Financial Time Series

Examples in Chapter 2

Hung Chen
Department of Mathematics
National Taiwan University
3/27/2002

OUTLINE

1. Capital Asset Pricing Model
2. Chapter 2.5
   - Returns on the S & P 500
   - Modeling the UK interest rate spread
   - Modeling returns on the FTA-All share index
3. Model Selection
   - Objective
   - AIC & BIC
4. ARIMA Modelling
   - Modeling the UK interest rate spread: $ARIMA(1, 1, 0)$
   - Modeling the dollar/sterling exchange rate
   - Modeling the FTA-All share index
5. Chapter 2.8: forecasting of financial time series

Capital Asset Pricing Model(CAPM)

- The investors in stock markets expect a return better than that offered on riskless investments like government debt.
- The standard equilibrium model for stock returns is CAPM.
- r_f: the return on a risk-free asset
- r_m: the returns on the overall market portfolio
- Let r_p denote a particular small portfolio's return and the associated risk σ_p, which is measured by the standard deviation of returns.
- CAPM postulates a linear relationship between the expected risk and return of holding a portfolio of financial assets.

$$r_p - r_f = (\sigma_p/\sigma_m) \cdot (r_m - r_f) \qquad (1)$$

where σ_m is the standard deviation of the returns on the overall market portfolio.

- $r_p - r_f$: the risk premium for portfolio p

- $r_m - r_f$: the overall market's risk premium Usually, it is estimated from the historic annual or monthly return.
- Write $r_p - r_f = y$, $r_m - r_f = x$, and $\beta = \sigma_p/\sigma_m$.
  Add an intercept term α and a stochastic error term u, the CAPM becomes the simple linear regression

$$y = \alpha + \beta x + u. \tag{2}$$

  - Portfolios having $\hat{\beta}s$ in excess of unity are relatively risky
  - Portfolios having $\hat{\beta}s$ less than unity are much less sensitive to market movement

Stock Return

- The predictability of price changes on commodity and stock
- Test the hypothesis that price changes (or logarithmic price changes) are independent. This hypothesis is called the random walk model.
- The random walk model:

$$P_t = P_{t-1} + a_t \tag{3}$$

  where P_t is the price observed at the beginning of time t and a_t is an error term which has zero mean and whose values are independent of each other.
  The price change, $\Delta P_t = P_t - P_{t-1} = a_t$ is independent of past price changes.

$$P_t = \sum_{i=1}^{t} a_i$$

  The random walk model implies that prices are indeed generated by Working's cumulation of purely random changes.
- Return on a stock from t to $t+1$ is defined as the sum of the dividend yield and the

capital gain.

$$r_{t+1} = \frac{P_{t+1} + D_t - P_t}{P_t} \qquad (4)$$

where D_t is the dividend paid during period t.

- Question: Are the returns a fair game?
  Is the expected return a constant?

$$E_t(r_{t+1}) = r,$$

where $E_t(\cdot)$ is the expectation conditional on information available at time t.

Ex. 2.1: Returns on the S & P 500

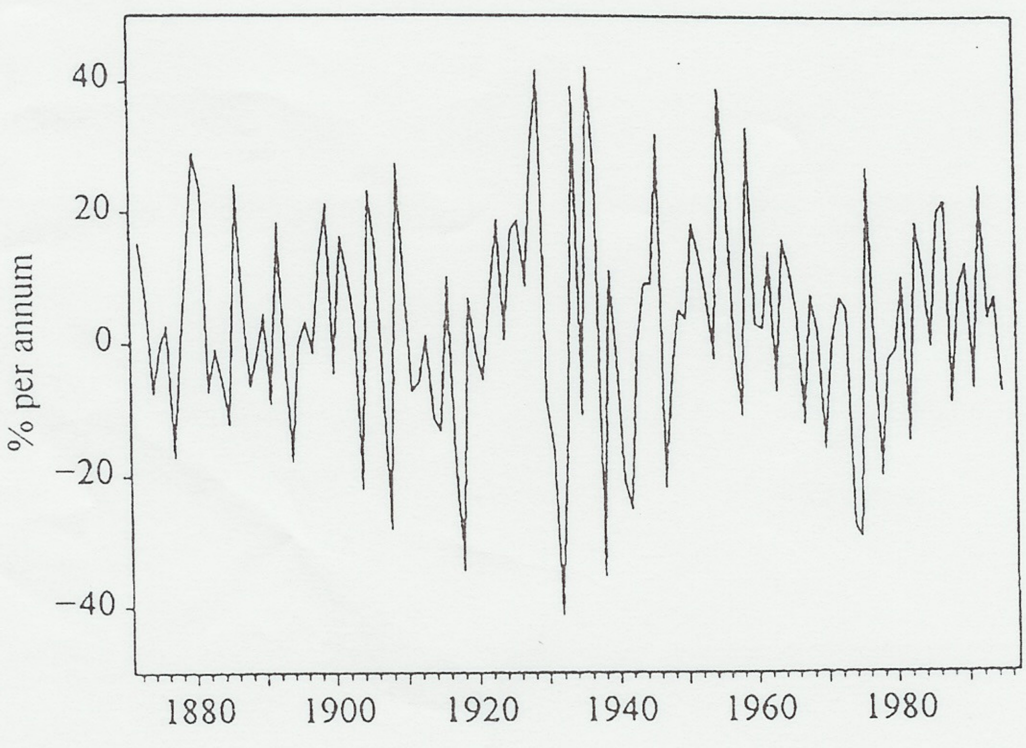
- Data: the real return on the annual Standard & Poor (S & P) 500 stock index for the US
- Figure 2.6: annual observations from 1872 to 1995
- Based on this plot, this series appears to be stationary about a constant mean, estimated to be 3.08 per cent per annum.
- Confirm by the SACF.
- Table 2.1: Give the SACF up to $k = 12$.
- None of SACF are individually significantly different from zero, thus suggesting that the series is white noise.
- Question: How do we calculate the level of significance for the above testing procedure?
- An alternative procedure on testing the hypothesis that $\{X_t\}$ is a white noise process. Proposal 1: Box and Pierce (1970)

$$Q^*(k) = T \sum_{i=1}^{k} r_i^2 \xrightarrow{L} \chi_k^2.$$

Table 2.1. *SACF of real S&P 500 returns and accompanying statistics*

k	r_k	$s.e.(r_k)$	$Q(k)$
1	0.043	0.093	0.24 [0.62]
2	−0.169	0.093	3.89 [0.14]
3	0.108	0.093	5.40 [0.14]
4	−0.057	0.094	5.83 [0.21]
5	−0.117	0.094	7.61 [0.18]
6	0.030	0.094	7.73 [0.26]
7	0.096	0.094	8.96 [0.25]
8	−0.076	0.096	9.74 [0.28]
9	−0.000	0.097	9.74 [0.37]
10	0.086	0.097	10.76 [0.38]
11	−0.038	0.099	10.96 [0.45]
12	−0.148	0.099	14.00 [0.30]

Note: Figures in [..] give $P(\chi_k^2 > Q(k))$.



Proposal 2: Ljung and Box (1978)

$$Q(k) = T(T+2)\sum_{i=1}^{k}(T-i)^{-1}r_i^2 \xrightarrow{L} \chi_k^2.$$

Purpose of Studying this Example:

- time series plot
- How do we test on a white noise process?

Programming:

- A regular time series is a sequence of observations obtained at regular intervals.
- It is characterized by four time parameters. They are (1) the time of the first observation (2) the interval between observation times, (3) the sampling rate, and (4) the time of the last observation.
- Let a be the vector containing the S&P 500 stock index from 1872 to 1995. We first use the function rts to create a regular time series object from a by

  $a.rts < -rts(a, start = 1856, deltat = 1,$
  $units = "years")$

- To get Figure 2.6, we just do the following:

$$tsplot(a.rts, xlab = \text{"S\&P 500 returns"},$$
$$ylab = \text{"\% per annum"})$$

- To estimate autocovariance, autocorrelation or partial autocorrelation, we can use the function acf. It is of the form

$$acf(a, lag.max = 12, type = \text{"}correlation\text{"},$$
$$plot = T).$$

Ex. 2.2: Modeling the UK interest rate spread

- Data: the spread between 20 Year UK Gilts and 91 day Treasury Bills
- spread: the difference between long and short interest rates
- Question: Model the spread
- Figure 2.7: monthly observations for the period from 1952 to 1995 ($T = 526$)
- Based on this plot, the spread seems to be considerably smoother than one would expect if it was a realization from a white-noise process.
- Table 2.2: Give the SACF and SPACF up to $k = 12$.
- SACF: all of whose values are positive and significant
- SPACF: $\hat{\phi}_{11}$ and $\hat{\phi}_{22}$ significant
- In practice, the order p of an AR time series is unknown. It must be specified using the data. Here we describe the first approach by using PACF.

- Tentative model: an $AR(2)$ process.
  Use OLS regression technique, it leads to

$$\begin{aligned} x_t &= 0.045(\pm.023) + 1.182(\pm.043)x_{t-1} \\ &\quad -0.219(\pm.043)x_{t-2} + \hat{a}_t \\ \hat{\sigma} &= 0.448 \end{aligned}$$

- Here we assume that SPACF of an $AR(p)$ process has the following nice properties:
  - $\hat{\phi}_{p,p}$ converges to ϕ_p as the sample size T goes to infinity.
  - $\hat{\phi}_{\ell,\ell}$ converges to zero for all $\ell > p$.
  - The asymptotic variance of $\hat{\phi}_{\ell,\ell}$ is T^{-1} for $\ell > p$.

Model Checking:

- Approach 1: If the model is adequate, then the residual series should behave as a white noise.
  Are the residuals $\hat{a}_t$ behave as white noise?
  Use either Q or Q^*. (df$= k - p - q$)
  Conclusion: No evidence of model inadequacy

Table 2.2. *SACF and SPACF of the UK spread*

k	r_k	$s.e.(r_k)$	$\hat{\phi}_{kk}$	$s.e.(\hat{\phi}_{kk})$
1	0.969	0.044	0.969	0.044
2	0.927	0.075	−0.217	0.044
3	0.884	0.094	0.011	0.044
4	0.844	0.109	0.028	0.044
5	0.803	0.121	−0.057	0.044
6	0.761	0.131	−0.041	0.044
7	0.719	0.139	−0.007	0.044
8	0.678	0.146	−0.004	0.044
9	0.643	0.152	0.057	0.044
10	0.613	0.157	0.037	0.044
11	0.586	0.162	0.008	0.044
12	0.560	0.166	−0.020	0.044

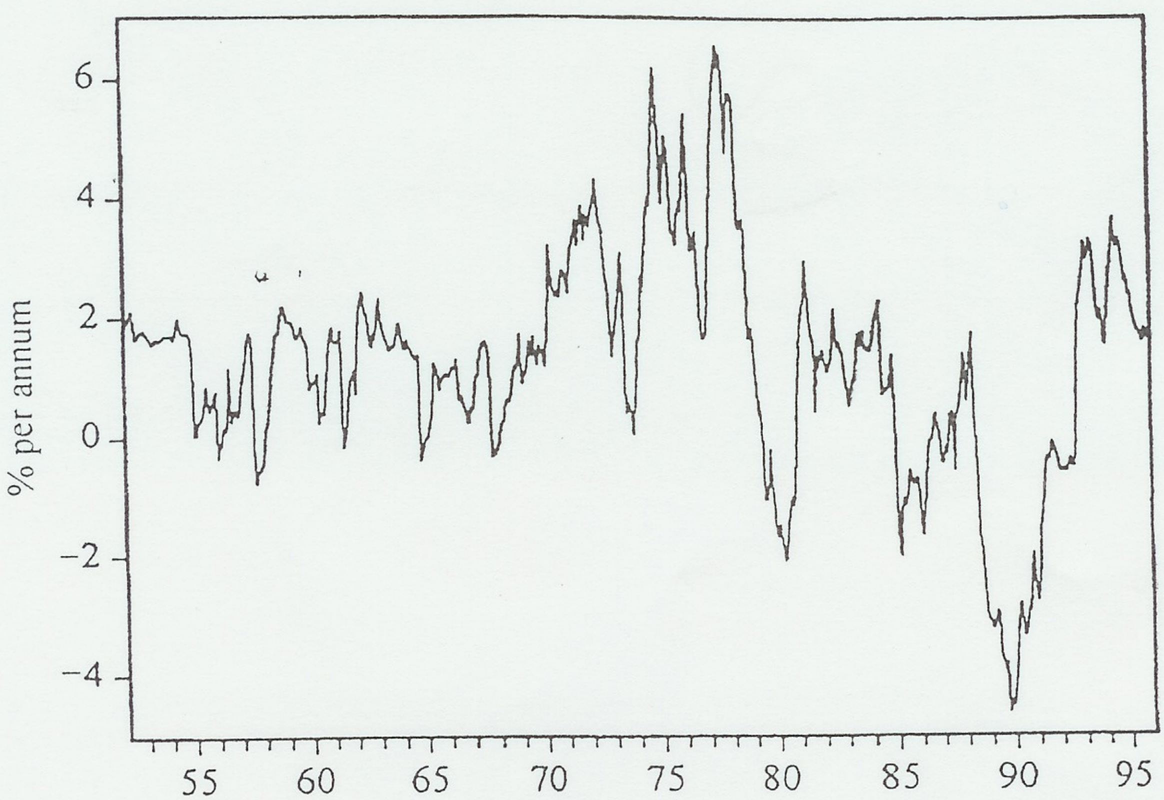


Figure 2.7 UK interest rate spread (monthly 1952.03–1995.12)

- For an $AR(p)$ model, the Ljung-Box statistic $Q(m)$ follows asymptotically a chi-square distribution with $m-p$ degrees of freedom.

- Approach 2: overfitting
  Construct Nested Models:
  Here we consider $AR(3)$ process or $ARMA(2,1)$ process.

$$
\begin{aligned}
x_t &= 0.044(\pm.023) + 1.185(\pm.044)x_{t-1} \\
&\quad -0.235(\pm.067)x_{t-2} + 0.013(\pm.044)x_{t-3} \\
&\quad +\hat{a}_t, \\
\hat{\sigma} &= 0.449 \\
x_t &= 0.046(\pm.025) + 1.137(\pm.196)x_{t-1} \\
&\quad -0.175(\pm.191)x_{t-2} + \hat{a}_t \\
&\quad +0.048(\pm.199)\hat{a}_{t-1}, \\
\hat{\sigma} &= 0.449
\end{aligned}
$$

In both models, the additional parameter is insignificant.
Confirm the adequacy an $AR(2)$ process.

Purpose of the Study:

- Familiarity of AR and ARMA

- Nested Model and Model Checking

- Unit Root
  - fitted mean: $\hat{\mu} = \hat{\theta}_0/(1 - \hat{\phi}_1 - \hat{\phi}_2) = 1.204$ with SE$= 0.529$
  - check stationarity:

    $$\hat{\phi}_1 + \hat{\phi}_2 = 0.963, -\hat{\phi}_1 + \hat{\phi}_2 = -1.402,$$

    and $-\hat{\phi}_2 = -0.219$.
  - roots: $\hat{g}_1 = 0.95$ and $\hat{g}_2 = 0.23$
  - closeness of $\hat{g}_1$ to unity?
- To be continued in Ex. 2.4, 3.1

Ex. 2.3: Modeling returns on the FTA-All share index

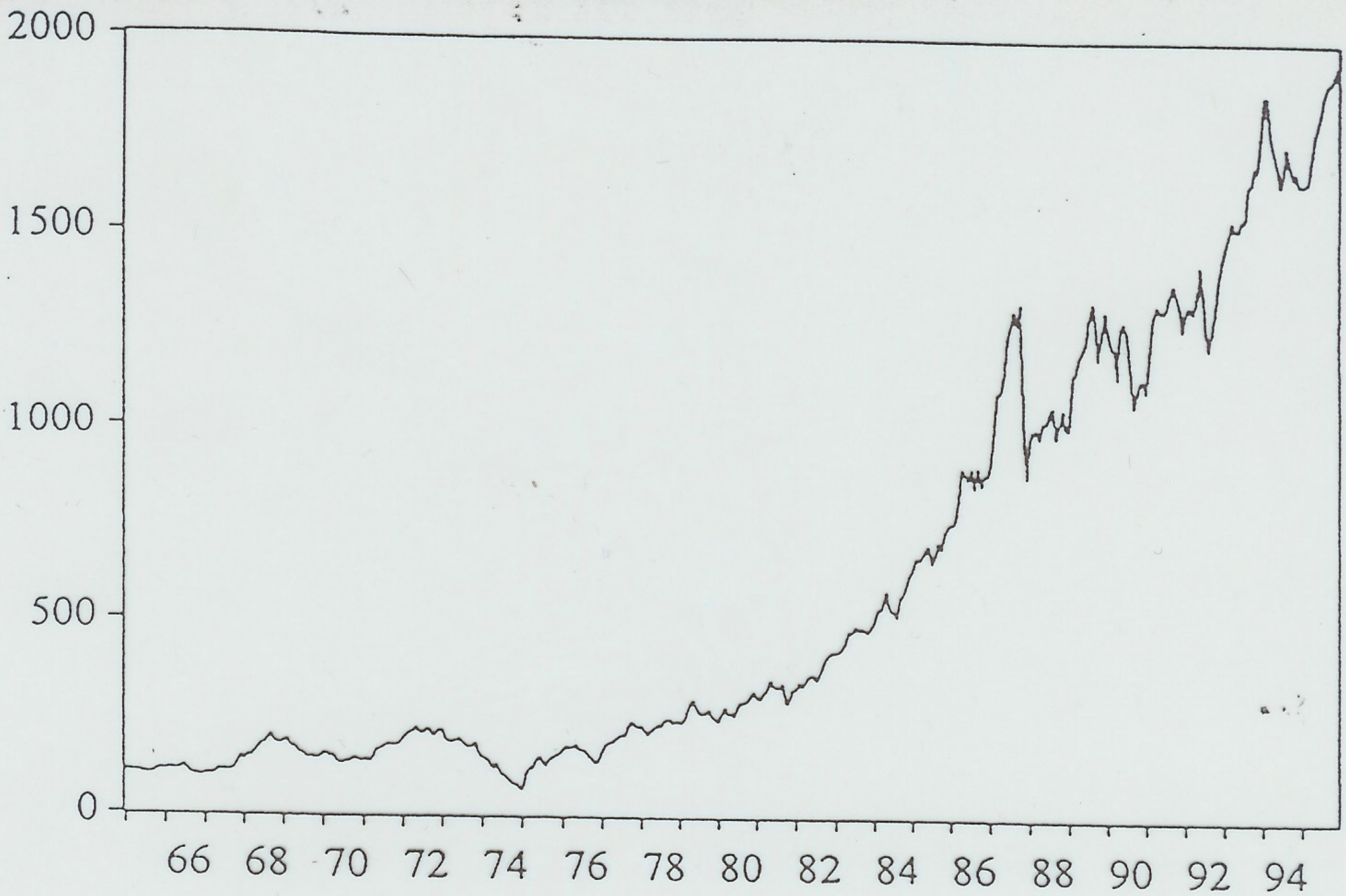
- Data: the broadest-based stock index in the UK:
  Financial Times Actuaries (FTA) All Share Index
- Question: Model its return
- Figure 2.14: monthly observations from January 1965 to December 1995 ($T = 371$)
- Figure 3.4: monthly dividend yield (D/P) from 1965 to 1995
- Based on the plot in Figure 2.14, this series exhibit a prominent upward, but not linear, trend, with pronounced and persistent fluctuations about it, which increase in variability as the level of the series increases.
- Use logarithmic transformation and the transformed observations are shown in Figure 2.14.
- Effect of taking logarithms: linearize the trend and stabilize the variance

Fit the ARMA model to the data

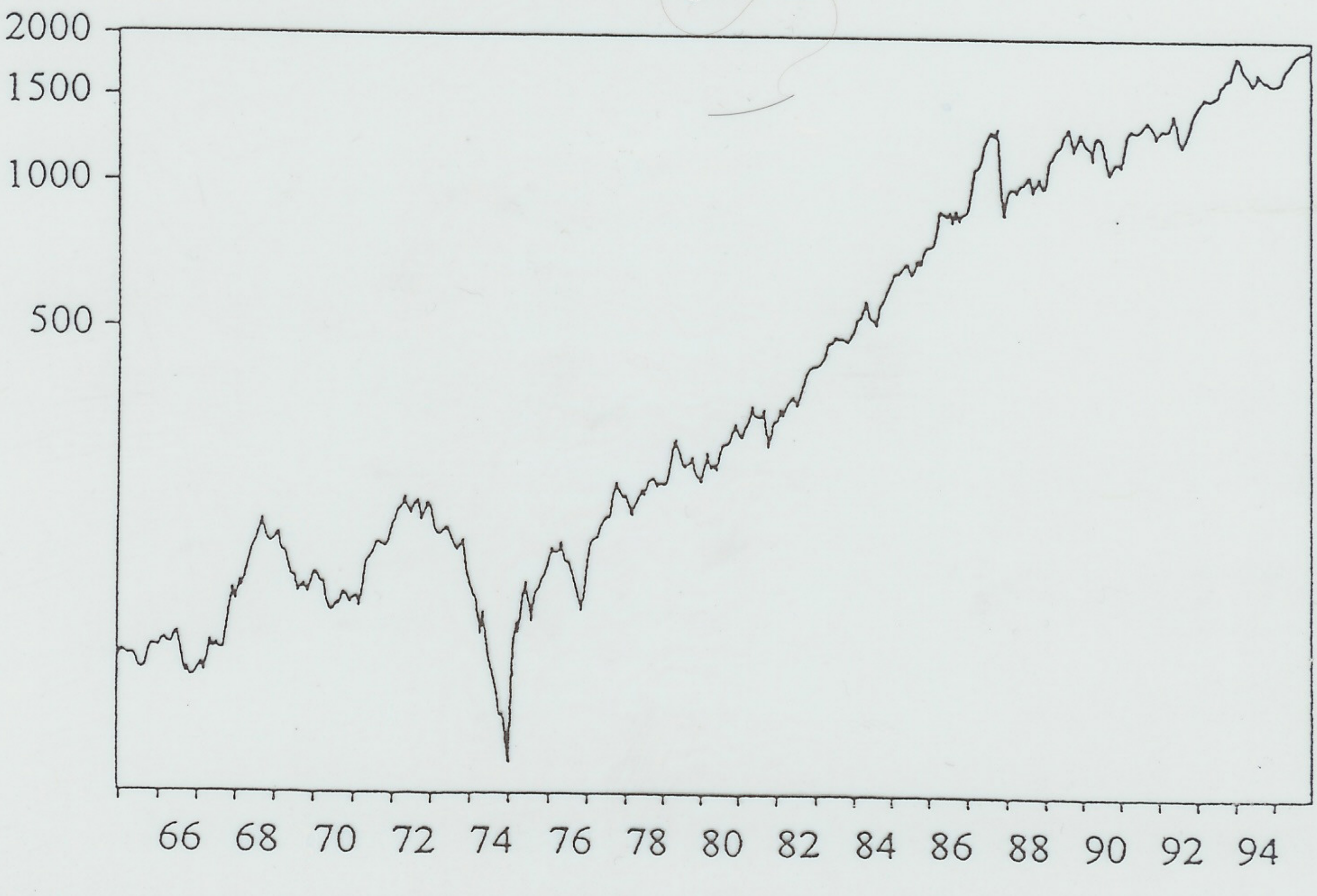
- $Q(12) = 26.4$, significance level: 0.009
- Table 2.3: Give the SACF and SPACF up to $k = 12$.
- Both r_k and $\hat{\phi}_{kk}$ at lags $k = 1$ and 2 are greater than two standard errors.
- THe ACF and PACF are not informative in determining the order of an ARMA model.
- What can be done now?
  Try the ARMA process.
- Based on the plot, we actually use a stationary process to model a non-stationary series. What is our best strategy without knowing the non-stationarity?
- Choose an appropriate model among the ARMA based on certain objective.
  - Now we introduce the second approach for model selection using information criterion function.
  - Consider Akaike Information Criterion (Akaike, 1974)

    $$AIC = \frac{-2}{T}\ln(\text{likelihood}) + \frac{2}{T}\times(\text{number of parameters})$$

Levels



Logarithms



where the likelihood function is evaluated at the MLE and T is the sample size. The second term of AIC is called the penalty function of the criterion.

– For a Gaussian $AR(\ell)$ model, AIC reduces to

$$AIC(\ell) = \ln(\hat{\sigma}_\ell^2) + \frac{2\ell}{T}$$

where $\hat{\sigma}_\ell^2$ is the MLE of σ_a^2.

- Specification of possible models: $\bar{p} = \bar{q} = 3$ based on the SACF and SPACF.

- AIC selects $ARMA(2,2)$:

$$\begin{aligned} x_t &= 1.57(\pm.10) - 1.054(\pm.059)x_{t-1} \\ &\quad -0.822(\pm.056)x_{t-2} + \hat{a}_t \\ &\quad +1.204(\pm.049)\hat{a}_{t-1} + 0.895(\pm.044)\hat{a}_{t-2}, \\ \hat{\sigma} &= 5.89 \end{aligned}$$

- BIC selects $MA(1)$:

$$\begin{aligned} x_t &= 0.55(\pm.04) + \hat{a}_t + 0.195(\pm.051)\hat{a}_{t-1}, \\ \hat{\sigma} &= 5.98 \end{aligned}$$

- Table 2.4: Give AIC and BIC for all possible models.

Model Selection

- Possible objectives:
  - Fit data well.
    Dubious approach: too many parameters
  - Find a true model. (Require consistency.)
    Drawback: What are we doing if the probability model underlying the series is not ARMA?
  - Find the best ARMA model for future prediction.
    Problem: Does it fit to our need?
- Akaike's (1974) Information Criteria:

$$AIC(p,q) = \log \hat{\sigma}^2 + 2(p+q)T^{-1}$$

- Schwarz's (1978) Bayesian Information Criteria

$$BIC(p,q) = \log \hat{\sigma}^2 + (p+q)T^{-1} \log T$$

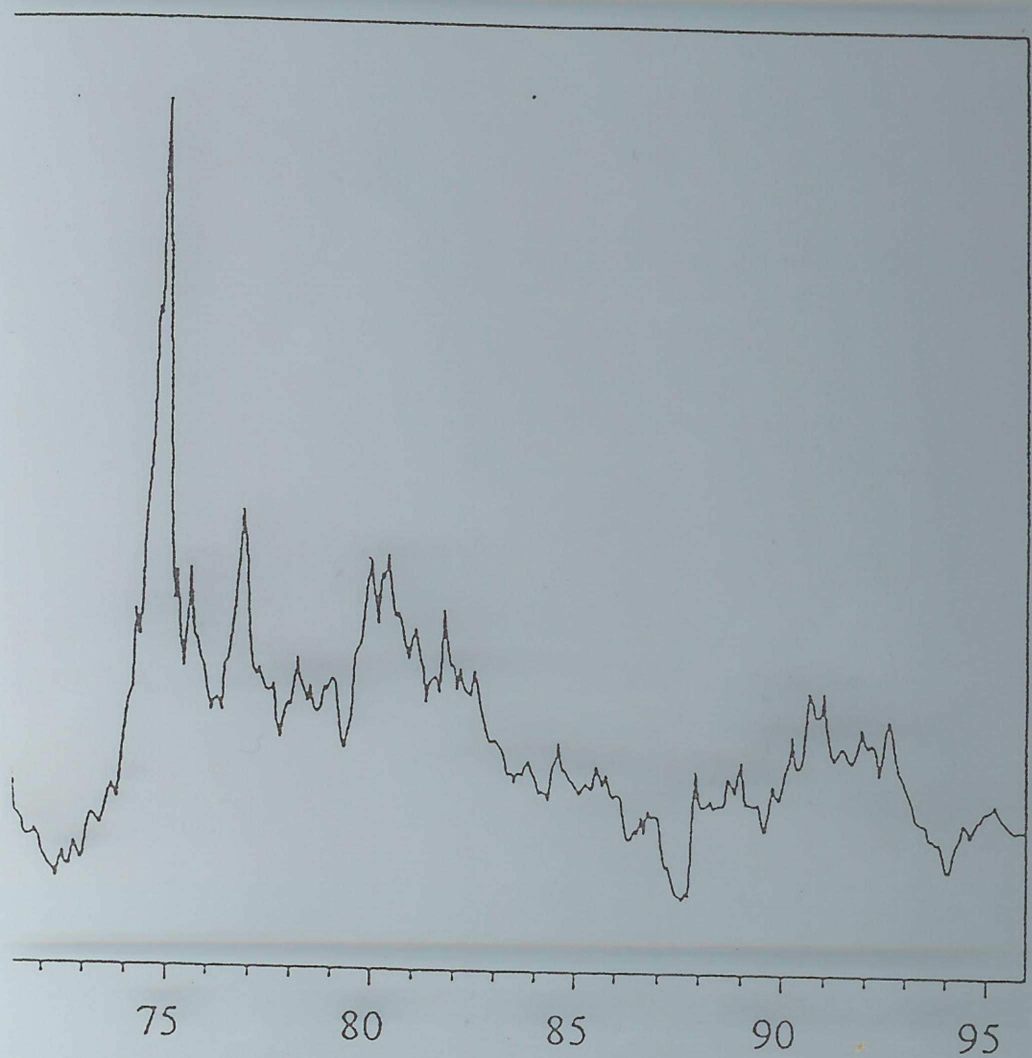
- General Criteria: estimated error variance $\hat{\sigma}^2$ plus a penalty adjustment involving the number of estimated parameters.
  - Usually, $\hat{\sigma}^2$ decreases as (p,q) increases.

Table 2.3. *SACF and SPACF of* FTA All Share *nominal returns*

k	r_k	$s.e.(r_k)$	$\hat{\phi}_{kk}$	$s.e.(\hat{\phi}_{kk})$
1	0.153	0.052	0.153	0.052
2	−0.068	0.053	−0.094	0.052
3	0.109	0.053	0.139	0.052
4	0.093	0.054	0.046	0.052
5	−0.066	0.054	−0.072	0.052
6	−0.017	0.054	−0.006	0.052
7	0.056	0.054	0.030	0.052
8	−0.022	0.054	−0.030	0.052
9	0.101	0.055	0.138	0.052
10	0.044	0.055	−0.017	0.052
11	−0.025	0.055	−0.014	0.052
12	0.019	0.055	0.017	0.052

Table 2.4. *Model selection criteria for nominal returns*

	q p	0	1	2	3
	0	−5.605	−5.629	−5.632	−5.633
AIC	1	−5.621	−5.631	−5.626	−5.629
	2	−5.622	−5.624	−5.649	−5.647
	3	−5.634	−5.629	−5.629	−5.646
	0	−5.594	−5.608	−5.601	−5.590
BIC	1	−5.600	−5.599	−5.584	−5.576
	2	−5.591	−5.582	−5.596	−5.583
	3	−5.591	−5.576	−5.565	−5.571



All Share dividend yield (monthly 1965–95)

– Penalty term will increase as (p, q) increases.
$AIC(p, q)$ uses $2(p+q)T^{-1}$ while $BIC(p, q)$ uses $(p + q)T^{-1} \log T$.

– Larger model has better chance with smaller model bias but the uncertainty of estimates of unknown parameters will increase. How do we balance them?

To be continued in Ex. 2.6, 3.2.

ARIMA modelling

Ex. 2.4: Modeling the UK interest rate spread

- Model it by an $I(1)$ process.
  Assume the first difference $w_t = \triangle x_t$ is stationary.
- Examine the behavior of the SACF and SPACF.
- Table 2.5 provides these estimates up to $k = 12$.
- Conclusion: Try an $AR(1)$ or an $MA(1)$ since both SACF and SPACF cut-off at $k = 1$.
- Try $AR(1)$:

$$\begin{aligned} w_t &= -0.0002(\pm.0198) + 0.201(\pm.043)w_{t-1} \\ &\quad +\hat{a}_t, \\ \hat{\sigma} &= 0.453 \end{aligned}$$

The residuals are effectively white noise ($Q(12) = 9.03$) and the mean of w_t is not significantly different from zero.
Model the spread as an $ARIMA(1,1,0)$ without drift.

SACF and SPACF of the first difference of the UK spread

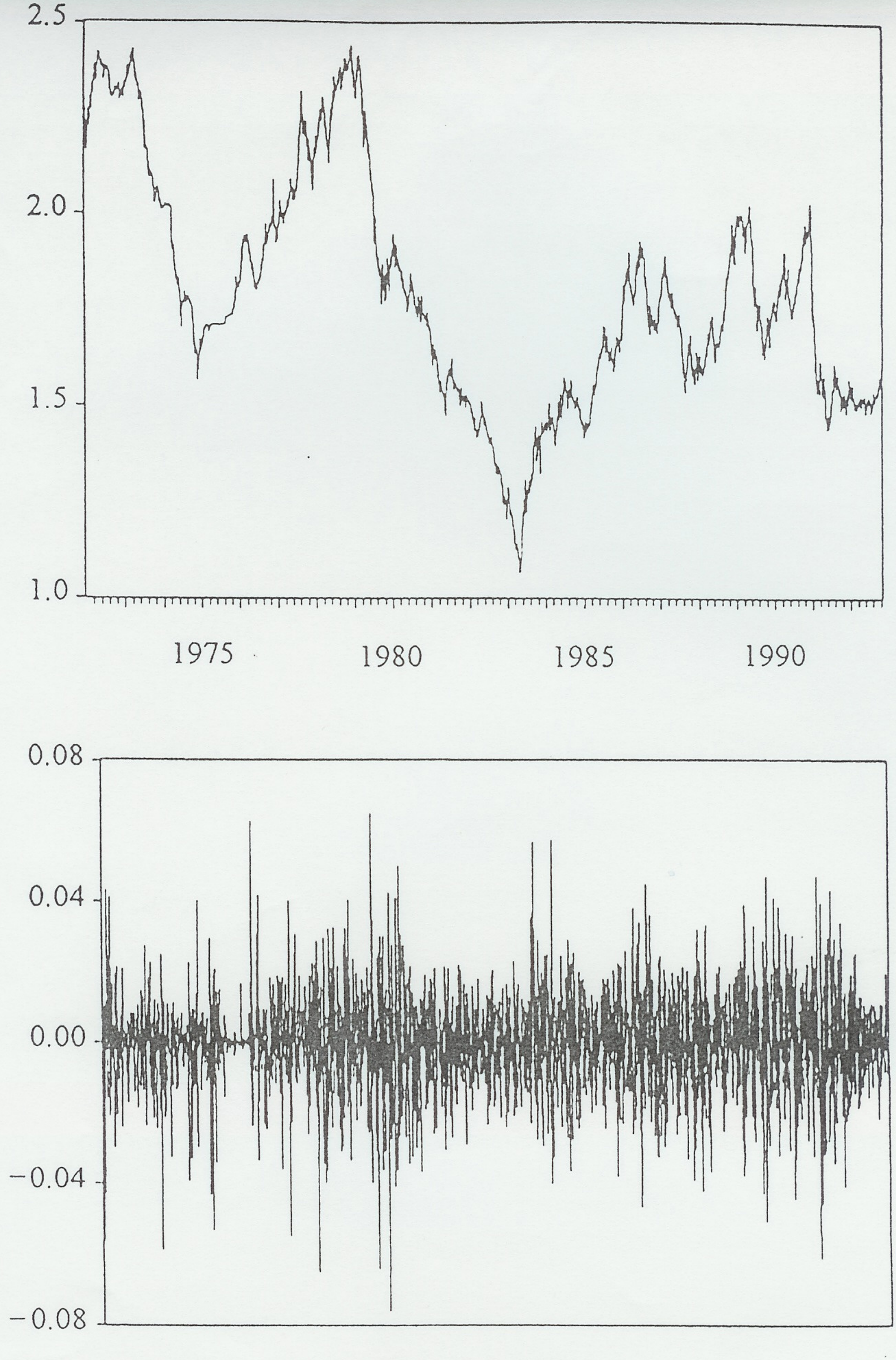
r_k	$s.e.(r_k)$	$\hat{\phi}_{kk}$	$s.e.(\hat{\phi}_{kk})$
0.201	0.044	0.201	0.044
0.006	0.045	−0.036	0.044
−0.053	0.045	−0.048	0.044
0.014	0.045	0.036	0.044
0.028	0.045	0.018	0.044
−0.006	0.045	−0.019	0.044
−0.028	0.045	−0.021	0.044
−0.088	0.046	−0.079	0.044
−0.087	0.046	−0.059	0.044
−0.049	0.046	−0.025	0.044
−0.006	0.046	0.000	0.044
0.017	0.046	0.015	0.044

- Try an $MA(1)$. (almost identical estimates) The ACF is useful in identifying the order of an MA model.
- Maximum likelihood estimation is commonly used to estimate an MA model. There are two approaches to evaluate the likelihood function of an MA model.
  - Approach 1: Assume that the initial shocks, i.e. a_t for $t \leq 0$, are zero. As such the shocks are computed recursively from the model, starting with $a_1 = r_1 - c_0$. Parameter estimates obtained by this approach are called the conditional maximum likelihood estimates.
  - Approach 2: Treat the initial shocks a_t, $t \leq 0$, as additional parameters of the model and estimate them jointly with other parameters. This approach is referred to as the exact likelihood method

Ex. 2.5: Modeling the dollar/sterling exchange rate

- Figure 13: Plots of daily observations of both the level and first differences of the dollar/sterling exchange rate from January 1974 to December 1994 (5192 observations).
- The SACF of the levels displays the slow, almost linear, decline. ($r_1 = 0.999$, $r_{10} = 0.989$, $r_{20} = 0.933$ and $r_{100} = 0.850$.) Model it by an $I(1)$ process.
- The differences are stationary.
- Examine the behavior of the SACF and SPACF. The only significant sample autocorrelation is $r_1 = 0.071$.
- Conclusion: Try an $AR(1)$ or an $MA(1)$ since both SACF and SPACF cut-off at $k = 1$.
  The R^2 statistic was less than 0.005.
  Refer to Ex 3.1 and 4.2 later on.

Levels



Differences

Figure 2.13 Dollar/sterling exchange rate (daily 1974–1994)

Ex. 2.6: Modeling the FTA-All share index

- Figure 14: Suggest a logarithmic transformation.
- Table 2.6: SACF and SPACF of the first difference
  The $\hat{\psi}_{kk}$s suggest an $AR(3)$ process.
- Fitted model:

$$\begin{aligned}
\triangle x_t &= 0.0069(\pm .0032) + 0.152(\pm .052)\triangle x_{t-1} \\
&\quad -0.140(\pm .052)\triangle x_{t-2} \\
&\quad +0.114(\pm .052)\triangle x_{t-3} + \hat{a}_t, \\
\hat{\sigma} &= 0.0603
\end{aligned}$$

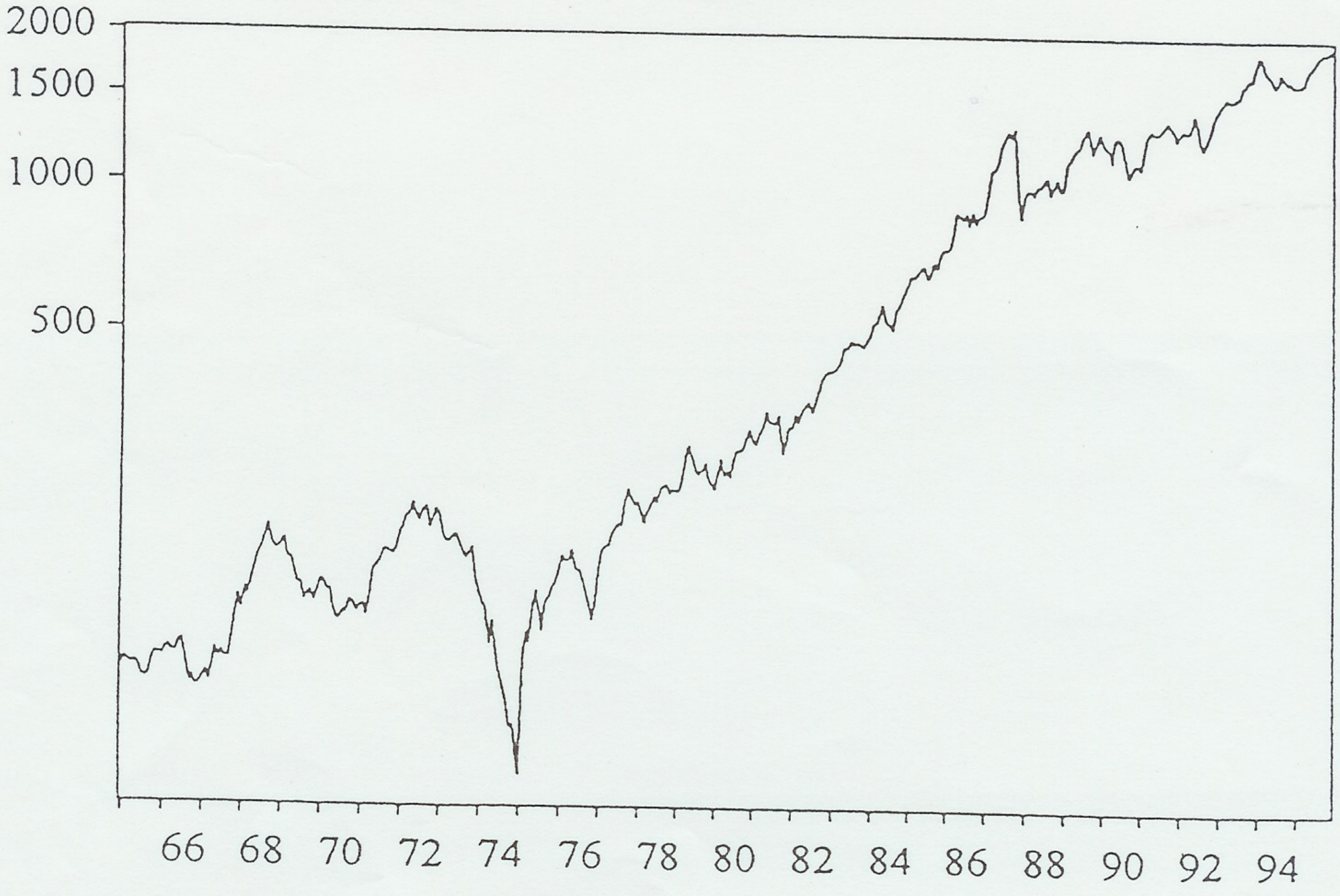
  Hence, $\hat{\mu} = 0.079$.

- $\triangle x_t$: the monthly growth of the index. (Annual mean growth rate: 9.5%.
  $x_t = \log(P_t)$ where P_t is the level of the index.
- Example 2.3 discusses the nominal return on the index. Consider

$$\begin{aligned}
r_t &= \frac{P_t + D_t - P_{t-1}}{P_{t-1}} \approx \log\left(\frac{P_t + D_t}{P_{t-1}}\right) \\
&= \log\left(\frac{P_t}{P_{t-1}}\right) + \log\left(1 + \frac{D_t}{P_t}\right) \approx \triangle x_t + \frac{D_t}{P_t}.
\end{aligned}$$

Levels

Logarithms



The nominal return is equal to the growth of the index plus the dividend yield.

- Facts:
  dividend yield: $ARMA(1, 3)$
  $\triangle x_t$: $AR(3)$
  nominal return: $ARMA(2, 2)$
  Granger and Morris (1976) discussed the sum of two independent $ARMA$ processes.

Ex. 2.7: ARIMA forecasting of financial time series

UK interest spread (Example 2.2)

- $AR(2)$ model:

$$\begin{aligned} x_t &= 0.045(\pm.023) + 1.182(\pm.043)x_{t-1} \\ &\quad -0.219(\pm.043)x_{t-2} + \hat{a}_t \\ \hat{\sigma} &= 0.448 \end{aligned}$$

- Last two observations: $x_{T-1} = 1.63$ and $x_T = 1.72$

- Forecasts:

$$\begin{aligned} f_{T,1} &= 1.182x_T - 0.219x_{T-1} = 1.676 \\ f_{T,2} &= 1.182f_{T,1} - 0.219x_T = 1.604 \\ f_{T,3} &= 1.182f_{T,2} - 0.219f_{T,1} = 1.529 \\ \cdots &= \cdots. \end{aligned}$$

  It can be shown that the forecasts eventually tend to 1.216, the sample mean of the spread.

- The ϕ-weight:

$$\begin{aligned} \phi_1 &= \psi_1 = 1.182 \\ \phi_2 &= \psi_1^2 + \psi_2 = 1.178 \end{aligned}$$

$$
\begin{aligned}
\phi_3 &= \psi_1^3 + 2\psi_1\psi_2 = 1.134 \\
\phi_4 &= \psi_1^4 + 3\psi_1^2\psi_2 + \psi_2^2 = 1.082
\end{aligned}
$$

The forecast error variances are

$$
\begin{aligned}
V(e_{T,1}) &= 0.448^2 = 0.201 \\
V(e_{T,2}) &= 0.448^2(1 + 1.182^2) = 0.482 \\
V(e_{T,3}) &= 0.448^2(1 + 1.182^2 + 1.178^2) = 0.761 \\
V(e_{T,4}) &= 0.448^2(1 + 1.182^2 + 1.178^2 \\
&\quad +1.134^2) = 1.019
\end{aligned}
$$

The forecast error variances converges to the sample variance of the spread, 3.53.

$ARIMA(0,1,1)$ model:

- $\hat{\theta} = 0.2$ and $\hat{\sigma} = 0.452$.
- Last observation: $x_T = 1.72$
- Final residual: $\hat{a}_T = 0.136$
- Forecasts:

$$
\begin{aligned}
f_{T,1} &= 1.72 - 0.2(0.136) = 1.693 \\
f_{T,h} &= f_{T,1} = 1.693.
\end{aligned}
$$

The forecasts will not converge to the sample mean.

- The forecast error variances are

$$\begin{aligned} V(e_{T,h}) &= 0.452^2[1 + 0.64(h-1)] \\ &= 0.204 + 0.131(h-1). \end{aligned}$$

The forecast error variances increase with h.