Financial Econometrics
Lecture 2: Analysis of Financial Time Series

Prof Hamed Ghoddusi
2019

Simple AR models: (Regression with lagged variables)

Motivating example: The growth rate of U.S. quarterly real GNP from 1947 to 1991. Recall that the model discussed before is

$$r_t = 0.005 + 0.35r_{t-1} + 0.18r_{t-2} - 0.14r_{t-3} + a_t, \hat{\sigma}_a = 0.01.$$

This is called an AR(3) model because the growth rate r_t depends on the growth rates of the past **three** quarters. How do we specify this model from the data? Is it adequate for the data? What are the implications of the model? These are the questions we shall address in this lecture.

AR(1) model

Another example: U.S. monthly unemployment rate.

1. Form: $r_t = \phi_0 + \phi_1 r_{t-1} + a_t$, where ϕ_0 and ϕ_1 are real numbers, which are referred to as "parameter" (to be estimated from the data in an application). For example,

$$r_t = 0.05 + 0.4r_{t-1} + a_t$$

2. Stationarity: necessary and sufficient condition $|\phi_1| < 1$. Why?

3. Mean: $E(r_t) = \frac{\phi_0}{1-\phi_1}$

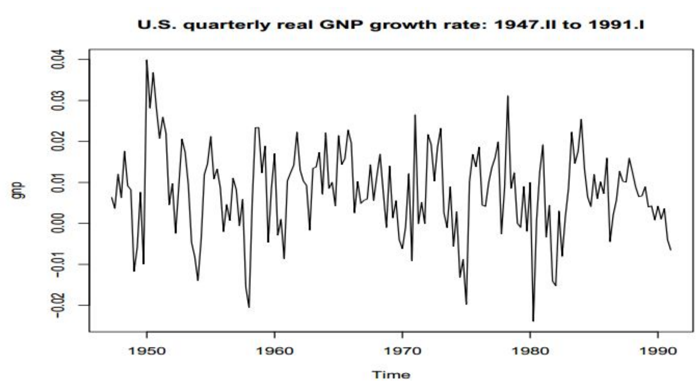


Figure: U.S. quarterly growth rate of real GNP: 1947-1991

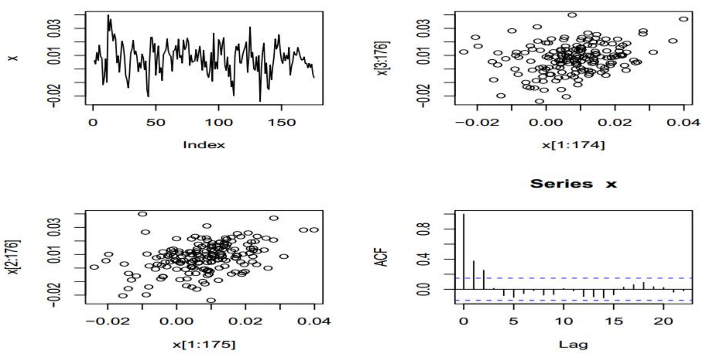


Figure: Various plots of U.S. quarterly growth rate of real GNP: 1947-1991

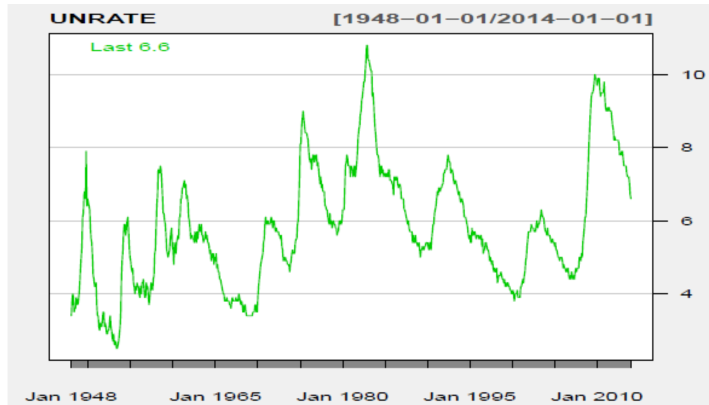


Figure: U.S. monthly unemployment rate (total civilian, 16 and older) from January 1948 to January 2014.

AR(1) model

4. Alternative representation: Let $E(r_t) = \mu$ be the mean of r_t so that $\mu = \phi_0/(1-\phi_1)$. Equivalently, $\phi_0 = \mu(1-\phi_1)$. Plugging in the model, we have

$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + a_t. \tag{1}$$

   This model also has two parameters (μ and ϕ_1). It explicitly uses the mean of the series. It is less commonly used in the literature, but is the model representation used in R.
5. Variance: $\mathrm{Var}(r_t) = \frac{\sigma_a^2}{1-\phi_1^2}$.
6. Autocorrelations: $\rho_1 = \phi_1, \rho_2 = \phi_1^2$, etc. In general, $\rho_k = \phi_1^k$ and ACF ρ_k decays exponentially as k increases,

AR(1) model

7. Forecast (minimum squared error): Suppose the forecast origin is n. For simplicity, we shall use the model representation in (1) and write $x_t = r_t - \mu$. The model then becomes $x_t = \phi_1 x_{t-1} + a_t$. Note that forecast of r_t is simply the forecast of x_t plus μ.
   a. 1-step ahead forecast at time n:
   $$\hat{x}_n(1) = \phi_1 x_n$$
   b. 1-step ahead forecast error:
   $$e_n(1) = x_{n+1} - \hat{x}_n(1) = a_{n+1}$$
   Thus, a_{n+1} is the $un-predictable$ part of x_{n+1}. It is the shock at time $n+1$!
   c. Variance of 1-step ahead forecast error:
   $$\mathrm{Var}[e_n(1)] = \mathrm{Var}(a_{n+1}) = \sigma_a^2.$$

AR(1) model

7. 2-step ahead forecast:

$$\hat{x}_n(2) = \phi_1 \hat{x}_n(1) = \phi_1^2 x_n.$$

8. 2-step ahead forecast error:

$$e_n(2) = x_{n+2} - \hat{x}_n(2) = a_{n+2} + \phi_1 a_{n+1}$$

9. Variance of 2-step ahead forecast error:

$$\mathrm{Var}[e_n(2)] = (1 + \phi_1^2)\sigma_a^2$$

which is greater than or equal to $\mathrm{Var}[e_n(1)]$, implying that uncertainty in forecasts increases as the number of steps increases.

AR(1) model

10. Behavior of multi-step ahead forecasts. In general, for the ℓ-step ahead forecast at n, we have

$$\hat{x}_n(\ell) = \phi_1^\ell x_n,$$

the forecast error

$$e_n(\ell) = a_{n+\ell} + \phi_1 a_{n+\ell-1} + \cdots + \phi^{\ell-1} a_{n+1},$$

and the variance of forecast error

$$\mathrm{Var}[e_n(\ell)] = (1 + \phi_1^2 + \cdots + \phi_1^{2(\ell-1)})\sigma_a^2.$$

In particular, as $\ell \to \infty$,

$$\hat{x}_n(\ell) \to 0, \quad i.e., \quad \hat{r}_n(\ell) \to \mu.$$

This is called the mean-reversion of the AR(1) process. The variance of forecast error approaches

$$\mathrm{Var}[e_n(\ell)] = \frac{1}{1-\phi_1^2}\sigma_a^2 = \mathrm{Var}(r_t).$$

AR(1) model

In practice, it means that for the long-term forecasts serial dependence is not important. The forecast is just the sample mean and the uncertainty is simply the uncertainty about the series.

8. A compact form: $(1 - \phi_1 B) r_t = \phi_0 + a_t$.
Half-life: A common way to quantify the *speed* of mean reversion is the half-life, which is defined as the number of periods needed so that the magnitude of the forecast becomes half of that of the forecast origin. For an AR(1) model, this mean

$$x_n(k) = \frac{1}{2} x_n.$$

Thus, $\phi_1^k x_n = \frac{1}{2} x_n$. Consequently, the half-life of the AR(1) model is $k = \frac{\ln(0.5)}{\ln(|\phi_1|)}$. For example, if $\phi_1 = 0.5$, the $k = 1$. If $\phi_1 = 0.9$, then $k \approx 6.58$.

AR(2) model

1. Form: $r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + a_t$, or

$$(1 - \phi_1 B - \phi_2 B^2) r_t = \phi_0 + a_t.$$

2. Stationarity condition: (factor of polynomial)
3. Characteristic equation: $(1 - \phi_1 x - \phi_2 x^2) = 0$
4. Mean: $E(r_t) = \frac{\phi_0}{1-\phi_1-\phi_2}$
5. Mean-adjusted format: Using $\phi_0 = \mu - \phi_1\mu - \phi_2\mu$, we can write the AR(2) model as

$$(r_t - \mu) = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t.$$

This form is often used in the finance literature to highlight the mean-reverting property of a stationary AR(2) model.

AR(2) model

6. ACF: $\rho_0 = 1, \rho_1 = \frac{\phi_1}{1-\phi_2}$,
$$\rho_\ell = \phi_1 \rho_{\ell-1} + \phi_2 \rho_{\ell-1}, \qquad \ell \geqslant 2.$$
7. Stochastic business cycle: if $\phi_1^2 + 4\phi_2 < 0$, then r_t shows characteristics of business cycles with average length
$$k = \frac{2\pi}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]},$$
where the cosine inverse is stated in radian. If we denote the solutions of the polynomial as $a \pm bi$, where $i = \sqrt{-1}$, then we have $\phi_1 = 2a$ and $\phi_2 = -(a^2 + b^2)$ so that
$$k = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2+b^2})}.$$
In R or S-Plus, one can obtain $\sqrt{a^2+b^2}$ using the command **Mod**.
8. Forecasts: Similar to AR(1) models

Discussion: (Reference only)

An AR(2) model can be written as an AR(1) model if one expands the dimension. Specifically, we have

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_2(r_{t-2} - \mu) + a_t$$
$$r_{t-1} - \mu = r_{t-1} - \mu, \qquad \text{(an identity.)}$$

Now, putting the two equations together, we have

$$\begin{bmatrix} r_t - \mu \\ r_{t-1} - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} r_{t-1} - \mu \\ r_{t-2} - \mu \end{bmatrix} + \begin{bmatrix} a_t \\ 0 \end{bmatrix}.$$

This is a 2-dimensional AR(1) model. Several properties of the AR(2) model can be obtained from the expanded AR(1) model.

Building an AR model

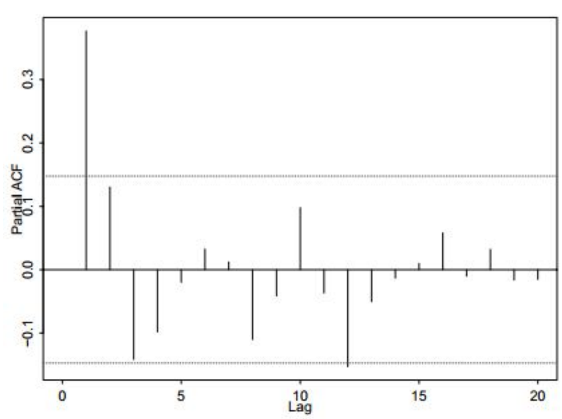
- Order specification

  1. Partial ACF: (naive, but effective)
     - Use consecutive fittings
     - See Text (p. 40) for details
     - **Key feature:** PACF cuts off at lag p for an AR(p)model.
     - Illustration: See the PACF of the U.S. quarterly growth rate of GNP.

  2. Akaike information criterion
  $$AIC(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{2\ell}{T},$$
  for an AR(ℓ) model, where $\tilde{\sigma}_\ell^2$ is the MLE of residual variance. Find the AR order with *minimum* AIC for $\ell \in [0, \dots, P]$.

  3. BIC criterion:
  $$BIC(\ell) = \ln(\tilde{\sigma}_\ell^2) + \frac{\ell \ln(T)}{T}.$$

Figure: Series : dgnp



Building an AR model

- Needs a constant term? Check the sample mean.
- Estimation: least squares method or maximum likelihood method
- Model checking:
  1. Residual: obs minus the fit, i.e. 1-step ahead forecast errors at each time point.
  2. Residual should be close to white noise if the model is adequate. Use Ljung-Box statistics of residuals, but degrees of freedom is $m - g$, where g is the number of AR coefficients used in the model.

Moving-average (MA) model

Model with finite memory!
Some daily stock returns have minor serial correlations and can be modeled as MA or AR models.

MA(1) model

- Form: $r_t = \mu + a_t - \theta_{a_{t-1}}$
- Stationarity: always stationary.
- Mean (or expectation): $E(r_t) = \mu$
- Variance: $\text{Var}(r_t) = (1 + \theta^2)\sigma_a^2$.
- Autocovariance:
  1. Lag 1: $\text{Cov}(r_t, r_{t-1}) = -\theta\sigma_a^2$
  2. Lag ℓ : $\text{Cov}(r_t, r_{t-\ell}) = 0$ for $\ell > 1$.
     Thus, r_t is not related to $r_{t-2}, r_{t-3}, \ldots$.
- ACF: $\rho_1 = \frac{-\theta}{1+\theta^2}, \rho_\ell = 0 for \ell > 1$.
  Finite memory! MA(1) models do not remember what happen two time periods ago.

MA(1) model

- Forecast (at origin $t = n$):
  1. 1-step ahead: $\hat{r}_n(1) = \mu - \theta_{a_n}$. Why? Because at time n, a_n is known, but a_{n+1} is not.
  2. 1-step ahead forecast error: $e_n(1) = a_{n+1}$ with variance σ_a^2.
  3. Multi-step ahead: $\hat{r}_n(\ell) = \mu$ for $\ell > 1$.
     Thus, for an MA(1) model, the multi-step ahead forecasts are just the mean of the series. Why? Because the model has memory of 1 time period.
  4. Multi-step ahead forecast error:
     $$e_n(\ell) = a_{n+l} - \theta a_{n+\ell-1}$$
  5. Variance of multi-step ahead forecast error: $(1 + \theta^2)\sigma_a^2 =$ variance of r_t.

MA(1) model

- Invertibility:
  - Concept: r_t is a proper linear combination of a_t and the past observations $r_{t-1}, r_{t-2}, \ldots$.
  - Why is it important? It provides a simple way to obtain the shock a_t.
    For an invertible model, the dependence of r_t on $r_{t-\ell}$ converges to zero as ℓ increases.
  - Condition: $|\theta| < 1$
  - Invertibility of MA models is the dual property of stationarity for AR models.

MA(2) model

- Form: $r_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$. or
$$r_t = \mu + (1 - \theta_1 B - \theta_2 B^2) a_t.$$
- Stationary with $E(r_t) = \mu$.
- Variance: $\mathrm{Var}(r_t) = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$.
- ACF: $\rho_2 \neq 0$, but $\rho_\ell = 0$ for $\ell > 2$
- Forecasts go the the mean after 2 periods.

Building an MA model

- Specification: Use sample ACF
Sample ACFs are all small after lag q for an MA(q) series. (See test of ACF.)
- Constant term? Check the sample mean.
- Estimation: use maximum likelihood method
  - Conditional: Assume $a_t = 0$ for $t \leqslant 0$
  - Exact: Treat a_t with $t \leqslant 0$ as parameters, estimate them to obtain the likelihood function.

  Exact method is preferred, but it is more computing intensive.
- Model checking: examine residuals (to be white noise) error 2
- Forecast: use the residuals as $\{a_t\}$ (which can be obtained from the data and fitted parameters) to perform forecasts.

Building an MA model

Model form in R: R parameterizes the MA(q) model as
$$r_t = \mu + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q},$$
instead of the usual *minus* sign in θ. Consequently, care needs to be exercised in writing down a fitted MA parameter in R. For instance, an estimate $\hat{\theta}_1 = -0.5$ of an MA(1) in R indicates the model is $r_t = a_t - 0.5a_{t-1}$.

Mixed ARMA model

A compact form for flexible models.
Focus on the ARMA(1,1) model for

1. simplicity
2. useful for understanding GARCH models in Ch. 3 for volatility modeling.

ARMA(1,1) model

- Form: $(1 - \phi_1 B) r_t = \phi_0 + (1 - \theta B) a_t$ or
$$r_t = \phi_1 r_{t-1} + \phi_0 + a_t - \theta_1 a_{t-1}.$$
A combination of an AR(1) on the LHS and an MA(1) on the RHS.
- Stationarity: same as AR(1)
- Invertibility: same as MA(1)

ARMA(1,1) model

- Mean: as AR(1), i.e. $E(r_t) = \frac{\phi_0}{1-\phi_1}$
- Variance: given in the text
- ACF: Satisfies $\rho_k = \phi_1 \rho_{k-1}$ for $k > 1$, but

$$\rho_1 = \phi_1 - [\theta_1 \sigma_a^2 / \operatorname{Var}(r_t)] \neq \phi_1.$$

  This is the difference between AR(1) and ARMA(1,1) models.
- PACF: does not cut off at finite lags.

Building an ARMA(1,1) model

- Specification: use EACF or AIC
- What is EACF? How to use it? [See text].
- Estimation: cond. or exact likelihood method

三角形顶点

- Model checking: as before
- Forecast: MA(1) affects the 1-step ahead forecast. Others are similar to those of AR(1) models.

Three model representations

- ARMA form: compact, useful in estimation and forecasting
- AR representation: (by long division)

$$r_t = \phi_0 + a_t + \pi_1 r_{t-1} + \pi_2 r_{t-2} + \dots$$

  It tells how r_t depends on its past values.
- MA representation: (by long division)

$$r_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots$$

  It tells how r_t depends on the past shocks.

Three model representations

- For a stationary series, ψ_i converges to zero as $i \to \infty$. Thus, the effect of any shock is transitory.
- The MA representation is particularly useful in computing variances of forecast errors.
- For a ℓ-step ahead forecast, the forecast error is

$$e_n(\ell) = a_{n+\ell} + \psi_1 a_{n+\ell-1} + \cdots + \psi_{\ell-1} a_{n+1}.$$

- The variance of forecast error is

$$\operatorname{Var}[e_n(\ell)] = (1 + \psi_1^2 + \cdots + \psi_{\ell-1}^2)\sigma_a^2.$$

Unit-root Nonstationarity

Random walk

- Form $p_t = p_{t-1} + a_t$
- Unit root? It is an AR(1) model with coefficient $\phi_1 = 1$.
- Nonstationary: Why? Because the variance of r_t diverges to infinity as t increases.
- Strong memory: sample ACF approaches 1 for any finite lag.
- Repeated substitution shows

$$p_t = \sum_{i=0}^{\infty} a_{t-i} = \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

Unit-root Nonstationarity

Random walk with drift

- Form $p_t = \mu + p_{t-1} + a_t, \quad \mu \neq 0$.
- Has a unit root
- Nonstationary
- Strong memory
- Has a time trend with slope μ. Why?

differencing

- 1st difference: $r_t = p_t - p_{t-1}$
  If p_t is the log price, then the 1st difference is simply the log return. Typically, 1st difference means the "change" or "increment" of the original series.
- Seasonal difference: $y_t = p_t - p_{t-s}$, where s is the periodicity, e.g. $s = 4$ for quarterly series and $s = 12$ for monthly series.
  If p_t denotes quarterly earnings, then y_t is the change in earning from the same quarter one year before.

Meaning of the constant term in a model

- MA model: mean
- AR model: related to mean
- 1st differenced: time slope, etc.

Practical implication in financial time series

Example: Monthly log returns of General Electrics (GE) from 1926 to 1999 (74 years)
Sample mean: 1.04%, std$(\hat{\mu}) = 0.26$
Very significant!
is about 12.45% a year $1 investment in the beginning of 1926 is worth

- annual compounded payment: $5907
- quarterly compounded payment: $8720
- monthly compounded payment: $9570
- Continuously compounded?

Unit-root test

Let p_t be the log price of an asset. To test that p_t is not predictable (i.e. has a unit root), two models are commonly employed:

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

The hypothesis of interest is $H_o : \phi_1 = 1$ vs $H_a : \phi_1 < 1$.
Dickey-Fuller test is the usual t-ratio of the OLS estimate of ϕ_1 being This is the DF unit-root test. The t-ratio, however, has a nonstandard limiting distribution.
Let $\triangle p_t = p_t - p_{t-1}$. Then, the augmented DF unit-root test for an AR(p) model is based on

$$\triangle p_t = c_t + \beta p_{t-1} + \sum_{i=1}^{p-1} \phi_i \triangle p_{t-i} + e_t.$$

The t-ratio of the OLS estimate of β is the ADF unit-root test statistic. Again, the statistic has a non-standard limiting distribution.

Exponential Smoothing Approach

Suppose the available data are $r_t, r_{t-1}, r_{t-2}, \ldots$ and we are interested in predicting r_{t+1}.

Intuition: The data r_t should be more relevant than r_{t-1} in predicting r_{t+1}. Similarly, r_{t-1} is more relevant than r_{t-2}, etc.

A simple formulation: Suppose we assign weight w to r_t, weight $w\theta$ with $0 < \theta < 1$ to r_{t-1}, weight $w\theta^2$ to the data r_{t-2}, etc. That is, we use an initial weight w and a discounting factor θ.

Simple fact: The weight must sum to 1. Why? Not to change the scale. Therefore,

$$\begin{aligned} 1 &= w + w\theta + w\theta^2 + w\theta^3 + \ldots \\ &= w[1 + \theta + \theta^2 + \theta^3 + \ldots] \\ &= \frac{w}{1-\theta} \end{aligned}$$

Exponential Smoothing Approach

Therefore, $w = 1 - \theta$. Consequently, the 1-step ahead prediction of r_{t+1} is

$$r_t(1) = (1-\theta)r_t + (1-\theta)\theta r_{t-1} + (1-\theta)\theta^2 r_{t-2} + (1-\theta)\theta^3 r_{t-3} + \ldots \tag{2}$$

Let a_{t+1} be the forecast error (or the innovation at time $t+1$), then we have

$$r_{t+1} = r_t(1) + a_{t+1}. \tag{3}$$

Now, suppose we have data $r_{t-1}, r_{t-2}, \ldots$ and are interested in forecasting r_t at time $t-1$. The same argument shows that

$$r_{t-1}(1) = (1-\theta)r_{t-1} + (1-\theta)\theta r_{t-2} + (1-\theta)\theta^2 r_{t-3} + \ldots \tag{4}$$

Exponential Smoothing Approach

Next, using Equation (2), we have

$$\begin{aligned} r_t(1) &= (1-\theta)r_t + \theta[(1-\theta)r_{t-1} + (1-\theta)\theta r_{t-2} + (1-\theta)\theta^2 r_{t-3} + \ldots] \\ &= (1-\theta)r_t + \theta r_{t-1}(1). \quad \text{[see Equation (4)].} \end{aligned}$$

Putting the prior result into Equation (3), we obtain

$$\begin{aligned} r_{t+1} &= (1-\theta)r_t + \theta r_{t-1}(1) + a_{t+1} \\ &= r_t - \theta[r_t - r_{t-1}(1)] \\ &= r_t - \theta a_t + a_{t+1}. \end{aligned}$$

In the above, we have use $r_t = r_{t-1}(1) + a_t$. Consequently, the exponential smoothing model is

$$r_{t+1} - r_t = a_{t+1} - \theta a_t,$$

Exponential Smoothing Approach

which is an ARIMA(0,1,1) model and can be written as

$$(1-B)r_t = (1-\theta B)a_t.$$

This shows that the exponential smoothing method is simply using an ARIMA(0,1,1) model with a positive θ, which is the discounting factor.

Updating: For a given discounting rate θ, it is easy to update the forecast via the exponential smoothing method, because

$$r_t(1) = (1-\theta)r_t + \theta r_{t-1}(1),$$

which means the new prediction is simply a weighted average of the new data r_t and the previous forecast $r_{t-1}(1)$. The weights are simply the initial weight and the discounting factor, respectively.