Financial networks: Empirics and theory

Anindya S. Chakrabarti[1]

[1]Economics area
Indian Institute of Management, Ahmedabad
anindyac@iima.ac.in

May 12, 2019

Financial networks

In this lecture, we will discuss financial networks.

- Most prominent example of large-scale network.
- Made news after 2007-08 crisis.
- Huge literature has been developed on the analysis of such networks.

General features

Emergence of patterns:

- Many agent system.
- Tipping points:
  - Rapid changes in behavior.
  - Sometimes slow build up towards catastrophic collapse.
- Rationality:
  - Herding: rational or not?
  - Diamond-Dybvig model: power of incorrect beliefs!
- Size, connectivity and heterogeneity (May-Wigner theorem).
- here we focus solely on the statistical aspect of the data.

Some of the most dramatic events of history

Looking back:

- Goes back 400 years!
- *I Can Calculate The Motions Of Heavenly Bodies, But Not The Madness Of People*!

Some of the most dramatic events of history

Looking back:

- Goes back 400 years!
- *I Can Calculate The Motions Of Heavenly Bodies, But Not The Madness Of People*: Isaac Newton (1720)

Tulip mania, Netherlands (1637)

Figure: First recorded speculative bubble!

Reference: Dutch catalog Verzameling van een Meenigte Tulipaanen

The great depression (1929-32)

Figure: Bank of United States in New York failed in 1931!

Reference: Library of Congress. New York World-Telegram & Sun Collection. http://hdl.loc.gov/loc.pnp/cph.3c17261

Black Monday (1987)

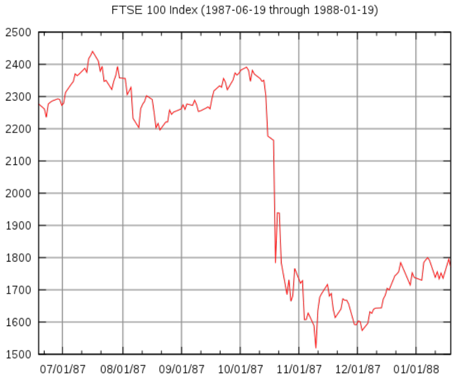


Figure: FTSE 100 index: Stock markets around the world crashed. Largest one-day percentage decline ever!

Reference: Wiki

Asian financial crisis (1997)

Figure: Fall of the ‘miracle economies’!

Reference: PatrickFlaherty (talk) Asian_Financial_Crisis.png: Bamse derivative work: Bluej100 (talk)

Global financial crisis (2007-09)

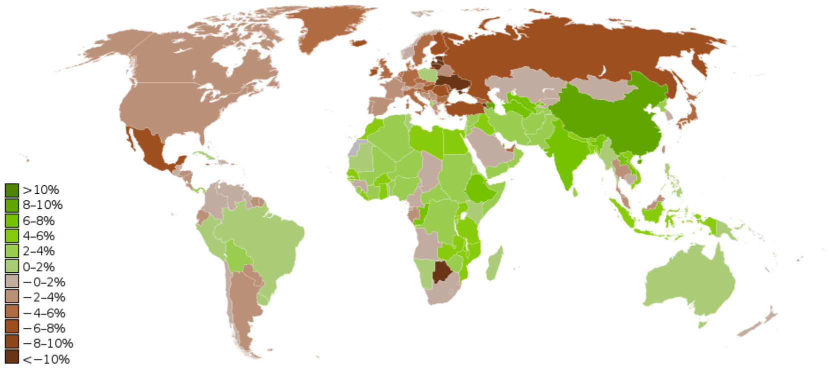


Figure: Growth rate of the countries worldwide.

Reference: Sbw01f, Kami888, Fleaman5000, Kami888derivative work: Mnmazur (talk) -

Gdp_real_growth_rate_2007_CIA_Factbook.PNG, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=10058473

Why stock markets crash?

No agreed upon answer that has any predictive value. Potential explanations:

- Informational story: Herding behavior.
- Behavioral factors: Over-optimism.
- Wrong idea about technology growth.
- Availability of easy credit.
- Miscalculation of risk: Systemic risk.
- ...

What does financial data look like?

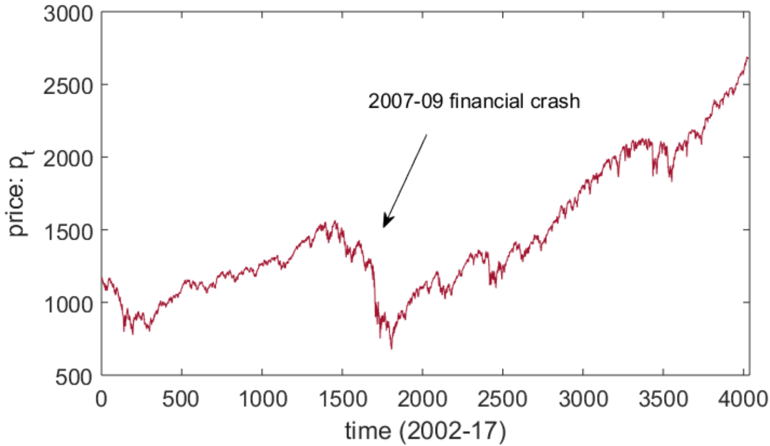


Figure: S&P 500 index. Growing trend with occasional downswings.

Do we care about price or return?

Return is the most important factor.

$$r_t = log(p_t) - log(p_{t-1}). \tag{1}$$

What does financial data look like?

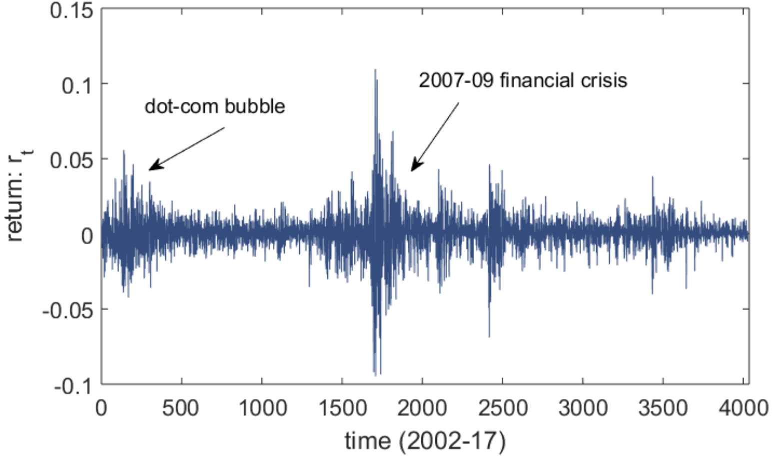


Figure: S&P 500 index fluctuations.

Some interesting properties

What can we read from the data?

- Lots of movement! Wild swings are observed.
- Average return is close to zero.
- level of return seems to have no relationship over time. A good return today does not indicate a good return tomorrow.
- Volatile periods tend to cluster.
- Market has ‘memory’ in volatility!

How to measure 'memory'?

We need some mathematical tools.

Definition (Autocorrelation)

Autocorrelation (or serial correlation) is the correlation of a time series with a delayed copy of itself as a function of delay (also called *lag*).

Formally, the expression is

$$R(\tau) = \frac{E\left((X_t - \mu)(X_{t+\tau} - \mu)\right)}{\sigma^2}. \quad (2)$$

Intuition: It is just like cross-correlation!

What does financial data look like?

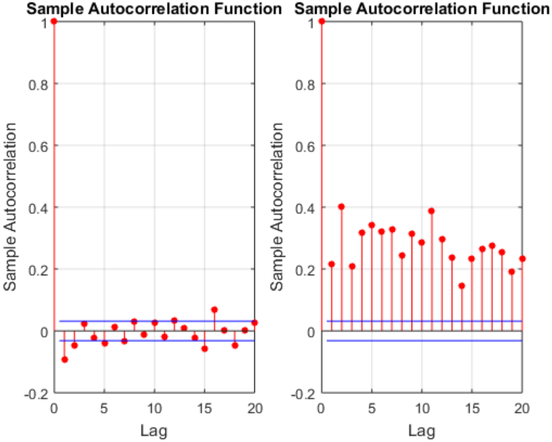


Figure: S&P 500 index: Autocorrelations (left: r_t, right: r_t^2).

What does financial data look like?

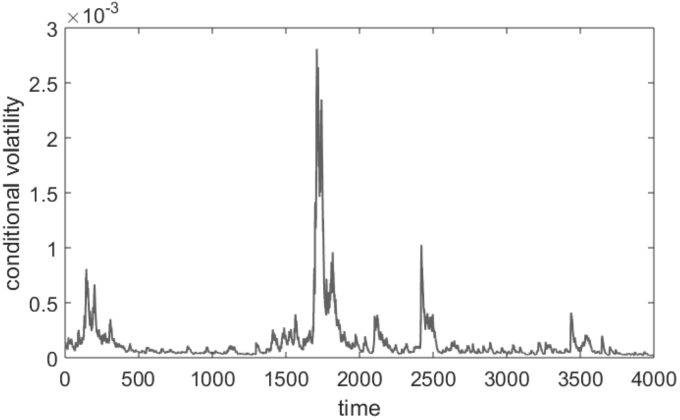


Figure: S&P 500 index: Underlying volatility.

Introduction to time series modeling

Models for stationary time series

Here we present and discuss some basic models of time series. In particular, we build a toolkit to analyze mutli-variate time series.

Definition (Weak stationarity)

$\{x_t\}$ is weakly stationary if $E(x_t)$, $E(x_t^2)$ are finite and covariance $E(x_t x_{t+k})$ is a function of k and not of t.

Definition (Strong stationarity)

$\{x_t\}$ is strongly stationary if the joint p.d.f. of $\{x_{t-k}, \ldots, x_{t+k}\}$ a function of k and not of t.

Persistence and volatility

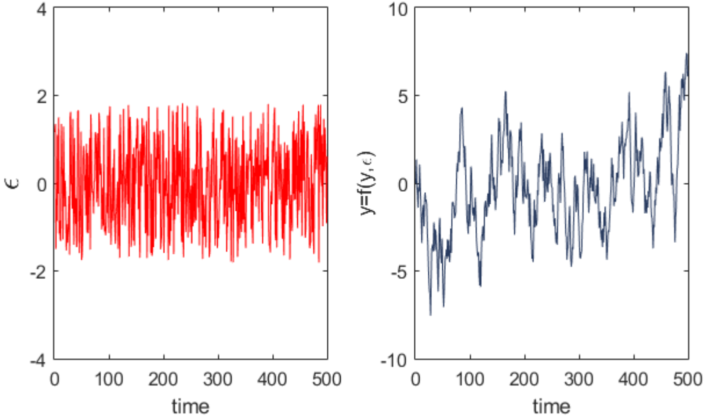


Figure: What is the difference between these two figures?

White Noise

The building block of time series models is the white noise process. Let us assume:

$$\varepsilon_t \sim \text{i.i.d } N(0, \sigma_\varepsilon^2). \tag{3}$$

Then, the implications of this assumption would be:

- $E(\varepsilon_t) = E(\varepsilon_t | \varepsilon_1, \varepsilon_2, ..., \varepsilon_{t-1}) = 0$
- $E(\varepsilon_t \varepsilon_{t-j}) = Cov(\varepsilon_t \varepsilon_{t-j}) = 0$
- $Var(\varepsilon_t) = Var(\varepsilon_t | \varepsilon_1, \varepsilon_2, ..., \varepsilon_{t-1}) = \sigma_\varepsilon^2$

White Noise

Basic ideas:

- White noise
- Lack of serial correlation
- Conditional homoskedasticity

Examples

Example of data: Quantities which non-trivially depend on their own history.

- GDP growth rates.
- firm size growth rates.
- temperature.
- ...

How to build an ARMA model?

Class of models created by taking linear combinations of white noise.

- $AR(1)$:

$$x_t = \phi x_{t-1} + \varepsilon_t \tag{4}$$

- $MA(1)$:

$$x_t = \theta \varepsilon_{t-1} + \varepsilon_t \tag{5}$$

- $AR(p)$:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \varepsilon_t \tag{6}$$

- $MA(q)$:

$$x_t = \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \tag{7}$$

How to build an ARMA model?

Most general form:

- $ARMA(p,q)$:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{8}$$

Without loss of generalization, we assume that $\bar{x} = 0$. If required, we can always introduce a constant term in the following way:

$$x_t - \bar{x} = \phi(x_{t-1} - \bar{x}) + \varepsilon_t \tag{9}$$

which again follows $AR(1)$ process with a constant. Nothing changes fundamentally.

Lag operators

Here, we introduce an operator that helps us to analyze and manipulate time series models very easily. We call it lag operator. The way it works is as follows:

$$\begin{aligned} Lx_t &= x_{t-1}, && (10) \\ L^2 x_t &= x_{t-2}, && (11) \\ &\vdots \\ L^n x_t &= x_{t-n}. && (12) \end{aligned}$$

Lag operators

We can also have a negative lag operator or a lead operator, which gives us

$$L^{-j}x_t = x_{t+j}. \tag{13}$$

Lag polynomial

Now, we can define a polynomial over the lag operators in the following way:

$$a(L) = \sum_{i=0}^{n} a_j L^j \tag{14}$$

where

$$a(l)x_t = a_0 x_t + a_1 x_{t-1} + \ldots + a_{n x_{t-n}}. \tag{15}$$

Lag polynomial

Some examples:

- $AR(1)$: $(1 - \phi_L) = \varepsilon_t$
- $MA(1)$: $x_t = (1 + \theta_L)\varepsilon_t$
- $ARMA(p, q)$: $a(L)x_t = b(L)\varepsilon_t$

Choice of representation

We have seen two representations, AR and MA. Which one is more desirable?

- Both are fine. It's more about convenience.
- For finding unconditional moments, MA process is desirable.
- If we need a representation of dependence on past values (which is more intuitive; e.g. higher gdp growth leads to higher gdp growth), then AR process is desirable.

$AR(1)$ to $MA(\infty)$ by recursive substitution

Here, we show that one can go from one representation to the other very easily. Let's say

$$x_t = \phi x_{t-1} + \varepsilon_t \quad \text{where } |\phi| < 1. \tag{16}$$

Using lag operator, we can expand on the expression:

$$\begin{aligned}
(1-\phi L)x_t &= \varepsilon_t \\
x_t &= \frac{\varepsilon_t}{(1-\phi L)} \\
&= (1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \ldots)\varepsilon_t \\
&= \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots \\
&= \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}.
\end{aligned} \tag{17}$$

$AR(1)$ to $MA(\infty)$ by recursive substitution

If the process started finite t periods back, we can expand it as:

$$\begin{aligned} x_t &= \phi x_{t-1} + \varepsilon_t \\ &= \phi^2 x_{t-2} + \phi \epsilon_{t-1} + \varepsilon_t \\ &= \phi^t x_0 + \sum_{j=0}^{t-1} \phi^j \varepsilon_{t-j}. \end{aligned} \tag{18}$$

If we assume that the process started infinite periods ago, then we have

$$x_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \tag{19}$$

ACF for ARMA process

Let us start with a simple example.

White Noise:

Consider $\varepsilon_t \sim \mathrm{iid}(0, \sigma_\epsilon^2)$.
$\gamma_0 = \sigma_\epsilon^2, \quad \rho_0 = 1$
$\gamma_j = 0, \quad \rho_0 = 0$ for $j > 0$.

ACF for MA(1) process

MA(1):

$x_t = \theta\varepsilon_{t-1} + \varepsilon_t$

$$
\begin{aligned}
\gamma_0 &= \mathrm{var}(\theta\varepsilon_{t-1} + \varepsilon_t) = (\theta^2 + 1)\sigma_\epsilon^2 && (20) \\
\gamma_1 &= E[(\theta\varepsilon_{t-1} + \varepsilon_t)(\theta\varepsilon_{t-2} + \varepsilon_{t-1})] \\
&= \theta\sigma_\epsilon^2 && (21) \\
\gamma_2 &= E[(\theta\varepsilon_{t-1} + \varepsilon_t)(\theta\varepsilon_{t-3} + \varepsilon_{t-2})] \\
&= 0. && (22)
\end{aligned}
$$

It can be easily shown that

$$\gamma_j = 0 \quad \forall\, j > 1. \tag{23}$$

ACF for MA(1) process

Clearly, we can write down the autocorrelation function now as

$$\rho_0 = 1 \tag{24}$$

$$\rho_1 = \frac{\theta}{1+\theta^2} \tag{25}$$

$$\rho_j = 0 \quad \forall\, j > 1. \tag{26}$$

ACF for AR(1) process

AR(1):

Consider a process $x_t = \phi x_{t-1} + \varepsilon_t$. Therefore,

$$\gamma_0 = \mathrm{var}(x_t) = \frac{\sigma_\epsilon^2}{1-\phi} \quad (27)$$

$$\gamma_1 = E(x_t x_{t-1}) = \frac{(\phi\sigma_\epsilon^2)}{(1-\phi)} = \phi\gamma_0 \quad (28)$$

$$\gamma_2 = E((\phi x_{t-1} + \varepsilon_t)x_{t-2}) = \phi^2 E(x_{t-2}^2) = \phi^2\gamma_0 \quad (29)$$

ACF for AR(1) process

We can easily derive the autocorrelation function as

$$\rho_1 = \phi \tag{30}$$
$$\rho_2 = \phi^2 \tag{31}$$
$$\vdots$$
$$\rho_j = \phi^j \quad \forall\, j > 0. \tag{32}$$

Multivariate ARMA Model

A two-variables example:

$$x_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}, \qquad \varepsilon_t = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{33}$$

Expectation of the error terms:

$$E(\varepsilon_t) = 0. \tag{34}$$

Variance-covariance matrix:

$$E(\varepsilon_t \varepsilon_t') = \begin{pmatrix} \sigma_{\epsilon_1}^2 & \sigma_{\epsilon_1 \epsilon_2} \\ \sigma_{\epsilon_2 \epsilon_1} & \sigma_{\epsilon_2}^2 \end{pmatrix} \tag{35}$$

Multivariate ARMA Model

If we assume that there is no correlation between error terms of different time series:

$$E(\varepsilon_t \varepsilon_t') = \begin{pmatrix} \sigma_{\epsilon_1}^2 & 0 \\ 0 & \sigma_{\epsilon_2}^2 \end{pmatrix}. \quad (36)$$

Lack of time-lagged correlation implies

$$E(\varepsilon_t \varepsilon_{t-j}') = 0, \quad (37)$$

for $j = 1, 2, \ldots$

Multivariate ARMA Model

The expression for a vector $AR(1)$ is $x_t = \phi x_{t-1} + \varepsilon_t$. In matrix form, this can be written as

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}. \tag{38}$$

This formulation is known as VAR model or *vector auto-regression model*.

Manipulating the interaction matrix

Let us consider a VAR process:

$$\Phi(L)X_t = \varepsilon_t \tag{39}$$

which can be written as

$$X_t = \Phi(L)^{-1}\varepsilon_t \tag{40}$$

Note that

$$\Phi(L) = \begin{pmatrix} \phi_{11}L & \phi_{12}L \\ \phi_{21}L & \phi_{22}L \end{pmatrix}, \tag{41}$$

implying

$$\Phi(L)^{-1} = \begin{pmatrix} \phi_{11}L^{-1} & \phi_{12}L^{-1} \\ \phi_{21}L^{-1} & \phi_{22}L^{-1} \end{pmatrix} \frac{1}{\phi_{11}\phi_{22} - \phi_{21}\phi_{12}}. \tag{42}$$

Vector autoregression model

Vector autoregression models are objects that follows the following structure:

$$
\begin{aligned}
X_{1t} &= \phi_{11}^{1} X_{1,t-1} + \phi_{12}^{1} X_{1,t-2} + ... + \phi_{21}^{1} X_{2,t-1} + ... + \epsilon_{1t} \\
X_{2t} &= \phi_{11}^{2} X_{1,t-1} + \phi_{12}^{2} X_{1,t-2} + ... + \phi_{21}^{2} X_{2,t-1} + ... + \epsilon_{2t} \\
&\vdots \\
X_{nt} &= \phi_{11}^{n} X_{1,t-1} + \phi_{12}^{n} X_{1,t-2} + ... + \phi_{21}^{n} X_{2,t-1} + ... + \epsilon_{nt}. \qquad (43)
\end{aligned}
$$

Why VAR representation is useful?

The idea is that any ARMA(p,q) process can be projected in to an AR(1) process which is essentially VAR. For an example, consider the following ARMA(2,1) process:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} \tag{44}$$

$$\begin{pmatrix} X_t \\ X_{t-1} \\ \epsilon_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_{t-1} \\ X_{t-2} \\ \epsilon_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} [\epsilon_t]. \tag{45}$$

Why VAR representation is useful?

This expression can be written as:

$$x_t = \Theta x_{t-1} + \Gamma w_t \tag{46}$$

where

$$\Gamma = \begin{pmatrix} \sigma_\epsilon \\ 0 \\ \sigma_\epsilon \end{pmatrix} \tag{47}$$

and w_t captures the normalized noise with $E(w_t w_t') = I$.

Granger Causality

Causality: If an event A takes place regularly after another event B, then the preceding event may cause the event that follows.

Definition

x_{1t} Granger causes x_{2t} if x_{1t} has a predictive component for x_{2t}, given past relaizations of x_{2t}.

Should not be confused with physical/true causality!

Granger Causality

Consider a bivariate VAR:

$$\begin{aligned} x_{1t} &= \theta_{11}(L)x_{1,t-1} + \theta_{12}(L)x_{2,t-1} + \epsilon_{1t} \\ x_{2t} &= \theta_{21}(L)x_{1,t-1} + \theta_{22}(L)x_{2,t-1} + \epsilon_{2t}. \end{aligned} \tag{48}$$

Granger Causality

Suppose x_1 Granger causes x_2, but not the other way round. Then

$$\begin{aligned} x_{1t} &= \theta_{11}(L)x_{1,t-1} + \epsilon_{1t} \\ x_{2t} &= \theta_{21}(L)x_{1,t-1} + \theta_{22}(L)x_{2,t-1} + \epsilon_{2t}. \end{aligned} \tag{49}$$

Granger causal network

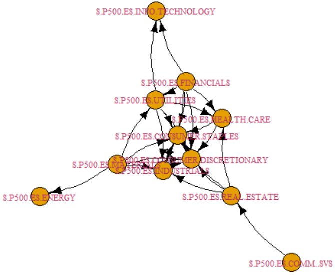


Figure: US sectoral data (2007-18)

Granger causal network

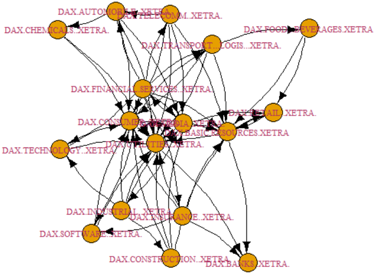


Figure: Germany sectoral data (2007-18)

Granger causal network

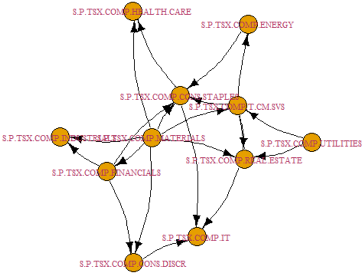


Figure: Canada sectoral data (2007-18)

Impulse response function

IRF is a very intuitive and easy way to understand the structure of a VAR model.

- It captures how shock propagates from one variable to another.
- Provides an idea about how shocks diffues over time.
- Useful for model validation.

Impulse response function: Basic idea

Consider a simple AR(1) process:

$$x_t = \phi x_{t-1} + \epsilon_t. \tag{50}$$

Imagine that $x_0 = 0$ and ϵ_t is given an unit shock. Then x responds as the following:

$$\begin{array}{rcccc} t: & 1 & 2 & 3 & 4\ldots \\ \epsilon_t: & 1 & 0 & 0 & 0\ldots \\ x_t: & 1 & \phi & \phi^2 & \phi^3\ldots \end{array} \tag{51}$$

Impulse response function: Basic idea

Note that by inversion, we get

$$x_t = (1 + \phi L + ... + \phi^n L^n + ...)\epsilon_t. \tag{52}$$

Therefore, the series of MA coefficients constitute the impulse response function.

IRF

The coefficients of MA representation of an ARMA process constitutes the corresponding IRF.

Impulse response function: Basic idea

As an example, consider an MA(2) process:

$$x_t = (1 + \gamma_1 L + \gamma_2 L^2)\epsilon_t. \tag{53}$$

The IRF is:

$$\begin{array}{lcccc} t: & 1 & 2 & 3 & 4\ldots \\ \epsilon_t: & 1 & 0 & 0 & 0\ldots \\ x_t: & 1 & \gamma_1 & \gamma_2 & 0\ldots \end{array} \tag{54}$$

Impulse response function

Generally, consider a VAR(1) process:

$$X_t = \Theta X_{t-1} + \Gamma \epsilon_t. \tag{55}$$

Then the impulse response function is given as

$$\begin{array}{rcccc} t: & 1 & 2 & 3 & 4\ldots \\ \epsilon_t: & 1 & 0 & 0 & 0\ldots \\ X_t: & \Gamma & \Theta\Gamma & \Theta^2\Gamma & \Theta^3\Gamma\ldots \end{array} \tag{56}$$

Predictions based on simple models

Consider an AR(1) process:

$$x_t = \theta x_{t-1} + \epsilon_t. \tag{57}$$

By repeated substitution, we can derive

$$E_t(x_{t+\tau}) = \theta^\tau x_t \tag{58}$$

and

$$var_t(x_{t+\tau}) = (1 + \theta^2 + \ldots + \theta^{2(\tau-1)})\sigma_\epsilon^2. \tag{59}$$

Predictions based on simple models

We note something very important here:

$$\lim_{\tau \to \infty} E_t(x_{t+\tau}) = E(x_t) \tag{60}$$

and

$$\lim_{\tau \to \infty} var_t(x_{t+\tau}) = var(x_t). \tag{61}$$

This is not an accidental outcome, we will see that this finding underlies the basic forecasting methodology.

Predictions based on more involved models

We have already seen that an ARMA model can be cast into an VAR(1) model.

1. VAR(1) models are easy to deal with.
2. All ARMA models have an VAR(1) representation.
3. Hence, the easiest way to forecast would be to convert all ARMA models into VAR(1) and develop the forecasting tools for VAR(1) process only.

Predictions based on VAR(1)

Consider a VAR(1) model:

$$x_t = \Theta x_{t-1} + \Gamma w_t. \tag{62}$$

Clearly, the conditional forecast is given by

$$E_t(x_{t+\tau}) = \Theta^{\tau} x_t. \tag{63}$$

Next, we will calculate the associated error variance.

Forecast error variance

Note that one-period ahead error is given by

$$x_{t+1} - E_t(x_{t+1}) = \Gamma w_{t+1}. \tag{64}$$

This implies

$$var_t(x_{t+1}) = \Gamma\Gamma'. \tag{65}$$

Similarly, two-periods ahead error is given by

$$\begin{aligned} x_{t+2} - E_t(x_{t+2}) &= \Theta x_{t+1} + \Gamma w_{t+1} - \Theta^2 x_t \\ &= \Gamma w_{t+1} + \Theta\Gamma w_t. \end{aligned} \tag{66}$$

This implies

$$var_t(x_{t+2}) = \Gamma\Gamma' + \Theta\Gamma\Gamma'\Theta \tag{67}$$

Forecast error variance

This way one can continue and show that

$$var_{t+\tau} = \sum_{j=0}^{\tau-1} \Theta^j \Gamma \Gamma' (\Theta^j)'. \tag{68}$$

Summary of predictive algorithm

Steps:

- Given an $ARMA(p, q)$ system, convert it into a $VAR(1)$ model:

$$x_t = \Theta x_{t-1} + \Gamma w_t. \tag{69}$$

- Use the following formulae:

$$E_t(x_{t+\tau}) = \Theta^{\tau} x_t \tag{70}$$

and

$$var_t(x_{t+\tau}) = \sum_{j=0}^{\tau-1} \Theta^j \Gamma \Gamma' (\Theta^j)'. \tag{71}$$

Exposition of a 4 variable VAR model (Forecast error contribution)

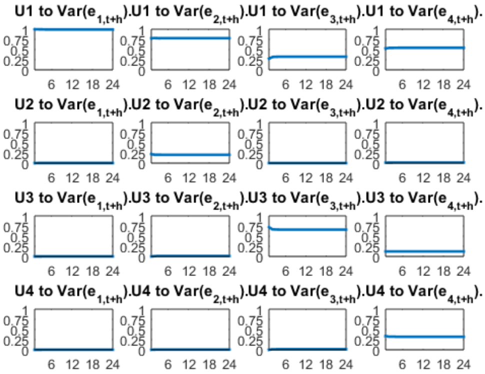


Figure: Forecast error: Contributions though spillover

Exposition of a 4 variable VAR model (FEVD)

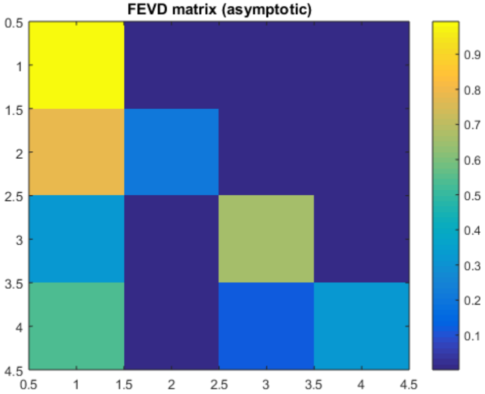


Figure: FEVD matrix of 4×4 asymptotic error variance contribution.

Exposition of a 4 variable VAR model (FEVD network)

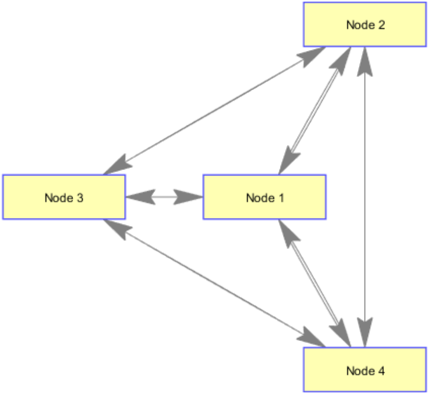


Figure: FEVD network (asymptotic)

How to model volatility clustering?

We have a model called GARCH (Generalized Autoregressive Conditional Heteroscedastic) that allows you to find out how volatile a market is.

A simple example of GARCH(1,1) is as follows:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha \sigma_{t-1}^2 + \beta r_{t-1}^2. \end{aligned} \tag{72}$$

where ϵ_t is an independent standard normal random variable (normality not necessary).

Revisit the latent volatility fit

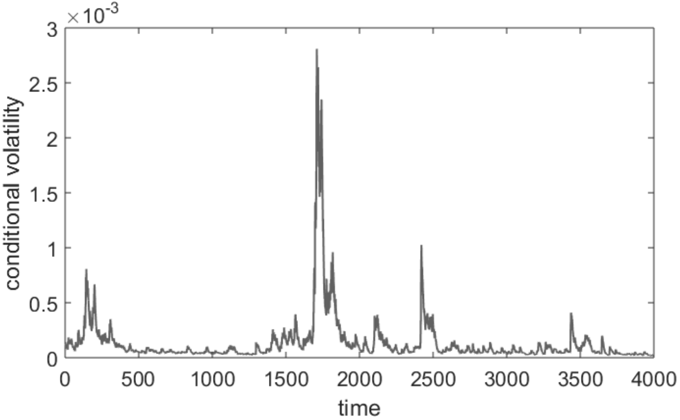


Figure: S&P 500 index: Underlying volatility.

Application of VAR model on latent volatility of individual stocks

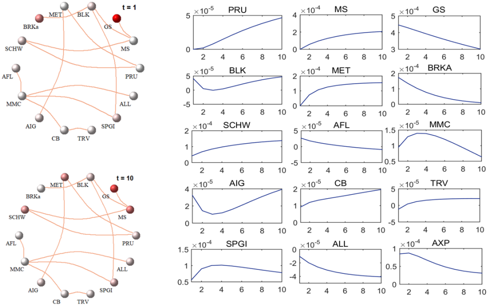
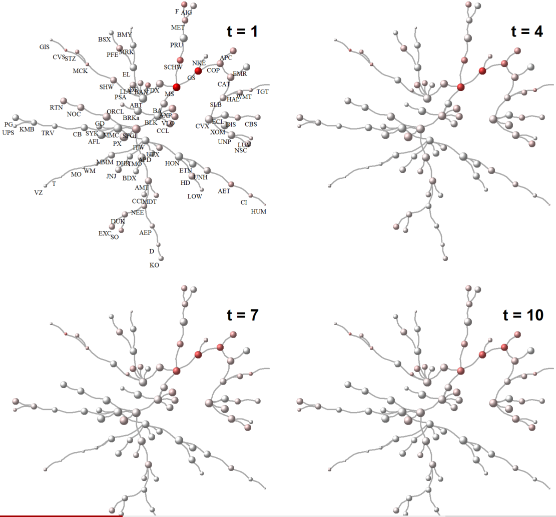


Figure: Shock spillover to financial firms from Goldman Sachs (2013-17). Source: Bansal, Kumar and Chakrabarti (2019)

Application of VAR model on latent volatility of individual stocks



GARCH(p, q) process

Consider a more generalized $GARCH(p,\ q)$ process:

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= w + \sum_{i=1}^{p} \alpha_i y_{t-i}^2 + \sum_{j=1}^{q} \beta_j y_{t-j}^2 \end{aligned} \tag{73}$$

No structural reason behind this type of models. This are only statistical in nature.

- $\{\varepsilon_i\}$ is IID, $E(\varepsilon_0)$, $E(\varepsilon_0^2) = 1$.
- $\sum_i \alpha_i + \sum_j \beta_j < 1$ for uniqueness and stationarity.

Think about the following.

Question

Why is the process stationary even when volatility is clearly time dependent?

GARCH(p, q) process

Define the information set:

$$\mathcal{F}_{t-1} = \sigma\{\varepsilon_{i\eta} : -\infty \leq i \leq t-1\}. \tag{74}$$

Then the unconditional first moment is

$$\begin{aligned} E(Y_t) &= E(\varepsilon_t \sigma_t) \\ &= E(E(\varepsilon_t \sigma_t | \mathcal{F}_{t-1})) \\ &= E(\sigma_t(E(\varepsilon_t | \mathcal{F}_{t-1}))) \\ &= 0. \end{aligned} \tag{75}$$

GARCH(p, q) process

By a similar logic, we can find the autocorrelation structure:

$$\begin{aligned} E(Y_t Y_{t+n}) &= E(Y_t \sigma_t \varepsilon_{t+n}) \\ &= E(E(Y_t \sigma_{t+n} \varepsilon_{t+n} | \mathcal{F}_{t+n-1})) \\ &= E(Y_t \sigma_{t+n}(E(\varepsilon_{t+n} | \mathcal{F}_{t+n-1}))) \\ &= 0 \end{aligned} \tag{76}$$

GARCH(p, q) process

Let us define:

$$Z_t = Y_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1). \tag{77}$$

Then we can write

$$E(Z_t) = \sigma_t^2(E(\varepsilon_t^2) - 1) = 0 \tag{78}$$

Therefore,

$$\begin{aligned} Y_t^2 &= \sigma_t^2 + Z_t && (79)\\ &= w + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j Y_{t-j}^2 + Z_t \\ &= w + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j (Y_{t-j}^2 - Z_{t-j}) + Z_t \end{aligned}$$

(80)

GARCH(p, q) process

rewrite the expression as:

$$Y_t^2 = w + \sum_{i=1}^{R} c_i Y_{t-i}^2 - \sum_{j=1}^{q} \beta_j Z_{t-j} + Z_t \tag{81}$$

where $c_i = \alpha_i + \beta_i$, $\alpha_i = 0$, $i > p$ and $\beta_i = 0$, $i > q$. By taking expectation, we get

$$\begin{aligned} E(Y_t^2) &= w + \sum_{i=1}^{R} (\alpha_i + \beta_i) E(Y_{t-i})^2 - \sum_{j=1}^{q} \beta_j E(Z_{t-j}) + E(Z_t) \\ &= w + \sum_{i=1}^{R} (\alpha_i + \beta_i) E(Y_{t-i})^2 \end{aligned} \tag{82}$$

GARCH(p, q) process

Utilizing stationarity $E(Y_t^2) = E(Y_{t+n}^2)$ (this is not the formal proof), we can derive

$$E(Y_t^2) = \frac{w}{1 - \sum_{i=1}^{R}(\alpha_i + \beta_i)}. \tag{83}$$

Thus the unconditional variance is finite.

GARCH(p, q) process

We can show that a GARCH process can accommodate a fat tail. For simplicity, consider GARCH(1, 1).

Assuming $E(\alpha_1\varepsilon_t^2 + \beta_1)^{q/2} > 1$ for some $q > 0$ (e.g. $\epsilon_k \sim N(0,1)$), for a given GARCH(1, 1)

$$\sigma_{t+1}^2 = w + \alpha_1 Y_t^2 + \beta_1\sigma_t^2, \tag{84}$$

we have

$$\begin{aligned} E(\sigma_{t+1}^q) &= E(w + (\alpha_1\varepsilon_t^2 + \beta_1)\sigma_t^2)^{q/2} \\ &\geq E((\alpha_1\varepsilon_t^2 + \beta_1)\sigma_t^2)^{q/2} \\ &= E((\alpha_1\varepsilon_t^2 + \beta_1)^{q/2})E(\sigma_t^2)^{q/2} \end{aligned} \tag{85}$$

GARCH(p, q) process

Assuming $E(\sigma_t^q)$ is finite, we can apply stationarity condition: $E(\sigma_t^q) = E(\sigma_{t+1}^q)$.

Then we get

$$E(\alpha_1 \varepsilon_t^2 + \beta_1)^{q/2} \leq 1, \tag{86}$$

which is a contradiction. This implies $E(\sigma_t^q)$ is not finite.

Some extra material

In the following, some complementary material is provided. Could be useful for some students who want to have an overview. Consult the textbooks in the references to have a complete picture.

Expanding series with lag polynomials

Let's consider an $AR(2)$ process given by

$$x_t = \phi_2 x_{t-2} + \phi_1 x_{t-1} + \varepsilon_t. \tag{87}$$

This can be rewritten as

$$(1 - \phi_1 L - \phi_2 L^2) x_t = \varepsilon_t \tag{88}$$

which in turn can be expressed as

$$x_t = \frac{\varepsilon_t}{(1 - \phi_1 L - \phi_2 L^2)}. \tag{89}$$

Lag polynomial rules

Here are some general rules about the manipulation of lag polynomials:

- $\alpha(L)\beta(L) = (\alpha_0 + \alpha_1 L + ..)(\beta_0 + \beta_1 L + ..)$
- $\alpha(L)\beta(L) = \beta(L)\alpha(L)$
- $\alpha(L)^2 = \alpha(L)\alpha(L)$
- If $\alpha(L) = (1 - \lambda_1 L)(1 - \lambda_2 L)...$, then $\alpha(L)^{-1} = (1 - \lambda_1 L)^{-1}(1 - \lambda_2 L)^{-1}...$

Autocovariance

We will denote autocovariance by

$$\gamma_j = Cov(x_t, x_{t-j}). \tag{90}$$

Note that here the time index t doesn't matter as covariance across j time points will be the same for all time points t.
The j-lag autocovariance can be written as

$$\gamma_j = E(x_t x_{t-j}). \tag{91}$$

Autocovariance

Note that 0-lag autocorrelation is just the variance:

$$\gamma_0 = Var(x_t). \tag{92}$$

Autocorrelation function

Now we can define the autocorrelation function (a.c.f.) as

$$\rho_j = \frac{\gamma_j}{\gamma_0}. \tag{93}$$

Calculating moments

Consider an $AR(1)$ process:

$$\begin{aligned} x_t &= \phi x_{t-1} + \varepsilon_t \\ &= \phi^2 x_{t-2} + \phi x_{t-1} + \varepsilon_t \\ &= \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \end{aligned}$$

We can calculate the first moment in two ways.

Method 1

We can directly take expectation: $E(x_t) = \sum_{j=0}^{\infty} \phi^j E(\varepsilon_{t-j}) = 0$.

Calculating moments

There is an indirect method which is sometimes very useful.

Method 2

For stationary process[a], the statistic is not evolving over time. Therefore,

$$E(x_t) = \phi E(x_{t-1}) + E(\varepsilon_t), \tag{94}$$

which can be written as

$$(1-\phi)E(x_t) = E(\varepsilon_t) \tag{95}$$

implying

$$E(x_t) = 0. \tag{96}$$

[a] We will define stationarity later.

Calculating moments

Homework:
Show that if

$$E(\varepsilon_t) = \mu, \tag{97}$$

then

$$E(x_t) = \frac{\mu}{1-\phi}. \tag{98}$$

Estimate a VAR model

Start from the state-space representation (companion form) of a k variable, p-lag VAR with T observations:

$$Y = X\beta + \epsilon. \tag{99}$$

Carry out a multi-variate least square estimation (same as MLE):

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{100}$$

Estimated covariance matrix (for OLS)

$$\hat{\Sigma} = \frac{1}{T - kp - 1}(Y - X\hat{\beta})(Y - X\hat{\beta})'. \tag{101}$$

One should apply GLS since the error terms here are not homoschedastic.

Relevant books and papers

References:

- J. Hamilton, Time Series Analysis, Princeton University Press (1994).
- R. Tsay, Analysis of Financial Time Series, Wiley (third edition; 2014).
- F. Dielbold and K. Yilmaz, Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring, Oxford University Press (2016).
- S. Kumar et al., Ripples on financial markets, (2019).