Innovations

- Time-series models are constructed as linear functions of fundamental forecasting errors e_t, also called **innovations** or **shocks**
- These basic building blocks satisfy
  - $\mathrm{E}e_t = 0$
  - $\mathrm{var}(e_t) = \mathrm{E}e_t^2 = \sigma^2$
  - Serially uncorrelated
  - These errors e_t are called **white noise**
- In general, if you see an error e_t, it should be interpreted as white noise. We will write
  - e_t is $\mathrm{WN}(0, \sigma^2)$

Unforecastable Innovations

- White noise processes are linearly unforecastable
- A stronger condition is unforecastable.
- The innovations e_t are **unforecastable** if
  - $E(e_t | \Omega_{t-1}) = 0$
  - This means the best forecast is zero
- For some purposes, we will assume the errors are unforecastable

Moving Average Processes

- Diebold, Chapter 7
- These models are linear functions of stochastic errors

MA(1) Process

- The **first-order moving average** process, or **MA(1)** process, is

$$y_t = e_t + \theta e_{t-1}$$

  where e_t is WN(0, σ^2)
- The MA coefficient θ controls the degree of serial correlation. It may be positive or negative.
- The innovations e_t impact y_t over two periods
  - An contemporaneous (same period) impact
  - A one-period delayed impact

Mean of MA(1)

- The unconditional mean of y_t is

$$
\begin{aligned}
E(y_t) &= E(e_t + \theta e_{t-1}) \\
&= E(e_t) + \theta E(e_{t-1}) \\
&= 0
\end{aligned}
$$

Variance of MA(1)

- The unconditional variance of y_t is

$$\begin{aligned}
&\operatorname{var}(y_t) = \operatorname{var}(e_t + \theta e_{t-1}) \\
&= \operatorname{var}(e_t) + \operatorname{var}(\theta e_{t-1}) + 2\operatorname{cov}(e_t, \theta e_{t-1}) \\
&= \sigma^2 + \theta^2 \sigma^2 + 0 \\
&= \left(1 + \theta^2\right)\sigma^2
\end{aligned}$$

- This is a function of both the innovation variance σ^2 and the MA coefficient θ.

Conditional Mean of MA(1)

- If the error is unforecastable E(e_t|$\Omega_{t\text{-}1}$) = 0 then the conditional mean of y_t is

$$
\begin{aligned}
E(y_t \mid \Omega_{t-1}) &= E(e_t + \theta e_{t-1} \mid \Omega_{t-1}) \\
&= E(e_t \mid \Omega_{t-1}) + \theta E(e_{t-1} \mid \Omega_{t-1}) \\
&= \theta e_{t-1}
\end{aligned}
$$

- This is the best forecast of y_t .
- The optimal forecast error is

$$
\begin{aligned}
y_t - E(y_t \mid \Omega_{t-1}) &= \left(e_t + \theta e_{t-1}\right) - \theta e_{t-1} \\
&= e_t
\end{aligned}
$$

Conditional Variance of MA(1)

- The conditional variance of y_t is

$$\begin{aligned}\operatorname{var}(y_t \mid \Omega_{t-1}) &= \operatorname{var}(y_t - E(y_t \mid \Omega_{t-1}) \mid \Omega_{t-1}) \\ &= \operatorname{var}(e_t \mid \Omega_{t-1}) \\ &= \sigma^2\end{aligned}$$

- The conditional variance, the forecast variance, and the innovation variance are all the same thing

Autocovariance of MA(1)

- The first autocovariance is

$$
\begin{aligned}
\gamma(1) &= E(y_t y_{t-1}) \\
&= E\big((e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2})\big) \\
&= E(e_t e_{t-1}) + \theta E(e_{t-1}^2) + \theta E(e_t e_{t-2}) + \theta^2 E(e_{t-1} e_{t-2}) \\
&= 0 + \theta E(e_{t-1}^2) + 0 + 0 \\
&= \theta \sigma^2
\end{aligned}
$$

Autocovariance of MA(1)

- The autocovariance for *k*>1 are

$$
\begin{aligned}
\gamma(k) &= E(y_t y_{t-k}) \\
&= E\big((e_t + \theta e_{t-1})(e_{t-k} + \theta e_{t-k-1})\big) \\
&= E(e_t e_{t-k}) + \theta E(e_{t-1} e_{t-k}) + \theta E(e_t e_{t-k-1}) + \theta^2 E(e_{t-1} e_{t-k-1}) \\
&= 0 + 0 + 0 + 0 \\
&= 0
\end{aligned}
$$

- Thus the autocovariance function is zero for k>1

Autocorrelations of MA(1)

- Since

$$\gamma(0) = \operatorname{var}(y_t) = (1+\theta^2)\sigma^2$$
$$\gamma(1) = \theta\sigma^2$$
$$\gamma(k) = 0, k \geq 2$$

then

$$\rho(1) = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$$
$$\rho(k) = 0, k \geq 2$$

- The autocorrelation function of an MA(1) is zero after the first lag.

First Autocorrelation

- The first autocorrelation has the same sign as θ

$$\rho(1) = \frac{\theta}{1+\theta^2}$$

- As θ ranges from -1 to 1, ρ(1) ranges from - ½ to ½

$$y_t = e_t + \theta e_{t-1}$$

- Θ<0 : negative autocorration

Lag Operator Notation

- Remember the lag operator *L*

$$Ly_t = y_{t-1}$$

- We can write the MA(1) as

$$\begin{aligned} y_t &= e_t + \theta e_{t-1} \\ &= e_t + \theta L e_t \\ &= (1 + \theta L) e_t \end{aligned}$$

or

$$y_t = \theta(L) e_t$$

where $\theta(L) = 1 + \theta L$ is a function of the lag operator.

Inversion of an MA(1)

- We can write an MA(1) in terms of lagged y_t

$$y_t = e_t + \theta e_{t-1}$$

- Rewrite as

$$e_t = y_t - \theta e_{t-1}$$

- Then lag this equation one period

$$e_{t-1} = y_{t-1} - \theta e_{t-2}$$

- Then combine

$$\begin{aligned} e_t &= y_t - \theta e_{t-1} \\ &= y_t - \theta\left(y_{t-1} - \theta e_{t-2}\right) \\ &= y_t - \theta y_{t-1} + \theta^2 e_{t-2} \end{aligned}$$

Inversion, Continued

- Do this again

$$e_{t-2} = y_{t-2} - \theta e_{t-3}$$

$$\begin{aligned} e_t &= y_t - \theta y_{t-1} + \theta^2 e_{t-2} \\ &= y_t - \theta y_{t-1} + \theta^2 \left(y_{t-2} - \theta e_{t-3} \right) \\ &= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 e_{t-3} \end{aligned}$$

- Repeat to infinity $e_t = y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 y_{t-3} + \cdots$
- Then

$$\begin{aligned} y_t &= \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} + \cdots + e_t \\ &= -\sum_{i=1}^{\infty} (-\theta)^i y_{t-i} + e_t \end{aligned}$$

Existence of Inverse

- This series converges (and the inversion exists) if $|\theta|<1$.
- Recall the lag operator expression

$$y_t = (1+\theta L)e_t$$

- We can write this as

$$(1+\theta L)^{-1} y_t = e_t$$

- This inversion is valid if $|\theta|<1$

Inversion of Lag Polynomial

- What does this mean? $(1+\theta L)^{-1} y_t = e_t$
- By taking a power series expansion (from calculus)

$$(1+\theta L)^{-1} = 1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \cdots$$

- This expansion converges if $|\theta|<1$
- Applying this expression

$$\begin{aligned}(1+\theta L)^{-1} y_t &= (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \cdots) y_t \\ &= y_t - \theta y_{t-1} + \theta^2 y_{t-2} - \theta^3 y_{t-3} + \cdots\end{aligned}$$

as needed

Optimal Forecast

- In the MA(1) model

$$y_t = e_t + \theta e_{t-1}$$

the optimal forecast is θe_{t-1} but the error is not directly observed.

- One approach is to use the autoregressive representation

$$E(y_t \mid \Omega_{t-1}) = -\sum_{i=1}^{\infty} (-\theta)^i y_{t-i}$$

- But this is cumbersome.

Recursive Forecast for MA(1)

- Another approach is to use the equation

$$e_t = y_t - \theta e_{t-1}$$

and realize that this gives a recursive formula to numerically compute the error

- Given θ, and given the initial condition e_0=0

$$e_1 = y_1 - \theta e_0$$
$$e_2 = y_2 - \theta e_1$$
$$\vdots$$

- This gives a recursive formula to compute all the errors.
- The out-of-sample forecast is $y_{T+1|T} = \theta e_T$

MA(q) Process

- The moving average process of order q, or MA(q), is

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

where e_t is WN(0, σ^2)

- We can write the equation as

$$\begin{aligned} y_t &= \left(1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q\right) e_t \\ &= \theta(L) e_t \end{aligned}$$

where θ(L) is a q'th order polynomial in L

Autocorrelations

- The first q autocorrelations of a MA(q) are non-zero, the autocorrelations above q are zero

Wold's Theorem

- If y_t is a zero-mean covariance stationary process, then it can be written as an infinite order moving average, also known as a **general linear process**

$$y_t = \sum_{i=0}^{\infty} \theta_i e_{t-i}$$
$$= \theta(L) e_t$$

where e_t is WN(0, σ^2)

Linear Process

$$y_t = \sum_{i=0}^{\infty} \theta_i e_{t-i}$$

$$= \theta(L) e_t$$

- Normalization: $\theta_0 = 1$
- Square summability

$$\sum_{i=0}^{\infty} \theta_i^2 < \infty$$

Interpretation of Wold's Theorem

- There is a best linear approximation for y_t in terms of is past values
- MA(q) may be a useful approximation

Mean and Variance

- Unconditional mean

$$E(y_t) = E(\sum_{i=0}^{\infty} \theta_i e_{t-i}) = 0$$

- Unconditional variance

$$\text{var}(y_t) = \text{var}(\sum_{i=0}^{\infty} \theta_i e_{t-i})$$

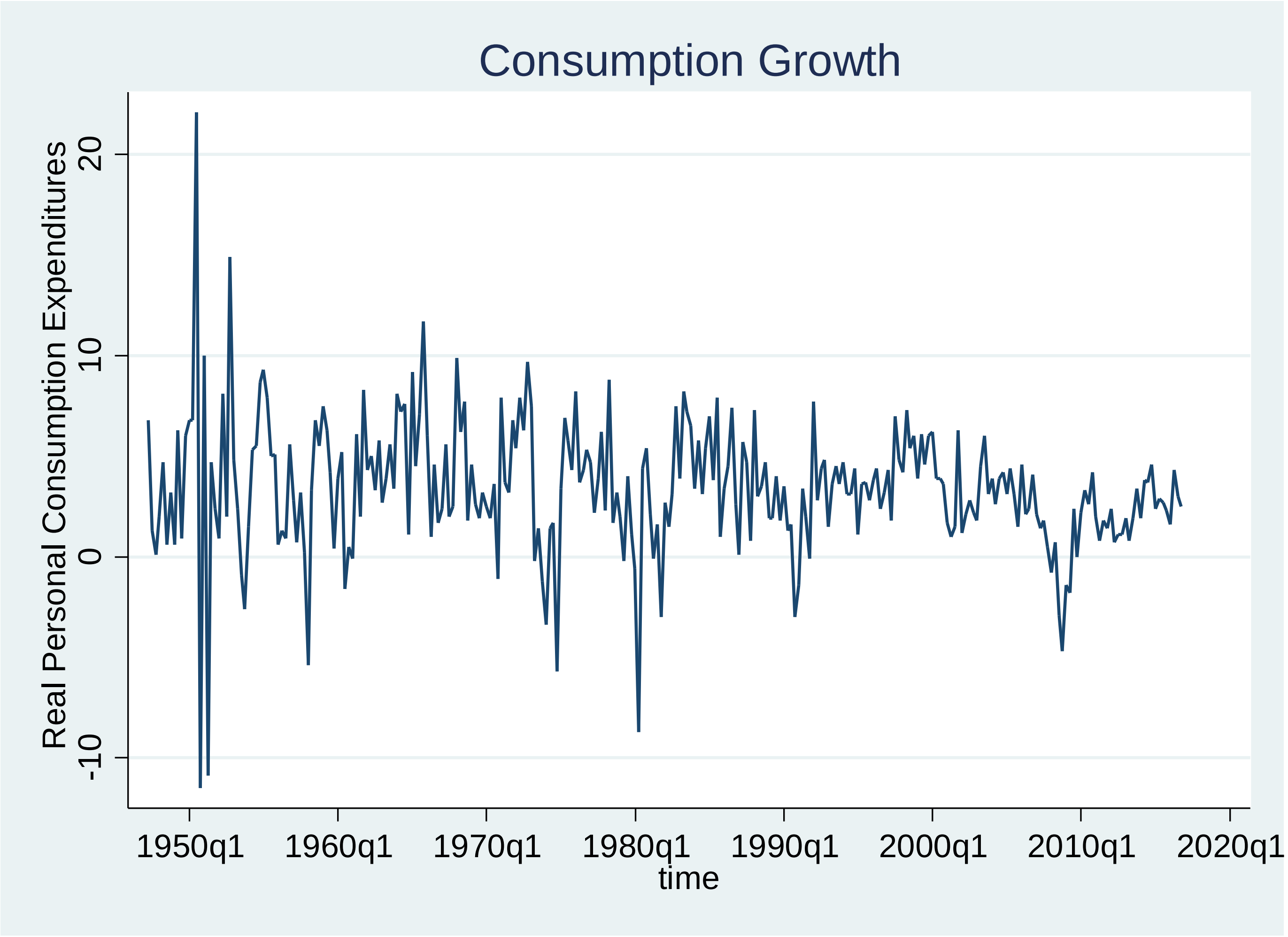
$$= \left(\sum_{i=0}^{\infty} \theta_i^2 \right) \sigma^2$$

Relevance of MA(q) Models

- MA(q) models help to build our understanding and intuition for serial dependence and autocorrelation
- But, not commonly used for forecasting
- To estimate in STATA, use command

  arima y, arima(0,0,q)

Quarterly Consumption Growth



MA(2) Model

- We will estimate a MA(2)

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

- Stata command

  arima y, arima(0,0,2)

MA(2) Estimation

- arima is a nonlinear optimizer, so the algorithm iterates until convergence

```
. arima pce, arima(0,0,2)

(setting optimization to BHHH)
Iteration 0:   log likelihood = -733.24995
Iteration 1:   log likelihood = -718.76714
Iteration 2:   log likelihood = -715.96169
Iteration 3:   log likelihood = -715.44537
Iteration 4:   log likelihood = -715.33103
(switching optimization to BFGS)
Iteration 5:   log likelihood =  -715.2785
Iteration 6:   log likelihood =  -715.2389
Iteration 7:   log likelihood = -715.23732
Iteration 8:   log likelihood =  -715.2373
```

MA(2) Estimates, cont.

- The estimated MA coefficients are shown as “L1” and “L2”. Note the MA(1) coef is small, the MA(2) coef is larger

ARIMA regression

Sample: 1947q2 - 2016q4 | Number of obs = 279
Wald chi2(2) = 90.51
Log likelihood = -715.2373 | Prob > chi2 = 0.0000

pce	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
pce						
_cons	3.368277	.270568	12.45	0.000	2.837974	3.898581
ARMA						
ma						
L1.	.0310889	.0377295	0.82	0.410	-.0428595	.1050373
L2.	.3652267	.041443	8.81	0.000	.2839999	.4464535
/sigma	3.139894	.0734445	42.75	0.000	2.995945	3.283843

Results

- MA(2) model for consumption growth

$$y = e_t + \underset{(0.04)}{0.03}\ e_{t-1} + \underset{(0.04)}{0.36}\ e_{t-2}$$

Autoregressive Processes

- The first-order autoregressive process, AR(1) is

$$y_t = \beta y_{t-1} + e_t$$

where e_t is WN(0, σ^2)

- Using the lag operator, we can write

$$(1 - \beta L) y_t = e_t$$

- If $\beta > 0$, y_{t-1} and y_t are positively correlated
- If $\beta < 0$, y_{t-1} and y_t are negatively correlated

Inversion

- By back-substitution

$$\begin{aligned} y_t &= \beta y_{t-1} + e_t \\ &= e_t + \beta\left(\beta y_{t-2} + e_{t-1}\right) \\ &= e_t + \beta e_{t-1} + \beta^2 e_{t-2} + \cdots \\ &= \sum_{i=0}^{\infty} \beta^i e_{t-i} \end{aligned}$$

  a general linear process with geometrically declining coefficients

  – This inversion requires that $|\beta|<1$
  – $|\beta|<1$ is required for stationarity

Importance of |β|<1

- If β=1 then

$$y_t = e_t + e_{t-1} + e_{t-2} + \cdots$$

does not converge, so the sum is not defined.

Mean and Variance

- By the formula for the unconditional mean and variance of a general linear process

$$E(y_t) = E(\sum_{i=0}^{\infty} \beta^i e_{t-i}) = 0$$

$$\begin{aligned} \text{var}(y_t) &= \text{var}(\sum_{i=0}^{\infty} \beta^i e_{t-i}) \\ &= \left(\sum_{i=0}^{\infty} \beta^{2i} \right) \sigma^2 \\ &= \frac{\sigma^2}{1-\beta^2} \end{aligned}$$

Another Variance Calculation

- Take variance of both sides of

$$y_t = \beta y_{t-1} + e_t$$

- Thus

$$\begin{aligned} \operatorname{var}(y_t) &= \operatorname{var}(\beta y_{t-1} + e_t) \\ &= \operatorname{var}(\beta y_{t-1}) + \operatorname{var}(e_t) \\ &= \beta^2 \operatorname{var}(y_{t-1}) + \sigma^2 \end{aligned}$$

- If y is variance stationary, we solve and find

$$\operatorname{var}(y_t) = \operatorname{var}(y_{t-1}) = \frac{\sigma^2}{1-\beta^2}$$

$|\beta|<1$

- If $|\beta|=1$ then

$$\operatorname{var}(y_t) = \frac{\sigma^2}{1-\beta^2}$$

is infinite

|β|=1

- We calculated that

$$\mathrm{var}(y_t) = \beta^2 \,\mathrm{var}(y_{t-1}) + \sigma^2$$

- When |β|=1, then

$$\mathrm{var}(y_t) = \mathrm{var}(y_{t-1}) + \sigma^2 > \mathrm{var}(y_{t-1})$$

  so the variance is increasing with *t*
- |β|=1 is inconsistent with variance stationarity.
- |β|<1 is necessary for stationarity.

Random Walk

- An AR(1) with β=1 is known as a random walk or unit root process

$$y_t = y_{t-1} + e_t$$

- By back-substitution

$$y_t = y_0 + \sum_{i=0}^{t} e_{t-i}$$

- The past never disappears. Shocks have permanent effects

Unit Root

- The random walk is called a **unit root** process because the lag operator 1-L has a “root” (intersection with the x-axis) at L=1
- It is called a **random walk** because it tends to wander without mean-reversion.
- If y_t is an AR(1) with a unit root (β=1) then its first difference $\Delta y_t = y_t - y_{t-1}$ is white noise

Assignments

- Read Diebold through Chapter 7.2
- Problem Set # 5
  - Due Tuesday (2/21)
- Read Chapter 4 from *The Signal and the Noise*
  - Reading Reflection
  - Due Thursday (2/16)