this case, the dealer is eventually ruined with probability one.
Need to change the bid and ask quotes to elicit an expected imbalance of buy and sell orders to push their inventories in the direction of their preferred long-run position.
Let consider:

$$\lambda^{buy} = Ae^{-\kappa(S^a - S)} \qquad \lambda^{sell} = Ae^{-\kappa(S - S^b)}$$

where S^a e S^b are the ask and bid price and S is the equilibrium price, so that:

$$S^a = S - \frac{1}{\kappa}\ln\frac{\lambda^{buy}}{A} \qquad S^b = S - \frac{1}{\kappa}\ln\frac{\lambda^{sell}}{A}$$

when $\lambda^{buy} = \lambda^{sell} \equiv \lambda$, profit is:

$$\text{Profit} = \lambda = \lambda(S^a - S^b) = -\frac{2}{\kappa}\lambda\ln\frac{\lambda}{A}$$

Profit is maximal if $\lambda = Ae^{-1}$, so the spread is:

$$\text{Spread} = S_a - S_b = \frac{2}{\kappa}$$

5.2 Avellaneda-Stoikov model

Avellaneda-Stoikov model

- The reference price S_t is modeled by a Brownian dynamics:

$$S_t = \sigma dW_t$$

- Bid and ask quotes are modeled by S_t^a and S_t^b
- $(N_t^b)_t$ and $(N_t^a)_t$ are two point processes modeling the number of assets that have been respectively bought and sold.
- The MM's inventory is:

$$q_t = N_t^b - N_t^a$$

- The intensity process of $(N_t^b)_t$ and $(N_t^a)_t$ are denoted by $(\lambda_t^b)_t$ and $(\lambda_t^a)_t$

$$(\lambda_t^b)_t = \Lambda^b(\delta_t^b) 1_{q_- < Q} \qquad (\lambda_t^a)_t = \Lambda^a(\delta_t^a) 1_{q_- < -Q}$$

where:

$$\delta_t^a = S_t - S_t^b \qquad \delta_t^b = S_t^a - S_t$$

where Λ^b and Λ^a are two positive nonincreasing functions.

- $Q \in \mathbb{N}$ is the maximum authorized invetory. MM stops posting a bid(ask) when $q_t = Q(q_t = -Q)$
- We assume:

$$\Lambda^b(\delta) = \Lambda^a(\delta) = Ae^{-\kappa\delta}$$

where $A > 0$ is the liquidity of the asset and $\kappa > 0$ characterises the price sensitivity of the market participants.

The amount of cash $dX_t = S_t^a dN_t^a - S_t^b dN_t^b = (S_t + \delta_t^a) dN_t^a - (S_t - \delta_t^b) dN_t^b$.
The goal of MM is to maximize the expected (CARA) utility functionmn at some determined time horizon T, s.t:

$$\mathbb{E}\left[-\exp\{-\gamma(X_T + q_T S_t - l(q_T))\}\right]$$

the term $l(q_t)$represents a penalization for the final inventory.
The solution requires tolls of stochastic optimization and we have to solve the associated Hamilton-Jacobi-Bellman equation. The system can be solvend numerically or exactly.

Avellaneda-Stoikov model solution

The optimal bid and quotes of the market maker in Avellaneda-Stoikov model are:

$$S^b(t,S,q) = S - \frac{1}{\kappa}\ln\left(\frac{v_q(t)}{v_{q+1}(t)}\right) - \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{\kappa}\right)$$
$$S^a(t,S,q) = S - \frac{1}{\kappa}\ln\left(\frac{v_q(t)}{v_{q-1}(t)}\right) - \frac{1}{\gamma}\ln\left(1+\frac{\gamma}{\kappa}\right)$$

Some general comment:

- In some situations it might look unnatural to consider a finite terminal time T
- when $T \to \infty$ the optimal spread is approximated by:

  $$\text{Spread} = S^a - S^b \simeq \frac{2}{\gamma}\ln\left(1+\frac{\lambda}{\gamma}\right) + \sqrt{\frac{\sigma^2\gamma}{2\kappa A}\left(1+\frac{\gamma}{\kappa}\right)^{1+\frac{\kappa}{\gamma}}}$$

  which does not depend on the inventory q
- The skewness of the strategy is:
  $$(S - S^b) - (S^a - S) \propto q$$
- The MM faces two risks: transiction uncertainty and price risk
- if $\sigma = 0$, there is a spread:
  $$\text{Spread} \simeq \frac{2}{\gamma}\ln\left(1+\frac{\gamma}{\kappa}\right) \xrightarrow{\gamma\to 0} \frac{2}{\kappa}$$

  that is a decreasing function of γ. Spread is $2/\kappa$ when $\gamma = 0$ even if $\sigma \neq 0$
- the skewness increases with σ and γ

Limitation of this model are:

- The model can be enriched by adding short term α, price drift, etc.
- Price is a continuous variable
- It is hard to apply in LOB market
- The function Γ are exponential
- It considers only one asset

VI

Modelling the Limit Order Book

Motivations for studying market microstructure include:

- **Market Simulators**: Developing market simulators to test and refine trading, market making, and execution strategies in a controlled environment before applying them to real markets.
- **Regulatory Design**: Informing the design of relevant regulations and policies by understanding how different market structures and mechanisms impact trading behavior and market outcomes.
- **Noise Modeling**: Creating realistic microstructure noise models to account for the inherent randomness and uncertainty in financial markets, which is crucial for accurate modeling and risk management.
- **Price Formation**: Investigating the interplay between trading strategies and market structure in the formation of asset prices, shedding light on how trading activities influence price movements.
- **Macro-Micro Connection**: Bridging the gap between microstructure-level analysis and macro-level variables such as market volatility, enabling a comprehensive understanding of market dynamics and their impact on broader financial conditions.

6.1 The zero intelligence model (Santa Fe model)

Several studies, including Daniels et al. (2003), Smith et al. (2003), Cont and De Larrard (2013), and Donier et al. (2015), have examined market microstructure using a common model framework. Key features of this model include:

- Order Book Grid: The order book is represented as a discrete price grid with a constant minimum price increment, dp (the tick size).
- Limit Order Placement: The placement of limit orders follows a Poisson process with a rate denoted as $\alpha(\Delta)$ per unit of price and unit of event time, where Δ represents the distance from the best available price.

- Market Orders: Market orders arrive at a rate of μ per unit of event time.
- Order Cancellations: Each existing limit order has the same probability, δ, per unit of event time to be canceled.
- Order Size: All orders, whether limit or market, have the same size, denoted as σ shares.

This modeling approach has been found surprisingly useful in providing testable predictions for certain short-term (Cont and De Larrard, 2013) or long-term (Farmer et al., 2005) properties of the order book. Moreover, it is easy to simulate and estimate parameters from data by counting events.

6.1.1 Asympotic depth

Because market order arrivals only inuence activity at the best quotes, very deep into the LOB the distribution of queue sizes reaches a stationary state that is independent of the distance from $m(t)$.
For $V \in \mathbb{N}$ (where V is in units of the lot size σ):

$$P_{st}(V) = e^{-V^*}\frac{(V^*)^V}{V!} \qquad V^* = \frac{\alpha}{\delta}$$

Two extreme cases are possible:

- **sparse LOB**: corrisponding to $V^* \ll 1$, where most price levels are empty, while the others are only weakly populated. This case corresponds to the behaviour of the LOB for very small-tick assets.
- **dense LOB**: corresponding to $V^* \gg 1$, where all price levels are populated with lagge number of orders. This corresponds to the behaviour of the LOB for large-tick assets

6.1.2 Estimating the spread

Let S the spread, i.e., the difference between the best ask and bid pric, the total flux of limit orders between the mid-point and $S/2$ is $\int_0^{S/2} \alpha(\Delta)d\Delta$ where Δ is the distance from the midpoint.
Let us analyse better:

- If S is sufficiently small so that α is approximately costant, one finds that his incoming flux is $\sim \alpha(0)S/2$.
- If $\mu > \alpha(0)S/2$, the rate of market order eats up the limit orders that appear within the spread completely, and the average volume present is close to zero.
- If $\mu \gg \delta$, the cancellation term can be safely neglected.
- Argument breaks down when $S \sim 2\mu/\alpha(0)$ which sets the typical position of the best price.
- The spread is therefore larger for larger market order rates, and smaller when the flow of limit order is larger, as expected intuitively.

- Scaling result for the spread has been derived more quantitatively when α and δ are indepntf od Δ. One finds for the average spread:

$$\mathbb{E}\left[S\right] = \frac{\mu}{\alpha} F\left(\frac{\delta}{\mu}\right)$$

  $F(u)$ is monotonically increasing function, we can approximate it as $F(u) \sim 0.28 + 1.86u^{3/4}$

- In the limit where cancellation can be neglected, we obtain $S \sim 0.28\mu/\alpha(0)$.

- With a similar arhuments, we show that volatility scales as:

$$\mathbb{E}\left[\sigma\right] \propto \mu^{5/2}\delta^{1/2}\rho^{-2}$$

6.1.3 Dimensional analysis

In order book we have three dimensions: price, share and time. from this we have:

- Three order flow rates: Limit orders arrival rate (α shares/ (time$\cdot$ price)), Market order arrival rate (μ shares/time), Limit order decay (δ 1/time)

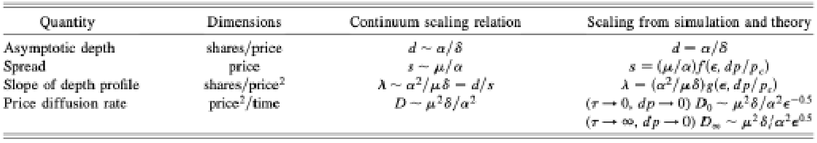
- Two dicreteness parameters: tocl size dp (price and order size) σ (shares)

- From order flows rate we derive:
  - number of shares $N_c = \mu/2\delta$
  - characteristic price interval $p_c = \mu/2\alpha$
  - characteristic time scale $t_c = 1/\delta$

- Nondimensional scale parameter $\epsilon = \sigma/N_c$ characterizing the granularity of the orders stored in the limit-order book. A nondimensional scale parameter based on tick size is constructed by dividing by the characteristic price dp/p_c

- Continuum limit as the tick size $dp \to 0$

TABLE I. Predictions of scaling of market properties vs order flow. The third column contains predictions from the continuum analysis, in which the discreteness parameters are ignored, and the fourth column gives more accurate predictions from theory and simulation. The functions f and g are the order of magnitude of one throughout the relevant ranges of variation of ϵ and dp/p_c.



Quantity	Dimensions	Continuum scaling relation	Scaling from simulation and theory
Asymptotic depth	shares/price	$d \sim \alpha/\delta$	$d = \alpha/\delta$
Spread	price	$s \sim \mu/\alpha$	$s = (\mu/\alpha)f(\epsilon, dp/p_c)$
Slope of depth profile	shares/price^2	$\lambda \sim \alpha^2/\mu\delta = d/s$	$\lambda = (\alpha^2/\mu\delta)g(\epsilon, dp/p_c)$
Price diffusion rate	price^2/time	$D \sim \mu^2\delta/\alpha^2$	$(\tau \to 0, dp \to 0)\ D_0 \sim \mu^2\delta/\alpha^2\epsilon^{-0.5}$ $(\tau \to \infty, dp \to 0)\ D_\infty \sim \mu^2\delta/\alpha^2\epsilon^{0.5}$

6.1.4 Model calibration

Estimating most of the parameters in this model is expected to be quite straightforward. However, one of the model's simplifying assumptions introduces a significant challenge when fitting the model. Specifically, the model assumes that the values of α and δ do not vary with increasing distance from the best quotes. To address this issue, Bouchaud et al. suggest restricting the estimation of the following parameters:

- α:should be estimated only for the set χ_{LO}, which includes limit orders that arrive either at the best quotes or within the spread.
- δ : should be estimated only for the set χ_C, which includes order cancellations that occur at the best quotes.
- μ: should be estimated only for the set χ_{MO}, which consists of market orders that, by definition, occur at the best quotes.

To perform these estimations, N_{LO}, N_C, and N_{MO} are counted, representing the number of limit order arrivals at the best quotes, cancellations at the best quotes, and market order arrivals within a specific time window. Importantly, these counts are conducted independently of the corresponding order signs, meaning that there is symmetry between buy and sell activities.
The lot size σ is estimated as:

$$\sigma = (N_{LO})^{-1} \sum_{x \in \chi_{LO}} \sigma_\chi$$

where σ_χ is the size (in shares) of LOs.
The total MO arrival rate per event $2\tilde{\mu}$ is estimates as:

$$\tilde{\mu} = (N_{MO} + N_{LO} + N_C)^{-1} \sum_{\chi \in \chi_{MO}} (\sigma_x / \sigma)$$

The total limit order arrival rate per event is:

$$2\tilde{\alpha}_{all} = \frac{N_{LO}}{N_{MO} + N_{LO} + N_C}$$

The limit order arrival rate in the ZI model is a rate per unit price, so to estimate $\tilde{\alpha}$ we divide $\tilde{\alpha}_{all}$ all by the mean number n of available price levels inside the spread and at the best quotes, measured only at the times of limit order arrivals.
The total cancellation rate per unit volume and per event is[1]:

$$2\tilde{\delta} = \frac{1}{N_{MO} + N_{LO} + N_C} \sum_{x \in \chi_C} \frac{\sigma_x}{\bar{V}} \qquad \bar{V} = \frac{\bar{V}_a + \bar{V}_b}{2}$$

[1] Note that the above estimation procedures all lead to rates per event rather than rates per unit time.

From Bouchauad et all, we estimated parameters: $\tilde{\lambda} \to \tilde{\alpha}, \tilde{\nu} \to \tilde{\delta}, v_0 \to \sigma$.
One of the limits is the presence of arbitrage opportunities

$$\mathcal{R}_\tau \equiv \mathbb{E}\left[\epsilon_t(m_{t+\tau} - m_t)\right] \simeq \frac{1}{2}P(V_{best} = v_0)\mathbb{E}\left[\text{first gap}\right]$$

"This simple observation has an important consequence: the Santa Fe model specification leads to profitable market-making strategies, even when the signature plot is flat (i.e. when prices are diffusive). [...] market-making is profitable on average if the mean bid-ask spread is larger than twice the long-term impact $\mathcal{R}_\infty$. In the Santa Fevmodel, the mean first gap is always smaller than the bid-ask spread. Therefore, $R_\infty < \mathbb{E}\left[spread\right] = 2$, so market-making is easy within this framework." "

The cancellation rate depends on past volatility:

$$\delta_t = \delta + a_k \left(\int_0^t \sqrt{2\beta} e^{-\beta(t-s)} dP_s \right)^2$$

Spread abruptly increases because of the feedback

6.2 Heavy traffic limit: Cont and de Larrad (2013)

For most liquid stocks, where both market orders and limit orders arrive at a high rate, the imbalance between limit orders (which increase the queue size) and market orders and cancellations (which decrease the queue size) is typically much smaller in magnitude. In this scenario, limit orders accumulate and disappear at same rate. Consequently, the behavior of the order queues follows a diffusive pattern, making it pertinent to consider the diffusion limit of the limit order book (similar brownian motion).
In the diffusion limit, the rescaled order book process for the two queues converges weakly to a Brownian motion in the positive orthant $\{x \geq 0, y \geq 0\}$ represent the volume at the best ask and best bid, respectively. These volumes are refreshed when the "walker" reaches one of the two axes, indicating queue depletion.
The probability that the next price move is an increase, given a queue of x shares on the bid side and y shares on the ask side, can be expressed as:

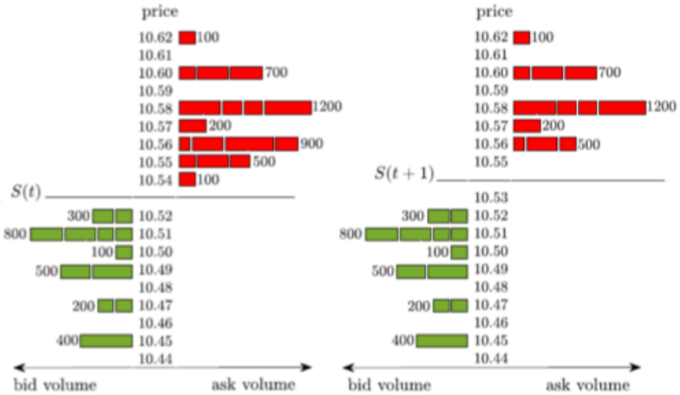
$$p_{up}(x, y) = \frac{1}{2} - \frac{\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}} \frac{y-x}{y+x}\right)}{2\arctan\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)}$$

where ρ is the correlation between order sizes at the bid and at the ask

6.3 Latent Limit Order Book

In modern electronic markets traders interact through a limit order book (LOB) in a continuous double auction.

- Continuous refers to time: at each moment traders can take an action on the LOB
- Double auction means that the LOB is divided into two sides: bid side (buyers) + ask side (sellers)



(a) order book at time t (b) order book at time $t+1$

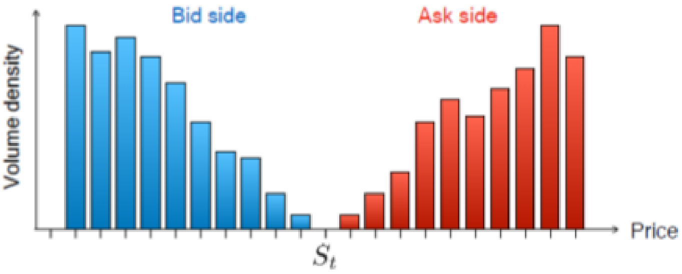
Let give some remarks:

- weak instantaneous liquidity ($<1\%$ of daily traded volume)
- The revealed LOB is short lived.
- Since the square-root impact law is an aggregate low-frequency phenomenon, the relevant object to consider cannot be the revealed LOB

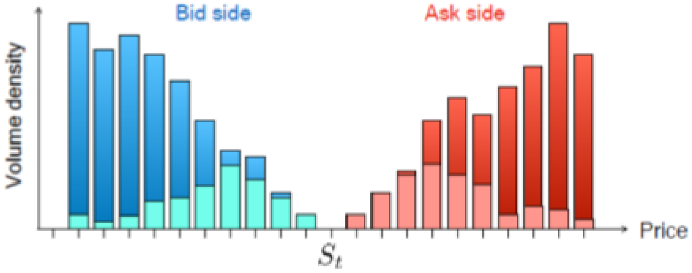
Most of the available liquidity is latent and it is only progressively reveales during the day (like mealting tip of an iceberg).
MM only act as small intermediaries between the much larger volume imbalances of low frequency actors that can only get resolved on large time scales.

6.3.1 Latent Limit Order Book scheme

1. The trade intentions are stored in the latent limit order book (LLOB)

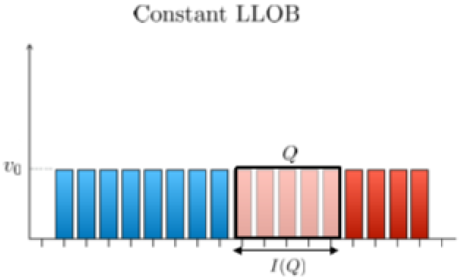


2. The trade intentions are stored in the latent limit order book (LLOB in dark color) and may or not materialise in the revealed limit order book (LOB in lighter color). Latent orders are revealed in the vicinity od the trade price S_t. No incentive i giving away private information too soon!

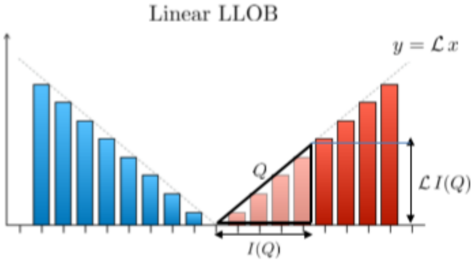


6.3.2 Simple Geometrical Considerations

Let consider a static world:

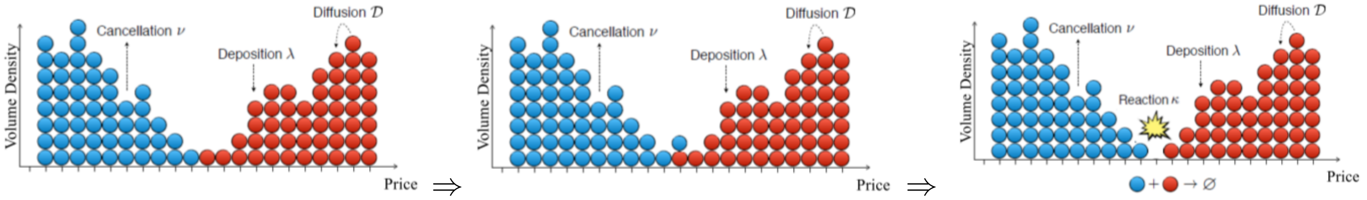


$Q = I(Q) \times v_0 \rightarrow I(Q) = Q/v_0$ linear market impact.
Let consider now a linear LLOB:

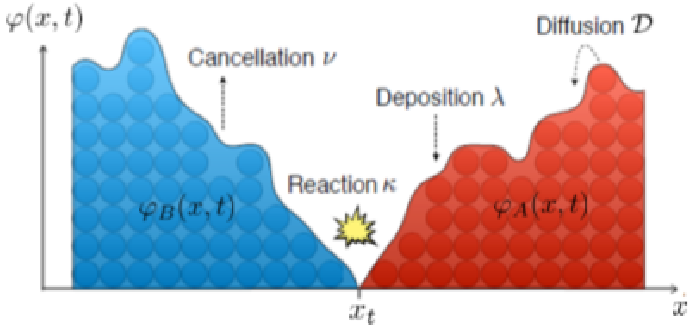


$Q = (I(Q) \times \mathcal{L}I(Q))/2 \rightarrow I(Q) = \sqrt{2Q\mathcal{L}}$ square root market impact
This model is not realistic, we need to include dynamics.

6.3.3 Coarse Grained Dynamics



A first dynamical approach has been proposed by J. Donier et al (Quantitaive Finance 2015):



$$\begin{aligned}
\partial_t\varphi_A(x,t) &= \mathcal{D}\partial_{xx}\varphi_A(x,t) - \nu\varphi_A(x,t) + \lambda\Theta(x_t - x) - R_{A,B}(x,t) \\
\partial_t\varphi_B(x,t) &= \underbrace{\mathcal{D}\partial_{xx}\varphi_B(x,t)}_{\text{Diffusion}} - \underbrace{\nu\varphi_B(x,t)}_{\text{Cancellation}} + \underbrace{\lambda\Theta(x_t - x)}_{\text{Deposition}} - \underbrace{R_{A,B}(x,t)}_{\text{Reaction}}
\end{aligned}$$

where $R_{A,B}(x,t) = \kappa\varphi_a(x,t)\varphi_B(x,t)$. The price equation for the transaction price x_t is:

$$\varphi_a(x_t,t) = \varphi_B(x_t,t)$$

Some consideration:

- in the limit $\kappa \to \infty$ no overlap between φ_A, φ_B
- reaction-diffusion problem in solvable considering the difference:

$$\begin{aligned}
\varphi(x,t) &:= \varphi_B(x,t) - \varphi_A(x,t) \\
\partial_t\varphi(x,t) &= \mathcal{D}\partial_{xx}\varphi(x,t) - \nu\varphi(x,t) + \lambda\text{sign}(x_t - x)
\end{aligned}$$

with boundary conditions:

$$\begin{aligned}
&\varphi(x_t,t) = 0 \quad \forall t \\
&\lim_{|x|\to\infty} \varphi(x,t) \neq \infty \\
&\varphi(x,t=0) = -\varphi(-x,t=0)
\end{aligned}$$

with $\varphi(x,t=0)$ the initial symmetric state of latent order book

We can evaluate the stationaruy LLOB solving:

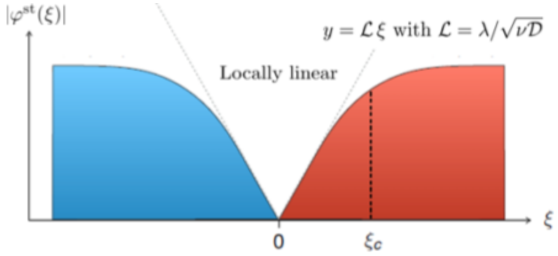
$$\partial_t \varphi^{st}(\xi, t) = 0 \rightarrow \mathcal{D}\partial_{xx}\varphi^{st}(\xi) - \nu\varphi^{st}(\xi) = \lambda$$

with $\xi = x - S_t$, we get:

$$\varphi^{st}(\xi) = -\frac{\lambda}{\nu}\text{sign}(\xi)(1 - e^{-|\xi|/\xi_c})$$

where $\xi_c := \sqrt{\mathcal{D}/\nu}$ and the market total turnover equal to:

$$J := \mathcal{D}\partial_\xi \varphi^{st}|_{\xi=0} = \mathcal{D}\mathcal{L}$$



Let evaluate the market impact: first of all Let us assume to execute the metaorder with volume:

$$Q = \int_0^T m(\tau)d(\tau) \qquad m(\tau) = \text{trading intensity rate}$$

Following Donier et al approach, in the infinit memory limit $\nu, \lambda \rightarrow 0$ with constant $\mathcal{L} \sim \lambda\nu^{-1/2}$, so the price dynamics is given solving:

$$\partial_t \varphi(x, t) = \mathcal{D}\partial_{xx}\varphi(x, t) + m(t)\delta(x - x_t)$$

(the last one is the extra current of buy/sell particles falling at the transition price $x = x_t$), with the boundary conditions:

$$\varphi(x, t = 0) = \varphi^{\text{st}}(x)$$
$$\lim_{x\rightarrow\infty} \partial_x \varphi(x, t) = -\mathcal{L}$$

Imposing $\varphi(x_t, t) = 0$ in the solution of the PDE:

$$\varphi(x, t) = -\mathcal{L}x + \int_0^t d\tau \frac{m(\tau)}{\sqrt{4\pi\mathcal{D}(t-\tau)}} e^{-\frac{(x-x_\tau)^2}{4\mathcal{D}(t-\tau)}}$$

we obtain the following self-consistent relation for the transaction pric:

$$x_t = \frac{1}{\mathcal{L}} \int_0^t d\tau \frac{m(\tau)}{\sqrt{4\pi\mathcal{D}(t-\tau)}} e^{-\frac{(x-x_\tau)^2}{4\mathcal{D}(t-\tau)}}$$

If impact is small $|y(s) - y(t)|^2 \ll D(t-s)$ the price dybamics becomes:

$$y(t) = \frac{1}{\mathcal{L}} \int_0^t \frac{dsm(s)}{\sqrt{4\pi D(t-s)}}$$

This makes explicit the fact that in this model market impact is transient!
In general the equation for x_t can be solved numerically and the price impact $I := x_T$ is described by:

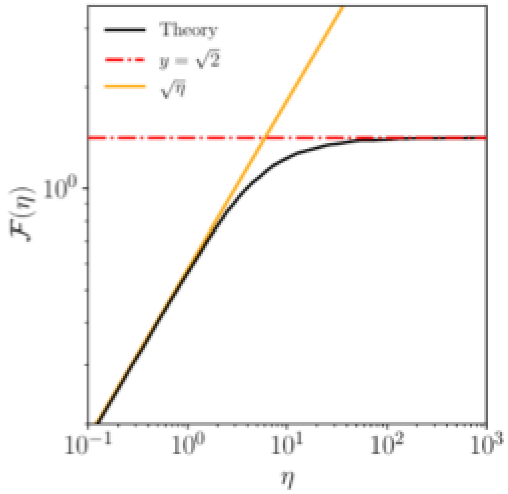
$$I(Q) = \frac{\mathcal{D}Q}{J}\mathcal{F}(\eta)$$

with $\eta = Q/(JT)$ the participation rate and:

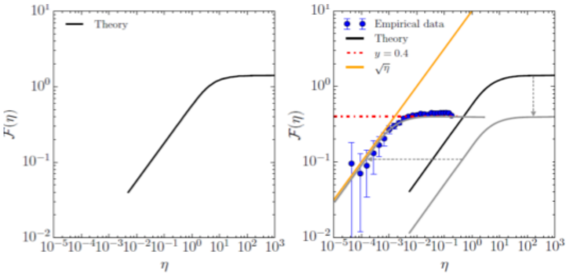
$$\mathcal{F} \sim \begin{cases} \sqrt{\eta/\pi} & \text{for } \eta \ll 1 \quad \text{small participation rate} \\ \sqrt{2} & \text{for } \eta \gg 1 \quad \text{large participation rate} \end{cases}$$

It follows that in the infinite memory LLOB model the market impact $I(Q)$:

- is linear in Q for small Q at fixed T
- crosses over to a square root for large Q
- is independent from T in the square root regime



Estimate it with empirical data:



We noticed that there is a good qualitative afreement but not quantitative.
The intuition is that the total market turnover J is actually dominated by HFTs/MM, while resistance to show metaorders can only be provided by slow participants(institutional investors). For this reason we introduce the LLOB model with two time-scales for market participants: fast and slow ones.
We consider slow agent:λ_s, ν_s and fast agents: λ_f, ν_f with $\lambda_s, \nu_s \ll \lambda_f, \nu_f/$
Two contributions to the latent order book:

$$\partial_t \varphi_s(x,t) = \mathcal{D}_s \partial_{xx} \varphi_s(x,t) - \nu_s \varphi_s(x,t) + \lambda_s(x_t - x) \quad \text{with } \nu_s T \to 0 \quad \text{(infinite memory)}$$
$$\partial_t \varphi_f(x,t) = \mathcal{D}_f \partial_{xx} \varphi_f(x,t) - \nu_f \varphi_f(x,t) + \lambda_f(x_t - x) \quad \text{with } \nu_f T \gg 1 \quad \text{(very short memory)}$$

Let $\xi = x - x_t$, the total market turnover is:

$$J = |\mathcal{D}_s \partial_\xi \varphi_s^{st} + \mathcal{D}_t \partial_\xi \varphi_f^{st}|_{\xi=0} = J_s + J_f \quad \text{with } J_f \gg J_s \to J \sim J_f$$

We focus on two different regime, where the metaorder intensity m_0 is:

- large compared to the average transaction rate of slow traders J_s (long time)
- but small compared to the total transaction rate of the market J (short time)

$$J_s \ll m_0 \ll J$$

We modify the equation including this information:

$$\partial_t \varphi_s(x,t) = \mathcal{D}_s \partial_{xx} \varphi_s(x,t) - \nu_s \varphi_s(x,t) + \lambda_s(x_t - x) + m_{s,t}\delta(x - x_t)$$
$$\partial_t \varphi_f(x,t) = \mathcal{D}_f \partial_{xx} \varphi_f(x,t) - \nu_f \varphi_f(x,t) + \lambda_f(x_t - x) + m_{f,t}\delta(x - x_t)$$

with $m_{f,t} + m_{s,t} = m_0 \quad \forall t$ (we added the fraction of the metaorder executed against the slow liquidity and the metaorder executed against the fast liquidity).
Price equation is

$$\varphi(x_t, t) = \varphi_s(x_t, t) + \varphi_f(x_t, t) = 0$$

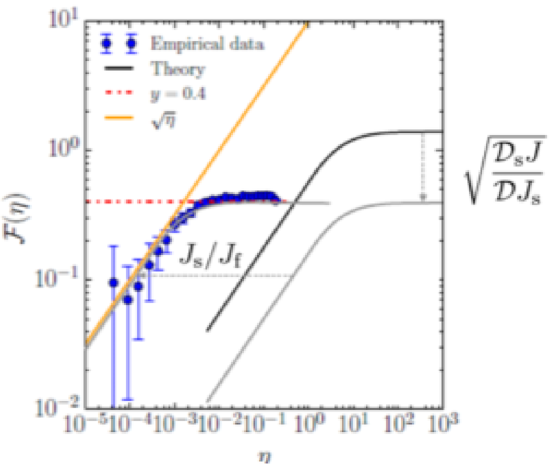
We can solve this model for $T > T^\dagger$ where:

- $T^\dagger = \nu_f^{-1} \eta^{*-2} \mathcal{D}_s / \mathcal{D}_f$
- $\eta^* = J_s / J_f$ For $T > T^\dagger$ the price impact is described by the scaling:

$$I(Q) = \sqrt{\frac{\mathcal{D}_s Q}{J_s}} \mathcal{F}\left(\frac{\eta}{\eta^*}\right)$$

We can recover the infinite memory LLOB rescaling both the axis:

$$I(Q) = \sqrt{\frac{\mathcal{D}Q}{J}} \mathcal{F}(\eta) \xrightarrow{\sqrt{\frac{\mathcal{D}_s Q}{\mathcal{D} J_s}}} \sqrt{\frac{\mathcal{D}_s Q}{\mathcal{D} J_s}} \mathcal{F}(\eta) \xrightarrow{\eta \to \frac{\eta}{\eta^*}} \sqrt{\frac{\mathcal{D}_s Q}{J_s}} \mathcal{F}\left(\frac{\eta}{\eta^*}\right)$$



Allowing at least two characteristic timescales for the liquidity (fast and slow) we empirically confirm the crossover from a linear to a square to a square root regime described by:

$$I(Q) \propto \sqrt{Q}\mathcal{F}(\eta) \sim \begin{cases} Q & \text{for } \eta \ll \eta^* \quad \text{'small participation rate regime'} \\ \sqrt{Q} & \text{for } \eta \gg \eta^* \quad \text{'large participation rate regime'} \end{cases}$$

with $\eta = Q/(JT)$, $T = t_{end} - t_{start}$ and:

$$\mathcal{F}(\eta) \sim \begin{cases} \sqrt{\eta/\pi} & \eta \ll \eta^* \\ c = 0.4 & \eta \ll \eta^* \end{cases}$$

Market impact should not be misconstrued as volatility, in particular the square root law has nothing to do with price diffusion, price changed Δs grow as the square root of duration $\sqrt{T}$.
Let give some final comments:

- In agreement with dynamical liquidity (earlier equation) the market impact is characterized by a crossover from a linear to a square root regime
- The market impact in the square-root regime is independent drom the execution duration T
- Market impact should not be misconstrued as volatility

6.4 Queue-reactive order book model

In their paper "Simulating and analyzing order book data: The queue-reactive model," Huang, W., Lehalle, C.A., and Rosenbaum, M. (2015) introduce three models for simulating the dynamics of the limit order book (LOB). Their central idea is that order flow still follows a Poissonian process, but the arrival rate is contingent on the current state of the LOB. As the order flow influences and modifies the state of the LOB, this interaction leads to order flow that is both auto-correlated (correlated with itself over time) and cross-correlated (correlated with other order flow components).

Queue-reactive order book

- The LOB is modeled as a $2K-$dimensional process, $X(t) = \{q_{-k}(t, \dots, q_{-1}(t), q_1(t), \dots, q_k(t))\}$, where $q_i(t)$ is the number of shares at price level i (ticks) from the reference price p_{ref}
- p_{ref} is equal to the midprice when the spread is equal to one tick or an odd number of ticks. When it is even, it is either:

$$p_{mid} + \frac{\text{tick}}{2} \qquad \text{or} \qquad p_{mid} - \frac{\text{tick}}{2}$$

  choosing the one which is the closest to the privious value of p_{ref}
- The arrival rates $\lambda_L^i(X(t)), \lambda_C^i(X(t))$, and $\lambda_M^i(X(t))$ of limit, cancellation and market order, respectively, at level $i \in \{-K, \dots, K\}$, depend on:
  - only of the target queue size (Model 1)
  - also of the size of the other queues (Model 2)
  - also the most recent chanfe in the reference price (Model 3)

n_{tot} shares must be executed in M periods. In each period we want to trade n_i shares.
While optimal execution (e.g Almgren-Chriss) tells us how much to trade in each period, it is silent about how to trade (e.g. limit vs market orders, when to cancel, etc) in each period. In the ith slice, both tactics post a limit order of size n_i at the best offer queue at the beginning of the period, and send a market order with all the remaining quantity to complete the execution of the target volume at the end time of the slice. In between:

- T1 (Fire and forget): When p_{mid} changes, cancel the limit order and send a market order at the opposite side with all the remaining volume if any.
- T2 (Pegging to the best): When the best offer price changes or our order is the only remaining orser at the best offer limit, cancel the order and repost all the remaining volume at the newly revealed best offer queue

Two types of order scheduling strategies:

- S1: A linear scheduling ($n_i = n_{tot}/M$), used for the VWAP benchmark
- S2: An exponential scheduling $n_i = n_{tot}(e^{-(i-1)/4} - e^{-i/4})$ with the arrival price S_0 as benchmark

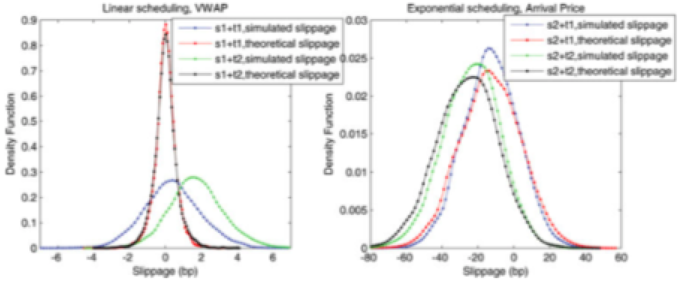
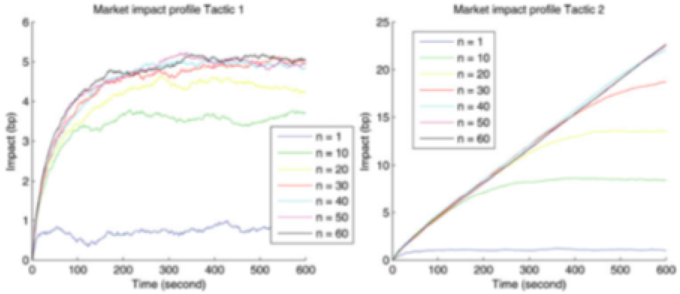


Figure 11. Simulation results for the tactics.



VII

Modelling microstructure noise

7.1 Introduction

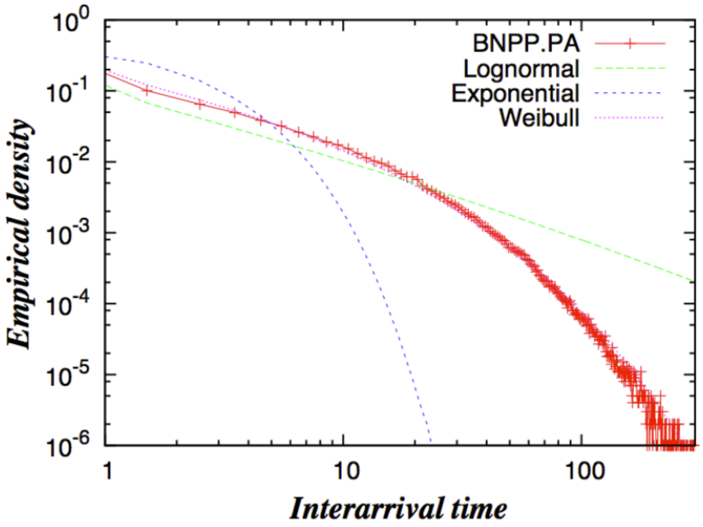
When modeling financial transaction data, particular attention is given to certain distinctive properties of such data, one of which is the irregular spacing in time between events. This characteristic leads to the modeling of financial time series as a point process, specifically referred to as a financial point process. In this context, the inter-event waiting times are classified into three categories:

- **Trade/Quote Duration Time**: This represents the time between two consecutive trade or quote arrivals.

- **Price Duration Time**: It measures the time it takes for cumulative absolute price changes of a given magnitude to occur.

- **Volume Duration Time**: This denotes the time it takes until a cumulative order volume of a specified size is traded.

By analyzing these inter-event waiting times and their characteristics, we can gain valuable insights into the dynamics of financial markets and develop models to better understand and predict market behavior.
From empirical data, Poissonian distribution does not fit inter-arrival trade times:



And analysing ACF, we notices a strong positive correlation, exists a memory.

Point process

A financial point process is a simple temporal marked point process, characterized by:

- $\{t_i\}_{i\in\{1,\ldots,n\}}$: random sequence of increasing event times, $0 := t_0 < t_2 < \ldots < t_n$
- N : cadlag counting process, $N_t := \sum_{1\geq 1} \mathbb{1}_{\{t_i \leq t\}}$ s.t. $\mathbb{E}\left[N_t\right] < \infty, \quad \forall t \geq 0$
- $x_i := t_i - t_{i-1}, i = 1, \ldots, n$:inter-event durations
- x: backward recurrence time, defined by $x(t) := t - t_{\breve{N}_t}$, where $\breve{N}_t := \sum_{i\geq 1} \mathbb{1}_{\{t_i < t\}}$
- $\mathcal{F}^N$: internal history of the point process
- $\mathcal{F}_t$: a more general filtration including covariates $\mathcal{F}_t^N \subseteq \mathcal{F}_t$

A counting process N_t is a submatringale,since:

$$\mathbb{E}\left[N_t|\{N_\tau : \tau \leq s\}\right] \geq N_s \qquad \forall s \leq t$$

Compensator

According to Doob-Meyer decomposition there is a unique predictable increasing process such that $\Lambda_0 = 0$, Λ_∞ is integrable, and $N_t = M_t + \Lambda_t$ with M an uniformly integrable $\mathcal{F}$-martingale. Λ is a $\mathcal{F}-$compensator of N.

In a Poisson process with rate λ, the compensator is $\Lambda_t = \lambda t$, since:

$$\mathbb{E}\left[N_t - \lambda t|\{N\tau : \tau \leq s\}\right] = 0$$

$N_t - \lambda t$ is an $\mathcal{F}-$ martingale

Intensity function

The $\mathcal{F}$-intensity function λ_t of N is a scalar positive $\mathcal{F}-$predictable process defined by:

$$\Lambda_t = \int_0^t \lambda_u du$$

It can also be defined by:

$$\mathbb{E}\left[N_t - N_s|\mathcal{F}\right] = \mathbb{E}\left[\int_s^t \lambda_u du|\mathcal{F}_s\right] \qquad \forall 0 \leq s \leq t$$

so:

$$\lambda_{t+} := \lim_{\Delta \to 0^+} \lambda_{t+\Delta} = \lim_{\Delta \to 0^+} \frac{1}{\Delta}\mathbb{E}\left[N_{t+\Delta} - N_t|\mathcal{F}_t\right]$$

we define the probability in unite per time to observe a count

Survivor function

Let X be the random variable describing the inter-event durations; $f(x)$ is its probability density function and $S(x)$ the survivor function:

$$S(x) := 1 - F(x) = 1 - Pr[X \geq x] = Pr[X > x]$$

Hazard function

The Hazard function $h(x)$ off a point process is defined as

$$h(x) := \frac{f(x)}{S(x)} = \lim_{\Delta \to 0} \frac{1}{\Delta} Pr[x \leq X < x + \Delta|X \geq x]$$

A point process can be equivalently defined by three differnet representation: intensity, duration and counting representation:

Intensity representation

$$Pr[(N_{t+\Delta} - N_t) = 1|\mathcal{F}_t] = \lambda\Delta + o(\Delta)$$
$$Pr[(N_{t+\Delta} - N_t) > 1|\mathcal{F}_t] = o(\Delta)$$

where $\lambda > 0$ is the constant Poisson rate

Duration representation

$$x_i \sim i.i.d(\text{Exp}(\lambda))$$

Counting representation

Let $N_{(a,b])} := N_b - N_a$ then:

$$Pr[N_{(a,b])} = k] = e^{-\lambda(b-s)} \frac{(\lambda(b-a))^k}{k!}$$

Let make an example for a non-homogeneous Poisson process:

- **intensity representation**:

$$Pr[(N_{t+\Delta} - N_t) = 1 | \mathcal{F}_t] = \lambda_t \Delta + o(\Delta)$$
$$Pr[(N_{t+\Delta} - N_t) > 1 | \mathcal{F}_t] = o(\Delta)$$

  where $\lambda_t > 0$ is the intensity function

- **duration representation**: $x_i \sim i.i.d(\text{Exp}(\Lambda_{t_i+1} - \Lambda_{t_i}))$
- **counting representation**: let $\Lambda(a,b) := \lambda_d - \Lambda_a = \int_a^b \lambda_s ds,$:

$$Pr[N_{(a,b]} = k] - e^{-\Lambda(a,b)} \frac{(\Lambda(a,b))^k}{k!}$$

Multivariate point process

A $K-$variate point process is a process $\{t_i, d_i\}_{i \in (1,\dots,n)}$ on $(0, \infty)$, where:

- $d_i \in \{1, \dots, K\}$ is a random variable representing K different types of events
- $\{t_i^k\}_{i \in \{1,\dots,n^k\}}, k = 1, \dots, K$ are the arrival times
- $N_t^k = \sum_{i \geq 1} \mathbb{1}_{\{t_i \leq t\}} \mathbb{1}_{\{d_i = k\}}$

Random Time Change Theorem

Given a K-varaite point process $N = (N^1, \dots, N^k)$ with event times $\{t_i^k\}_{i\in\{1,\dots,n^k\}}$, $k = 1, \dots, K$ and compensators $\Lambda = (\Lambda^1, \dots, \Lambda^k)$ s.t $\Lambda^k_\infty = \infty, \forall k = 1, \dots, K$. Then the K scalat point processes with event times $\{\Lambda^k_{t_i^k}\}_{i\in\{1,\dots,n^k\}}, k = 1, \dots, K$ are independent homogeneoud Poisson processes with unit intensity.

The random time change theorem plays an important role in order to construct diagnostic tests for point process models or to simulate point processes, in fact: we estimate with Λ_t, using theorem we obtain Λ from λ, I plot a qqplot and evaluate if process is poisson process.

7.2 Model

Models of the discrete time duration process $\{x_i\}_{i=1,\dots,n}$ observable at the event times $\{t_i\}_{i=1,\dots,n}$ obtained parametrizing either the conditional distribution function $F(x_i|\mathcal{F}_{t_{i-1}})$ or the hazard rate $h(x_i|\mathcal{F}_{t_{i-1}}\ )$

Proportional Hazard model (Cox 1972)

$$h(x|z;\theta) = h_0(x|\gamma_1)g(z,\gamma_2)$$

where $\theta = (\gamma_1, \gamma_2)$ is the baselime hazard rate and g is a function of covariates z and parameters γ_2

The baseline hazard rate h_0 can be parametrized by a Weibull distribution with parameters λ and p

$$h_0(x|\gamma_1) = \lambda(\lambda x)^{p-1}$$

which is the hazard rate associated with the pdf:

$$f(x) = \begin{cases} \lambda p(\lambda x)^{p-1} e^{-(\lambda x)^p} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

we can distinguish three cases:

- $p = 1 \to h_0(x|\gamma_1) = \lambda$ (exponential, no memory, it is not dependent on x)
- $p > 1 \to \partial_x h_0(x|\gamma_1) > 0$ (positive duration dependence)
- $p < 1 \to \partial_x h_0(x|\gamma_1) < 0$ (negative duration dependence)

7.2.1 Autoregressive Conditional Duration

Let us denote by $x_i := \frac{t_i - t_{i-1}}{s(t_i)}$ the inter-event duration normalized by a seasonality factor $s(t_i)$.
Engle and Ruseel in 1997 proposed that $x_i = \Psi_i \epsilon_i$ where ϵ_i are i.i.d positive random variables with $\mathbb{E}[\epsilon_i] = 1$, so:

$$\Psi_i = \mathbb{E}[x_i|\mathcal{F}_{t_i - 1}]$$

is the conditional duration mean.
The model can be rewritten in the terms of the intenisty as:

$$\lambda(t|\mathcal{F}_t) = \lambda_\epsilon \left(\frac{x(t)}{\Psi_{\breve{N}(t)+1}} \right) \frac{1}{\Psi_{\breve{N}(t)+1}}$$

where $\lambda_\epsilon(s)$ is the hazard function of the errore term

Exponential ACD model (Dufour and Engle 2000)

$$\epsilon_i \sim i.i.d(Exp(1)) \qquad \rightarrow \qquad \lambda(t|\mathcal{F}_t) = \frac{1}{\Psi_{\breve{N}_{t+1}}}$$

Typically Ψ_i is choosen to be a function of the past information set $\mathcal{F}_{t_i-1}$ as $\Psi_i = \Psi(\Psi_{i-1}, \ldots, \Psi_{i-q}, x_{i-1}, \ldots, x_{i-p})$

Linear ACD model

The generic ACD(r,s) model is:

$$\psi_i = \omega + \sum_{j=1}^{r} \alpha_j x_{i-j} + \sum_{j=1}^{s} \beta_j \Psi_{i-j}$$

Similar to GARCH models, the process $\eta_i = x_i - \Psi_i$ is a Martingale difference sequence ($\mathbb{E}\left[\eta_i|\mathcal{F}_{i-1}\right] = 0$) and the ACD can be written as:

$$x_i = \omega + \overbrace{\sum_{j=1}^{\max(r,s)} (\alpha_j + \beta_j) x_{i-j}}^{\text{autoregressive}} - \underbrace{\sum_{j=1}^{s} \beta_j \eta_{i-j} + \eta_j}_{\text{mean average}}$$

This is an ARMA representation with non Gaussian innovations.
Taking expections of both sides od the last expression and assuming stationarity:

$$\mathbb{E}\left[x_i\right] = \frac{\omega}{1 - \sum_{j=1}^{\max(r,s)} (\alpha_j + \beta_j)}$$

hence we need to assume that $\sum_{j=1}^{\max(r,s)} (\alpha_j + \beta_j) < 1$.

7.2.2 EACD(1,1) model

$$x_i = \Psi_i \epsilon_i \qquad \epsilon_i \sim i.i.d(Exp(1)) \qquad \Psi_i = \omega + \alpha_1 x_{i-1} + \beta_i \Psi_{i-1}$$

Evaluating the expected value of each element, we obtain:

$$\mathbb{E}\left[\epsilon_i\right] = 1 \qquad Var[\epsilon_i] = 1 \qquad \mathbb{E}\left[\epsilon_i^2\right] = 2$$

We have:

$$\mathbb{E}\left[x_i\right] = \mathbb{E}\left[\mathbb{E}\left[\psi_i \epsilon_i | \mathcal{F}_{i-1}\right]\right] = \mathbb{E}\left[\psi_i\right] \qquad \mathbb{E}\left[\psi_i\right] = \omega + \alpha_1 \mathbb{E}\left[x_{i-1}\right] + \beta_i \mathbb{E}\left[\psi_{i-1}\right]$$

under weak stationarity:

$$\mu_x := \mathbb{E}\left[x_i\right] = \mathbb{E}\left[\psi_i\right] = \frac{\omega}{1-\alpha_1-\beta_1}$$

Moreover $\mathbb{E}\left[x_i^2\right] = \mathbb{E}\left[\mathbb{E}\left[\psi_i^2 \epsilon_i^2 | \mathcal{F}_{i-1}\right]\right] = 2\mathbb{E}\left[\psi_i^2\right]$ With similar arguments we can show that:

$$Var[x_i] = \mu_x^2 \frac{1-\beta_1^2-2\alpha_1\beta_1}{1-\beta_1^2-2\alpha_1\beta_1-2\alpha_1^2} \tag{VII.1}$$

$$\rho_1 := Corr[x_i, x_{i-1}] = \frac{\alpha_1(1-\beta_1^2-\alpha\beta)}{1-\beta_1^2-2\alpha_1\beta_1} \tag{VII.2}$$

$$\rho_k := Corr[x_i, x_{i-k}] = (\alpha - 1 + \beta_1)\rho_{k-1} \qquad \forall k \geq 2 \tag{VII.3}$$

From the first we can evaluate that the duration process is covariance stationary if:

$$\beta_1^2 + 2\alpha_1\beta_1 + 2\alpha_1^2 < 1$$

From the third we notice that ACF decreases at a geometric rate (while real data can also exhibit a hyperbolic ACF)

7.2.3 Estimation of ACD models

For an ACD(r,s) model, setting $i_0 = \max(r, s)$, the likelihood of the durations, $x_1, \ldots, x_T$ is:

$$f(\mathbf{x}_T|\theta) = \left[\prod_{i=i_0+1}^{T} f(x_i|\mathcal{F}_{i-1}, \theta)\right] \times f(\mathbf{x}_{i_0}|\theta)$$

Conditional likelihood is much easier to handle and essentially equivalent for large T.
For a WACD the log-likelihood is:

$$\ell(\mathbf{x}|\theta, \mathbf{x}_{i_0}) = \sum_{i=i_0+1}^{T} p \ln\left[\Gamma\left(1+\frac{1}{p}\right)\right] + \ln\left(\frac{p}{x_i}\right) + p\ln\left(\frac{x_i}{\Psi_i}\right) - \left[\frac{\Gamma\left(1+\frac{1}{p}\right)x_i}{\Psi_i}\right]^p$$

where $\Psi_i = \omega + \sum_{j=1}^{r} \alpha_j x_{i-j} + \sum_{j=1}^{s} \beta_j \psi_{i-j}$.
When $p = 1$ we obtain the conditional log-likelihood for a EACD(r,s) model

7.2.4 Logarithmic ACD model

From ACD model is difficult to allow Ψ_i to depend on functions of covariates without violating the non-negativity restriction. Bauwens and Giot (200) propose a class of logarithmic ACD models, where no parametric restriction are needed to ensure positiveness of the process:

$$\ln \Psi_i = \omega + \beta_1 \ln \Psi_{i-1} + \alpha_1 g(\epsilon_{i-1}) \qquad \omega > 0, \alpha, \beta \geq 0$$

$$\begin{aligned} &\text{type } I \qquad g(\epsilon_{i-1}) = \ln \epsilon_{i-1} \\ &\text{type } II \qquad g(\epsilon_{i-1}) = \epsilon_{i-1} \end{aligned}$$

The duration process is covariance-stationary iff:

$$\beta < 1 \qquad \mathbb{E}\left[\epsilon_i e^{\alpha g(\epsilon_i)}\right] < \infty \qquad \mathbb{E}\left[e^{2\alpha g(\epsilon_i)}\right] < \infty$$

7.2.5 Autoregressive Conditional Intensity model (Russel, 1999)

Let $\lambda(t) = (\lambda^1(t), \ldots, \lambda^K(t))'$, for $k = 1, \ldots, K$:

$$\lambda^k(t) = \Phi^k_{\breve{N}(t)+1} \lambda^k_0(t) s^k(t)$$

where:

- λ^k_0 is a baseline (determinstic) intensity
- $s^k(t)$ is a seasonali component
- $\Phi^k_i = \exp(\tilde{\Phi}^k_i + z'_{\tilde{t}_j}\gamma_k)$
- z_i ate vectors of covariates observable at arrival time t_i with parameter vector γ^k

The vector $\tilde{\Phi}_i = (\tilde{\Phi}^1_i, \ldots, \tilde{\Phi}^k_i)'$ is parametrized as:

$$\tilde{\Phi}_i = \sum_{k=1}^{K} (A^k \epsilon_{i-1} + B^k \tilde{\Phi}_{i-1}) y^k_{k-1}$$

where $y^k_i = \mathbb{1}_{\{t_i = t^*_j\}}$ and $A^k = (a^k_1, \ldots, a^k_K)'$ and $B^k = \{b^k_{ij}\}_{i,j=1,\ldots,K}$ and scalar innovation term ϵ_i:

$$\epsilon_i = \sum_{k=1}^{K} \left(1 - \int_{t^k_{N^k_{t_i}-1}}^{t^k_{N^k_{t_i}}} \lambda^k(u) du \right) y^k_i$$

The fundamental principle of the ACI model is that at each event occurrence at time t_i, all K processes are updated based on the realization of the integrated intensity relative to the most recent process. Importantly, the impact of this innovation on the K processes can vary, and it also depends on the type of the most recent point in the process sequence.

7.2.6 Hawkes processes

A different dynamic intensity model is obtained by specifying $\lambda(t)$ as a(linear) self-exciting process given by:

$$\lambda(t) = \mu + \int_0^t \phi(t-s)dN(s) = \mu + \sum_{t_i<t} \phi(t-t_i)$$

where $\phi(s)$ is a non-negative weight function and $\int_0^t \phi(s)dN(s)$ is the stochastic Stieltjes integral of the process ϕ with respect to the counting process $N(t)$

Hawkes process

Hawkes process belong to the classs of mutually-exciting process, it is defined through its intensity function as:

$$\lambda_i(t) = \mu_i + \sum_{j=1}^{D} \int_{\infty}^{t} \phi_{ij}(t-s)dN_j(s) = \mu_i + \sum_{i=1}^{D} \sum_{t_j<t} \phi_{ij}(t-t_j)$$

where:

- $dN_j(s) = \sum_{t_j<s} \delta(s-t_j)ds$
- μ_i is a positive constant baseline intenisty
- the kernels $\phi_{ij}(t)$ are positive and causal functions in L_1

Denoting by $\mathbf{\Phi}(t)$ the kernel matrix:

$$\mathbf{\Phi}(t) = \begin{pmatrix} \phi_{11}(t) & \dots & \phi_{1D}(t) \\ \vdots & \ddots & \vdots \\ \phi_{D1}(t) & \dots & \phi_{DD}(t) \end{pmatrix}$$

The process $N(t)$ has asymptotically statonary increments and $\lambda(t)$ is asymptotically stationary if the spectral radius of the matrix is:

$$||\mathbf{\Phi}|| = \{||\phi_{ij}||\} < 1$$

Utility of Hawkes processes:

- flexible and versatile tool to investigate mutual interaction (excitation);
- the linear structure of $\lambda(t)$ allows to compute many properties analytically;
- quantities have a clear interpretation;
- availability of parametric and non-parametric estimation tools

Hawkes processes, introduced more than four decades ago, have found significant utility in modeling earthquake data. In recent years, they have gained popularity in mathematical finance and econometrics, with various applications:

- Standard Applications: Hawkes processes are applied to model the arrival times of trades, buy-sell market orders, and mid-price changes in financial markets. These applications are found in research by Bowsher (2007) and Bauwens and Hautsch (2009).

- Additional Examples: Other examples include modeling limit and market order flow in continuous double auction markets (Muni-Toke and Pomponio, 2012), studying the arrival of trades-through orders in a limit order book, and developing price impact models to replicate strong microscopic mean reversion and the Epps effect. Filimonov and Sornette (2012) introduced a measure based on Hawkes processes, aiming to provide insights into the level of endogeneity and its potential as a predictor of market micro-instabilities. However, Bouchaud and collaborators (2013) challenged this claim in a recent paper.

- Contagion Modeling: Ait-Sahalia et al. (2012) propose a model for asset returns capable of capturing crisis periods characterized by contagion. In this model, the jump diffusion component of asset dynamics is described using a class of multi-dimensional Hawkes models. They discuss an estimation methodology based on the Generalized Method of Moments, although their analysis is limited to pairs of assets.

- These applications demonstrate the versatility of Hawkes processes in modeling various aspects of financial markets and risk contagion.

VIII

Networks

8.1 Introduction

The structure of a network can be descibed by a $n \times n$ matrix: the (weighted) adjacency matrix $A = \{\mathbf{A}_{ij}\}$. If two nodes i and j are not joined by a link, $A_{ij} = 0$. Otherwise, $A_{ij} = 1$ (for binary networks) or $A_{ij} \in \mathbb{R} - \{0\}$ (for weighted networks). For binary directed networks $A_{ij} = 1$ if there is an edge from i to j. For undirected networks the adjacency matrix is symmetric, $\mathbf{A} = \mathbf{A}^T$. Usually, $A_{ii} = 0$ for any i (no self-edges).

Walk, trail, path, length

- A walk is any sequence of edges which joins a sequence of vertices
- A trail is a walk in which all edges are distinct
- A path is a trail in which all vertices (and therefore also all edges) are distinct
- The length of a walk/trial/path is the number of edges traversed along the path

The total number of path of length 2 from i to j is:

$$N_{ij}^{(2)} = \sum_{k=1}^{n} A_{ik} A_{kj} = (A^2)_{ij}$$

and in general the number of walks connecting $i \to j$ of length r is $(A^r)_{ij}$

Cycles

A cycle is any walk/trial/path $i \to i$ (initial node and final one are the same). The total numer of cycles of length r is:

$$L_r = \sum_{i=1}^{n} (A^r)_{ii} = \text{Tr}[A^r]$$

For symmetric networks, the adjacency matrix is symmetric, eigenvalues are real and it can be diagonalized as:

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$$

where Λ is the diagonal matrix of eigenvalues λ_i and U is the orthogonal matrix of eigenvectors (arranged in columns). Then $A^r = (U\Lambda U^T)^r = U\Lambda^r U^T$ and:

$$L_r = \text{Tr}[\mathbf{U}\boldsymbol{\Lambda}^r\mathbf{U}^T] = \text{Tr}[\boldsymbol{\Lambda}^r] = \sum_{i=1}^{n} \lambda_i^r$$

That relation is true from any networks (proof using Schur decomposition)

Acyclic Directed Networks or Directed Acyclic Graph (DAG)

A cycle in a directed network is a closed loop of edges with the arrows on each edge pointing in the same direction around the loop. Directed networks without cycles are acyclic. One can always depict a DAG in a ordered way - vertices running from top to bottom, all edges point downward.

By using the sorting of nodes in the previous representation, the adjacency matrix becomes (strictly) upper triangular.

Eigenvalues acyclic network

The eigenvalues of the adjacency matrix are all zero if and only if the network is acyclic, then the total number L_r of cycles of length r is:

$$L_r = \sum_{i=1}^{n} \lambda_i^r$$

A tree is a connected, undirected network that contains no closed loops. (Connected means that every vertex is reachable from every other via some path).
A network consisting of more than one componenet, each being a tree, is a forest.

Condition for existence of tree

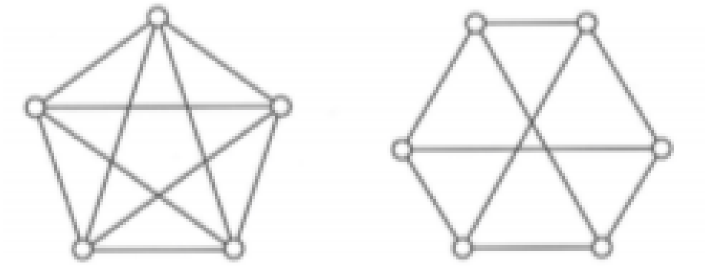
A network is a tree if and only if it is connected and has $n - 1$ edges.

A planar network is a network that can be drawn on a plane without having any edges cross. Two important examples of non-planar networks

K_5 is the complete graph with 5 vertices. UG is the complete bipartite graph on two groups of three vertices.

Kuratowski's theorem

Every non-planar networks contains at least one subgraph that is an expansion of K_5 or UG



(a) K_5

(b) UG

The (binary) network is represented by the $n \times m$ incidence matrix B (the equivalent of the adjacency matrix).One-mode projection is a unipartite network where nodes are linked if they share at least one node of the second type. Nodes of type A connected to the same node of type B form a clique in the projected network.

Weighed adjacency Matrix Bipartite network

In bipartite network:

$$A_{ij} = \sum_{k=1}^{m} B_{ki} B_{kj} = \sum_{k=1}^{m} B_{ik}^{T} B_{kj}$$

However the diagonal elements are $A_{ij} = \sum_{k=1}^{m} B_{ki}^2 \neq 0$ and should be set to zero by hand. The ptjer projecton gives $P = BB^T$

8.2 Graph Laplacian

Graph Laplacian

The graph Laplacian is the matrix is $\mathbf{L} = \mathbf{D} - \mathbf{A}$

We can write a diffusion equation as:

$$\frac{d\vec{\psi}}{dt} + cL\vec{\psi} = 0$$

similar to $\partial_t \vec{\psi} + c\lambda^2 \vec{\psi} = 0$.
The solution of the diffusion equation can be found by setting:

$$\vec{\psi}(t) = \sum_i a_i(t)\vec{v}_i$$

where $\vec{v}_i$ are the eigenvectors of L assocaited to the eigenvalues λ_i. Substituting we get:

$$\dot{a}_i + c\lambda_i a_i = 0$$

with solution:

$$a_i(t) = a_i(0)e^{-c\lambda_i t}$$

The diffusion is governed by $\{\lambda_i\}$.
Let us analyse the eigenvalues of graph Laplacian:

- $\lambda_i \in \mathbb{R}$ for any i ($\mathbf{L}$ is symmetric if network is undirected) and $\lambda_i \geq 0 \quad \forall i$ (it can be proof through an edge incidence matrix $\mathbf{B}$). From this we see that diffusion over a network is never an exploding process.
- At least one eigenvalue is zero. This can be seen as a conservative law.
- If the network is divided in k disconnected components, the adjency matrix and the Laplacian are block diagonal: in similar way of above, for each block leads to other k eigenvalues equal to zero (never exist a link between a group and another one)
- The second largest eigenvalue of the graph Laplacian is non-zero $\iff$ the network is connected. This eigenvalue is called algebraic connectivity or spectral gap of the network.

8.3 Random walk on a network

A random walk on a network is a walk defined by a sequence of random steps, it is useful finding the probability $p_i(t)$ that the walker is at node i at time t.
The possible steps from generic j, it is the uniform probability $1/k_j$ of taking a step along one of the k_j

edges incident to j times the probability to be in j at time $t-1$. Summer over all possible j we find a recursive formula:

$$p_i(t) = \sum_j \frac{A_{ij}}{k_j} p_j(t-1)$$

In the limit $t \to \infty$ the stationary probability is:

$$\vec{p} = \mathbf{A}\mathbf{D}^{-1}\vec{p} \to (\mathbf{I} - \mathbf{A}\mathbf{D}^{-1})\vec{p} = (\mathbf{D} - \mathbf{A})\mathbf{D}^{-1}\vec{p} = \mathbf{L}\mathbf{D}^{-1}\vec{p} = 0$$

In case network is connected $\mathbf{D}^{-1}\vec{p} = a\vec{1} \to \vec{p} = a\mathbf{D}\vec{1} \to p_i = ak_i$, then:

$$p_i = \frac{k_i}{\sum_i k_i} = \frac{k_i}{2m}$$

The probability of a random walk is proportional to the degree.

8.3.1 Mean first passage time

The first passage time for a random walk from a vertex u to a vertex v is the number of steps before a walk starting at u reaches v. First passage time is a random variable, we want to find the mean value. To evaluate it, we model an absorbing random walk: if the walk reaches v it remains there forever.
Let $p_v(t)$ be the probability that the walker is at v (absorbed) at time t.
$p_v(t)$ is also the probability that the walk has a first passage time to v that is less than or equal to t. For this, the probability that the walk has a first passage time to v exactly at time t is $p_v(t) - p_v(t-1)$. Until now, we can express the mean first passage time as:

$$\tau_v = \mathbb{E}\left[t\right] = \sum_{t=0}^{\infty} t[p_v(t) - p_v(t-1)]$$

$\forall i \neq v$:

$$p_i(t) = \sum_j \frac{A_{ji}}{k_j} p_j(t-1) = \sum_{j \neq v} \frac{A_{ij}}{k_j} p_j(t-1)$$

because $a_{iv} = 0$ (absorbing state). If $i \neq b$, there are no terms in A_{vj} in the sum either thus we can write:

$$\vec{p}'(t) = \mathbf{A}'\mathbf{D}'^{-1}\vec{p}'(t-1) = [\mathbf{A}'\mathbf{D}'^{-1}]\vec{p}'(0)$$

where we have used:

$$\sum_{t=0}^{\infty} t[\mathbf{M}^{t-1} - \mathbf{M}^t] = [\mathbf{I} - \mathbf{M}]^{-1}$$

We transformed a stochastic problem in a deterministic one.
Since

$$[\mathbf{I} - \mathbf{A}'\mathbf{D}'^{-1}]^{-1} = \mathbf{D}'[\mathbf{D}' - \mathbf{A}']^{-1} = \mathbf{D}'\mathbf{L}'^{-1}$$

obtaining:

$$\tau_v = \vec{1}^T\mathbf{D}'\mathbf{L}'^{-1}\vec{p}'(0)$$

$\mathbf{L}'$ is the graph Laplacian where the $v-th$ row and column are removed and is called reduced Laplacian.

8.4 Random Graph

A random graphs define a probability space or networks ensemble $(\Theta, \mathcal{A}, \mathbb{P})$, with:

- Θ the sample space for graphs: $\Theta \in \{0,1\}^{n\times n}$
- $\mathcal{A}$ the $\sigma-$algebra on Θ
- $\mathbb{P} : \mathcal{A} \to [0,1]$ some properly defined probability measure

8.4.1 Erdös-Rèny

That is a random graph defines by two variables $G(n.m)$, m is the average number of edges.
The average node degree is $\mathbb{E}\left[k\right] = 2m/n$. However it is difficult to work with $G(n,m)$ becaus it is difficult, first of all counting the Γ graphs with ecactly n nodes and m edges.
It is better working with $G(n,p)$ where each edge is Bernoulli random variable with probability p.
In this graph:

$$\mathbb{P}(G) = p^m(1-p)^{\binom{n}{2}-m}$$

The mean value of m is:

$$\mathbb{E}\left[m\right] = \sum_{m=0}^{\infty} m\mathbb{P}(m) = \binom{n}{2}p$$

The mean degree is:

$$\mathbb{E}\left[k\right] = \sum_{m=0}^{\infty} \frac{2m}{n}\mathbb{P}(m) = \ldots = (n-1)p$$

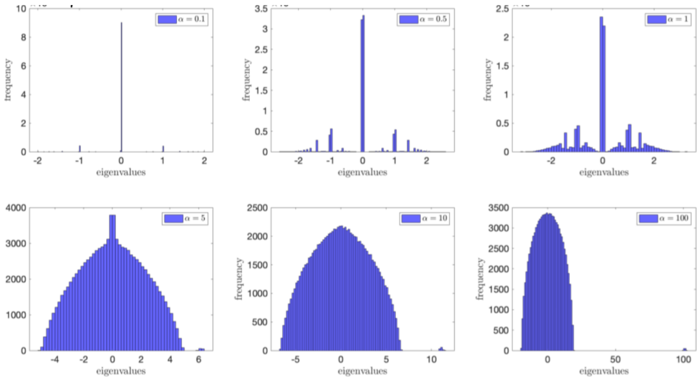
we define the mean degree with c.
It is possible show that degree distribution is a Poisson distribution:

$$p_k \sim e^{-c}\frac{c^k}{k!}$$

In this graph, the expected clustering coefficient is given by:

$$\mathbb{E}\left[C_i\right] = p = \frac{c}{n-1}$$

Let us analyse the eigenvalues: plotting the eigenvalues histogram for a random graph at different p:



8.4.2 Wigner matrices

Wigner matrices

Let Y_n a $n \times n$ symmetric (or Hermitian) matrix with $\{Y_{ij}\}_{j>i,i=1,\dots,n}$ and $\{Y_{ii}\}_{i=1,\dots,n}$ two sets of i.i.d random variables with $\mathbb{E}\left[Y_{ij}\right] = 0$ for any i, j. Then the matrix:

$$X_n = n^{-1/2} Y_n$$

with $\mathbb{E}\left[X_{ij}\right] = \sigma^2 < \infty$ is a Wigner matrix

Eigenvalues distribution Wigner matrix

For a Wigner matrix satisfying the Lindeberf's condition (th central limit → converge to a distribution) the sequence of eigenvalue Empricial Density Function (EDF) F^W converges weakly to the Wigner semicircle law in almost sure sense, i.e. $\forall f$ bounded and continuous:

$$\int f(x) F^W(x) dx \xrightarrow{\text{a.s.}} \int f(x) \mu_{sc}(x, \sigma^2) dx$$

where:

$$\mu_{sc}(x, \sigma^2) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}$$

The semicircle law for $X_n = n^{-1/2} Y_n$ implies that the distribution of eigenvalues of Y_n is scaling as:

$$\rho(\lambda) = \frac{1}{2\pi\sigma^2 n}\sqrt{4n\sigma^2 - \lambda^2}$$

8.5 Random Graph eigenvalues

In order to study the distribution of the eigenvalues, we need to normalize the adjacency matrix $\mathbf{A}$:

$$\hat{\mathbf{A}} = (np(1-p))^{-1/2}\mathbf{A} \rightarrow \mathbb{E}\left[\hat{A}_{ij}\right] = (np(1-p))^{-1/2} p, \qquad Var[\hat{A}_{ij}] = n^{-1}$$

Than $\hat{\mathbf{A}} = \bar{\mathbf{A}} + \tilde{\mathbf{A}}$ with:

- $\hat{A}_{ij} = (np(1-p))^{-1/2} p$ mean of A_{ij}
- $\tilde{A}_{ij}$ are i.i.d random variables with zero mean and variance $1/n$.
- The Wigner's semicircle law hold for $\tilde{\mathbf{A}}$ and the Wigner theorem ensures that $||\tilde{\mathbf{A}}||_2 = 2$

To determine the eigenvcalues of $\mathbf{A}$ we use two lemmas:

Lemma 1: Inequality Distribution

If $F^{X_1}(x)$ and $F^{X_2}(x)$ are eigenvalue EDF of X_1 and X_2 symmetric matrices of size n, then:

$$|F^{X_1}(x) - F^{X_2}(x)| \leq \frac{\text{rank}(X_1 - X_2)}{n}$$

So since $\hat{\mathbf{A}}$ has tank 1 for any n, when $n \to \infty$ the two eigenvalues EDFs vecome equal, thus the EDF of $\hat{\mathbf{A}}$ is the semicircle law.

Bauer-Fike theorem

For symmetric matrices with $\hat{\mathbf{A}} - \hat{\mathbf{A}} = \tilde{\mathbf{A}}$, it is:

$$|\lambda_i(\hat{\mathbf{A}}) - \lambda(\bar{\mathbf{A}}) \leq ||\tilde{\mathbf{A}}||_2 = 2$$

This theorem is usegul for the largest eigenvalue. Since $\lambda(\hat{A}) = np/\sqrt{np(1-p)} = \sqrt{np/(1-p)}$, it is:

$$\left|\lambda_1(\hat{\mathbf{A}}) - \sqrt{\frac{np}{1-p}}\right| \leq 2$$

and:

$$\lambda_1(\hat{\mathbf{A}}) \xrightarrow[n\to\infty]{\text{a.s.}} \sqrt{\frac{np}{1-p}}$$

Finally, remembering that $\mathbf{A} = \sqrt{n(1-p)}\hat{\mathbf{A}}$, the spectrum of the adjacency matrix $G(n,p)$ is composed by large eigenvalue:

$$\lambda_1(A) = np$$

and a semicircle part with support $[-2\sqrt{np(1-p)}, 2\sqrt{np(1-p)}]$

8.6 Clustering and partition in networks

In a random graphs we have seen $G(n,p)$ with a mean clustering coefficient $C = \frac{c}{n-1} \to 0$ for large networks ($n \to \infty$), however real-world networks display often some modularity structure (communities). In this case adjacency matrix show a block structure.
Let be n the number of nodes, q the number of groups and let assume that each node belongs to one of the q groups.

8.6.1 Stochastic Block Models (SBMs)

We consider labels generation: $i \to g_i \in \{1, \dots, q\}$ with some probability $\mathbb{P}(g_i = a) = n_a$(for example $n_a = 1/q, \forall a = 1, \dots, q)$.

Define an affinity matrix (link probability within and between groups):

$$p = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

Links generation $\mathbb{P}(A_{ij} = 1) = p_{g_i g_j}$.
Let consider 2×2 case:

$$p = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

- if $p_{11}, p22 > p_{12}$ the network has a modular structural
- if $p_{12} > p_{11}, p_{22}$ the network has a bipartite structural
- if $p_{11} > p_{12} > p_{22}$ the network has a core-periphery

Case direct network: if $\mathbb{E}\left[\# \text{ forward link}\right] \geq \mathbb{E}\left[\# \text{ backward link}\right]$ the network has a hierarchical structure .

The Bauer-Fike theorem is a useful tool in the context of a symmetric Stochastic Block Model (SBM) with M equal-sized blocks. It allows us to establish that the largest M eigenvalues of the SBM are determined by the affinity matrix p. Here's how it works: consider a symmetric SBM with M equal-sized blocks.

The mean adjacency matrix of this SBM can be expressed as $\mathbb{E}\left[A\right] = p \otimes J_k$, where: $k = n/M$ represents the number of elements per block and J_k is a $k \times k$ matrix consisting of all ones. Now, let's consider the eigenvalues of a Kronecker product ($\otimes$). The eigenvalues of the Kronecker product of two matrices are the product of the eigenvalues of the individual matrices..

In this case, the Kronecker product involves $p \otimes J_k$. The largest eigenvalue of this product is $z_1 n/M$, where:z_1 is the largest eigenvalue of the affinity matrix p.

Also in this case, let consider the case with two blocks:

$$p = \begin{pmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{pmatrix}$$

and the two eigenvalues are:

$$z_{1,2} = \frac{p_{11} + p_{22} \pm \sqrt{(p_{11} - p_{22})^2 + 4p_{12}^2}}{2}$$

The first eigenvector has two components of the same sign and using Bauer-FIke we can conclude that the largest eigenvalue of the adjacency matrix of a SBM with two equal sized blocks is in the limit $n \to \infty$:

$$\lambda_1 = \frac{p_{11} + p_{22} + \sqrt{(p_{11} - p_{22})^2 + 4p_{12}^2}}{4} n$$

IX

PCA metric and Granger Causality

9.1 Introducion

In previous chapters we introduced the concept of system risk, we characterized it as:

- Involves the whole financial system;
- Threatens the stability and the well-functioning;
- Undermine public confidence;
- It is related to contagion

we understood that linkages play an important role.
Some studies tried:

- CoVaR: measures Value-at-Risk of financial institutions conditional on the entire set of institutions' poor performance;
- Systemic Expected Shortfall: measures expected loss to each institution conditional on the entire set of institutions' poor performance;
- Distressed Insurance Premium (DIP): measures the insurance premium required to cover distressed losses in the banking system

All these studies had some criticism, in fact: they are based on the magnitude of losses during periods when many institutions are simultaneously distressed. In particular they are related to market volatility: possible underestimation of systemic risk during period of prosperity and growth. They are not good as early warning indicators.
An idea to overcome this problem is trying to explain the variance of the returns in terms of common uncorrelated factor. When returns are driven by few common factors they will move more closely together.
To do this, we define:

- R_i: stock return of institution i, $i = 1, \dots N$;
- $R_S = \sum_i R_i$: aggregate return of the system;
- $\mu_i = \mathbb{E}\left[R_i\right]$ and $\sigma_i^2 = \text{Var}[R_i]$;

We can define the total risk of the system as:

Total risk of the system

$$\sigma_s^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_i \sigma_j \mathbb{E}\left[z_i z_j\right], \qquad z_k \equiv \frac{(R_k - \mu_k)}{\sigma_k}$$

We find in the uncorrelated factors as the orthogonal eigenvectors of the corelation matrix the uncorrelated factors that we are looking for. Each eigenvector identifies a portfolio and the corresponding eigenvalue is related to the fraction of total variance explained.
Larger eigenvalues explain more the variation of the system. We can reduce the dimensionality of the problem by focus on the n largest eigenvalues.
We noticed that more returns are correlated and larger the portion of the total volatility captured by the first n components. We define:

$$\Omega = \sum_{k=1}^{N} \lambda_k, \quad \omega_n = \sum_{k=1}^{n} \lambda_k, \quad h_n = \frac{\omega_n}{\Omega}$$

and the periods of increased interconnectedness can be characterized by:

$$h_n > \text{ some treshold } H$$

Now we can introduce N variables η_k s.t.:

$$\mathbb{E}\left[\eta_k \eta_l\right] = \lambda_l \delta_{kl}$$

we can express the standardized returns as:

$$z_i = \sum_{i=1}^{N} L_{ik} \eta_k$$

so that:

$$\mathbb{E}\left[z_i z_j\right] = (L \Lambda L^T)_{ij} = \sum_{k=1}^{N} L_{ik} L_{jk} \lambda_k$$

and:

$$\sigma_S^2 = \sigma^T L \Lambda L^T \sigma = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sigma_i \sigma_j L_{ik} L_{jk} \lambda_k$$

We can introduce als a measure of exposure of the single institution:

$$PCAS_{i,n} = \frac{1}{2}\frac{\sigma_i^2}{\sigma_S^2}\frac{\partial^2\sigma_S^2}{\partial\sigma_i^2}\Big|_{h_n \geq H} = \sum_{k=1}^{n}\frac{\sigma_i^2}{\sigma_S^2}L_{ik}^2\lambda_k\Big|_{h_n \geq H}$$

This gives both he contribution and the exposure of the i-th institution to the overall risk of the system given a common component across the returns of all institutions.

9.2 Granger Causality

Granger Causality (Granger,1969)

Test whether the past information on a time series y is statistically useful in predicting the future of another time series x,better than using only the past information on x.

In the original paper, the information on past realization of the two time series defines the information set, which is also called the Universe. The information set may include also the information on other variables. Granger causality include different forms:

- Granger causality in mean (or linear) that is the most widely used.
- Granger causality in variance, useful for systemic risk.
- Granger in causality in tail: how an extreme event in stock A imply an extreme event in stock B

9.2.1 Linear Granger causality

Linear Granger causality

Given two zero-mean, stationary, time series R_i^t and R_t^j, let's assume:

$$R_{t+1}^i = \sum_{k=1}^{L} a_k^i R_{t-k}^i + \sum_{k=1}^{L} b_k^{ij} R_{t-k}^j + e_{t+1}^i$$
$$R_{t+1}^j = \sum_{k=1}^{L} a_k^j R_{t-k}^j + \sum_{k=1}^{L} b_k^{ji} R_{t-k}^j + e_{t+1}^j$$

where e_{t+1}^i, e_{t+1}^j are two uncorrelated white noise process.

We say that j Granger-cause i if $b^{ij} \neq 0$ and vice-versa. If both $b^{ij}, b^{ji} \neq 0$, there is a feedback relationship.
In notation:

$$(j \to i) = \begin{cases} 1 & \text{if } j \text{Granger causes } i \\ 0 & \text{otherwise} \end{cases}$$

Test for Linear Granger causality

The test is based on the Granger-Sargent statics:

$$GS = \frac{(R_2 - R_1)/L}{R_1/(N-2L)}$$

where R_1 is the residual sum of squares from:

$$R^i_{t+1} = \sum_{k=1}^{L} a^i_k R^i_{t-k} + \sum_{k=1}^{L} b^{ij}_k R^j_{t-k} + e^i_{t+1}$$

and R_2 are the residual sum of squares from:

$$R^i_{t+1} = \sum_{k=1}^{L} a^i_k R^i_{t-k}$$

The statistics GS follows an F-distribution with L and $N - 2L$ degrees of freedom

So x Granger causes y is past values of x helps to forecast y, given past y. This idea can be easily applied in VAR framework;

$$\begin{bmatrix} y_{1,t} \\ y_{2,y} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \phi_{11,i} & \phi_{12,i} \\ \phi_{21,i} & \phi_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \varepsilon_t$$

Then $y_{2,t}$ not Granger causes $y_{1,t}$ if:

$$\phi_{12,i} = 0 \quad i = 1, 2, \ldots, p$$

and that happens if:

$$\begin{bmatrix} y_{1,t} \\ y_{2,y} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \phi_{11,i} & 0 \\ \phi_{21,i} & \phi_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \varepsilon_t$$

These are just restrictions on VAR parameters that can be easily tested with standard join F-test. Denoting $RSS \equiv SS$ of the Restricted model and $USS \equiv$ the SS of the Unrestricted one:

$$F = \frac{(RSS - USS)/p}{USS/(T - 2p - 1)}$$

Under general condition $pF \to \chi^2_p$.
We can define networks based on Granger Causality:

- **Degree of Granger Causality**

  $$DGC = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{i \neq j} (j \to i)$$

  systemic risk is high when $DGC \geq K$ for a certain treshold K

- **Number of connections**:

  $$\# \text{ Out: } (j \to S)|_{DGC \geq K} = \frac{1}{N-1} \sum_{i \neq j} (j \to i)|_{DGC \geq K}$$
  $$\# \text{ In: } (S \to j)|_{DGC \geq K} = \frac{1}{N-1} \sum_{i \neq j} (i \to j)|_{DGC \geq K}$$
  $$\# \text{ In + Out: } (j \leftrightarrow S)|_{DGC \geq K} = \frac{1}{2(N-1)} \sum_{i \neq j} (i \to j) + (j \to i)|_{DFC \geq K}$$

- **Sector-conditional connection**:

  $$\# \text{ Out-to-Other}{=}\frac{1}{(M-1)N/M} \sum_{i \neq j} \sum_{\beta \neq \alpha} (j|\alpha) \to (i|\beta)|_{DGC \geq K}$$
  $$\# \text{ In-from-Other}{=}\frac{1}{(M-1)N/M} \sum_{i \neq j} \sum_{\beta \neq \alpha} (i|\beta) \to (j|\alpha)|_{DGC \geq K}$$
  $$\# \text{ In+Out-Other}{=}\frac{1}{2(M-1)N/M} \sum_{i \neq j} \sum_{\beta \neq \alpha} ((i|\beta) \to (j|\alpha)) + ((j|\alpha) \to (i|\beta))$$

- **Eigenvatore centrality**: It includes the idea: " THe more my conncetions are importnat, the more I am". So the centrality of node i, v_i, is defined by:

  $$v_i = \sum_{t \in \text{connections}_i} v_t = \sum_{j=1}^{N} A_{i,j} v_j$$

  where A is the adjacency matrix:

  $$[A]_{ij} = (j \to i)$$

  the vector of the centralities is the eigenvector of A that satisfies:

  $$Av = v$$

  (the one corresponding to eigenvalue 1)

- **Closeness** j is causally $C-$connected to i if there exists a sequence of nodes $k_1, \ldots, k_{C-1}$. C_{ij} =length of the shortest $C-$connection:

  $$C_{ij} = \min_{C} \{C \in [1, N-1] : (j \xrightarrow{C} i)\}$$

Closeness for institution j:

$$C_{jS}|_{DGC\geq K} = \frac{1}{N-1} \sum_{i \neq j} C_{ji}(j \xrightarrow{C} i)|_{DGC \geq K}$$

Granger-causality-network can be a toll for identifying dynamic linkages between different parts of the financial system.
Applied it to empirical data, we conclude that:

- Metrics for connectedness both ”instantaneous” (PCA) and ”lagged” (Granger-causality) are robust and have forecasting power
- It is possible to extract valuable information from ”simple” objects like returns and volatilities. Easier to obtain than e.g. credit relationships
- They are interesting tools to explore the dynamic relationships between different sectors.

9.3 Volatility estimation and Granger-causality in tails

Volatilities σ_{it} are estimated through a GARCH(1,1) model:

$$\begin{aligned} R_{i,t} &= \mu_i + \sigma_{i,t}\epsilon_{i,t}, \qquad \epsilon_{i,t} \sim WN(0,1) \\ \sigma^2_{i,t} &= \omega_i + \alpha_i(R_{i,t} - \mu_i)^2 + \beta_u \sigma^2_{i,t-1} \end{aligned}$$

where $\mu_i, \alpha_i, \omega_i, \beta_i$ are the parameters of the model.
Due to a market distress, most investors rebalancing of portfolios toward safer type of assets. At the financial turmoil, investors liquidate risky assets + purchase safe ones: flight-to-quality. So identifying and anticipating flight-to-quality is of great importance in the context of early-warning and monitoring of systemic risk.
Fire sales spillovers due to assets' illiquidity and common portfolio holdings are definitely one of the main drivers of systemic risk. Shared investments create a significant overlap of portfolios between couples of financial institutions. Fire sales move prices due to the finite liquidity of assets and to market impact. Finally, leverage management amplifies such feedbacks.

Simplified model flight-to-quality

two assets a and b with $\sigma_a < \sigma_b$ and $\mathbb{E}\left[r_a\right] < \mathbb{E}\left[r_b\right]$

$\omega =$ percentage of total assets A invested in the safe asset a. $1 - \omega$ is the percentage invested in the risky asset b.

Bank is VaR-constrained:

$$\begin{aligned} \max_{A,\omega} \quad & A\boldsymbol{\mu}'\boldsymbol{\omega} \\ \text{s.t.} \quad & \alpha A\sqrt{\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}} \leq E \end{aligned}$$

Profit-maximizer

A profit-maximizer that allocates the available resources accorging to the early setting, always reacts to quity drops with a flight-to-quality, in formula:

$$\frac{d\omega}{dE} < 0$$

In literature review we have:

- Unusual capital flows to proxy flights, both to liquidity and to quality: data on order flow are needed.
- In periods of market distress characterized by a high level of uncertainty, investors require liquidity rather than quality.
- We define an econometric measure of flight-to-quality based on easily available daily market prices: information on the order flow is now inferred from data.
- This allows us to considerably extend the time span of our analysis
- Identify periods of financial turbulence by abnormal levels of Granger inter-connectedness among equities of hedge funds, banks, broker/dealers, and insurance companies.
- We adopt a bipartite network of equities and bonds: investigate the effect of crises on sovereign debt
- We adopt the Granger-causality test in the tails by Hong et al., 2009: suited to describe events pertaining to a crisis, hence of extraordinary nature.

Let consider chains of events in which a large negative equity drop of a bank causes a significant variation (positive or negative) of a given sovereign bond yield. For this we use Granger-causality test for tail dependence:

- $\{Y_{1,t}\}_{t=1}^{T}$ and $\{Y_{2,t}\}_{t=1}^{T}$ be two stationary (not integrated) time series.
- Consider the $\alpha-$Value-at-Risk $\{V_{i,t}^{(\alpha)}\}_{t=1}^{T}$ for each series:

$$\text{Prob}[Y_{i,t} \leq -V_{i,t}^{(\alpha)} | \Omega_t] = \alpha, \quad i = 1, 2$$

- We introduce Cavial Model (Engle and Manganelli, 2004):

$$V_{i,t}^{(\alpha)} = \beta_1 + \beta_2 V_{i,t-1}^{(\alpha)} + \beta_3 Y_{i,t-1}^{+} + \beta_4 Y_{i,t-1}^{-}$$

  This model is a GARCH one
- Tail events are identified by the sequence of backtesting expections:

$$Z_{i,t}^{(-)} = \mathbb{1}_{\{Y_{1,t} < -v_{1,t}^{(\alpha)}\}}$$

From empirical data we obtain:

- $2007 - 2008$ financial crisis: strong increase in sovereign bond purchases which is not accompanied by a substantial rise in sovereign bond selling.
- Eurozone crisis: contemporaneous presence of both generalized distress buying and distress selling of sovereign bonds.

9.3.1 Econometric Measure of Flight-to-Quality

Each bond for each time-window is associated with a S&P ratings from AAA to SD. We have different definition of goodness for bond:

Weak, Strong definition Bond

- Weak definition:

$$\begin{aligned} \text{Good} &= \{\text{AAA,AA,A}\} \\ \text{Bad} &= \{\text{BBB, BB, B,...,SD}\} \end{aligned}$$

- Strong definition:

$$\begin{aligned} \text{Good} &= \{\text{AAA}\} \\ \text{Bad} &= \{\text{AA, A,...,SD}\} \end{aligned}$$

The quality indicator function is:

$$\mathbf{1}_{i \in \text{Good}}(t) = \begin{cases} 1 & \text{Bond } \imath \text{is in the Good class in time-window } t \\ 0 & \text{otherwise} \end{cases}$$

Applying this to empirical data, we obtain the following results:

- A methodology suited for studying periods of financial distress: Granger-causality in tails of distributions imply focus on events of extraordinary nature.
- Bipartite networks of risk spillover between major banks and government bonds.
- Simple economic interpretation of centrality measures: indicators of distressed selling and distressed buying.
- Eurozone crisis: major banks across the world chased top-quality bonds. Distressed selling on non-AAA rated bonds and distressed buying on AAA.
- Out-of-sample forecast: early warning indicators of systemic risk. Turn on the red alarm when the information contained in the network of risk spillover improves the forecast of bond quality measures.

9.3.2 VDAR and Granger causality in tail

Let $\{X_t\}_{t=1,\dots,T}$ and $\{Y_t\}_{t=1,\dots,T}$ be the binary time series representing the occurrences of extreme events.

Bivariate VDAR(p)

$$\begin{cases} X_t = V_t^1((1-J_t^1)X_{t-\tau_t^{11}} + J_t^1 Y_{t-\tau_t^{12}}) + (1-V_t^1)Z_t^1 \\ Y_t = V_t^2(J_t^2 X_{t-\tau_t^{21}} + (1-J_t^2)Y_{t-\tau_t^{22}}) + (1-V_t^2)Z_t^2 \end{cases}$$

where $X_t, Y_t \in \{0,1\}\ \forall t,\ V_t^i \sim \mathcal{B}(\nu_i) \in [0,1]\ \forall i = 1,2,\ \ J_t^i \sim \mathcal{B}(\lambda_i)$ with $\lambda_1 \in [0,1]\ \forall i = 1,2$ and $\tau_t^{i,j} \sim \mathcal{M}(\gamma_{ij,1}, \dots, \gamma_{ij,p})$with $\sum_{s=1}^{p} \gamma_{ij,s} = 1$.
The marginals Z_t^1 and Z_t^2 are also Bernoulli random variables with distribution $\mathcal{B}(\chi_1)$ and $\mathcal{B}(\chi_2)$, respectively with $\chi_1, \chi_2 \in [0,1]$

The model parameters can be estimated via Maximum Likelihood. Some comments about this model:

- The proposed test is bivariate and can give rise to spurious causalities
- Moreover the Hong et al method is sensitive to autocorrelation in extreme events which may result in spurious detections (see below)
- We propose a new parametric method to cope with this two problems.
- The idea is to use linear autoregressive models for discrete variables

n Statistical Inference, there exist many regularization methods that force the estimation algorithm to infer a less complex model by putting some parameters to zero, when not statistically significant. The two most widely used types of regularization are the so-called L1 (i.e. LASSO)and L2 regularization.