Lecture 5 - Forecasting with Time Series Models

Pedro Valls[1]

[1]Sao Paulo School of Economics - FGV
CEQEF-FGV

July 7, 2024

FGV EESP
CENTRO DE ESTUDOS
QUANTITATIVOS EM
ECONOMIA E FINANÇAS

Outline

Introduction

- [Box and Jenkins, 1976] popularized the use of univariate time series models for forecasting
- The key ideia is to exploit the past behavior of a time series to forecast its future values.
- The future sould be rather similar to the past - weak stationary.
- Any weakly stationary stochastic process can always be represented as an infinite sum of white noise which has zero mean, is uncorrelated over time and has a constant variance. - Wold decomposition theorem .
- The Wold decomposition can be approximated by another one where the variable of interest depends on a finite number of its own lags, possibly combined with a finite number of lags of an uncorrelated and homoskedastic process - ARMA representation.

Representation

- A time series process is strictly stationary when

$$D\{y_t, \cdots, y_{t+T}\} = D\{y_{t+k}, \cdots, y_{t+T+k}\} \quad \text{for all } t, T \text{ and } k \tag{1}$$

- where $D(\cdot)$ indicates the joint density.
- The time series weakly stationary if

$$\begin{aligned} E(y_t) &= E(y_{t+h}) \quad \text{for all } t, k \\ Var(y_t) &= Var(y_{t+h}) \quad \text{for all } t, k \\ Cov(y_t, y_{t-m}) &= Cov(y_{t+k}, y_{t-m+k}) \quad \text{for all } t, k, m \end{aligned} \tag{2}$$

Representation

- A weakly stationary process can be represented as

$$
\begin{aligned}
y_t &= \varepsilon_t + c_1\varepsilon_{t-1} + c_2\varepsilon_{t-2} + \cdots \\
&= \sum_{i=0}^{\infty} c_i\varepsilon_{t-i} = \sum_{i=0}^{\infty} c_i L^i \varepsilon_t \\
&= c(L)\varepsilon_t \qquad (3)
\end{aligned}
$$

- where L is the lag operator: $L\varepsilon_t = \varepsilon_{t-1}$ and $L^i\varepsilon_t = \varepsilon_{t-i}$, $c_0 = 1$ and the error $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$
- The representation (3) is knwon as Wold decomposition

Representation

- Model (3) has an infinite number of parameters.
- However we can approximate $c(L)$ via a ratio of two finite polynomials

$$c(L) = \frac{\psi(L)}{\phi(L)} \tag{4}$$

- where $\psi(L) = 1 - \psi_1 L - \psi_2 L^2 - \cdots - \psi_q L^q$ and $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$
- Because it is weakly stationary all the roots of $\phi(L)$ are outside the unit circle and we can rewrite (3) as

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \psi_1 \varepsilon_{t-1} - \cdots - \psi_q \varepsilon_{t-q} \tag{5}$$

- this is a $ARMA(p,q)$ representation

Representation

- Can use Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to identify the order of AR and MA
- The ACF is defined as

$$ACF(k) = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t)}\sqrt{Var(y_{t-k})}} = \frac{\gamma(k)}{\gamma(0)} \tag{6}$$

- And the PACF as specific coefficients in the following regression
  - PACF(1): coefficient of y_{t-1} in the regression of y_t on y_{t-1}
  - PACF(2): coefficient of y_{t-2} in the regression of y_t on y_{t-1}, y_{t-2}
  - $\vdots$
  - PACF(k): coefficient of y_{t-k} in the regression of y_t on $y_{t-1}, y_{t-2}, \cdots, y_{t-k}$

Autoregressive processes

- Assuming that $c(L)$ in (3) is invertible, i.e. has all the roots outside the unit circle, then (3) can rewrite as

$$y_t = \sum_{j=1}^{\infty} \phi_j y_{t-j} + \varepsilon_t \tag{7}$$

- assuming that the process is weakly stationary (7) can be approximate by a finite order, for example order p as

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{8}$$

- The $AR(p)$ can be inverted as a $MA(\infty)$
- The ACF for an $AR(1)$ decays geometrically if $\phi_1 > 0$ and it is a damped sine if $\phi_1 < 0$.
- the PACF for an $AR(1)$ is zero after the firat lag.
- Properties of $AR(p)$ for $p \geq 2$ can be seen in my lectures notes for Econometrics III.

Moving average processes

- The $q-$th order moving average process is defined as

$$y_t = \varepsilon_t - \psi_1 \varepsilon_{t-1} - \cdots - \psi_q \varepsilon_{t-q} \tag{9}$$

- and this process is always weakly stationary and the ACF is zero after lag q.
- When the MA is invertible, that is all roots of the MA polynomial are outside the unit circle, it can be written as an $AR(\infty)$, therefore the PACF decays geometrically
- Properties of $MA(q)$ for $q \geq 2$ can be seen in my lectures notes for Time Series Econometrics Lecture Notes.

Integrated processes

- An integrated process y_t is a non stationary process such that $(1-L)^d$ is stationary.
- d is the order of integration and is denoted $I(d)$
- The most common integrated process is the Random Walk (RW)

$$y_t = y_{t-1} + \varepsilon_t \tag{10}$$

- and (10) can be written as

$$(1-L)y_t = \Delta y_t = \varepsilon_t \Longrightarrow y_t = \frac{1}{1-L}\varepsilon_t = \varepsilon_t + \varepsilon_{t-1} + \cdots \tag{11}$$

- the effect of a shock do not decay over time.

ARIMA Processes

- An $ARIMA(p, d, q)$ process is given by

$$\phi(L)\Delta^d y_t = \psi(L)\varepsilon_t \tag{12}$$

- where $\Delta^d = (1 - L)^d$
- and $\phi(L)$ and $\psi(L)$ satisfy the stationarity and invertibility conditions, so that the ARMA model for $w_t = (1 - L)^d y_t$ is stationary and invertible.

Model Specification

- To determine d, p and q use ACF and PACF and also unit root tests.
- In order to test sample autocorrelation used Box-Pierce or Ljung-Box statistics to check if the autocorrelations and partial autocorrelations are zero.
- Testing for ARCH can use Box-Pierce or Ljung-Box in the square residuals.
- Also can use Information Criteria to determine the order of the *ARMA* model

Estimation

- The objective function to be minimized is the usual resdidulas sum of squared.
- MLE can also be used when it is assumed normality, Student t, GED, and Student-t Skewed for the errors. It is possible to obtain exact MLE for AR, MA and ARMA models.

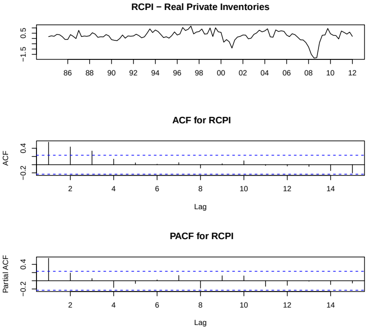
Diagnostic checking

- Test residuals for non serial correlation, homoscedasticity and parameter stability.
- The parameter stability can be tested using RLS

Modeling US inventory

- Modeling Quarterly time series of the change in real private inventories (RCPI) in US using data for the period 1985-2012
- The following comand in R are used
- Modeling US Inventories

Modeling US inventory



Modeling US inventory I

- ADF test for the entire sample
- The following comand in R are used
- Modeling US Inventories

  The results are:

- `Unit Root Test for entire sample`
- `Augmented Dickey-Fuller Test`
- `data:  arima.inven[, rcpi]`
- `Dickey-Fuller = -3.8792, Lag order = 1, p-value = 0.01762`
- `alternative hypothesis:  stationary`
- `Unit Root Test for estimation sample up to 2002Q4`
- `Augmented Dickey-Fuller Test`

Modeling US inventory II

- `data:  arima.inven[, rcpi[1:72]]`
- `Dickey-Fuller = -3.2235, Lag order = 1, p-value = 0.09099`

Modeling US inventory - best ARMA using BIC estimation period

	MA0	MA1	MA2	AR
1	56.28	43.55	43.21	0
2	**34.99**	36.91	40.41	1
3	36.72	40.90	43.44	2
4	40.75	43.78	41.05	3
5	43.08	47.07	45.30	4
6	46.88	51.14	49.34	5
7	51.12	55.20	53.57	6
8	53.83	56.35	52.83	7
9	55.71	59.56	56.82	8
10	58.70	62.47	63.62	9
11	61.17	65.06	66.81	10
12	64.18	66.58	68.07	11

Modeling US inventory - best ARMA using AIC estimation period

	MA0	MA1	MA2	AR
1	60.28	45.29	42.68	0
2	36.73	36.39	37.62	1
3	36.19	38.11	38.39	2
4	37.97	38.72	**33.74**	3
5	38.03	39.76	35.73	4
6	39.56	41.56	37.51	5
7	41.55	43.36	39.47	6
8	41.99	42.25	36.46	7
9	41.61	43.19	38.19	8
10	42.34	43.84	42.74	9
11	42.55	44.17	43.66	10
12	43.29	43.43	42.66	11

Modelling US Inventories, best model using AIC, ARMA(3,2)

```
Call:
arima(x = arima.inven[1:72, rcpi], order = c(3, 0, 2))
```

Coefficients:

	ar1	ar2	ar3	ma1	ma2	intercept
	0.2665	-0.6047	0.5558	0.1910	1.000	0.3114
s.e.	0.1023	0.0813	0.1035	0.0464	0.067	0.0862

sigma^2 estimated as 0.07185 loglikelihood = -9.87 aic = 33.74

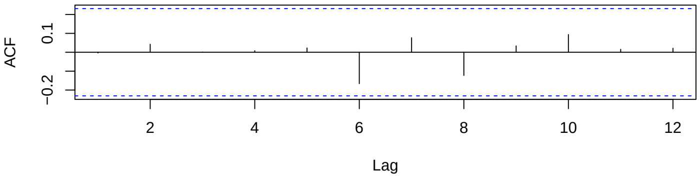
Where the roots for the AR polynomial are $\phi_1 = 0.65$, $\phi_2 = -0.19 + 0.90i$, and $\phi_3 = -0.19 - 0.90i$ and for MA polynomial are $\theta_1 = -0.10 + 1.00i$, and $\theta_2 = -0.10 - 1.00i$. Since ϕ_2 and ϕ_3 are close to θ_1 and θ_2 it is better to use the model chosen by BIC, i.e. $AR(1)$.

Modelling US Inventories, best model using BIC, AR(1)

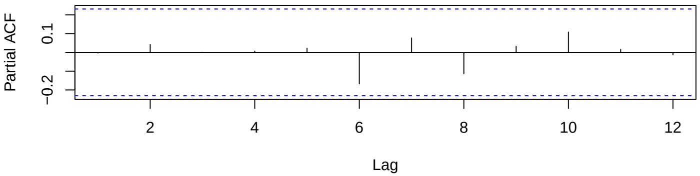
```
Call:
arima(x = arima.inven[1:72, rcpi], order = c(1, 0, 1))

Coefficients:
         ar1   intercept
      0.5428      0.3071
s.e.  0.0973      0.0758

sigma^2 estimated as 0.08928    loglikelihood = -15.36    aic = 36.71
```

Diagnostic Checking - ACF and PACF for residuals

ACF for Residual AR(1) for RCPI



PACF for Residual AR(1) for RCPI



Diagnostic Checking - Breusch-Godfrey serial correlation LM test

- `arima.inven[1:72, eps := as.numeric(arma.fit$rcpi$residuals)]`
- `arima.inven$eps1 <- lag(arima.inven$eps, 1)`
- `arima.inven$eps2 <- lag(arima.inven$eps, 2)`
- `summary(lm(eps   eps1 + eps2, data = arima.inven[3:72]))`

Diagnostic Checking - Breusch-Godfrey serial correlation LM test

The results are:

Table: Breusch-Godfrey serial correlation LM test

	Dependent variable:
	eps
eps1	-0.085
	(0.122)
eps2	0.108
	(0.121)
Constant	0.003
	(0.037)
Observations	70
R^2	0.021
Adjusted R^2	-0.008
Residual Std. Error	0.306 (df = 67)
F Statistic	0.712 [0.4943] (df = 2; 67)
Note:	*p<0.1; **p<0.05; ***p<0.01

Diagnostic Checking - Jarque Bera test

- `jarque.bera.test(arma.fit$rcpi$residuals)`

The results are:

Jarque Bera -test;	Null hypothesis:	Normality
data: arma.fit$rcpi$residuals		
Chi-squared = 2.945	df = 2	p-value =0.229

Diagnostic Checking - White test

- `white.test(arma.fit$rcpi$residuals))`

The results are:

White -test;	Null hypothesis:	Homoscedasticity
data: arma.fit$rcpi$residuals		
Chi-squared = 2.775	df = 2	p-value =0.250

Forecasting with known parameters

- The optimal forecast of y_{T+h} in the *MSFE* sense is

$$\widehat{y}_{T+h} = E(y_{T+h}|y_T, y_{T-1}.\cdots, y_1) \tag{13}$$

- The optimal linear forecast for $ARIMA(p, d, q)$ models which coincide with $E(y_{T+h}|y_T, y_{T-1}.\cdots, y_1)$ if we assume that $\{\varepsilon_t\}$ is normally distributed.
- We also assume that the parameters of the *ARIMA* model are known

Forecasting with known parameters - General Formula I

- Start by defining $\Delta^d y_t = \omega_t$ so that ω_t is an $ARMA(p, q)$ that is

$$\omega_t = \phi_1 \omega_{t-1} + \cdots + \phi_p \omega_{t-p} + \varepsilon_t - \psi_1 \varepsilon_{t-1} - \cdots - \psi_q \varepsilon_{t-q}$$

- The one-step ahead prediction is given by

$$\begin{aligned} \widehat{\omega}_{T+1} &= E(\omega_{T+1}|I_T) = \phi_1 \omega_T + \cdots + \phi_p \omega_{T-p+1} - \psi_1 \varepsilon_T - \\ &\cdots - \psi_q \varepsilon_{T-q+1} \end{aligned} \tag{14}$$

- Similarly

$$\begin{aligned} \widehat{\omega}_{T+2} &= E(\omega_{T+2}|I_T) = \phi_1 \widehat{\omega}_{T+1} + \cdots + \phi_p \omega_{T-p+2} - \psi_2 \varepsilon_T \\ &- \cdots - \psi_q \varepsilon_{T-q+1} \\ &\vdots \\ \widehat{\omega}_{T+h} &= E(\omega_{T+h}|I_T) = \phi_1 \widehat{\omega}_{T+h-1} + \cdots + \phi_p \widehat{\omega}_{T+h-p} - \psi_h \varepsilon_T \\ &- \cdots - \psi_q \varepsilon_{T-q+h} \end{aligned} \tag{15}$$

- where $\widehat{\omega}_{T-j} = \omega_{T-j}$ if $j \leq 0$ and there is no MA component for $h > q$.
- if $d = 1$ we have

$$
\begin{aligned}
\widehat{\omega}_{T+1} &= \widehat{y}_{T+1} - y_T \Longrightarrow \widehat{y}_{T+1} = y_T + \widehat{\omega}_{T+1} \\
\widehat{\omega}_{T+2} &= \widehat{y}_{T+2} - \widehat{y}_{T+1} \Longrightarrow \widehat{y}_{T+2} = \widehat{y}_{T+1} + \widehat{\omega}_{T+2} \\
&\Longrightarrow \widehat{y}_{T+2} = y_T + \widehat{\omega}_{T+1} + \widehat{\omega}_{T+2} \\
&\vdots \\
\widehat{\omega}_{T+h} &= \widehat{y}_{T+h} - \widehat{y}_{T+h-1} \Longrightarrow \widehat{y}_{T+h} = \widehat{y}_{T+h-1} + \widehat{\omega}_{T+h} \\
&\Longrightarrow \widehat{y}_{T+h} = y_T + \widehat{\omega}_{T+1} + \cdots + \widehat{\omega}_{T+h} \qquad (16)
\end{aligned}
$$

Forecasting with known parameters - AR(1) Model I

- Start with an $AR(1)$ process:

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{17}$$

- Equation (14) simplifies to

$$\begin{aligned} \widehat{y}_{T+1} &= \phi y_T \\ \widehat{y}_{T+2} &= \phi \widehat{y}_{T+1} = \phi^2 y_T \\ &\vdots \\ \widehat{y}_{T+h} &= \phi^h y_T \end{aligned} \tag{18}$$

Forecasting with known parameters - AR(1) Model II

- Since

$$\begin{aligned} y_{T+1} &= \phi y_T + \varepsilon_{T+1} \\ y_{T+2} &= \phi^2 y_T + \varepsilon_{T+2} + \phi\varepsilon_{T+1} \\ &\vdots \\ y_{T+h} &= \phi^h y_T + \varepsilon_{T+h} + \phi\varepsilon_{T+h-1} + \cdots + \phi^{h-1}\varepsilon_{T+1} \end{aligned} \tag{19}$$

- using (18) and (19) the forecast errors are given by:

$$\begin{aligned} e_{T+1} &= \varepsilon_{T+1} \\ e_{T+2} &= \varepsilon_{T+2} + \phi\varepsilon_{T+1} \\ &\vdots \\ e_{T+h} &= \varepsilon_{T+h} + \phi\varepsilon_{T+h-1} + \cdots + \phi^{h-1}\varepsilon_{T+1} \end{aligned} \tag{20}$$

Forecasting with known parameters - AR(1) Model III

- and their variances

$$
\begin{aligned}
Var(e_{T+1}) &= \sigma_\varepsilon^2 \\
Var(e_{T+2}) &= (1+\phi^2)\sigma_\varepsilon^2 \\
&\vdots \\
Var(e_{T+h}) &= (1+\phi^2+\cdots+\phi^{2(h-1)})\sigma_\varepsilon^2
\end{aligned} \tag{21}
$$

- and we also have

$$
\begin{aligned}
\lim_{h\longrightarrow\infty} \widehat{y}_{T+h} &= 0 = E(y_t) \\
\lim_{h\longrightarrow\infty} Var(e_{T+h}) &= \frac{1}{1-\phi^2}\sigma_\varepsilon^2 = Var(y_t)
\end{aligned}
$$

Forecasting with known parameters - MA(1) Model I

- Consider the $MA(1)$ process given by:

$$y_t = \varepsilon_t - \psi\varepsilon_{t-1} \tag{22}$$

- Equation (14) simplifies to

$$\begin{aligned}
\widehat{y}_{T+1} &= \psi\varepsilon_{t-1} \\
\widehat{y}_{T+2} &= 0 \\
&\vdots \\
\widehat{y}_{T+h} &= 0
\end{aligned} \tag{23}$$

Forecasting with known parameters - MA(1) Model II

- Since

$$\begin{aligned} y_{T+1} &= \varepsilon_{T+1} - \psi\varepsilon_T \\ y_{T+2} &= \varepsilon_{T+2} - \psi\varepsilon_{T+1} \\ &\vdots \\ y_{T+h} &= \varepsilon_{T+h} - \psi\varepsilon_{T+h-1} \end{aligned} \tag{24}$$

- using (23) and (24) the forecast errors are given by:

$$\begin{aligned} e_{T+1} &= \varepsilon_{T+1} \\ e_{T+2} &= \varepsilon_{T+2} - \psi\varepsilon_{T+1} \\ &\vdots \\ e_{T+h} &= \varepsilon_{T+h} - \psi\varepsilon_{T+h-1} \end{aligned} \tag{25}$$

Forecasting with known parameters - MA(1) Model III

- with variances

$$
\begin{aligned}
Var(e_{T+1}) &= \sigma_\varepsilon^2 \\
Var(e_{T+2}) &= (1+\psi^2)\sigma_\varepsilon^2 \\
&\vdots \\
Var(e_{T+h}) &= (1+\psi^2)\sigma_\varepsilon^2
\end{aligned} \tag{26}
$$

- and we also have

$$
\begin{aligned}
\lim_{h\longrightarrow\infty} \widehat{y}_{T+h} &= 0 = E(y_t) \\
\lim_{h\longrightarrow\infty} Var(e_{T+h}) &= (1+\psi^2)\sigma_\varepsilon^2 = Var(y_t)
\end{aligned}
$$

Forecasting with known parameters - Random Walk I

- Consider the Random Walk process given by:

$$y_t = y_{t-1} + \varepsilon_t \tag{27}$$

- Equations (14) and (15) simplify to

$$\widehat{y}_{T+h} = y_T \tag{28}$$

- and equation (20) is given by

$$e_{T+h} = \varepsilon_{T+h} + \varepsilon_{T+h-1} + \cdots + \varepsilon_{T+1} \tag{29}$$

- Therefore the variance of the forecast error is given by

$$Var(e_{T+h}) = h\sigma_\varepsilon^2 \tag{30}$$

- From these expressions it follows that

$$\begin{aligned} \lim_{h \longrightarrow \infty} \widehat{y}_{T+h} &= y_T \\ \lim_{h \longrightarrow \infty} Var(e_{T+h}) &= \infty \end{aligned}$$

Forecasting with known parameters - additional comments I

- The $MA(\infty)$ representation is given by

$$y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \tag{31}$$

- The one-step ahead prediction is given by

$$\widehat{y}_{T+1} = \sum_{j=1}^{\infty} c_j \varepsilon_{T+1-j} \tag{32}$$

Forecasting with known parameters - additional comments II

- Similarly

$$
\begin{aligned}
\widehat{y}_{T+2} &= \sum_{j=2}^{\infty} c_j \varepsilon_{T+2-j} \\
&\vdots \\
\widehat{y}_{T+h} &= \sum_{j=h}^{\infty} c_j \varepsilon_{T+h-j} \qquad (33)
\end{aligned}
$$

Forecasting with known parameters - additional comments III

- Since

$$
\begin{aligned}
y_{T+1} &= \sum_{j=0}^{\infty} c_j \varepsilon_{T+1-j} \\
y_{T+2} &= \sum_{j=0}^{\infty} c_j \varepsilon_{T+2-j} \\
&\vdots \\
y_{T+h} &= \sum_{j=0}^{\infty} c_j \varepsilon_{T+h-j} \qquad (34)
\end{aligned}
$$

Forecasting with known parameters - additional comments IV

- using (33) and (34) the forecast errors are given by:

$$
\begin{aligned}
e_{T+1} &= \varepsilon_{T+1} \\
e_{T+2} &= \varepsilon_{T+2} + c_1 \varepsilon_{T+1} \\
&\vdots \\
e_{T+h} &= \varepsilon_{T+h} + c_1 \varepsilon_{T+h-1} + \cdots + c_{h-1} \varepsilon_{T+1} \\
\Longrightarrow \quad & e_{T+h} = \sum_{j=0}^{h-1} c_j \varepsilon_{T+h-j} \qquad (35)
\end{aligned}
$$

- which implies that when using an optimal forecast, the $h-$step ahead forecast erros is serially correlated and can be represented by a $MA(h-1)$ as given by equation (35).

Forecasting with known parameters - additional comments V

- Moreover

$$
\begin{aligned}
E(e_{T+h}) &= 0 \\
Var(e_{T+h}) &= \sigma_{\varepsilon}^2 \sum_{j=0}^{h-1} c_j^2 \\
\lim_{h \longrightarrow \infty} Var(e_{T+h}) &= \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} c_j^2 = Var(y_t)
\end{aligned}
$$

- from which is also follows that

$$
Var(e_{T+h+1}) - Var(e_{T+h}) = \sigma_{\varepsilon}^2 c_h^2 \geq 0
$$

- so the forecast error variance increases monotonically with the forecast horizon.

Forecasting with known parameters - additional comments VI

- If the error ε_t are normally distributed so is the forecast error and in particular

$$\frac{y_{T+h} - \widehat{y}_{T+h}}{\sqrt{Var(e_{T+h})}} \sim N(0,1) \tag{36}$$

- and it can be used to construct $(1-\alpha)\%$ interval forecasts as

$$\left(\widehat{y}_{T+h} - z_{\alpha/2}\sqrt{Var(e_{T+h})} : \widehat{y}_{T+h} + z_{\alpha/2}\sqrt{Var(e_{T+h})}\right) \tag{37}$$

- where $z_{\alpha/2}$ are critical values from the standard normal.
- Consider $\widehat{y}_{T+h}$ and $\widehat{y}_{T+h+k}$ i.e. forecasts of y_{T+h} and y_{T+h+k} made using information up to time T.

Forecasting with known parameters - additional comments VII

- Using (35) it can be shown that

$$
\begin{aligned}
E\left(e_{T+h}e_{T+h+k}\right) &= E\left(\left(\sum_{j=0}^{h-1} c_j\varepsilon_{T+h-j}\right)\left(\sum_{j=0}^{h+k-1} c_j\varepsilon_{T+h+k-j}\right)\right) \\
&= E\left(\sum_{j=0}^{h-1} c_j c_{j+k}\varepsilon^2_{T+h-j} + cross - product\right) \\
&= \sigma^2_\varepsilon \sum_{j=0}^{h-1} c_j c_{j+k} \qquad (38)
\end{aligned}
$$

- so the forecast errors for different horizons are correlated

Forecasting with known parameters - additional comments VIII

- From (35) and since ε_t is white noise and the predictor $\widehat{y}_{T+h}$ as an estimator, it follows that

$$Cov(\widehat{y}_{T+h}, e_{T+h}) = E\left(\left(\sum_{j=h}^{\infty} c_j \varepsilon_{T+h-j}\right)\left(\sum_{j=0}^{h-1} c_j \varepsilon_{T+h-j}\right)\right) = 0$$

- Therefore

$$Var(y_{T+h}) = Var(\widehat{y}_{T+h}) + Var(e_{T+h})$$

- and

$$Var(y_{T+h}) \geq Var(\widehat{y}_{T+h})$$

- the forecast is always less volatile than the actual realized value.

Forecasting with estimated parameters I

- If we use consistent parameter estimators, the optimal forecasts formulas remain valid.
- The complication is an increase in the variance of the forecast error due to the estimation uncertainty
- The first case is a stationary $AR(1)$ with drift

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t \tag{39}$$

- and the parameters μ and ϕ have to be estimated by $\widehat{\mu}$ and $\widehat{\phi}$ and the forecst error for $h = 1$ is

$$\begin{aligned}
e_{T+1} &= y_{T+1} - \widehat{y}_{T+1} \\
&= \mu + \phi y_T + \varepsilon_{T+1} - (\widehat{\mu} + \widehat{\phi} y_T) \\
&= \varepsilon_{T+1} + (\mu - \widehat{\mu}) + (\phi - \widehat{\phi}) y_T \\
&= \varepsilon_{T+1} + (\theta - \widehat{\theta})' \mathbf{x}_T
\end{aligned} \tag{40}$$

- where

$$\mathbf{x}_T = \begin{pmatrix} 1 \\ y_T \end{pmatrix} \quad \text{and} \quad (\theta - \widehat{\theta}) = \begin{pmatrix} \mu - \widehat{\mu} \\ \phi - \widehat{\phi} \end{pmatrix} \tag{41}$$

- and

$$Var(e_{T+1}) = \sigma_\varepsilon^2 + \mathbf{x}_T' Var(\widehat{\grave{}})\mathbf{x}_T \tag{42}$$

- where

$$\begin{aligned} Var(\widehat{\theta}) &= Var\begin{pmatrix} \widehat{\mu} \\ \widehat{\phi} \end{pmatrix} = \sigma_\varepsilon^2 E\begin{bmatrix} T & \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} y_t & \sum_{t=1}^{T} y_t^2 \end{bmatrix}^{-1} \\ &\simeq T^{-1}\begin{bmatrix} \sigma_\varepsilon^2 + \mu^2\frac{(1+\phi)}{(1-\phi)} & -\mu(1+\phi) \\ -\mu(1+\phi) & (1-\phi^2) \end{bmatrix} \end{aligned} \tag{43}$$

Forecasting with estimated parameters III

- see [Clements and Hendry, 1998], and (43) is known as the approximate forecats error variance.
- For the h-steps ahead prediction we have

$$
\begin{aligned}
\widehat{y}_{T+2} &= \widehat{\mu} + \widehat{\phi}\widehat{y}_{T+1} = \widehat{\mu} + \widehat{\phi}(\widehat{\mu} + \widehat{\phi} y_T) = \widehat{\mu}(1+\widehat{\phi}) + \widehat{\phi}^2 y_T \\
\widehat{y}_{T+3} &= \widehat{\mu} + \widehat{\phi}\widehat{y}_{T+2} = \widehat{\mu} + \widehat{\phi}(\widehat{\mu}(1+\widehat{\phi}) + \widehat{\phi}^2 y_T) \\
&= \widehat{\mu}(1+\widehat{\phi}+\widehat{\phi}^2) + \widehat{\phi}^3 y_T \\
&\vdots \\
\widehat{y}_{T+h} &= \widehat{\mu} + \widehat{\phi}\widehat{y}_{T+h-1} \\
&= \widehat{\mu} + \widehat{\phi}(\widehat{\mu}(1+\widehat{\phi}+\cdots+\widehat{\phi}^{h-2}) + \widehat{\phi}^{h-1} y_T) \\
&\Longrightarrow \widehat{y}_{T+h} = \widehat{\mu}(1+\widehat{\phi}+\cdots+\widehat{\phi}^{h-1}) + \widehat{\phi}^h y_T \\
&\Longrightarrow \widehat{y}_{T+h} = \widehat{\mu}\frac{1-\widehat{\phi}^h}{1-\phi} + \widehat{\phi}^h y_T \qquad (44)
\end{aligned}
$$

Forecasting with estimated parameters IV

- Therefore the forecast error estimated is given by

$$
\begin{aligned}
\widehat{e}_{T+2} &= y_{T+2} - \widehat{y}_{T+2} \\
&= \mu + \phi(\mu + \phi y_T + \varepsilon_{T+1}) + \varepsilon_{T+2} - (\widehat{\mu}(1+\widehat{\phi}) + \widehat{\phi}^2 y_T) \\
&= \mu(1+\phi) + \phi^2 y_T + \phi\varepsilon_{T+1} + \varepsilon_{T+2} - \widehat{\mu}(1+\widehat{\phi}) - \widehat{\phi}^2 y_T \\
&= (\mu - \widehat{\mu}) + (\mu\phi - \widehat{\mu}\widehat{\phi}) + (\phi^2 - \widehat{\phi}^2) y_T + \varepsilon_{T+2} + \phi\varepsilon_{T+1}
\end{aligned}
$$

Forecasting with estimated parameters V

$$
\begin{aligned}
\widehat{e}_{T+3} &= y_{T+3} - \widehat{y}_{T+3} \\
&= \mu + \phi y_{T+2} + \varepsilon_{T+3} - (\widehat{\mu}(1 + \widehat{\phi} + \widehat{\phi}^2) + \widehat{\phi}^3 y_T) \\
&= \mu + \phi(\mu + \phi(\mu + \phi y_T + \varepsilon_{T+1}) + \varepsilon_{T+2}) \\
&\quad + \varepsilon_{T+3} - (\widehat{\mu}(1 + \widehat{\phi} + \widehat{\phi}^2) + \widehat{\phi}^3 y_T) \\
&= (\mu - \widehat{\mu}) + (\mu\phi - \widehat{\mu}\widehat{\phi}) + (\mu\phi^2 - \widehat{\mu}\widehat{\phi}^2) \\
&\quad + (\phi^3 - \widehat{\phi}^3) y_T + \varepsilon_{T+3} + \phi\varepsilon_{T+2} + \phi^2\varepsilon_{T+1} \\
&= \sum_{i=0}^{2} (\mu\phi^i - \widehat{\mu}\widehat{\phi}^i) + (\phi^3 - \widehat{\phi}^3) y_T + \sum_{i=0}^{2} \phi^i \varepsilon_{T+3-i} \\
&\vdots
\end{aligned}
$$

$$
\widehat{e}_{T+h} = \sum_{i=0}^{h-1} (\mu\phi^i - \widehat{\mu}\widehat{\phi}^i) + (\phi^h - \widehat{\phi}^h) y_T + \sum_{i=0}^{h-1} \phi^i \varepsilon_{T+h-i} \qquad (45)
$$

- and

$$
\begin{aligned}
Var(\widehat{e}_{T+h}) \quad = \quad & E\left[\sum_{i=0}^{h}(\mu\phi^{i}-\widehat{\mu}\widehat{\phi}^{i})\right]^{2}+Var\left[(\phi^{h}-\widehat{\phi}^{h})\right]y_{T}^{2} \\
& +\sigma_{\varepsilon}^{2}\frac{1-\phi^{2h}}{1-\phi^{2}} \\
& +2E\left[\sum_{i=0}^{h-1}(\mu\phi^{i}-\widehat{\mu}\widehat{\phi}^{i})(\phi^{h}-\widehat{\phi}^{h})\right]y_{T}^{2}
\end{aligned}
\tag{46}
$$

Now we will evaluate the effects of the presence of a unit root by setting $\phi = 1$ in (39) we already shown que

$$
\begin{aligned}
y_{T+h} &= \mu h + y_T \\
e_{T+h} &= \sum_{i=0}^{h-1} \varepsilon_{T+h-i} \\
Var(e_{T+h}) &= h\sigma_\varepsilon^2
\end{aligned}
$$

- Now using the estimated parameters from an $AR(1)$ with drift, without imposing the unit root, in the forecast error in (45) but now imposing $\phi = 1$ becomes

$$
\widehat{e}_{T+h} = (\mu - \widehat{\mu})h + (1 - \widehat{\phi}^h)y_T + \sum_{i=0}^{h-1} \varepsilon_{T+h-i}
$$

- The idea of multi-steps (or direct) estimation is to estimate the parameters that will be used in forecasting by minimizing the same loss function as in the forecast period.
- Let us consider the $AR(1)$

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{47}$$

- so that

$$y_{T+h} = \phi^h y_T + \sum_{i=0}^{h-1} \phi^i \varepsilon_{T+h-i} \tag{48}$$

- The standard forecast was given by

$$\widehat{y}_{T+h} = \widehat{\phi}^h y_T \tag{49}$$

Multi-steps (or direct) estimation II

- where

$$\widehat{\phi} = \arg\min_{\phi} \sum_{t=1}^{T} (y_t - \phi y_{t-1})^2 = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2} \tag{50}$$

- and

$$E(y_{T+h} - \widehat{y}_{T+h}) = (\phi^h - E(\widehat{\phi}^h))y_T \tag{51}$$

- The forecast $\widehat{y}_{T+h}$ is also called "iterated" as it can be derived by replacing the unknown future values of y with their forecasts for $T+1, \ldots, T+h-1$.
- The alternative forecast is

$$\widetilde{y}_{T+h} = \widetilde{\phi}_h y_T \tag{52}$$

Multi-steps (or direct) estimation III

- where

$$\widetilde{\phi}_h = \arg\min_{\phi_h} \sum_{t=1}^{T} (y_t - \phi_h y_{t-h})^2 = \frac{\sum_{t=1}^{T} y_t y_{t-h}}{\sum_{t=1}^{T} y_{t-h}^2} \tag{53}$$

- and

$$E(y_{T+h} - \widetilde{y}_{T+h}) = (\phi^h - E(\widetilde{\phi}_h))y_T \tag{54}$$

- the forecast $\widetilde{y}_{T+h}$ is labeled "direct" since it is derived from a model where the target variable y_{T+h} is directly related to the available information set at time T.
- The relative performance of the two forecasts $\widehat{y}_{T+h}$ and $\widetilde{y}_{T+h}$ in terms of bias and efficiency depends on the bias and efficiency of the alternative estimaators of ϕ_h - $\widehat{\phi}^h$ and $\widetilde{\phi}_h$.

- When the model is correctly specified both estimators of ϕ_h are consistent but $\widehat{\phi}^h$ is more efficient than $\widetilde{\phi}_h$ since it coincides with the MLE
- When the model is mis-specified the ranking could change

Multi-steps (or direct) estimation - model mis-specification I

- The DGP is a $MA(1)$:

$$y_t = \varepsilon_t + \psi \varepsilon_{t-1} \tag{55}$$

- with $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, but the chosen model for y_t is an $AR(1)$

$$y_t = \phi y_{t-1} + v_t \tag{56}$$

- where $v_t \sim WN(0, \sigma_v^2)$.
- Wish to comare standard and direct estimation based forecasts assuming $h = 2$ and using $MSFE$ comparison criterion.

Multi-steps (or direct) estimation - model mis-specification II

- Standard estimation yields

$$\widehat{\phi} = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$$

- and can be approximated by

$$E(\widehat{\phi}) \simeq \frac{\psi}{1+\psi^2} = \phi$$

- Then

$$\widehat{y}_{T+2} = \widehat{\phi}^2 y_T \quad \text{and} \quad E(\widehat{y}_{T+2}) \simeq \phi^2 y_T$$

Multi-steps (or direct) estimation - model mis-specification III

- and the estimated $MSFE$ is given by

$$\begin{aligned}\widehat{MSFE} &= E[(y_{T+2} - \widehat{\phi}^2 y_T)^2 | y_T] \\ &\simeq (1+\psi^2)\sigma_\varepsilon^2 + (Var(\widehat{\phi}^2) + \phi^4) y_T^2\end{aligned}$$

- In the case of direct estimation we have

$$\widetilde{\phi}_2 = \frac{\sum_{t=2}^{T} y_t y_{t-2}}{\sum_{t=2}^{T} y_{t-2}^2} = \frac{\sum_{t=2}^{T} (\varepsilon_t + \psi\varepsilon_{t-1})(\varepsilon_{t-2} + \psi\varepsilon_{t-3})}{\sum_{t=2}^{T} y_{t-2}^2} \simeq 0$$

- so that

$$\widetilde{y}_{T+2} = \widetilde{\phi}_2 y_T \simeq 0$$

Multi-steps (or direct) estimation - model mis-specification IV

- and

$$\begin{aligned} \widetilde{MSFE} &= E[(y_{T+2} - \widetilde{y}_{T+2})|y_T] \\ &\simeq (1+\psi^2)\sigma_\varepsilon^2 + Var(\widetilde{\phi}_2)y_T^2 \end{aligned}$$

- for some values of the parameters it is possible that

$$\widetilde{MSFE} \leq \widehat{MSFE}$$

- Difficult to characterize the trade-off between bias and estimation in multi-period forecasts(see [Clements and Hendry, 1996] and [?])
- [Marcellino et al., 2006] compare empirical iterated and direct forecasts from linear univariate and bivariate models. The iterated forecasts typically outperform the direct forecasts.

Forecasting US inventories: h-steps vs 1-step

- The following comand in R are used
- Modeling US Inventories

	1 step ahead	2 steps ahead, iterated	1-to-x steps ahead
RMSFE	0.2737092	0.3364507	0.3239460
MAE	0.2381975	0.2858003	0.2913178

- The following comand in R are used
- Modeling US Inventories

	1 step ahead	2 steps ahead, iterated	2 steps ahead, direct	1-to-x steps ahead
RMSFE	0.1458	0.3781	0.2737	0.2218
MAE	0.1254	0.3218	0.2382	01848

Permanent-transitory decomposition

- Sometimes is of interest to decompose a process y_t into two components: permanent and transitory components.
- The permanent component captures the long-run - the trend
- The transitory component measures the short term deviation form the trend.

Permanent-transitory decomposition - Beveridge & Nelson I

- A weakly stationary process can be written as a $MA(\infty)$ and if $y_t \sim I(d)$ then $\Delta^d y_t$ is weakly stationary
- Let $d = 1$ then

$$\Delta y_t = \mu + c(L)\varepsilon_t \quad \text{and} \quad \varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2) \tag{57}$$

- Let define the following polynomial in L

$$d(L) = c(L) - c(1) \tag{58}$$

- Since $d(1) = 0$, 1 is root of $d(L)$ and can be rewritten as

$$d(L) = \widetilde{c}(L)(1 - L) \tag{59}$$

Permanent-transitory decomposition - Beveridge & Nelson II

- Combining equation (58) and (59) we have

$$c(L) = \widetilde{c}(L)(1 - L) + c(1) \tag{60}$$

- and

$$\Delta y_t = \mu + \widetilde{c}(L)\Delta\varepsilon_t + c(1)\varepsilon_t \tag{61}$$

- and integrating both sides of (61) we obtain a representation for y_t, i.e.

$$y_t = \underbrace{\mu t + c(1)\sum_{j=1}^{t}\varepsilon_j}_{\substack{\text{trend} \\ \text{(permanent component)} \\ \text{(PC)}}} + \underbrace{\widetilde{c}(L)\varepsilon_t}_{\substack{\text{cycle} \\ \text{(transitory component)} \\ \text{(CC)}}}$$

Permanent-transitory decomposition - Beveridge & Nelson III

- The permanent component is a random walk with drift:

$$PC_t = PC_{t-1} + \mu + c(1)\varepsilon_t \tag{62}$$

- Variance of the trend innovation is

$$c^2(1)\sigma_\varepsilon^2 \tag{63}$$

- which is larger (smaller) than the innovation in y_t if $c(1)$ is larger (smaller) than one.
- The innovation in the cyclical component is

$$\widetilde{c}(0)\varepsilon_j \tag{64}$$

Permanent-transitory decomposition - Beveridge & Nelson IV

- Since

$$\widetilde{c}(L) = \frac{c(L) - c(1)}{1 - L} \tag{65}$$

- then

$$\widetilde{c}(0) = c(0) - c(1) = 1 - c(1) \tag{66}$$

- therefore the innovation in the cyclical component is

$$(1 - c(1))\varepsilon_t \tag{67}$$

Beveridge & Nelson decomposition - an example I

- Let the $ARIMA(1,1,1)$ model given by

$$\Delta y_t = \phi \Delta y_{t-1} + \varepsilon_t - \psi \varepsilon_{t-1} \tag{68}$$

- and we want to derive the Beveridge & Nelson (BN) decomposition.
- The MA representation for Δy_t we have

$$\begin{aligned} c(L) &= \frac{1-\psi L}{1+\phi L} \qquad\qquad c(1) = \frac{1-\psi}{1+\phi} \\ \widetilde{c}(L) &= \frac{c(L)-c(1)}{1-L} = \frac{(\phi+\psi)}{(1+\phi L)(1+\phi)} \end{aligned}$$

- It follows that the BN decomposition is given by

$$y_t = PC + CC = \frac{1-\psi}{1+\phi}\sum_{j=1}^{t}\varepsilon_j + \frac{(\phi+\psi)}{(1+\phi L)(1+\phi)}\varepsilon_t \tag{69}$$

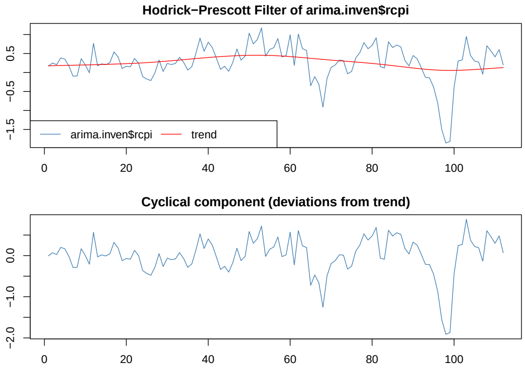
Permanent-transitory decomposition - The Hodrick-Prescott I

- The permanent component is obtained as

$$\min_{PC} \underbrace{{}_{t=1}^{T} \left(y_t - PC_t\right)^2}_{\text{Variance of CC}} + \lambda {}_{t=2}^{T-1} \left[\left(PC_{t+1} - PC_t\right) - \left(PC_t - PC_{t-1}\right)\right]^2 \tag{70}$$

- The bigger is λ, the smoother is the trend.
- In practice
  - $\lambda = 100$ if data is annual
  - $\lambda = 16000$ if data is quarterly
  - $\lambda = 144000$ if data is monthly
  - $\lambda = 0$ then $PC_t = y_t$

Permanent-transitory decomposition - The Hodrick-Prescott



Exponential Smoothing I

- Decomposes a time series into a "level" component and an umpredictable residual component
- Once the level at the end of the estimation sample is obtained, y_T^L it is used as a forecast for y_{T+h} , $h > 1$
- If y_t is an i.i.d. process with non-zero mean, y_T^L is estimated as the sample mean
- If y_t is persistent then the more recent observations should receive a greater weigh
- Hence

$$y_T^L = \sum_{t=1}^{T-1} \alpha(1-\alpha)^i y_{T-i} \tag{71}$$

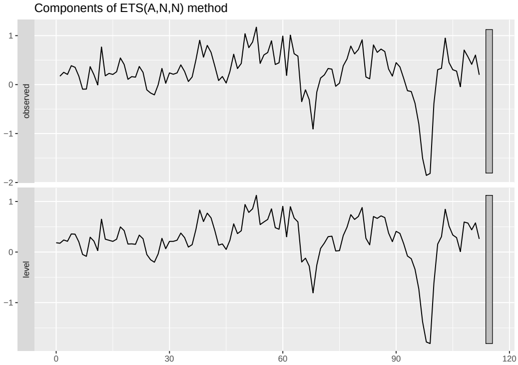
- with $0 < \alpha < 1$ and

$$y_{T+h} = y_T^L \quad \text{for all } h \tag{72}$$

- We can rewrite (71) as

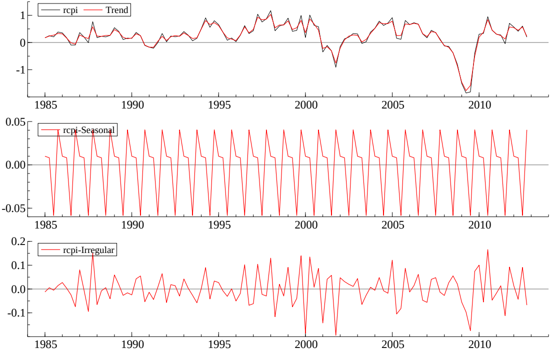
$$y_T^L = \alpha y_T + (1-\alpha) y_{T-1}^L \tag{73}$$

- with starting condition $y_1^L = y_1$.
- The following comand in R are used
- Modeling US Inventories
- For more about ETS see [Hyndman and Athanasopoulos,]



Components of ETS(A,N,N) method
observed
level
1
0
−1
−2
1
0
−1
0
30
60
90
120

UC STAMP



rcpi
Trend
1
0
-1
1985
1990
1995
2000
2005
2010
rcpi-Seasonal
0.05
0.00
-0.05
1985
1990
1995
2000
2005
2010
rcpi-Irregular
0.2
0.1
0.0
-0.1
1985
1990
1995
2000
2005
2010

Reference I

Box, G. E. P. and Jenkins, G. M. (1976).
Time Series Analysis: Forecasting and Control.
Holden-Day.

Clements, M. P. and Hendry, D. F. (1996).
Multi-step estimation for forecasting.
Oxford Bulletin of Economics and Statistics, 58:657–684.

Clements, M. P. and Hendry, D. F. (1998).
Forecasting Economic Time Series.
Cambridge University Press.

Hyndman, R. J. and Athanasopoulos, G.
Forecasting: principles and practice.
OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 22/04/2022., 3rd edition edition.

Reference II

Marcellino, M. G., Stock, J. H., and Watson, M. W. (2006).
A compariosn of direct and iterated multistep ar methos for forecasting macroeconomic time series.
Journal of Econometrics, 18:427–443.