Financial Econometrics Econ 40357
ARIMA (Auto Regressive Integrated Moving Average) Models
Part 1.

N.C. Mark

University of Notre Dame and NBER

August 26, 2020

Overview: Univariate, parametric time-series models

- Time series are interesting when there is dependence over time.
- Strategy is to develop **model** that describes the time-series data.
  - Familiar story in econometrics. Can characterize properties of theoretical model.
  - If data conforms to model, use properties of model to generate infererence, say something about the real world.
- Time-series models: Describe how the current state depends on past states. Then use the current state to **predict** future states.
- Estimation and prediction. Prediction $\Leftrightarrow$ forecast.

What we learn in this segment

- Parametric models, often found to be useful for modeling stationary time series. So called ARIMA models.
- AutoRegressive Integrated Moving Average
- Key features are knowing about
  - First two moments (mean and variance)
  - Autocovariance, autocorrelation: Characterizing dependence over time
  - Conditional expectation. To be used as forecasting model.
  - Estimation, using estimated models to forecast
  - How to evaluate the forecasts.

Covariance stationarity (again)

The time series $\{y_t\}_{t=1}^{T}$ is covariance (weakly) stationary if the mean, variance, and autocovariances of the process are constant.

$$
\begin{aligned}
E\left(y_t\right) &= \mu \\
E\left(y_t-\mu\right)^2 &= \sigma^2 \\
E\left(y_t-\mu\right)\left(y_{t-k}-\mu\right) &= \gamma_k
\end{aligned}
$$

Conditional expectation

Q: What function F minimizes the mean square prediction error,

$$E\left[y_{t+1} - F\left(y_{t+1} | I_t\right)\right]^2$$

A:

$$E\left(y_{t+1} | I_t\right) = \int y_{t+1} p\left(y_{t+1} | I_t\right) dy_{t+1}$$

where $p\left(y_{t+1} | I_t\right)$ is conditional pdf of y_{t+1}, I_t is the available information at t.

- **Important result: Conditional expectation** is **minimum mean-square error predictor**. It's the **best!**
- Think of **fitted value** of regression as conditional expectation. **Systematic** part of regression also called **projection**.
- **Notational Convention** $E_t(X_{t+k}) \equiv E(X_{t+k} | I_t)$

The white noise process (again)

- Stochastic (random) nature of the world
- White noise is the basic building block of all time series

$$y_t = \sigma\epsilon_t, \quad \epsilon_t \overset{iid}{\sim} (0, 1)$$

- These are random shocks, no dependence over time, representing purely unpredictable events. It's a model of news.
- We didn't say they are normally distributed. In time-series, it doesn't matter because all inference is asymptotic
- By itself, it is uninteresting, because there is no dependence over time.
- Next, I show you how we build in dependence.

Moving Average models

- An MA(k) process. y_t is correlated with y_{t-k} and possibly $y_{t-1}, \ldots, y_{t-k+1}$.
- The MA(1). Example might be daily returns with slow moving capital. News occurs today. High frequency traders pounce, institutional investors, move later in the day. Retail investors don't know until they see the nightly Bloomberg report.

The MA(1) model

- Let y_t be the observations

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

where $\epsilon_t \overset{iid}{\sim} (0, \sigma_\epsilon^2)$. Shift time index back one period,

$$y_{t-1} = \mu + \epsilon_{t-1} + \theta\epsilon_{t-2}$$

- Calculate the mean of y_t

$$E(y_t) = E(\mu + \epsilon_t + \theta\epsilon_{t-1}) = \mu$$

- Calculate the variance of y_t

$$
\begin{aligned}
\sigma_y^2 &= Var\left(y_t\right) = E\left(y_t - \mu\right)^2 = E\left(\epsilon_t + \theta\epsilon_{t-1}\right)^2 && (1)\\
&= E\left(\epsilon_t^2 + \theta^2\epsilon_{t-1}^2 + 2\theta\epsilon_t\epsilon_{t-1}\right) && (2)\\
&= E\left(\epsilon_t^2\right) + E\left(\theta^2\epsilon_{t-1}^2\right) + E\left(2\theta\epsilon_t\epsilon_{t-1}\right) && (3)\\
&= \sigma_\epsilon^2 + \theta^2\sigma_\epsilon^2 + 2\theta\underbrace{E\left(\epsilon_t\epsilon_{t-1}\right)}_{0} && (4)\\
&= \left(1 + \theta^2\right)\sigma_\epsilon^2 && (5)
\end{aligned}
$$

- Calculate the auto covariance function

$$\begin{aligned}
\gamma_1 &= Cov\left(y_t, y_{t-1}\right) = E\left(\tilde{y}_t, \tilde{y}_{t-1}\right) && (6)\\
&= E\left(\epsilon_t + \theta\epsilon_{t-1}\right)\left(\epsilon_{t-1} + \theta\epsilon_{t-2}\right)\\
&= E\left(\epsilon_t\epsilon_{t-1} + \theta\epsilon_{t-1}^2 + \theta\epsilon_t\epsilon_{t-2} + \theta^2\epsilon_{t-1}\epsilon_{t-2}\right)\\
&= \theta\sigma_\epsilon^2 && (7)
\end{aligned}$$

Autocorrelation

$$\rho\left(y_t, y_{t-1}\right) = \mathrm{Corr}(y_t, y_{t-1}) = \frac{\gamma_1}{\sigma_y\sigma_y} = \frac{\gamma_1}{\left(1+\theta\right)\sigma_\epsilon^2}$$

and for any $k > 1$,

$$\gamma_k = Cov\left(y_t, y_{t-k}\right) = 0.$$

MA(1) process is covariance stationary and displays one period dependence (memory).

MA(1) Forecasting formula

- Use the fact that the conditional expectation (projection), the fitted value of model (regression), is the **optimal** forecast. One period ahead forecast

$$E_t\left(y_{t+1}\right) = E_t\left(\mu + \epsilon_{t+1} + \theta\epsilon_t\right) = \mu + \theta\epsilon_t$$

$$E_t(y_{t+2}) = E_t\left(\mu + \epsilon_{t+1} + \theta\epsilon_{t+1}\right) = \mu$$

And for any $k \geq 2$, the model has no forecasting power

$$E_t(y_{t+k}) = \mu$$

The MA(2) model

- Observations correlated with (exhibit dependence) at most 2 lags of itself.

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

- I assign you to verify the following

$$\begin{aligned} E\left(y_t\right) &= \mu \\ Var\left(y_t\right) &= \left(1 + \theta_1^2 + \theta_2^2\right) \sigma_\epsilon^2, \\ Cov\left(y_t, y_{t-1}\right) &= \left(\theta_1 + \theta_1 \theta_2\right) \sigma_\epsilon^2 \\ Cov\left(y_t, y_{t-2}\right) &= \theta_2, \\ Cov\left(y_t, y_{t-k}\right) &= 0 \quad \text{for} \quad k > 2 \end{aligned}$$

- **MA(2) Forecasts**
- One-step ahead forecast

$$
\begin{aligned}
E_t\left(y_{t+1}\right) &= E_t\left(\mu + \epsilon_{t+1} + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1}\right) \\
&= \mu + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1}
\end{aligned}
$$

- Two-step ahead forecast

$$
E_t\left(y_{t+2}\right) = E_t\left(\mu + \epsilon_{t+2} + \theta_1 \epsilon_{t+1} + \theta_2 \epsilon_t\right) = \mu + \theta_2 \epsilon_t
$$

- Three-step ahead forecast

$$
E_t(y_{t+3}) = E\left(\mu + \epsilon_{t+3} + \theta_1 \epsilon_{t+2} + \theta_2 \epsilon_{t+1}\right) = \mu
$$

Hence for any $k \geq 3$, $E_t(y_{t+k}) = \mu$.

How to Estimate MA models?

- There are no independent variables, so you can't run least squares regression.
- We do something called maximum likelihood estimation.
- Illustrate the idea with the MA(1) model.

Maximum Likelihood Estimation of MA(1)

- The ϵ_t are random variables. Let's assume they are drawn from a normal distribution, $N\left(0, \sigma_\epsilon^2\right)$. The marginal probability density function (pdf) for ϵ_t is

$$f_1\left(\epsilon_t\right) = \frac{1}{\sigma_\epsilon \sqrt{2\pi}} e^{-\frac{\epsilon_t^2}{2\sigma_\epsilon^2}}$$

  The joint pdf for $\epsilon_1, \epsilon_2, ..., \epsilon_t, \epsilon_{t+1}, ..., \epsilon_T$ is the product of the $f_1\left(\right)$, because the $\epsilon' s$ are independent.

$$f\left(\epsilon_T, \epsilon_{T-1}, ..., \epsilon_1\right) = \left(\frac{2}{\sigma_\epsilon \sqrt{2\pi}}\right)^T e^{-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^T \epsilon_t^2}$$

Maximum Likelihood Estimation of MA(1)

- Notice that $\epsilon_t = y_t - \mu - \theta\epsilon_{t-1}$, $\epsilon_{t-1} = y_{t-1} - \mu - \theta\epsilon_{t-2}, \epsilon_{t-2} = y_{t-2} - \mu - \theta\epsilon_{t-3}, \ldots$ This means

$$
\begin{aligned}
\epsilon_t &= y_t - \mu - \theta\left(y_{t-1} - \mu - \theta\left(y_{t-2} - \mu - \theta\left(\ldots\right)\right)\right) \\
\epsilon_{t-1} &= y_{t-1} - \mu - \theta\left(y_{t-2} - \mu - \theta\left(y_{t-3} - \mu - \theta\left(\ldots\right)\right)\right) \\
&\vdots
\end{aligned}
$$

- Substitute these back into the joint pdf, and we get a function of the y_t, which I won't write out specifically.

$$
f\left(y_T, y_{T-1}, \ldots, y_1 | \mu, \theta, \sigma_\epsilon^2\right)
$$

  This is now a function of the **data.** By substitution of the MA(1) model into the joint pdf, we've transformed the pdf into a function of the data. This is called a **likelihood** function. PDFs are for random variables. Likelihood functions are for data.

- Maximum likelihood estimation is done by asking the computer to search and those $\mu, \eta, \sigma_\epsilon^2$ that maximizes $f\left(\right)$.

Let's apply MA(1) to daily stock returns

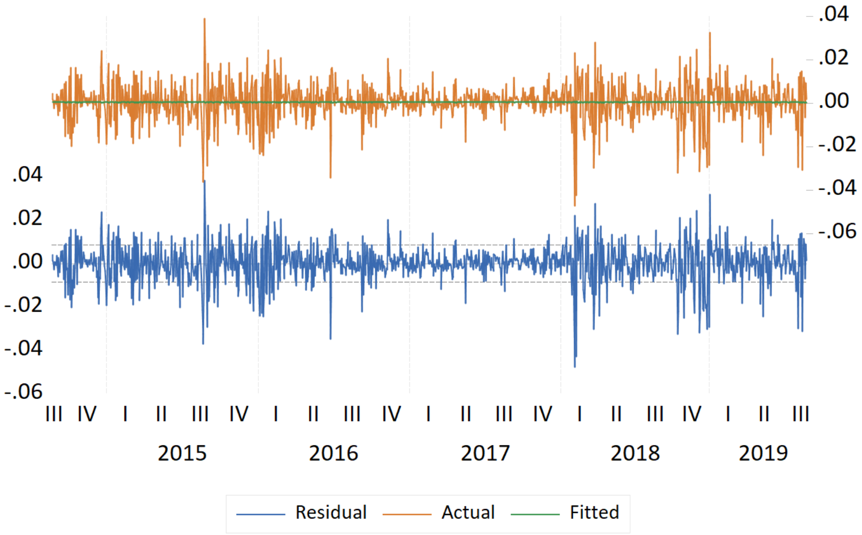
Eviews/ARIMA_Models.wf1

Code: equation eqma1.ls(optmethod=opg) djiaret c ma(1)

Dependent Variable: DJIARET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 09/12/19 Time: 10:50
Sample: 8/25/2014 8/22/2019
Included observations: 1208
Convergence 28 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000385	0.000260	1.480358	0.1390
MA(1)	-0.008067	0.019399	-0.415869	0.6776
SIGMASQ	7.25E-05	1.90E-06	38.04024	0.0000
R-squared	0.000058	Mean dependent var		0.000384
Adjusted R-squared	-0.001601	S.D. dependent var		0.008516
S.E. of regression	0.008523	Akaike info criterion		-6.689660
Sum squared resid	0.087529	Schwarz criterion		-6.677003
Log likelihood	4043.555	Hannan-Quinn criter.		-6.684894
F-statistic	0.035126	Durbin-Watson stat		1.965005
Prob(F-statistic)	0.965485			
Inverted MA Roots	.01			

Let's apply MA(1) to daily stock returns



Let's apply MA(5) to daily stock returns

equation eqma5.ls(optmethod=opg) djiaret c ma(1) ma(2) ma(3) ma(4) ma(5)

Dependent Variable: DJIARET
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 09/12/19 Time: 10:50
Sample: 8/25/2014 8/22/2019
Included observations: 1208
Convergence achieved after 36 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000378	0.000266	1.421888	0.1553
MA(1)	-0.005196	0.019996	-0.259843	0.7950
MA(2)	-0.037908	0.021856	-1.734453	0.0831
MA(3)	0.067974	0.020404	3.331478	0.0009
MA(4)	-0.019959	0.022182	-0.899776	0.3684
MA(5)	-0.050492	0.025925	-1.947578	0.0517
SIGMASQ	7.18E-05	1.94E-06	36.98844	0.0000
R-squared	0.008933	Mean dependent var		0.000384
Adjusted R-squared	0.003981	S.D. dependent var		0.008516
S.E. of regression	0.008499	Akaike info criterion		-6.691267
Sum squared resid	0.086752	Schwarz criterion		-6.661733
Log likelihood	4048.525	Hannan-Quinn criter.		-6.680146
F-statistic	1.804113	Durbin-Watson stat		1.971303
Prob(F-statistic)	0.094952			
Inverted MA Roots	.54	.19-.55i	.19+.55i	-.46-.25i
	-.46+.25i			

Let's apply MA(5) to daily stock returns