ECO374 Winter 2019
Forecasting and Time Series Econometrics

Tianyu Du

April 15, 2019

- GitHub: `https://github.com/TianyuDu/Spikey_UofT_Notes`
- Website: `TianyuDu.com/notes`

Contents

1 Introduction and Statistics Review

1.1 Definitions

Definition 1.1. Given random variable X, the k^{th} **non-central moment** is defined as

$$\mathbb{E}[X^k] \tag{1.1}$$

Definition 1.2. Given random variable X, the k^{th} **central moment** is defined as

$$\mathbb{E}[(X - \mathbb{E}[X])^k] \tag{1.2}$$

Remark 1.1. Moments of order higher than a certain k may not exist for certain distribution.

Definition 1.3. Given the **joint density** $f(X, Y)$ of two *continuous* random variables, the **conditional density** of random Y conditioned on X is

$$f_{Y|X}(y|x) = \frac{f_{Y,X}(y, x)}{f_X(x)} \tag{1.3}$$

Definition 1.4. Given *discrete* random variables X and Y, the **conditional density** of Y conditioned on X is defined as

$$P(Y = y|X = x) = \frac{P(Y = y \wedge X = x)}{P(X = x)} \tag{1.4}$$

1.2 Multiple Linear Regression

Assumption 1.1. Assumptions on linear regression on time series data:

(i) **Linearity**

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u \tag{1.5}$$

(ii) **Zero Conditional Mean**

$$\mathbb{E}[u|X_1, X_2, \ldots, X_k] = 0 \tag{1.6}$$

(iii) **Homoscedasitcity**

$$\mathbb{V}[u|X_1, X_2, \ldots, X_k] = \sigma_u^2 \tag{1.7}$$

(iv) **No Serial Correlation**

$$Cov(u_t, u_s) = 0 \ \forall t \neq s \in \mathbb{Z} \tag{1.8}$$

(v) **No Perfect Collinearity**

(vi) **Sample Variation in Regressors**

$$\mathbb{V}[X_j] > 0 \ \forall j \tag{1.9}$$

Theorem 1.1 (Gauss-Markov Theorem)**.** Under assumptions 1.1, the OLS estimators $\hat{\beta}_j$ are *best linear unbiased estimators* of the unknown population regression coefficients β_j.

Remark 1.2. The *no serial correlation* assumption is typically not satisfied for time series data. And the *linearity* assumption is also too restrictive for time series featuring complex dynamics. Hence, for time series data we typically use other models than linear regression with OLS.

2 Statistics and Time Series

2.1 Stochastic Processes

Definition 2.1. A **stochastic process** (or **time series process**) is a family (collection) random variables indexed by $t \in \mathcal{T}$ and defined on some given probability space $(\Omega, \mathcal{F}, P)$.

$$\{Y_t\}_{t \in \mathcal{T}} = Y_1, \dots, Y_T \tag{2.1}$$

so *each Y_t in $\{Y_t\}$ collection is associated with a density function, conditioned on time t.*

Definition 2.2. The function from $\mathcal{T}$ to $\mathbb{R}$ which assigns to each point in time $t \in \mathcal{T}$ the realization of the random variable Y_t, y_t is called a **realization**[1] of the stochastic process.

$$\{y_t\} = y_1, \dots, y_T \tag{2.2}$$

Such realization is called is a **time series**.

Definition 2.3. A **time series model** or a **model** for the observations, $\{y_t\}$, is a specification of the *joint distribution* of $\{Y_t\}$ for which $\{y_t\}$ is a realization.

Assumption 2.1. The **ergodicity** assumption requires the observations cover in principle all possible events.

Definition 2.4. A stochastic process $\{Y_t\}$ is **first order strongly stationary** if all random variables $Y_t \in \{Y_t\}$ has the *exactly same probability density function.*

Definition 2.5. A stochastic process $\{Y_t\}$ is **first order weakly stationary** if

$$\forall t \in \mathcal{T},\ \mu_{Y_t} \equiv \mathbb{E}[Y_t] = \bar{\mu} \tag{2.3}$$

Definition 2.6. A stochastic process $\{Y_t\}$ is **second order weakly stationary**, or **covariance stationary** if all random variables $\{Y_t\}$ have the same mean and variance. And the covariances do not depend on t. That's, for all $t \in \mathcal{T}$,

(i) $\mathbb{E}[Y_t] = \mu\ \forall t$ constant;

(ii) $\mathbb{V}[Y_t] = \sigma^2 < \infty\ \forall t$ constant (lag 0 covariance);

(iii) $Cov(Y_t, Y_s) = Cov(Y_{t+r}, Y_{s+r})\ \forall t, s, r \in \mathbb{Z}$.

2.2 Auto-correlations

Definition 2.7. Let $\{Y_t\}$ be a stochastic process with $\mathbb{V}[Y_t] < \infty\ \forall t \in \mathcal{T}$, the **auto-covariance function** is defined as

$$\begin{aligned}\gamma_Y(t,s) &:= Cov(Y_t, Y_s) && (2.4)\\ &= \mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_s - \mathbb{E}[Y_s])] && (2.5)\\ &= \mathbb{E}[Y_tY_s] - \mathbb{E}[Y_t]\mathbb{E}[Y_s] && (2.6)\end{aligned}$$

[1] It's also called a **trajectory** or an **outcome**

Lemma 2.1. If $\{Y_t\}$ is stationary, then the auto-covariance function only depends on *the lag between two inputs*, and does not depend on specific time point t. We can write the $h \in \mathbb{Z}$ degree auto-covariance as

$$\gamma_Y(h) := \gamma_X(t, t+h)\ \forall t \in \mathcal{T} \tag{2.7}$$

Proposition 2.1. By the symmetry of covariance,

$$\gamma_Y(h) = \gamma_Y(-h) \tag{2.8}$$

Definition 2.8. The **auto-correlation coefficient** of order k is given by

$$\rho_{Y_t, Y_{t-k}} = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{\mathbb{V}[Y_t]}\sqrt{\mathbb{V}[Y_{t-k}]}} \tag{2.9}$$

Definition 2.9. Let $\{Y_t\}$ be a covariance stationary process and the **auto-correlation function** (ACF) is a mapping from *order* of auto-correlation coefficient to the coefficient $\rho_Y : k \to \rho_{Y_t, Y_{t-k}}$, defined as

$$\rho_Y(k) := \frac{\gamma(k)}{\gamma(0)} = corr(Y_{t+k}, Y_t) \tag{2.10}$$

notice the choice of t does not matter, by definition of covariance stationary process.

Proposition 2.2. Note that

$$\rho_k = \rho_{-k} = \rho_{|k|} \tag{2.11}$$

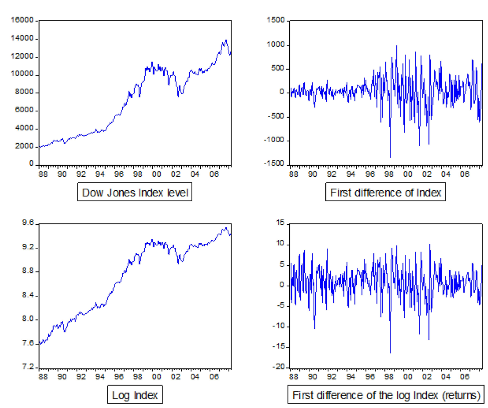
so the ACF for stationary process can be simplified to a mapping

$$\rho : k \to \rho_{|k|} \tag{2.12}$$

Remark 2.1. Strong stationarity is difficult to test so we will focus on weak(covariance) stationarity only.

Proposition 2.3 (Transformations)**.** For a non-stationary stochastic process $\{Y_t\}$,

- $\{\Delta Y_t\}$ becomes *first order weakly stationary*;
- and $\{\Delta \log(Y_t)\}$ becomes *second order weakly stationary (covariance stationary).*



Remark 2.2 (Interpretation of PACF)**.** Partial auto-correlation $r(k)$ only measures correlation between two variables Y_t and Y_{t+k} while *controlling intermediate variables* $(Y_{t+1}, \ldots, Y_{t+k-1})$.

Remark 2.3. Partial auto-correlation can be interpreted as the estimated coefficients when regressing Y_t on its lagged values.

Remark 2.4. AR and MA signatures on ACF and PACF plot.[2]

processes	ACF (ρ)	PACF (r)
AR(p)	Declines exponentially (monotonic or oscillating) to zero	$r(h) = 0 \ \forall h > p$
MA(q)	$\rho(h) = 0 \ \forall h > q$	Declines exponentially (monotonic or oscillating) to zero

2.3 Test for Auto-correlation

Multiple Auto-correlation Test To test single auto-correlation with

$$H_0 : \rho_k = 0 \tag{2.13}$$

we can use usual t-statistic.

Multiple Auto-correlation Test While testing the joint hypothesis

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0 \tag{2.14}$$

we are using the **Ljung-Box Q-statistic**:

$$Q_k = T(T+1) \sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{T-j} \sim \chi_k^2 \tag{2.15}$$

2.4 Causality and Invertibility (Optional)

2.4.1 Causality

Definition 2.10 (Causality)**.** An ARMA(p, q) process $\{Y_t\}$ with

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t \tag{2.16}$$

is called **causal with respect to** $\{\varepsilon_t\}$ if there exists a sequence $\Psi \equiv \{\psi_j\}$ with the property $\sum_{j=0}^{\infty} |\psi_j| < \infty$ such that

$$Y_t = \Psi(L)\varepsilon_t \ \text{ with } \psi_0 = 1 \tag{2.17}$$

where $\Psi(L) \equiv \sum_{j=0}^{\infty} \psi_j L^j$. The above equation is referred to as the **causal representation** of $\{Y_t\}$ with respect to $\{\varepsilon_t\}$.

Proposition 2.4. By the definition of causality, a pure MA process (which is stationary) is naturally a causal representation with respect to it's own error term $\{\varepsilon_t\}$.

[2] Zero here means statistically insignificant.

Theorem 2.1. Let $\{Y_t\}$ be an ARMA(p,q) process with

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t \tag{2.18}$$

such that polynomials $\Phi(z)$ and $\Theta(z)$ have no common roots.
Then $\{Y_t\}$ is causal with respect to $\{\varepsilon_t\}$ if and only if all roots of $\Phi(z)$ are **outside** the unit circle. The coefficients Ψ are then uniquely defined by identity

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)} \tag{2.19}$$

2.4.2 Invertibility

Definition 2.11 (Invertibility)**.** An ARMA(p,q) process for $\{Y_t\}$ satisfying

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t \tag{2.20}$$

is called **invertible** with respect to $\{\varepsilon\}$ if and only if there exists a sequence Π with the property $\sum_{j=0}^{\infty} |\pi_j| < \infty$ such that

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j L^j Y_t \tag{2.21}$$

Proposition 2.5. By the definition of invertibility, a stationary auto-regressive process is naturally invertible with respect to it's own error term $\{\varepsilon_t\}$.

Theorem 2.2. Let $\{Y_t\}$ be an ARMA(p,q) process with

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t \tag{2.22}$$

Then $\{Y_t\}$ is invertible with respect to $\{\varepsilon_t\}$ if and only if all roots of $\Theta(z)$ are **outside** the unit circle. And the coefficients of $\Pi \equiv \{\pi_j\}$ are then uniquely determined by the relation

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\Phi(z)}{\Theta(z)} \tag{2.23}$$

3 Forecasting Tools

3.1 Information Set

Definition 3.1. For stochastic process $\{Y_t\}$, the **information set** I_t is the *known time series* up to time t. It takes the form of a n-tuple $(y_{t_1}, y_{t_2}, \ldots, y_{t_n})$ such that $t_n \leq t$ so the information set is *certain* at time t.

Definition 3.2. A **forecast** $f_{t,h}$ is the image of a *time series model* g under given information set I_t. Specifically,

$$f_{t,h} = g(I_t) \in \mathbb{R} \tag{3.1}$$

3.2 Forecast Horizon

Remark 3.1. Covariance stationary processes are **short-memory processes**. More recent observation contains information far more relevant for the future than older information. Therefore, as a result, the forecast $f_{t,h}$ converges when $h \to \infty$.

Remark 3.2. Non-stationary processes are **long-memory processes** and older information is as relevant for the forecast as more recent information (e.g., including trends).

3.2.1 Forecasting Environments: *Recursive*

(i) *Re-train* and predict with *updated* information set;

(ii) Advantageous if model is *stable over time*;

(iii) Not robust to *structural break*.

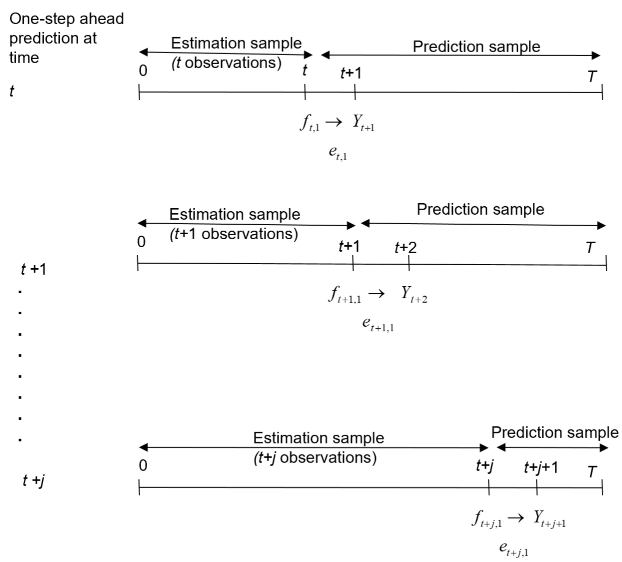


Figure 3.1: Recursive Forecasting Scheme

3.2.2 Forecasting Environments: *Rolling*

(i) *Re-train* and predict with *updated but fixed-size* information set;

(ii) Robust against *structural breaks*;

(iii) Not fully exploit information available.

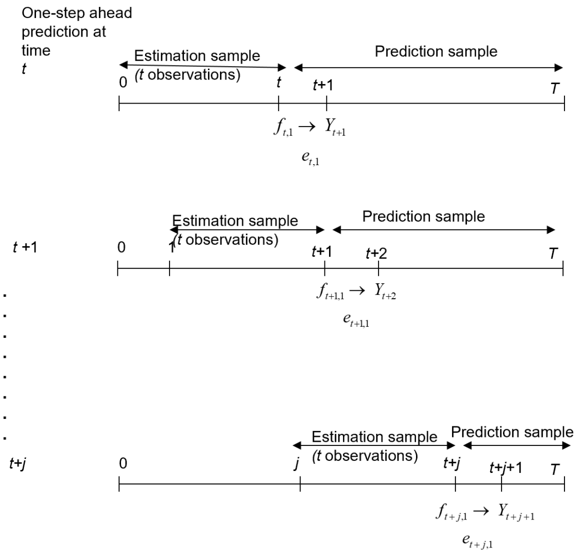


Figure 3.2: Rolling Forecasting Scheme

3.2.3 Forecasting Environments: ***Fixed***

(i) *One estimation* and forecast with *fixed-size but updated* information set.

(ii) Computationally cheap.

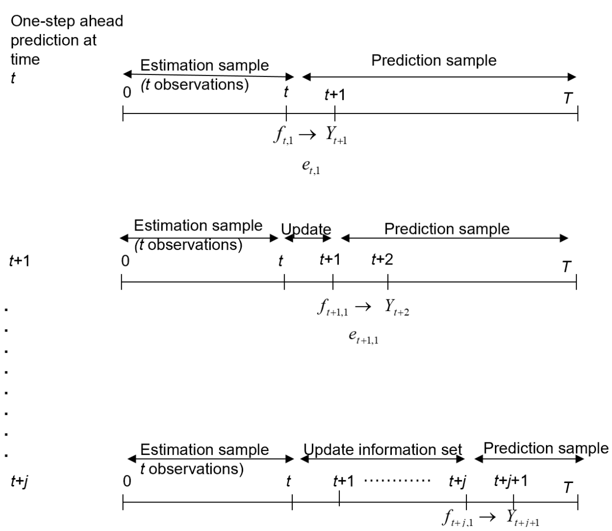


Figure 3.3: Fixed Forecasting Scheme

3.3 Loss Function

Definition 3.3. A loss function $L(e)$ is a real-valued function defined on the space of *forecast errors*, $\mathcal{E}$, and satisfies the following properties

(i) $L(e) = 0 \iff |e| = 0$;

(ii) $\forall e \in \mathbb{R}, L(e) \geq 0$[3];

(iii) L is *monotonically increasing* in the *norm* of forecast error.

Example 3.1 (Symmetric Loss Functions)**.**

$$L(e) = ae^2, \;\; a > 0 \tag{3.2}$$
$$L(e) = a \; \left|e\right|, \;\; a > 0 \tag{3.3}$$

Example 3.2 (Asymmetric Loss Functions)**.**

$$L(e) = \exp(ae) - ae - 1, \;\; a > 0 \text{ Lin(ear)-ex(ponential) Function} \tag{3.4}$$
$$L(e) = a\left|e\right|\mathbb{1}\{e \geq 0\} + b\left|e\right|\mathbb{1}\{e < 0\} \text{ Lin-lin Function} \tag{3.5}$$

3.4 Optimal Forecast

Definition 3.4. Based on information set I_t, the optimal forecast for future value y_{t+h} is the $f^*_{t,h}$ minimize the expected loss function

$$\mathbb{E}[L(y_{t+h} - f_{t,h})|I_t] = \int L(y_{t+h} - f_{t,h}) f(y_{t+h}) \; dy_{t+h} \tag{3.6}$$

Assumption 3.1. Assuming the forecast $f(y_{t+h}|I_t)$ follows

$$f(y_{t+h}|I_t) \sim \mathcal{N}(\mathbb{E}[Y_{t+h}|I_t], \mathbb{V}[Y_{t+h}|I_t]) \tag{3.7}$$

Proposition 3.1. Given symmetric quadratic L, the optimal forecast $f^*_{t,h}$ is

$$f^*_{t,h} := \mu_{t+h|t} = \mathbb{E}[Y_{t+h}|I_t] \tag{3.8}$$

Proof.

$$\min_{f_{t,h} \in \mathbb{R}} L \equiv \int (y_{t+h} - f_{t,h})^2 f(y_{t+h}|I_t) \; dy_{t+h} \tag{3.9}$$
$$\frac{\partial L}{\partial f_{t,h}} = -2 \int (y_{t+h} - f_{t,h}) f(y_{t+h}|I_t) \; dy_{t+h} = 0 \tag{3.10}$$
$$\implies \int (y_{t+h} - f_{t,h}) f(y_{t+h}|I_t) \; dy_{t+h} = 0 \tag{3.11}$$
$$\implies \int y_{t+h} f(y_{t+h}|I_t) \; dy_{t+h} = f_{t,h} \int f(y_{t+h}|I_t) \; dy_{t+h} \tag{3.12}$$
$$\implies f_{t,h} := \mu_{t+h|t} := \mathbb{E}[y_{t+h}|I_t] \tag{3.13}$$

■

Remark 3.3. Above proof for optimal forecasting does not depends on the specific distribution of Y_t.

[3]Since forecasting here can be considered as an optimization process, with L as the objective function. It's fine for L not satisfying the non-negativity condition. However, by convention, we assume L to be non-negative.

4 Moving Average Process

4.1 Wold Decomposition Theorem

Theorem 4.1 (Wold Decomposition Theorem)**.** Every covariance stationary stochastic process $\{Y_t\}$ with and finite positive variance has an unique linear representation.

$$Y_t = V_t + \underbrace{\sum_{j=0}^{\infty} \psi_j L^j \varepsilon_t}_{\text{MA}(\infty)} = V_t + \Psi(L)\varepsilon_t \tag{4.1}$$

where

(i) $\{V_t\}$ is a *deterministic component* (e.g. trend or cycle);

(ii) $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ is the *stochastic component*;

(iii) $\psi_0 = 1$[4] and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$;

(iv) $\mathbb{E}[\varepsilon_t, V_s] = 0 \ \forall t, s \in \mathcal{T}$.

Definition 4.1. The stochastic component $\{\varepsilon_t\}$ in the decomposition is called **random shocks** or **innovations**.

Lemma 4.1. Given $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, then for all $\varepsilon > 0$ there exists a natural number J such that

$$\sum_{j=J}^{\infty} \psi_j^2 < \varepsilon \tag{4.2}$$

Corollary 4.1. By above lemma, assuming $V_t = 0$, we can approximate the decomposition by a linear combination of finite innovations.

$$Y_t \approx \hat{Y}_t = \sum_{j=0}^{n} \psi_j L^j \varepsilon_t \tag{4.3}$$

and the approximation is *accurate in Euclidean norm*, that's,

$$\mathbb{E}[Y_t - \sum_{j=0}^{n} \psi_j L^j \varepsilon_t]^2 \to 0 \text{ as } n \to \infty \tag{4.4}$$

Corollary 4.2. The Wold decomposition guarantees that there always exists a finite linear model that can represent the dynamics of a covariance stationary process.

4.2 Moving Average Process

Definition 4.2. The **Moving Average** of order q with deterministic trend, MA(q), process is defined by the following stochastic difference equation

$$Y_t = \mu + \Theta(L)\varepsilon_t = \mu + \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \text{ where } \theta_0 = 1, \theta_q \neq 0 \tag{4.5}$$

where $\{\varepsilon_t\}$ is the series of innovations.

[4] We can always normalize ψ_0 to 1.

Lemma 4.2. The infinite lag polynomial in $\Psi(L)$ can be approximated by

$$\Psi(L) \approx \frac{\Theta_q(L)}{\Phi_p(L)} \tag{4.6}$$

Proof Idea. Taylor's Series. ■

Remark 4.1. MA process is always *causal* by definition.

Definition 4.3. MA(1) process takes the form of

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} \tag{4.7}$$

Unconditional Moments of MA(1) where $\mu :=$ unconditional mean.

$$\mathbb{E}[Y_t] = \mathbb{E}[\mu + \varepsilon_t + \theta\varepsilon_{t-1}] = \mu \tag{4.8}$$

$$\mathbb{V}[Y_t] = \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})^2] = \mathbb{V}[\varepsilon_t] + \theta^2\mathbb{V}[\varepsilon_{t-1}] = (1+\theta^2)\sigma_\varepsilon^2 \tag{4.9}$$

Auto-covariance

$$\gamma_0 = \mathbb{V}[Y_t] = (1+\theta^2)\sigma_\varepsilon^2 \tag{4.10}$$

$$\gamma_1 = \mathbb{E}[(Y_t - \mu)(Y_{t-1} - \mu)] = \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})] = \theta\sigma_\varepsilon^2\gamma_k = 0\ \forall k > 1 \tag{4.11}$$

Auto-correlation

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta}{1+\theta^2} \tag{4.12}$$

$$\rho_k = 0\ \forall k > 1 \tag{4.13}$$

Definition 4.4. A MA(1) process is **invertible** if $|\theta| < 1$, so that it can be written as an AR(∞) process.

Inverting. Let

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} \tag{4.14}$$

where $|\theta| < 1$. Then,

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} \implies Y_t - \mu = (1+\theta L)\varepsilon_t \tag{4.15}$$

$$\implies \frac{Y_t - \mu}{1 - (-\theta L)} = \varepsilon_t \implies \varepsilon_t = (Y_t - \mu)\sum_{j=0}^{\infty}(-\theta L)^j \tag{4.16}$$

■

Equivalence note that for MA(1) process,

$$r_1 = \rho_1 = \frac{\theta}{1+\theta^2} \tag{4.17}$$

and for any θ, $\frac{1}{\theta}$ will generate the same auto-correlation. We always choose the invertible MA representation with $|\theta| < 1$.

Proposition 4.1. For any process, $\rho_1 = r_1$.

Remark 4.2. If the MA process is invertible, we can always find an autoregressive representation in which the present is a function of the past innovations.

Remark 4.3. The value of ε_t is retrieved from dataset $\{y_t\} \subset I_t$ using the AR representation from the invertibility of MA process.

4.3 Forecasting with MA(1)

4.3.1 Forecasting with Horizon $h = 1$

Point estimate

$$f_{t,1} = \mathbb{E}[Y_{t+1}|I_t] = \mathbb{E}[\mu + \theta\varepsilon_t + \varepsilon_{t+1}|I_t] = \mu + \theta\varepsilon_t \tag{4.18}$$

Forecasting error

$$e_{t,1} = Y_{t+1} - f_{t,1} = \varepsilon_{t+1} \tag{4.19}$$

Forecasting uncertainty

$$\sigma^2_{t+1|t} = \mathbb{V}[Y_{t+1}|I_t] = \mathbb{E}[\varepsilon^2_{t+1}|I_t] = \sigma^2_\varepsilon \tag{4.20}$$

Density forecast assuming *normality* of ε. Confidence interval can be computed using the the density forecast.

$$\mathcal{F} \equiv \mathcal{N}(\mu_{t+1|t}, \sigma^2_{t+1|t})\mu_{t+1|t} = \mu + \theta\varepsilon_t\sigma^2_{t+1|t} = \sigma^2_\varepsilon \tag{4.21}$$

4.3.2 Forecasting with Horizon $h = 2$

Point estimate

$$f_{t,2} = \mathbb{E}[Y_{t+2}|I_t] = \mathbb{E}[\mu + \theta\varepsilon_{t+1} + \varepsilon_{t+2}|I_t] = \mu \tag{4.22}$$

Forecasting error

$$Y_{t+2} - f_{t,2} = \theta\varepsilon_{t+1} + \varepsilon_{t+2} \tag{4.23}$$

Forecasting Uncertainty

$$\sigma^2_{t+2|t} \equiv \mathbb{V}[Y_{t+2}|I_t] = (1+\theta^2)\sigma^2_\varepsilon \tag{4.24}$$

Density forecast

$$\mathcal{F} = \mathcal{N}(\mu, (1+\theta^2)\sigma^2_\varepsilon) \tag{4.25}$$

Remark 4.4. Since for any $h > 1$, MA(1) only generates the unconditional mean μ as the point estimate, we say MA(1) process is **short memory**.

4.4 Properties of MA(2) Process

Model

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} \tag{4.26}$$

Unconditional Moments

$$\mathbb{E}[Y_t] = \mu\mathbb{V}[Y_t] = (1+\theta_1^2+\theta_2^2)\sigma^2_\varepsilon \tag{4.27}$$

Auto-covariance

$$\gamma_0 = \mathbb{V}[Y_t] = (1+\theta_1^2+\theta_2^2)\sigma_\varepsilon^2 \tag{4.28}$$
$$\gamma_1 = (\theta_1+\theta_1\theta_2) \tag{4.29}$$
$$\sigma_\varepsilon^2\gamma_2 = \theta_2\sigma_\varepsilon^2 \tag{4.30}$$

Auto-correlation

$$\rho_1 \equiv \frac{\gamma_1}{\gamma_0} = \frac{\theta_1+\theta_1\theta_2}{1+\theta_1^2+\theta_2^2} \tag{4.31}$$
$$\rho_2 \equiv \frac{\gamma_2}{\gamma_0} = \frac{\theta_2}{1+\theta_1^2+\theta_2^2} \tag{4.32}$$

Optimal Forecasting

$$f_{t,1} = \mu + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1} \tag{4.33}$$
$$f_{t,2} = \mu + \theta_2\varepsilon_t \tag{4.34}$$
$$f_{t,h} = \mu \ \forall h > 2 \tag{4.35}$$

5 Auto-Regression Process and Seasonality

5.1 AR Process

Definition 5.1. An **auto-regressive** model of order p is taken in the form of

$$Y_t = c + \sum_{j=1}^{p} \phi_j L^j Y_t + \varepsilon_t \tag{5.1}$$

or equivalently

$$\Phi_p(L)Y_t = c + \varepsilon_t \tag{5.2}$$

Remark 5.1. In AR(1) process, ACF decays exponentially to zero, <u>faster for smaller ϕ</u>.

Definition 5.2. An **auto-regressive** process of order 1 takes the form of *stochastic difference equation*:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \tag{5.3}$$

where ϕ is called the **persistence parameter**.

Proposition 5.1. AR(1) process is stationary if and only if $|\phi| < 1$ (*all roots outside the unit circle*).

ACF and PACF

$$\rho_1 = r_1 = \phi \tag{5.4}$$
$$r_k = 0 \ \forall k > 1 \tag{5.5}$$

5.1.1 Forecasting with AR(1) and $h = 1$

Assumption 5.1. While examining the optimal forecast in this section, we are assuming the loss function is *symmetric*.

Point estimate

$$f_{t,1} = \mathbb{E}[Y_{t+1}|I_t] = c + \phi Y_t \tag{5.6}$$

Forecast variance(uncertainty)

$$\mathbb{V}[Y_{t+1}|I_t] = \mathbb{V}[c + \phi Y_t + \varepsilon_{t+1}|I_t] = \sigma_\varepsilon^2 \tag{5.7}$$

Density forecast

$$\mathcal{F} = \mathcal{N}(c + \phi Y_t, \sigma_\varepsilon^2) \tag{5.8}$$

5.1.2 Forecasting with AR(1) and $h = s > 1$

Point estimate

$$f_{t,s} = \mathbb{E}[Y_{t+s}|I_t] = c + \mathbb{E}[\phi Y_{t+s-1}|I_t] = (1 + \phi + \cdots + \phi^{s-1})c + \phi^s Y_t \tag{5.9}$$

Forecasting uncertainty

$$\mathbb{V}[Y_{t+s}|I_t] = \mathbb{V}[\varepsilon_{t+s} + \phi\varepsilon_{t+s-1} + \cdots + \phi^{s-1}\varepsilon_{t+1}|I_t] = \sum_{j=0}^{s-1} \phi^{2j}\sigma_\varepsilon^2 \tag{5.10}$$

Remark 5.2. *ACF and PACF are estimated functions subject to sampling error, so the estimated ACF and PACF from sample might be different from their theoretical values.*

5.1.3 Forecasting with AR(1) and $h \to \infty$

Assumption 5.2. For this subsection, assuming the AR(1) process is stationary, that's, $|\phi| < 1$.

Point Estimate

$$\lim_{h\to\infty} f_{t,h} = \frac{c}{1-\phi} \tag{5.11}$$

Forecasting Uncertainty

$$\lim_{h\to\infty} \mathbb{V}[Y_{t+h}|I_t] = \frac{\sigma_\varepsilon^2}{1-\phi^2} \tag{5.12}$$

Remark 5.3. The convergences demonstrated above suggest auto-regressive process is still a **short memory** process.

5.1.4 Forecasting with AR(2) process

Definition 5.3. AR(2) process

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{5.13}$$

Unconditional Moments

$$\mathbb{E}[Y_t] = c + \phi_1\mathbb{E}[Y_{t-1}] + \phi_2\mathbb{E}[Y_{t-2}] \tag{5.14}$$

$$\implies \mu_Y = \frac{c}{1-\phi_1-\phi_2} \tag{5.15}$$

Auto-covariance and Auto-correlation

$$\rho_1 = r_1 \tag{5.16}$$
$$r_2 = \phi_2 + \text{ sampling error} \tag{5.17}$$

Optimal Forecasts $h = 1$

$$f_{t,1} = \mathbb{E}[Y_{t+1}|I_t] = \mathbb{E}[c + \phi_1 Y_t + \phi_2 Y_{t-1} + \varepsilon_{t+1}|I_t] \tag{5.18}$$
$$= c + \phi_1 Y_t + \phi_2 Y_{t-1} \tag{5.19}$$

$$e_{t,1} = \varepsilon_{t+1} \tag{5.20}$$

$$\sigma^2_{t+1|t} = \mathbb{V}[Y_{t+1}|I_t] = \sigma^2_\varepsilon \tag{5.21}$$

Optimal Forecasts $h = 2$

$$f_{t,2} = \mathbb{E}[Y_{t+2}|I_t] = \mathbb{E}[c + \phi_1 Y_{t+1} + \phi_2 Y_t + \varepsilon_{t+2}|I_t] = c + \phi_1 f_{t,1} + \phi_2 Y_t \tag{5.22}$$

$$e_{t,2} = Y_{t+2} - f_{t,2} = \phi_1(Y_{t+1} - f_{t,1}) + \varepsilon_{t+2} = \phi_1 e_{t,1} + \varepsilon_{t+2} \tag{5.23}$$

$$\sigma^2_{t+2|t} = \mathbb{V}[Y_{t+2}|I + t] = \phi_1^2 \sigma^2_{t+1|t} + \sigma^2_\varepsilon = (1 + \phi_1^2)\sigma^2_\varepsilon \tag{5.24}$$

Optimal Forecasts $h = s > 2$

$$f_{t,s} = \mathbb{E}[Y_{t+s}|I_t] = c + \phi_1 f_{t,s-1} + \phi_2 f_{t,s-2} \tag{5.25}$$

$$e_{t,s} = \phi_1(Y_{t+s-1} - f_{t,s-1}) + \phi_2(Y_{t+s-2} - f_{t,s-2}) + \varepsilon_{t+s} \tag{5.26}$$
$$= \phi_1 e_{t,s-1} + \phi_2 e_{t,s-2} + \varepsilon_{t+s} \tag{5.27}$$

$$\sigma^2_{t+s|t} = \mathbb{V}[Y_{t+s}|I_t] = \mathbb{V}[e_{t,s}|I_t] = \phi_1 \sigma^2_{t+s-1|t} + \phi_2 \sigma^2_{t+s-2|t} + \sigma^2_\varepsilon \tag{5.28}$$

Remark 5.4. AR(2) is still classified as *short memory processes* as

$$\lim_{s\to\infty} f_{t,s} = \mu \tag{5.29}$$
$$\lim_{s\to\infty} \sigma^2_{t+2|t} = \sigma^2_Y \tag{5.30}$$

5.1.5 AP(p) process

Definition 5.4. Let $\Phi(L)Y_t = c + \varepsilon_t$ be an AR(p) process with trend c, then $\Phi(\cdot)$ is called the **characteristic polynomial** of this stochastic process.

Theorem 5.1. An autoregressive process is stationary <u>if and only if</u> all roots of its characteristic polynomial are outside the unit circle on $\mathbb{C}$.

Forecasting with AR(p) Process we apply a recursive scheme, or **chain rule of forecasting**, in which the we use the *forecasted* values to make prediction on even further values.

Example 5.1 (AP(p) chain rule of forecasting)**.**

$$f_{t,1} = c + \sum_{j=1}^{p} \phi_j L^j Y_{t+1} \tag{5.31}$$

$$f_{t,2} = c + \phi_1 f_{t,1} + \sum_{j=2}^{p} \phi_j L^j Y_{t+2} \tag{5.32}$$

$$f_{t,3} = c + \phi_1 f_{t,2} + \phi_2 f_{t,1} + \sum_{j=3}^{p} \phi_j L^j Y_{t+3} \tag{5.33}$$

$$f_{t,s} = c + \sum_{j=1}^{p} \phi_j f_{t,s-j} \ \forall s > p \tag{5.34}$$

5.2 Procedures of Forecasting with Autoregressive Models

(i) **Estimate** $\hat{\Phi}$ and $\hat{\sigma}_\varepsilon^2$ and $\hat{\mu}$ (unconditional mean).

(ii) Calculate $\hat{c}$(intercept) from $\hat{\Phi}$ and $\hat{\mu}$.

$$\hat{c} = \hat{\mu}(1 - \sum_{j=1}^{p} \hat{\phi}_j) \tag{5.35}$$

(iii) Construct forecast $f_{t,h}$ with **chain rule of forecasting**.

(iv) **Density forecast**, under normality assumption of ε is

$$\mathcal{N}(f_{t,h}, \hat{\sigma}^2_{t+h|t}) \tag{5.36}$$

(v) 95% confidence interval of forecasting is

$$(f_{t,h} \pm 1.96 \times \hat{\sigma}_{t+h|t}) \tag{5.37}$$

5.3 Seasonality

5.3.1 Deterministic Seasonality

Definition 5.5. The seasonality is **deterministic** if the seasonal component regressors are always exactly predictable.

Remark 5.5. To handle deterministic seasonality, just add those indicator terms into the regression.

5.3.2 Stochastic Seasonality

Definition 5.6. The seasonality is **stochastic** if the seasonal component is driven by random variables.

Definition 5.7. A seasonal AP(p) model, S-AR(p), is defined by

$$Y_t = c + \phi_s Y_{t-s} + \phi_{2s} + Y_{t-2s} + \cdots + \phi_{ps} Y_{t-ps} + \varepsilon_t \tag{5.38}$$
$$\Phi_p(L^s) Y_t = c + \varepsilon_t \tag{5.39}$$

where s refers to the **data frequency**. Such model seeks to explain the **dynamics across seasons**.

Definition 5.8. Characteristics of realizations from S-AR(p) process:

(i) *ACF decays slowly with spikes at multiples of s.*

(ii) PACF only spikes at multiples of s.

Definition 5.9. A seasonal MA(q) model, S-MA(q), is given by

$$Y_t = \mu + \Theta_q(L^s)\varepsilon_t \tag{5.40}$$

Remark 5.6. Characteristics of realizations from S-MA(q) process:

(i) ACF only spikes at multiples of s.

(ii) *PACF decays slowly with spikes at multiples of s.*

Proposition 5.2 (Combing ARMA and S-ARMA)**.** Given ARMA

$$\Phi_p(L) Y_t = c + \Theta_q(L)\varepsilon_t \tag{5.41}$$

and S-ARMA

$$\Phi'_p(L^{s_1}) Y_t = c + \Theta'_q(L^{s_2})\varepsilon_t \tag{5.42}$$

The combined model is given by multiplying the lag polynomials

$$\Phi_p(L)\Phi'_p(L^{s_1}) Y_t = c + \Theta_q(L)\Theta'_q(L^{s_2})\varepsilon_t \tag{5.43}$$

6 Model Assessment and Asymmetric Loss

6.1 Model Assessment

Definition 6.1. Akaike information criterion(AIC) of a model with k parameters is defined as

$$AIC := -2\ln(\mathcal{L}) + 2k \tag{6.1}$$

Definition 6.2. Bayes information criterion(BIC)/**Schwarz information criterion**(SIC) of a model with k parameters and fitted on the sample with size N is defined as

$$BIC := -2\ln(\mathcal{L}) + 2\ln(N)k \tag{6.2}$$

Definition 6.3. Given time series data sample with size T, and use the *recursive scheme* starting from $t < T$, with forecasting horizon h, given sequence of *ground truth*

$$\mathscr{Y} = (y_j)_{j=t+h}^{T} \tag{6.3}$$

we can construct a sequence of forecast

$$\mathscr{F} = (f_{t,h}, f_{t+1,h}, f_{t+2,h}, \ldots, f_{T-h,h}) \tag{6.4}$$

and a sequence of forecasting errors

$$\mathscr{E} = (e_{j,t})_{j=t+h}^{T} \tag{6.5}$$

then,

$$\text{MSE} \equiv \frac{1}{|\mathscr{F}|}\sum_{e\in\mathscr{E}} e^2 \tag{6.6}$$

$$\text{MAE} \equiv \frac{1}{|\mathscr{F}|}\sum_{e\in\mathscr{E}} ||e|| \tag{6.7}$$

$$\text{MAPE} \equiv \frac{1}{|\mathscr{F}|}\sum_{(y,e)\in(\mathscr{Y},\mathscr{E})} ||\frac{e}{y}|| \tag{6.8}$$

6.2 Asymmetric Loss

Definition 6.4 (Log-Normal Distribution)**.** Let X be a Gaussian random variable with mean μ and variance σ^2. Define $Y \equiv \exp(X)$, then Y follows **log-normal distribution**, with

$$\mathbb{E}[Y] = \exp(\mu + \frac{\sigma^2}{2}) \tag{6.9}$$

Example 6.1. Consider the Lin-ex loss function

$$L(e) = \exp(ae) - ae - 1 \tag{6.10}$$

then the expected loss for h step forecasting made at t is

$$\mathbb{E}[L(e_{t,h})|I_t] \tag{6.11}$$

$$= \mathbb{E}[\exp(a(y_{t+h} - f_{t,h})) - a(y_{t+h} - f_{t,h}) - 1|I_t] \tag{6.12}$$

$$= \mathbb{E}[\exp(ay_{t+h})\exp(-af_{t,h})|I_t] - \mathbb{E}[ay_{t+h}|I_t] + af_{t,h} - 1 \tag{6.13}$$

$$= \exp(-af_{t,h})\mathbb{E}[\exp(ay_{t+h})|I_t] - a\mathbb{E}[y_{t+h}|I_t] + af_{t,h} - 1 \tag{6.14}$$

$$\tag{6.15}$$

To find the optimal forecasting, take the FOC

$$\frac{\partial \mathbb{E}[L(e_{t,h})|I_t]}{\partial f_{t,h}} = 0 \tag{6.16}$$

$$\implies -a\exp(-af_{t,h})\mathbb{E}[\exp(ay_{t+h})|I_t] + a = 0 \tag{6.17}$$

$$\implies \exp(-af_{t,h})\mathbb{E}[\exp(ay_{t+h})|I_t] = 1 \tag{6.18}$$

$$\implies -af_{t,h} + \log(\mathbb{E}[\exp(ay_{t+h})|I_t]) = 0 \tag{6.19}$$

$$\implies f_{t,h} = \frac{1}{a}\log(\mathbb{E}[\exp(ay_{t+h})|I_t]) \tag{6.20}$$

Assuming $y_{t+h} \sim \mathcal{N}(\mu_{t+h|t}, \sigma^2_{t+h|t})$,

$$\implies f_{t,h} = \frac{1}{a}\log(\exp(a\mathbb{E}[y_{t+h}|I_t] + \frac{a^2\sigma^2_{t+h|t}}{2})) \tag{6.21}$$

$$\implies f_{t,h} = \mathbb{E}[y_{t+h}|I_t] + \frac{a\sigma^2_{t+h|t}}{2} \tag{6.22}$$

So if $a < 0$, the penalty on making negative error is higher than positive error, and the optimal forecast would be *pushed down* to less than the conditional mean.

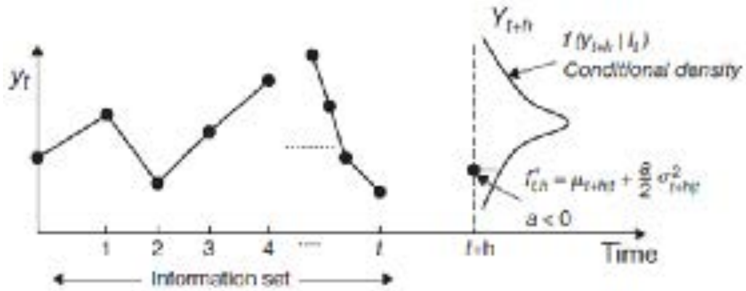


Figure 6.1: Illustration of the Optimal Forecast with Lin-Ex loss and $a < 0$

Forecasting Assuming AR(1) process.
Error with $h = 1$

$$e_{t+1,t} = c + \phi Y_t + \varepsilon_{t+1} - f_{t,1} = \varepsilon_{t+1} - \frac{a\sigma^2_{t+1|t}}{2} \tag{6.23}$$

$$\mathbb{E}[e_{t+1,t}] = -\frac{a\sigma^2_{t+1|t}}{2} \tag{6.24}$$

$$\mathbb{V}[e_{t+1,t}|I_t] = \sigma^2_\varepsilon \tag{6.25}$$

Error with $h = 2$

$$e_{t+2,t} = Y_{t+2} - f_{t,2} \tag{6.26}$$

$$= c + \phi(Y_{t+1}) + \varepsilon_{t+2} - f_{t,2} \tag{6.27}$$

$$= c + \phi c + \phi\varepsilon_{t+1} + \phi Y_t + \varepsilon_{t+2} - f_{t,2} \tag{6.28}$$

$$= c + \phi c + \phi\varepsilon_{t+1} + \phi Y_t + \varepsilon_{t+2} - \mathbb{E}[Y_{t+2}|I_t] - \frac{a\sigma^2_{t+2|t}}{2} \tag{6.29}$$

$$= \underbrace{\phi\varepsilon_{t+1} + \varepsilon_{t+2} - \frac{a\sigma^2_{t+2|t}}{2}}_{\text{forecasting error is not mean-zero}} \tag{6.30}$$

Note that for every $h > 0$, the term $\sigma_{t+h|t}$ is constant conditioned on I_t.

$$\mathbb{E}[e_{t+2,t}|I_t] = -\frac{a\sigma^2_{t+2|t}}{2} \tag{6.31}$$

$$\mathbb{V}[e_{t+2,t}|I_t] = (1 + \phi^2)\sigma^2_\varepsilon \tag{6.32}$$

7 Trends

Definition 7.1. Time series is called **non-stationary** if it contains a **trend**.

7.1 Deterministic Trend

Definition 7.2. Deterministic trends take form of

$$Y_t = g(t) + \varepsilon_t \tag{7.1}$$

where f is a function capturing the *shape of trend*.

Remark 7.1. The error term in trend modelling is always additive.

Example 7.1. Common deterministic trends

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \varepsilon_t \text{ (polynomial)} \tag{7.2}$$
$$Y_t = e^{\beta_0 + \beta_1 t + \beta_2 t^2 + \cdots} + \varepsilon_t \text{ (exponential)} \tag{7.3}$$
$$Y_t = \frac{1}{\beta_0 + \beta_1 t + \beta_2 t^2 + \cdots} + \varepsilon_t \text{ (logistic/inverse-polynomial)} \tag{7.4}$$

Remark 7.2. With trend in mean, the unconditional mean depends on time t, and thus the process is not covariance stationary, we call such stochastic process **trend-stationary**. For such process, the first order

- Unconditional mean $= \mathbb{E}[Y_t] \equiv g(t)$
- Variance $\mathbb{V}[Y_t] = \sigma_\varepsilon^2$
- Auto-covariance $\gamma_k = 0 \ \forall k \neq 0$
- Auto-correlation $\rho_k = 0 \ \forall k \neq 0$
- Optimal forecast $f_{t,h} = g(t+h) \ \forall h$
- Forecast error $e_{t,h} = \varepsilon_{t+h} \ \forall h$
- Density forecast $f_{t,h} \sim \mathcal{N}(g(t+h), \sigma_\varepsilon^2) \ \forall h$

Remark 7.3. After accounting for the non-stationary trend, the remainder of the model is stationary and we could replace ε_t with any ARMA model specification.

7.2 Model Selection

Remark 7.4. The specification for the deterministic trend is selected based on minimization of the AIC/BIC.

7.3 Stochastic Trend

Definition 7.3. A **stochastic trend** is the result of *accumulation over time of random shocks/innovations* of the form

$$Y_t = \sum_{j=0}^{t-1} \varepsilon_{t-j} \tag{7.5}$$

where $\varepsilon_{t-j} \sim \mathcal{N}(0, \sigma_\varepsilon^2) \ \forall j$. Note that such model is equivalent to an AR(1) model with $\phi_0 = 0, \phi_1 = 1$.

Example 7.2 (Random walk without a drift).

$$Y_t = Y_{t-1} + \varepsilon_t = \sum_{\tau=0}^{t} \varepsilon_\tau + Y_0 \tag{7.6}$$

$$\mu = 0 \tag{7.7}$$

$$\sigma_Y^2 = t\sigma_\varepsilon^2 \tag{7.8}$$

$$\gamma_k = (t-k)\sigma_\varepsilon^2 \tag{7.9}$$

$$\rho_k = \frac{t-k}{\sqrt{t(t-k)}} = \sqrt{\frac{t-k}{t}} \to 1 \text{ as } t \to \infty \tag{7.10}$$

Remark 7.5. AR(2) model with $\phi_1 + \phi_2 = 1$ can also generate a random walk. This can be generalized to AR(p) for any order p. Moreover, any non-stationary AR(p) process generates a random walk.

Example 7.3 (Random walk without a drift).

$$Y_t = c + Y_{t-1} + \varepsilon_t = ct + \sum_{\tau=0}^{t} \varepsilon_\tau + Y_0 \tag{7.11}$$

$$\mu = ct \tag{7.12}$$

$$\sigma_Y^2 = t\sigma_\varepsilon^2 \tag{7.13}$$

$$\gamma_k = (t-k)\sigma_\varepsilon^2 \tag{7.14}$$

$$\rho_k = \frac{t-k}{\sqrt{t(t-k)}} = \sqrt{\frac{t-k}{t}} \to 1 \text{ as } t \to \infty \tag{7.15}$$

Remark 7.6. Difference between deterministic trend model and random walk.

model	deterministic trend	random walk
mean	changing	constant
variance	constant	changing

7.4 Unit Root Test

Dickey-Fuller test for AR(1)

$$H_0 : \phi = 1 \text{ non-stationary} \tag{7.16}$$

$$H_1 : \phi < 1 \text{ stationary} \tag{7.17}$$

and the test statistic

$$z = \frac{\hat{\phi} - 1}{\sigma_{\hat{\phi}}} \tag{7.18}$$

follows **Dickey-Fuller distribution**.

Remark 7.7. Dickey-Fuller test is most useful in small samples ($T < 100$).

- Type I: if non-stationary.

$$H_0 : Y_t = Y_{t-1} + \varepsilon_t \tag{7.19}$$
$$H_1 : Y_t = \phi Y_{t-1} = \varepsilon_t \tag{7.20}$$

- Type II: if *unconditional mean* presents.

$$H_0 : Y_t = Y_{t-1} + \varepsilon_t \tag{7.21}$$
$$H_1 : Y_t = c + \phi Y_{t-1} = \varepsilon_t \tag{7.22}$$

- Type III: if *linear trend* presents.

$$H_0 : Y_t = c + Y_{t-1} + \varepsilon_t \tag{7.23}$$
$$H_1 : Y_t = c + at + \phi Y_{t-1} = \varepsilon_t \tag{7.24}$$

7.5 Optimal Forecast

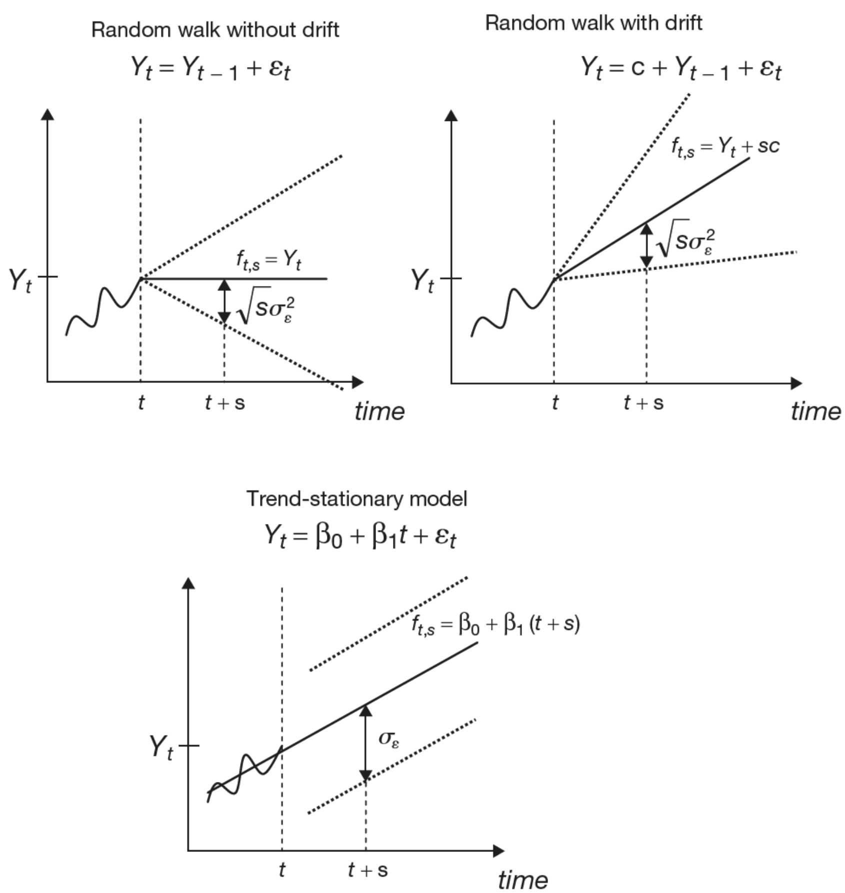


Figure 7.1: Optimal forecast for different trends

8 Vector Auto-regression

8.1 VAR Model

Definition 8.1.

$$\mathbf{V}_t := \begin{pmatrix} Y_t \\ X_t \end{pmatrix} \tag{8.1}$$

$$\mathbf{V}_t = \boldsymbol{\gamma}_0 + \Gamma_1 \mathbf{V}_{t-1} + \dots \Gamma_p \mathbf{V}_{t-p} + \boldsymbol{\varepsilon}_t \tag{8.2}$$

where Γ_j are 2×2 matrices, and $\boldsymbol{\gamma}_0$ and $\boldsymbol{\varepsilon}_t$ are vectors.

Remark 8.1. Note that the error terms (components in ε_t) can be correlated.

Remark 8.2. We assume the error terms follows multivariate Gaussian distribution

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{8.3}$$

Remark 8.3. Model selection based on ACF and PACF can become quite cumbersome in the multivariate environment. Therefore we will not consider the MA component, and use information criteria (AIC) to select p in the VAR model specification.

8.2 Granger Causality

Definition 8.2. The process $\{X_t\}$ is said to **Granger cause** the process $\{Y_t\}$ if lagged values of X_t help predicting Y_t beyond the information contained in the lagged values of Y_t.

Remark 8.4. Granger causality measures predictive power (statistical association) and has nothing to do with the concept of causality

Remark 8.5 (Granger Causality Test)**.** partial F-test for the VAR(p) model

$$Y_t = c_1 + \alpha_{11} Y_{t-1} + \cdots + \alpha_{1p} Y_{t-p} \tag{8.4}$$

$$+ \beta_{11} X_{t-1} + \cdots + \beta_{t-p} X_{t-p} + \varepsilon_{1t} \tag{8.5}$$

$$X_t = c_2 + \alpha_{21} Y_{t-1} + \cdots + \alpha_{2p} Y_{t-p} \tag{8.6}$$

$$+ \beta_{21} X_{t-1} + \cdots + \beta_{2p} X_{t-p} \tag{8.7}$$

- X_t does not Granger cause Y_t:

$$H_0 := \beta_{11} = \beta_{12} = \cdots = \beta_{1p} = 0 \tag{8.8}$$

- Y_t does not Granger cause X_t:

$$H_0 := \alpha_{21} = \alpha_{22} = \cdots = \alpha_{2p} = 0 \tag{8.9}$$

8.3 Impulse-Response Function

Definition 8.3. The **Impulse-Response Function**(IRF) quantifies the progression of the unit-sized shock through the VAR system.

	ε_{1t}	ε_{2t}
Y_{t+s}	$\frac{\partial Y_{t+s}}{\partial \varepsilon_{1t}} \approx \frac{\Delta Y_{t+s}}{\Delta \varepsilon_{1t}}$	$\frac{\partial Y_{t+s}}{\partial \varepsilon_{2t}} \approx \frac{\Delta Y_{t+s}}{\Delta \varepsilon_{2t}}$
X_{t+s}	$\frac{\partial X_{t+s}}{\partial \varepsilon_{1t}} \approx \frac{\Delta X_{t+s}}{\Delta \varepsilon_{1t}}$	$\frac{\partial X_{t+s}}{\partial \varepsilon_{2t}} \approx \frac{\Delta X_{t+s}}{\Delta \varepsilon_{2t}}$

Remark 8.6 (Compute IRF)**.** Given $\varepsilon_{11} = 1$ (initial impulse), and $\varepsilon_{21} = 0$, with

$$\varepsilon_{1t} = \varepsilon_{2t} = 0 \ \forall t > 1 \tag{8.10}$$

8.4 Forecasting

Use chain rule of forecasting.

- Optimal forecast;
- Forecast errors;
- Forecast error variance.

$$\sigma^2_{Y,t+2|t} = (1 + \alpha^2_{11})\sigma^2_{\varepsilon_1} + \beta^2_{11}\sigma^2_{\varepsilon_2} + 2\alpha_{11}\beta_{11}Cov(\varepsilon_1, \varepsilon_2) \tag{8.11}$$

9 Vector Error Correction Model

9.1 Cointegration

Definition 9.1. For any two unit-root process $\{x_t\}$ and $\{y_t\}$, we say they are **cointegrated** is there is a linear combination $z_t = y_t - \alpha_0 - \alpha x_t$ that is stationary. The process $\{z_t\}$ is called the **disequilibrium error**. *If y_t and x_t are cointegrated, then they share the same stochastic trend (common long-run information).*

Remark 9.1. The notion of cointegration only apply to unit-root processes.

9.2 Tests for Cointegrations

Remark 9.2. Augmented Dickey Fuller(ADF) Test for the cointegration.

(i) $\hat{z}_t$ is constructed using OLS of y_t on x_t;

(ii) Test the stationarity of $\{\hat{z}_t\}$ using ADF test.

$$H_0 : \text{ no cointegration } (z_t \text{ has a unit root/ non-stationary}) \tag{9.1}$$
$$H_1 : \text{ cointegration presents } (z_t \text{ is stationary}) \tag{9.2}$$

Remark 9.3. Johansen Cointegration Test

$$\mathbf{Y}_t := \begin{pmatrix} x_t \\ y_t \end{pmatrix} \tag{9.3}$$

- if $r = 0$ then there is <u>no</u> cointegrating relationship in $\mathbf{Y}_t$;
- if $r = 1$ then there is <u>one</u> cointegrating relationship in $\mathbf{Y}_t$.

Reject H_0 when test statistic exceeds the critical value given in the test output.

9.3 Vector Error Correction for Short-term Dynamics

Definition 9.2.

$$z_t = \log Y_t - \alpha_0 - \alpha \log X_t \tag{9.4}$$
$$\Delta x_t = \gamma_1 z_{t-1} + \varepsilon_{1,t} \tag{9.5}$$
$$\Delta y_t = \gamma_2 z_{t-1} + \varepsilon_{2,t} \tag{9.6}$$

where γ_1, γ_2 are called the **adjustments coefficients**, and $\varepsilon_{1,t}, \varepsilon_{2,t}$ are white noise components.

Definition 9.3. VEC model can be extended to include **temporal dependence** and **cross-correlation**

$$
\begin{aligned}
\Delta x_t &= \gamma_1 z_{t-1} + \beta_{11}\Delta x_{t-1} + \beta_{12}\Delta x_{t-2} + \dots && (9.7)\\
&\quad + \phi_{11}\Delta y_{t-1} + \phi_{12}\Delta y_{t-2} + \dots + \varepsilon_{1,t} && (9.8)\\
\Delta y_t &= {\color{red}\gamma_2 z_{t-1}} + \underbrace{\beta_{21}\Delta x_{t-1} + \beta_{22}\Delta x_{t-2} + \dots}_{\text{cross correlation}} && (9.9)\\
&\quad + \underbrace{\phi_{21}\Delta y_{t-1} + \phi_{22}\Delta y_{t-2} + \dots}_{\text{temporal dependence}} + \varepsilon_{2,t} && (9.10)
\end{aligned}
$$

Remark 9.4. Note that the VEC model has the same form as the VAR model plus the error correcction terms $\gamma_1 z_{t-1}$ and $\gamma_2 z_{t-1}$. The number of lags in the model can be selected by information criteria.

Theorem 9.1. Consider two stochastic processes, $\{x_t\}$ and $\{y_t\}$, each containing a unit root (i.e. non-stationary). If $\{x_t\}$ and $\{y_t\}$ are cointegrated, then:

(i) There exists a linear combination of y_t and x_t such as $z_t = y_t - \alpha_0 - \alpha x_t$ that is a stationary process.

(ii) There exists an error correction representation given

$$
\begin{aligned}
\Delta x_t &= \gamma_1 z_{t-1} + \beta_{11}\Delta x_{t-1} + \beta_{12}\Delta x_{t-2} + \dots && (9.11)\\
&\quad + \phi_{11}\Delta y_{t-1} + \phi_{12}\Delta y_{t-2} + \dots + \varepsilon_{1,t} && (9.12)\\
\Delta y_t &= \gamma_2 z_{t-1} + \beta_{21}\Delta x_{t-1} + \beta_{22}\Delta x_{t-2} + \dots && (9.13)\\
&\quad + \phi_{21}\Delta y_{t-1} + \phi_{22}\Delta y_{t-2} + \dots + \varepsilon_{2,t} && (9.14)
\end{aligned}
$$

9.4 Forecasting

Example 9.1.

$$
\begin{aligned}
\Delta x_t &= \gamma_1 z_{t-1} + \varepsilon_{1,t} && (9.15)\\
\Delta y_t &= \gamma_2 z_{t-1} + \varepsilon_{2,t} && (9.16)\\
z_{t-1} &= y_{t-1} - \alpha_0 - \alpha x_{t-1} && (9.17)\\
\implies x_t - x_{t-1} &= \gamma_1(y_{t-1} - \alpha_0 - \alpha x_{t-1}) + \varepsilon_{1,t} && (9.18)\\
\implies x_t &= \gamma_1 y_{t-1} + (1-\gamma_1\alpha)x_{t-1} - \gamma_1\alpha_0 + \varepsilon_{1,t} && (9.19)\\
\implies f_{t,1}^X &= \gamma_1 y_t + (1-\gamma_1\alpha)x_t - \gamma_1\alpha_0 && (9.20)\\
f_{t,2}^X &= \underbrace{\gamma_1 f_{t,1}^Y + (1-\gamma_1\alpha) f_{t,1}^X - \gamma_1\alpha_0}_{\text{Chain rule of forecasting}} && (9.21)
\end{aligned}
$$

10 Volatility I

10.1 Higher-order Moments

Definition 10.1. The third order moment, **Skewness**, of X is defined as

$$k_3 := \mathbb{E}[(X - \mathbb{E}[X])^3] \tag{10.1}$$

Definition 10.2. The **excess Kurtosis** of defined as

$$k_4 := \underbrace{\mathbb{E}[(X - \mathbb{E}[X])^4]}_{\text{Kurtosis}} - \underbrace{3\mathbb{V}[X]^2}_{\text{Kurtosis of Gaussian}} \tag{10.2}$$

Kurtosis measures the thickness of the tails of the density of X, and consequently the peakedness of the middle, relative to the Normal distribution with $k_4 = 0$.

(i) **Leptokurtic**, $k_4 > 0$, *Heavy Tails*;

(ii) **Platykurtic**, $k_4 < 0$, *Thin Tails.*

10.2 Moving Average on $\{r_t\}$

Definition 10.3. Modelling conditional variance

$$\hat{\sigma}^2_{t|t-1} = \frac{1}{n} \sum_{i=1}^{n} (r_{t-i} - \mu)^2 \tag{10.3}$$

Higher n will yield a smoother estimate.

10.3 Simple Exponential Smoothing (SES) on $\{y_t\}$

Model for the mean μ

$$\mu_T = \alpha y_T + (1-\alpha)\mu_{T-1} \text{ (smooth equation)} \tag{10.4}$$

$$\iff \mu_T = \mu_{T-1} + \alpha(\underbrace{y_T - \mu_{T-1}}_{\text{error}}) \text{ (error correction form)} \tag{10.5}$$

where $\alpha \in (0, 1)$ is a **smoothing constant**.

Choosing Parameter α using *numerical methods*(e.g. grid search) to minimize the *in-sample prediction error*:

$$\alpha^* = \underset{\alpha \in (0,1)}{\operatorname{argmin}} \underbrace{\sum_{t=1}^{T} (y_t - \mu_{t-1})^2}_{SSE} \tag{10.6}$$

Recursive Expansion

$$\mu_T = \alpha y_T + (1-\alpha)\mu_{T-1} \tag{10.7}$$

$$= \alpha y_T + (1-\alpha)[\alpha y_{T-1} + (1-\alpha)\mu_{T-2}] \tag{10.8}$$

$$= \alpha \sum_{k=0}^{T-1} (1-\alpha)^k y_{T-k} + \underbrace{(1-\alpha)^{T-1}\mu_0}_{=0} \tag{10.9}$$

$$\tag{10.10}$$

and in general μ_0 is initialized to zero.

10.4 Exponentially Weighted Moving Average (EWMA)

Remark 10.1. By replacing the objective of prediction in SES (μ) with the variance of series, we can build a EWMA model.

Model for the variance $(r_t - \mu)^2$

$$\hat{\sigma}^2_{t|t-1} = (1-\lambda)\sum_{k=0}^{t-1} \lambda^k (r_{t-k} - \mu)^2 \tag{10.11}$$

Error Correction Form (Recursive Definition)

$$\hat{\sigma}^2_{t|t-1} = \hat{\sigma}^2_{t-1|t-2} + (1-\lambda)((r_{t-1} - \mu)^2 - \hat{\sigma}^2_{t-1|t-2}) \tag{10.12}$$

11 Volatility II

11.1 Heteroskedasticity

Definition 11.1. Given model

$$r_t = \mu_{t|t-1} + \varepsilon_t \tag{11.1}$$

where $\mu_{t|t-1}$ can be modelled using an ARMA model, and

$$\varepsilon_t = \sigma_{t|t-1} z_t \tag{11.2}$$

and z_t satisfies $\mathbb{E}[z_t] = 0$ and $\mathbb{V}[z_t] = 1$ and *i.i.d.*

Remark 11.1.

$$\sigma^2_{t|t-1} = \mathbb{E}[\varepsilon_t^2 | I_t] \tag{11.3}$$

where ε_t is **heteroskedastic**, which means it has non-constant conditional variance.

Remark 11.2. Note here only the *conditional* variance is heteroskedastic, but the unconditional variance is constant $\mathbb{E}[\mathbb{E}[\varepsilon_t^2 | I_t]] = \sigma_\varepsilon^2$. This is similar to the unconditional mean (μ) of stochastic process $\{y_t\}$ in ARMA, which is constant, and the conditional mean of $\{y_t\}$ ($\mu_{t|t-1}$), which is non-constant.

11.2 Autoregressive Conditional Heteroskedasticity (ARCH)

Definition 11.2. ARCH(p) model

$$\sigma^2_{t|t-1} = \omega + \alpha_1 \varepsilon^2_{t-1} + \cdots + \alpha_p \varepsilon^2_{t-p} \tag{11.4}$$

Remark 11.3. In contrast to AR models, here the dependent variable is not ε_t^2 but instead its conditional expectation (since we already assumed the white noise distribution of ε_t).

Remark 11.4. Sufficient conditions for $\sigma^2_{t|t-1} > 0$ are

$$\begin{cases} \omega > 0 \\ \alpha_i \geq 0 \ \forall i \in \{1, \ldots, p\} \end{cases} \tag{11.5}$$

Remark 11.5. ARCH(1) forecasting (*apply the chain rule of forecasting*)

$$\begin{aligned}
\sigma^2_{t+h|t} &= \omega + \alpha\sigma^2_{t+h-1|t} &(11.6)\\
&= \omega + \alpha\omega + \alpha^2\sigma^2_{t+h-2|t} &(11.7)\\
&= \omega(1+\alpha+\cdots+\alpha^{h-2}) + \alpha^{h-1}\sigma^2_{t+1|t} &(11.8)\\
&\to \frac{\omega}{1-\alpha} \text{ as } h\to\infty &(11.9)
\end{aligned}$$

11.3 Generalize Autoregressive Conditional Heteroskedasticity (GARCH)

Definition 11.3. GARCH(p,q)

$$\begin{aligned}
\sigma^2_{t|t-1} &= \omega + \overbrace{\alpha_1\varepsilon^2_{t-1} + \cdots + \alpha_p\varepsilon^2_{t-p}}^{\text{ARCH}(p)\text{ Part}} &(11.10)\\
&\quad \underbrace{+\,\beta_1\sigma^2_{t-1|t-2} + \cdots + \beta_q\sigma^2_{t-q|t-q-1}}_{\text{Generalized Past Conditional Part}} &(11.11)
\end{aligned}$$

Remark 11.6. GARCH(p,q) typically needs smaller total number of parameters $(\alpha_i,\beta_j)_{i=1,\ldots,p;j=1,\ldots,q}$ than the number of $ARCH(\tilde{p})$ parameters $(\alpha_i)_{i=1,\ldots,\tilde{p}}$ required for a comparable model fit.

Proposition 11.1. GARCH(1,1) is equivalent to an ARCH(∞).

Proof.

$$\begin{aligned}
\sigma^2_{t|t-1} &= \overbrace{\omega + \alpha\varepsilon^2_{t-1} + \beta\sigma^2_{t-1|t-2}}^{\text{GARCH(1,1)}} &(11.12)\\
&= \omega + \alpha\varepsilon^2_{t-1} + \beta\omega + \beta\alpha\varepsilon^2_{t-2} + \beta^2\sigma^2_{t-1|t-2} &(11.13)\\
&= \underbrace{\omega(1+\beta+\beta^2+\cdots) + \alpha\sum_{i=1}^{\infty}\beta^{i-1}\varepsilon^2_{t-i}}_{\text{Expand Recursively to }(t-\infty)\text{ Period}} &(11.14)\\
&= \frac{\omega}{1-\beta} + \alpha\underbrace{\sum_{i=1}^{\infty}\beta^{i-1}}_{\text{Persistence}}\varepsilon^2_{t-i} &(11.15)
\end{aligned}$$

■

Definition 11.4. The **persistence** of GARCH(1,1) process is defined as

$$\alpha\sum_{i=1}^{\infty}\beta^{i-1} = \frac{\alpha}{1-\beta} \tag{11.16}$$

measures how permanent the innovations ε^2_t are in the conditional variance. When $\frac{\alpha}{1-\beta} = 1$, the process is an **integrated GARCH** process, its *unconditional* variance does not exist (equals ∞).

Proof.

$$\mathbb{E}[\sigma^2_{t|t-1}] = \mathbb{E}[\frac{\omega}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1} \varepsilon^2_{t-i}] \tag{11.17}$$

$$\implies \sigma^2_\varepsilon = \frac{\omega}{1-\beta} + \sigma^2_\varepsilon \tag{11.18}$$

$$\implies 0\sigma^2_\varepsilon = \frac{\omega}{1-\beta} \neq 0 \tag{11.19}$$

$$\implies \sigma^2_\varepsilon = \infty \tag{11.20}$$

■

12 Volatility III: Applications

12.1 Risk Management

Definition 12.1. The α**-Value at Risk (VaR)**, $r_t^{VaR(\alpha)}$, is the α-quantile of r_t. Let F be the cumulative density function of returns, then

$$r_t^{VaR(\alpha)} := F^{-1}(\alpha) \tag{12.1}$$

Remark 12.1 (Interpretation)**.** $r_t^{VaR(\alpha)}$ is the value in the domain of r_t such that *the possibility of obtaining an equal or smaller value than* r_t *is* α%.

$$\mathbb{P}[r_t \leq r_t^{VaR(\alpha)}] = \alpha \tag{12.2}$$

Model Setup

$$r_t = \mu_{t|t-1} + \sigma_{t|t-1} z_t \tag{12.3}$$

$$z_t \sim \mathcal{N}(0,1) \tag{12.4}$$

$$\mu_{t|t-1} \sim \text{ARMA} \tag{12.5}$$

$$\sigma_{t|t-1} \sim \text{GARCH} \tag{12.6}$$

$$\implies r_t \sim \mathcal{N}(\mu_{t|t-1}, \sigma^2_{t|t-1}) \tag{12.7}$$

$$\implies r_t^{VaR(\alpha)} = \mu_{t|t-1} + \Phi^{-1}(\alpha)\sigma_{t|t-1} \tag{12.8}$$

Definition 12.2. The **expected shortfall** measures the average value of such loss,

$$ES(\alpha) := \mathbb{E}[r_t | r_t < r_t^{VaR(\alpha)}] \tag{12.9}$$

Lemma 12.1 (Expected of Truncated Normal Distribution)**.**

$$X \sim \mathcal{N}(\mu, \sigma^2) \tag{12.10}$$

$$\implies \mathbb{E}[X|a < X < b] = \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \tag{12.11}$$

$$= \mu + \sigma \frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} \tag{12.12}$$

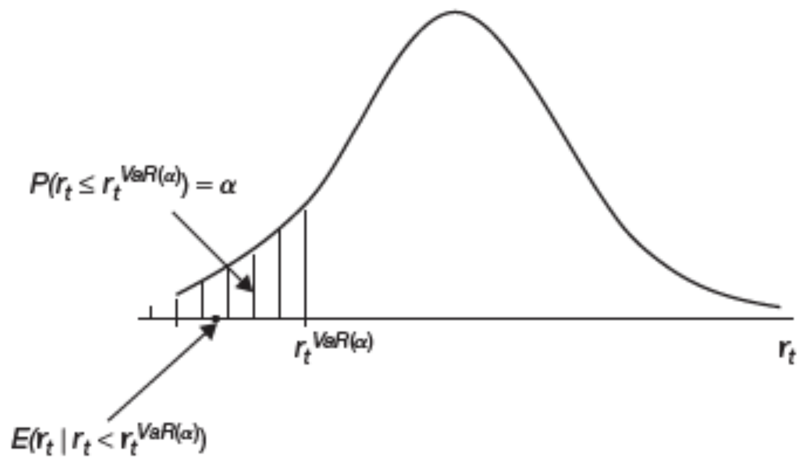


Figure 12.1: VaR and Expected Shortfall

Proposition 12.1. Computing the *expected shortfall* with normality assumption.

$$r_t \sim \mathcal{N}(\mu_{t|t-1}, \sigma^2_{t|t-1}) \tag{12.13}$$

$$\implies \mathbb{E}[r_t | r_t < r_t^{VaR(\alpha)}] = \mu_{t|t-1} + \sigma_{t|t-1} \frac{-\phi\left(\frac{r_t^{VaR(\alpha)} - \mu_{t|t-1}}{\sigma_{t|t-1}}\right)}{\Phi\left(\frac{r_t^{VaR(\alpha)} - \mu_{t|t-1}}{\sigma_{t|t-1}}\right)} \tag{12.14}$$

$$= \mu_{t|t-1} - \sigma_{t|t-1} \frac{\phi(\Phi^{-1}(\alpha))}{\alpha} \tag{12.15}$$

with R-code `coef <- dnorm(pnorm(alpha))/alpha`.

12.2 Portfolio Allocation

The Problem Suppose there are two assets (can be easily extend to the general case, let N denote the set of assets). Returns associated with those assets have mean μ_i and variance σ_i^2 for each $i \in N$. For given level of $\overline{\mu}_p$, one has to choose $(w_i)_{i \in N}$ such that the risk, σ_p^2, is minimized.

Assumption 12.1.

$$Cov(r_i, r_j) = 0 \ \forall i \neq j \in N \tag{12.16}$$

Solution

$$\min_{(w_i)_{i \in N}} \sigma_p^2 \equiv \sum_{i \in N} w_i^2 \sigma_i^2 \tag{12.17}$$

$$s.t. \ \overline{\mu}_p = \sum_{i \in N} w_i \mu_i \tag{12.18}$$

$$\implies w_i^* = \frac{\mu_i / \sigma_i^2}{\sum_{j \in N} \mu_j^2 / \sigma_j^2} \overline{\mu}_p \tag{12.19}$$

Remark 12.2. The means and variances of assets can be estimated *conditionally on t* using the hybrid $ARMA - GARCH$ model.

12.3 Asset Pricing: Classical Capital Asset Pricing Model

Definition 12.3. A **market portfolio** is a theoretical bundle of investments that includes every type of asset available in the world financial market, with each asset weighted in proportion to its total presence in the market.

Proposition 12.2.

$$\mathbb{E}[r_i] = r_f + \beta_i \mathbb{E}[r_m - r_f] \tag{12.20}$$

$$\beta_i := \frac{Cov(r_i, r_m)}{Var(r_m)} \equiv \frac{\rho_{im}\sigma_i\sigma_m}{\sigma_m^2} \tag{12.21}$$

and if $\beta_i > 1$, the asset i is more *risky* than the market risk. In practice, β_i is estimated with ρ_{im} estimated directly from samples and $\sigma^2_{i,t|t-1}, \sigma^2_{mz,t|t-1}$ estimated using $GARCH$.

13 Nonlinear Models

13.1 Threshold Auto-Regression (TAR)

Definition 13.1. Given data structure containing **main/target series**

$$\{y_t\} \tag{13.1}$$

and **threshold/activation variable series**

$$\{x_t\} \tag{13.2}$$

To construct a $TAR(p)$ model with r regimes,

(i) Partitioning the range of X_t (typically $\mathbb{R}$) into r *disjoint* sets $\{R_i\}_{i=1}^r$.

(ii) Construct model using indicator functions

$$Y_t = \sum_{i=1}^{r} \mathbb{1}\{x_t \in R_i\} \underbrace{\left(\phi_{i0} + \sum_{j=1}^{p} \phi_{ij} Y_{t-j} + \varepsilon_{it}\right)}_{\text{Regim } i \text{ AR}(p) \text{ Model}} \tag{13.3}$$

Definition 13.2. If the threshold/activation variable in a TAR model is any *lagged variation of the main/target series*, then the TAR model is called **self-exciting**.

13.2 Test Significance of Regimes

Remark 13.1. To test the significance of multiple regimes, we use conventional hypothesis testing for significance of coefficients in multiple regression models.

Example 13.1. To test the significance (whether it is necessary to include it) of two regimes (but with same error term) here

$$Y_t = \mathbb{1}\{Y_{t-1} \geq 0\}\left(\phi_0 + \phi_1 Y_{t-1}\right) + \mathbb{1}\{Y_{t-1} < 0\}\left(\phi_0' + \phi_1' Y_{t-1}\right) + \varepsilon_t \tag{13.4}$$

$$D_t := \mathbb{1}\{Y_{t-1} \geq 0\} \tag{13.5}$$

$$\implies Y_t = D_t\left(\phi_0 + \phi_1 Y_{t-1}\right) + (1 - D_t)\left(\phi_0' + \phi_1' Y_{t-1}\right) + \varepsilon_t \tag{13.6}$$

$$\implies Y_t = \phi_0' + \phi_1' Y_{t-1} + \underbrace{(\phi_0 - \phi_0')}_{\Delta\phi_0} D_t + \underbrace{(\phi_1 - \phi_1')}_{\Delta\phi_1} D_t Y_{t-1} + \varepsilon_t \tag{13.7}$$

And test the hypothesis (equivalently, test the *joint significance* of coefficients of D_t and $D_t Y_{t-1}$ (partial F test)

$$
\begin{aligned}
H_0 &:= \Delta\phi_0 = 0 \wedge \Delta\phi_1 = 0 \text{ model is linear} && (13.8)\\
H_1 &:= \neg H_0 \text{ model is non-linear} && (13.9)
\end{aligned}
$$

13.3 Smooth Transition Autoregressive Model (STAR)

Definition 13.3. A **smooth transition autoregressive model**(STAR) is defined by two autoregression regimes and a *bounded and continuous* (activation function) $\mathcal{G}(s_t, \gamma, c)$. Where

(i) γ denotes the **speed parameter**;

(ii) c denotes **threshold**;

(iii) $\mathcal{G}$ is bounded and continuous in **threshold variables** s_t.

$$
Y_t = \underbrace{\left(\phi_0 + \sum_{j=1}^{p} \phi_j L^j Y_t\right)}_{\text{base model}} + \underbrace{\left(\phi'_0 + \sum_{j=1}^{p} \phi'_j L^j Y_t\right) \mathcal{G}(s_t, \gamma, c)}_{\text{activation}} + \varepsilon_t \tag{13.10}
$$

Example 13.2 (STAR(1) without intercept)**.**

$$
Y_t = \phi_0 Y_{t-1} + \phi_1 Y_{t-1} \mathcal{G}(s_t, \gamma, c) + \varepsilon_t \tag{13.11}
$$

Example 13.3 (Logistic STAR)**.**

$$
\mathcal{G}(s_t, \gamma, c) = \frac{1}{1 + \exp(-\gamma(s_t - c))} \tag{13.12}
$$

Example 13.4 (Exponential STAR)**.**

$$
\mathcal{G}(s_t, \gamma, c) = 1 - \exp(-\gamma(s_t - c)^2) \tag{13.13}
$$

Remark 13.2. STAR is effectively a TAR with infinitely many regimes.

13.4 Markov Switching Model

Definition 13.4. MS models the *probability of various regimes to happen.* A Markov switching model consists both

(i) **State space** $\Omega := \{\omega_i\}_{i=1}^n$;

(ii) **Outcome** functions depend on state realized at time t, s_t;

$$
Y_t \sim \mathcal{N}(\mu_i, \sigma^2) \text{ if } s_t = \omega_i \tag{13.14}
$$

(iii) A $n \times n$ **transition matrix** T defined as

$$
T_{i,j} := \mathbb{P}[s_t = \omega_j | s_{t-1} = \omega_i] \tag{13.15}
$$

14 Appendix

A list of hypothesis tests in this course:

(i) t test for auto-correlation $H_0 : \rho_k = 0$.

(ii) Ljung-Box test for auto-correlations $H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0$.

(iii) ADF for unit root, $z = \frac{\hat{\phi}-1}{\sigma_{\hat{\phi}}}$, H_0 : non-stationary/unit root exists three types.

(iv) Partial F test for testing Ganger causality.

(v) ADF test for testing co-integration, 2 step procedure:

  (1) construct $\hat{z}$ from OLS procedure;

  (2) ADF test for the stationarity of $\hat{z}$.

(vi) Johansen co-integration test: $r = 0$ no integrating relationship, $r = 1$ one integrating relationship.

(vii) TAR validity: partial F test with dummy variables.