

Forecasting and Time Series Econometrics

ECO374 Winter 2019

Tianyu Du

February 12, 2019

Contents

1	Introduction and Statistics Review	1
2	Statistics and Time Series	2
2.1	Stochastic Processes	2
2.2	Auto-correlations	3
3	Forecasting Tools	5
3.1	Information Set	5
3.2	Forecast Horizon	5
3.2.1	Forecasting Environments: <i>Recursive</i>	5
3.2.2	Forecasting Environments: <i>Rolling</i>	6
3.2.3	Forecasting Environments: <i>Fixed</i>	6
3.3	Loss Function	7
3.4	Optimal Forecast	7
4	Moving Average Process	8

1 Introduction and Statistics Review

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 variables X and Y , the **conditional density** of Y conditioned on X is defined as

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

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

(i) **Linearity**

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u \quad (1.5)$$

(ii) **Zero Conditional Mean**

$$\mathbb{E}[u|X_1, X_2, \dots, X_k] = 0 \quad (1.6)$$

(iii) **Homoscedasitcity**

$$\mathbb{V}[u|X_1, X_2, \dots, X_k] = \sigma_u^2 \quad (1.7)$$

(iv) **No Serial Correlation**

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

(v) **No Perfect Collinearity**

(vi) **Sample Variation in Regressors**

$$\mathbb{V}[X_j] > 0 \quad \forall j \quad (1.9)$$

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 (1.1). 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 (Ω, \mathcal{F}, P) .

$$\{Y_t\} = Y_1, \dots, Y_T \quad (2.1)$$

Definition 2.2 (1.2). The function $t \rightarrow y_t$ 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** or a **trajectory** or an **outcome** of the stochastic process.

Definition 2.3. An *outcome* of a stochastic process

$$\{y_t\} = y_1, \dots, y_T \quad (2.2)$$

is a **time series**.

Definition 2.4 (1.3). A **time series model** or a **model** for the observations (data), $\{y_t\}$, 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 2.5. A stochastic process $\{Y_t\}$ is **first order strongly stationary** if all random variables $Y_t \in \{Y_t\}$ has the *same probability density function*.

Definition 2.6 (1.7). A stochastic process $\{Y_t\}$ is **strictly stationary** if for all $h, n \geq 1$, (X_1, \dots, X_n) and $(X_{1+h}, \dots, X_{n+h})$ have the same distribution.

Definition 2.7. A stochastic process $\{Y_t\}$ is **first order weakly stationary** if

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

Definition 2.8. 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) $\mathbb{E}[Y_t] = \mu \forall t$
- (ii) $\mathbb{V}[Y_t] = \sigma^2 < \infty \forall t$
- (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 2.9. 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

$$\gamma_Y(t, s) \equiv Cov(Y_t, Y_s) \quad (2.4)$$

$$= \mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_s - \mathbb{E}[Y_s])] \quad (2.5)$$

$$= \mathbb{E}[Y_t Y_s] - \mathbb{E}[Y_t] \mathbb{E}[Y_s] \quad (2.6)$$

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

$$\gamma_Y(h) \equiv \gamma_X(t, t + h) \forall t \in \mathcal{T} \quad (2.7)$$

Proposition 2.1. By the symmetry of covariance,

$$\gamma_Y(h) = \gamma_Y(-h) \quad (2.8)$$

Definition 2.10. The **auto-correlation coefficient** of order k is given by

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

Definition 2.11. Let $\{Y_t\}$ be a *stationary process* and the **auto-correlation function** (ACF) is a mapping from *order* of auto-correlation coefficient to the coefficient $\rho_Y : k \rightarrow \rho_{Y_t, Y_{t-k}}$, defined as

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

Proposition 2.2. Note that

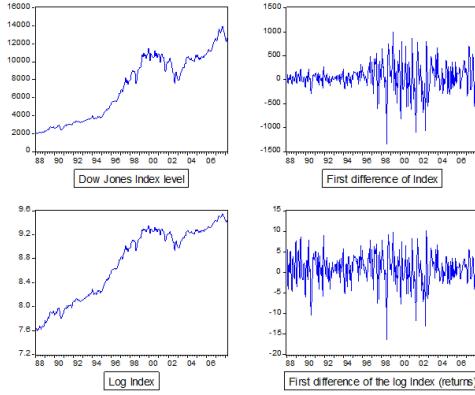
$$\rho_k = \rho_{-k} = \rho_{|k|} \quad (2.11)$$

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

$$\rho : k \rightarrow \rho_{|k|} \quad (2.12)$$

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

Proposition 2.3. 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*.



Definition 2.12 (1.8). A stochastic process $\{Y_t\}$ is called a **Gaussian process** if all *finite dimensional* distribution from the process are multivariate normally distributed. That's

$$\forall n \in \mathbb{Z}_{>0}, \forall (t_1, \dots, t_n) \in \mathcal{T}^n, (Y_{t_1}, \dots, Y_{t_n}) \sim \mathcal{N}(\mu, \Sigma) \quad (2.13)$$

Notation 2.1. Consider the problem of forecasting Y_{T+1} from observations $\{Y_t\}_{t=1}^T$, the *best linear predictor* is denoted as

$$\mathbb{P}_T Y_{T+1} = \sum_{i=1}^T a_i L^i Y_{T+1} \quad (2.14)$$

And Y_{T+1} can be expressed as

$$Y_{T+1} = \mathbb{P}_T Y_{T+1} + Z_{T+1} \quad (2.15)$$

where Z_{T+1} denotes the forecast error which is *uncorrelated* with X_T, \dots, X_1 .

Definition 2.13 (3.3). The **partial auto-correlation function** (PACT) $\alpha(h)$ with $h \in \mathbb{Z}_{\geq 0}$ of a *stationary* process is defined as

$$\alpha(0) = 1 \quad (2.16)$$

$$\alpha(1) = \text{corr}(Y_2, Y_1) = \rho(1) \quad (2.17)$$

$$\alpha(h) = \text{corr}\left(Y_{h+1} - \mathbb{P}(Y_{h+1}|1, Y_2, \dots, Y_h), X_1 - \mathbb{P}(Y_1|1, Y_2, \dots, Y_h)\right) \quad (2.18)$$

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 $(Y_{t+1}, \dots, Y_{t+k-1})$.

Remark 2.3. Properties of ACF and PACF

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

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

$$H_0 : \rho_k = 0 \quad (2.19)$$

we can use usual t-statistic. While testing the joint hypothesis

$$H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0 \quad (2.20)$$

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

$$Q_k = T(T + 1) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{T - j} \sim \chi_k^2 \quad (2.21)$$

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 .

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

$$f_{t,h} = g(I_t) \in \Omega \quad (3.1)$$

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.

Remark 3.2. Non-stationary processes are **long-memory processes** and older information is as relevant for the forecast as more recent information.

3.2.1 Forecasting Environments: *Recursive*

- (i) *Updating* with *flexible* information set.
- (ii) Advantageous if model is *stable over time*.
- (iii) Not robust to *structural break*.

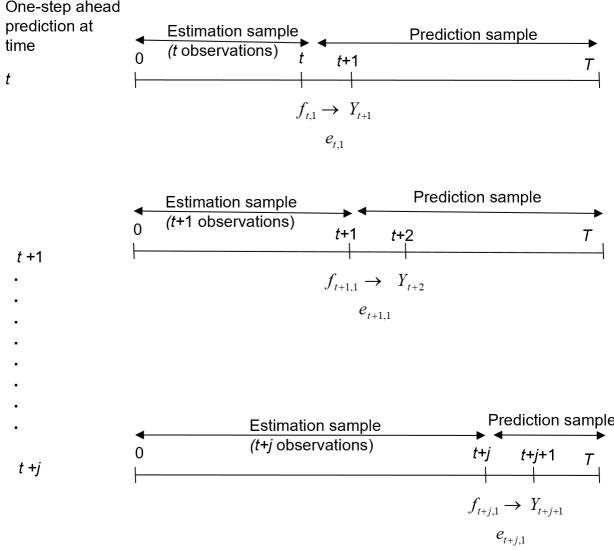


Figure 3.1: Recursive Forecasting Scheme

3.2.2 Forecasting Environments: *Rolling*

- (i) *Updating* with *fixed-size* information set.
- (ii) Robust against *structural breaks*.
- (iii) Not fully exploit information available.

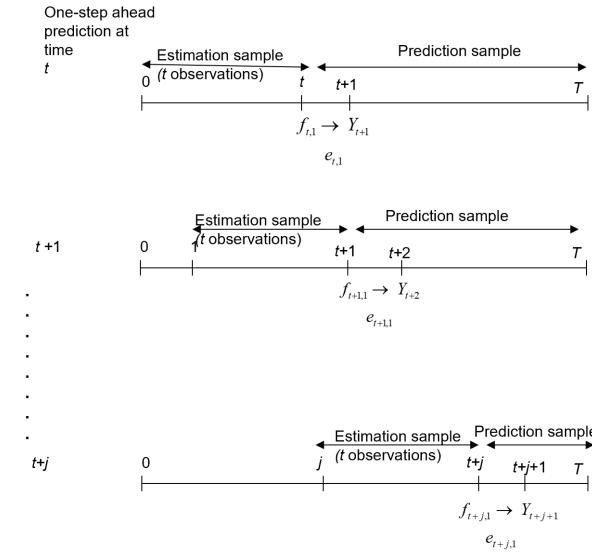


Figure 3.2: Rolling Forecasting Scheme

3.2.3 Forecasting Environments: *Fixed*

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

(ii) Computationally cheap.

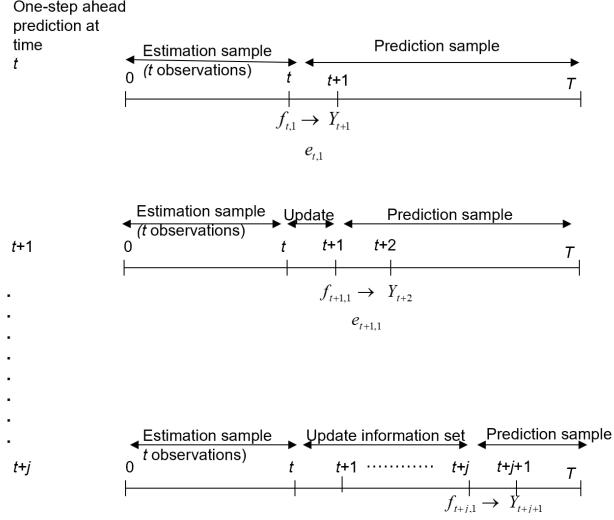


Figure 3.3: Fixed Forecasting Scheme

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) $\forall e \in \mathcal{E}, L(e) \geq 0$ ¹;
- (iii) L is *monotonically increasing* in the norm of forecast error.

Example 3.1 (Symmetric Loss Functions with $\mathcal{E} = \mathbb{R}$).

$$L(e) = ae^2, \quad a > 0 \quad (3.2)$$

$$L(e) = a|e|, \quad a > 0 \quad (3.3)$$

Example 3.2 (Asymmetric Loss Functions with $\mathcal{E} = \mathbb{R}$).

$$L(e) = \exp(ae) - ae - 1, \quad a > 0 \quad \text{Linex Function} \quad (3.4)$$

$$L(e) = a|e| \mathbb{I}(e \geq 0) + b|e| \mathbb{I}(e < 0) \quad \text{Lin-lin Function} \quad (3.5)$$

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**

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

¹Since 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.

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

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

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

$$\mu_{t+h|t} \equiv \mathbb{E}[Y_{t+h}|I_t] \quad (3.8)$$

Proof.

$$\min_{f_{t,h} \in \mathbb{R}} L \equiv \int (y_{t+h} - f_{t,h})^2 f(y_{t+h}|I_t) dy_{t+h} \quad (3.9)$$

$$\frac{\partial L}{\partial f_{t,h}} = -2 \int (y_{t+h} - f_{t,h}) f(y_{t+h}|I_t) dy_{t+h} = 0 \quad (3.10)$$

$$\implies \int (y_{t+h} - f_{t,h}) f(y_{t+h}|I_t) dy_{t+h} = 0 \quad (3.11)$$

$$\implies \int y_{t+h} f(y_{t+h}|I_t) dy_{t+h} = f_{t,h} \int f(y_{t+h}|I_t) dy_{t+h} \quad (3.12)$$

$$\implies f_{t,h} = \mathbb{E}[y_{t+h}|I_t] \equiv \mu_{t+h|t} \quad (3.13)$$

■

4 Moving Average Process