UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

Master's
Applied Time-Series Econometrics 2018

Session 2: Box Jenkins ARIMA modelling and forecast evaluation techniques

Lecturer:
Nico Katzke
nicokatzke@sun.ac.za

Department of Economics

DEPARTMENT OF ECONOMICS

Plan for the day

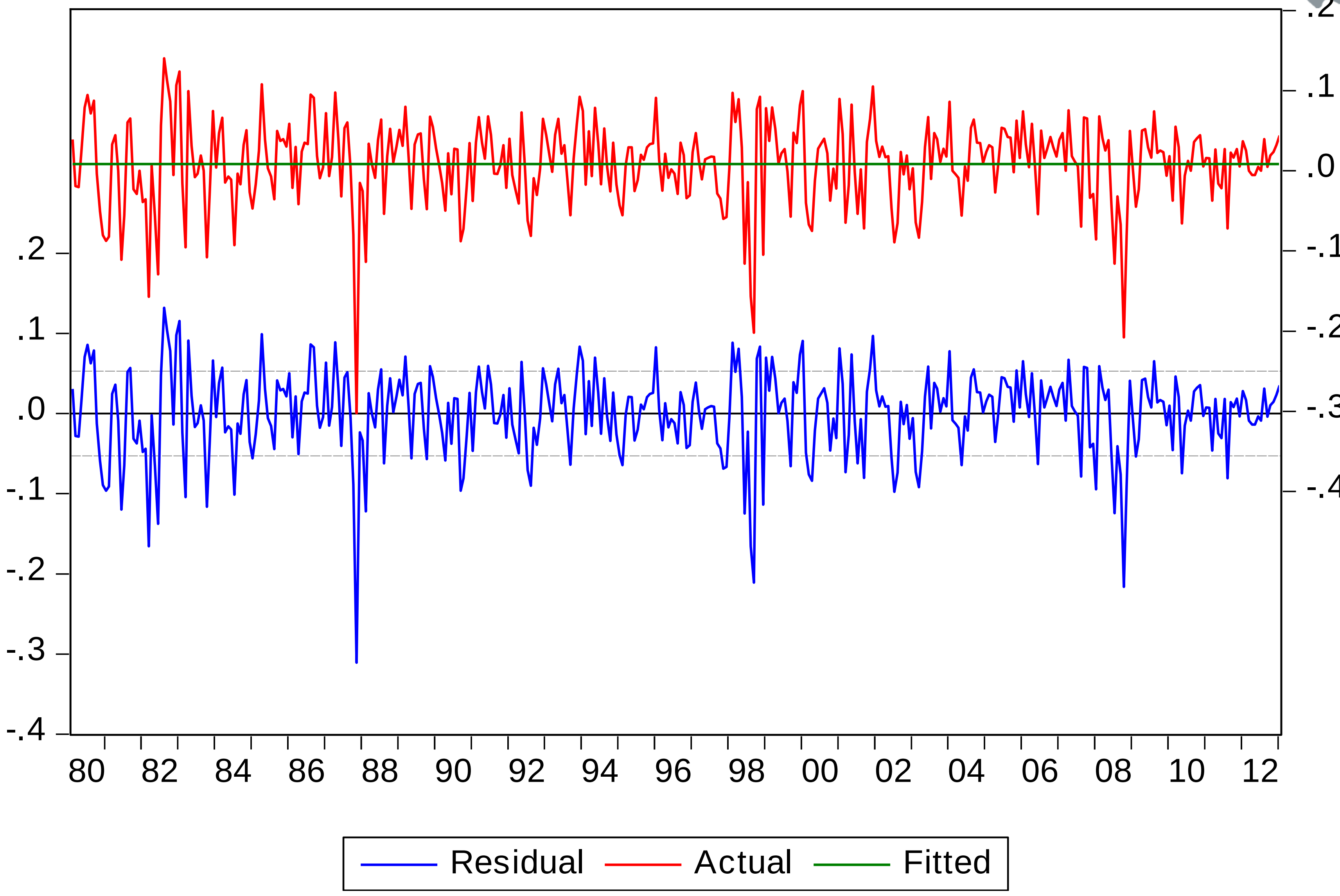
- We will be discussing Box and Jenkin's ARIMA family of models.
- Then we will touch on how we use it to evaluate impulse responses, and use it to forecast and how we can evaluate the forecasts.

Motivation for ARIMA

- After differencing our typically strongly persistent time-series (remember – this implicitly assumes a stochastic trend), we are left with a series that has its time-dependence removed to a large degree.
- But in many (even most) cases, the residuals still have strong persistence embedded in them.

- Thus although the ADF test may be rejected and the possibility of a second unit root can be eliminated (very few if any macroeconomic series behave explosively): we might likely still have <u>strong influences from past behaviour</u>…

Residuals of Dlog(Alsi)



Motivation for controlling persistence

- The *momentum* or persistence that we see at times in the residual series of the standard mean model of dlog(Y) tells us that there are still **some remaining serial autocorrelation** in the series…

- We can now control for this by including autoregressive components, called AR-parameters.

The hope is then that controlling for past dependence – we will remove any statistically visible patterns in the residual series prior to forecasting!

Conditions for stationarity

- Let: $L = lag\ operator$ so that: $L^k(y_t) = y_{t-k}$

- If we assume a **zero mean** stationary y_t, the **AR(p)** can be written as:

$$y_t = \sum_{i=1}^{P} \theta_i \,.\, L^i(y_t) + \varepsilon_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \varepsilon_t$$

$$y_t - \sum_{i=1}^{P} \theta_i \,.\, L^i(y_t) = \boldsymbol{y_t(1 - \theta_1 L^1 - \theta_2 L^2 - \cdots - \theta_P L^P) = \varepsilon_t}$$

Now, for the process to be **stationary**, we need to be able to write:

$$\boldsymbol{y_t = \left(1 - \theta_1 L^1 - \theta_2 L^2 - \cdots - \theta_P L^P\right)^{-1} \varepsilon_t}$$

This is known as the Wold Decomposition.

[Note that for stationarity, $\boldsymbol{\theta_p \rightarrow 0\ as\ P \rightarrow \infty}$]

Conditions for stationarity

- Now, for the characteristic equation's optimal solution (setting the FD equal to 0):

$$(1 - \theta_1 z^1 - \theta_2 z^2 - \cdots - \theta_P z^P) = 0$$

- The values of the $\boldsymbol{z's}$ that solve this solution must lie **outside the unit circle** for the series to be regarded as stationary: (i.e. $|z| > 1$)

- **Thus, if we have an AR(1) series, and the following optimal point is fitted:**

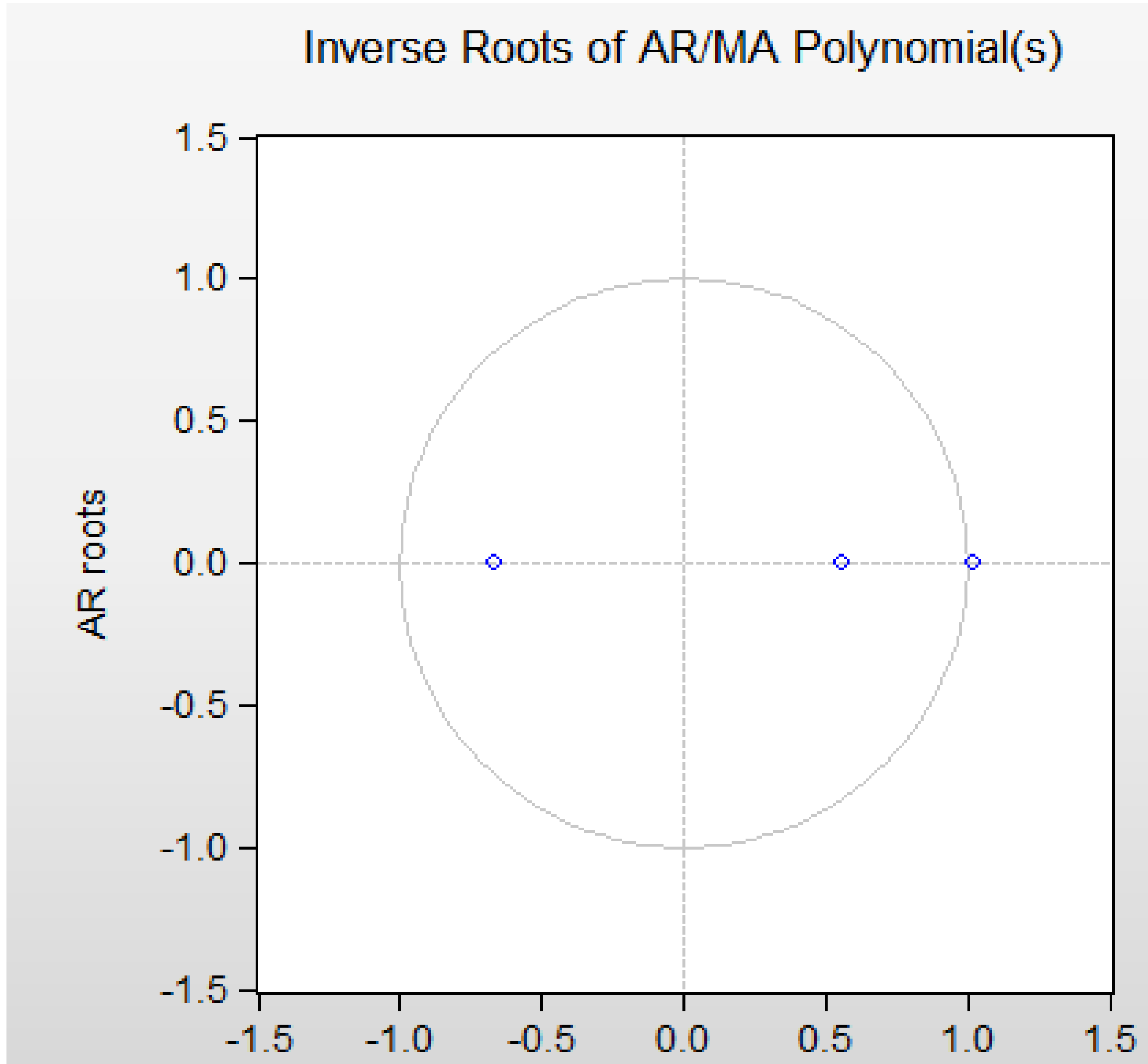
$$y_t = y_{t-1} + \varepsilon_t$$

- $(1 - \theta_1 z^1) = (1 - 1.z^1) = 0 \quad \therefore\ z = 1$, which violates the **stationarity principle**, as all solutions must be **outside** (not ***on*** or ***in*** the) unit circle: $z > 1$

a formal proof of this result is omitted for brevity

In Eviews: Non-stationary roots

- Eviews provides us with the ***Inverse of the Roots*** – which should then of course lie **within** the unit circle



Autoregressive process : Restrictions on β

- <u>Thus suppose we have three processes :</u>

- $y_t = 0.03 + 0.58y_{t-1} + u_t \rightarrow \boldsymbol{Stationary}\ AR - process$

- $y_t = 0.03 + y_{t-1} + u_t \rightarrow$ **Not–Stationary** $AR - process$ called a **Random Walk** process with a **drift**: **(known as a unit root process).**

- $y_t = 0.03 + 1.32\ y_{t-1} + u_t \rightarrow \boldsymbol{Explosive}\ AR - process$ (**not stationary). This process makes no practical sense, as previous shocks to the system tends to impact future values more and more…).**

- This implies it is important to test whether an AR(1)-process has a **unit root** (i.e. the $\beta = 1$). Such a series is **<u>non-stationary</u>**, as the $\boldsymbol{E}(.)$ **and** $\boldsymbol{V}(.)$ **<u>do not exist!</u>**
- Conducting formal unit-root tests require a different approach than the standard t-test.
  - This follows as such a time series model does not satisfy the conditions required for the test (even assymptotically it fails)

Characteristics of a stationary AR(P)

- Consider the following **stationary WN** AR(1) model:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$$
$$with\ \varepsilon_t \sim N(0, \sigma^2)$$

- The unconditional mean (i.e. not conditional on the past) would be:

$$E(y_t) = E(\mu + \phi_1 y_{t-1} + \varepsilon_t) = \mu + E(\phi_1 y_{t-1}) + 0$$
$$E(y_t) = \mu + E(\phi_1(\mu + \phi_1 y_{t-2} + \varepsilon_{t-1})) = \mu + \phi_1 \mu + \phi_1^2 E(y_{t-2})$$
$$E(y_t) = \mu + \phi_1 \mu + \cdots + \phi_1^N \mu + \phi_1{}^N E(y_{t-N})$$
$$E(y_t) = \mu(1 + \phi_1 + \phi_1^2 \ldots + \phi_1^N + \phi_1{}^N E(y_{t-N})$$

$\underline{NOW: if\ the\ model\ is\ stationary\ we\ must\ have}$:

$|\phi_1| < 1$ **so that**: $\phi_1^N \to 0\ as\ N \to \infty$ **and thus** $\phi_1^N E(y_{t-N}) \to 0$

Now, a basic algebra rule is that if $|x| < 1$, the series: $(1 + x + x^2 + x^3 + \cdots) \to \frac{1}{x}$

Thus, the unconditional mean of $\boldsymbol{y_t \to \ E(y_t) = \mu/(1 - \phi_1)}$

Characteristics of a stationary AR(P)

- This can be stated more generally that the unconditional mean of an AR(p) process is:

$$\boldsymbol{E(y_t) = \frac{\mu}{1 - \phi_1 - \phi_2 \ldots - \phi_P}}$$

If we assume the **mean** to be **zero** (as we can transform the data to have a zero mean) the **unconditional variance** of $AR(1)$ can be shown to be:

$$y_t = \frac{\sigma^2}{1 - \phi_1^2}$$

Autocovariances and Autocorrelations

- Remember that the **autocovariance** suggests how y_t varies with its lags. Thus the first order Autocovariance is given by:

$$\gamma_1 = cov(y_t, y_{t-1}) = E[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]$$

Second order Autocovariance:

$$\gamma_2 = cov(y_t, y_{t-2}) = E[y_t - E(y_t)][y_{t-2} - E(y_{t-2})]$$

The **Autocorrelation** is thus a **standardised version** (between -1 & 1) of the above.

We divide the Autocovariance (γ_P) by the Variance (γ_0):

Fist order Autocorr: $\rho_1 = \frac{\gamma_1}{\gamma_0}$, Second order Autocorr: $\rho_2 = \frac{\gamma_2}{\gamma_0}$

Kth order Autocorr: $\rho_K = \frac{\gamma_K}{\gamma_0}$

Autocorrelations of AR(1)

- It can then further be shown that the k^{th} **Autocovariance** of the AR(1) model as given previously (with a **mean of zero** for simplicity) is:

$$\gamma_K = \phi_1^K.(\frac{\sigma^2}{1-\phi_1^2})$$

Now, recall that the unconditional (full sample) variance : $\gamma_0 = (\frac{\sigma^2}{1-\phi_1^2})$

The **Autocorrelation** is then given by:

1st order Autocorr: $$\rho_1 = \frac{\gamma_1}{\gamma_0} = \phi_1^1.(\frac{\sigma^2}{1-\phi_1^2})/\gamma_0 = \phi_1^1.\frac{(\frac{\sigma^2}{1-\phi_1^2})}{(\frac{\sigma^2}{1-\phi_1^2})} = \boldsymbol{\phi_1}$$

Kth order Autocorr: $$\rho_K = \frac{\gamma_k}{\gamma_0} = \boldsymbol{\phi_1^K}$$

From the above, as $0 < \rho_k < 1$, it implies that $\phi_k \to 0 \; as \; k \to \infty$

Moving Averages

- Suppose we have a series y_t which seems stationary and has an unconditional mean of μ.
- It is then often the case that we see (at times) momentum in the residual series – i.e. a pattern of residuals emerging.
- Left untreated – such periods of residual autocorrelation may render parameter estimates biased and untrustworthy and may at times distort the ability of the model to forecast.
  - Such residual momentum then also violates the assumption of WN residuals needed for our model to be useful… How to proceed?

Moving Averages

- Note that controlling for autocorrelation using an autoregressive (AR) structure – essentially imposes an infinite future impact of today's residuals on future values of Y_t (albeit at an **exponentially decreasing rate of importance**).

- Moving Averages, on the other hand, allow the impact of a residual to be limited to, e.g. only the next observation.
- This is, intuitively, more fitting when we know that certain lags have definitive and contained impacts on contemporaneous values

Moving Averages

- We can then simply define y_t in terms of its past shocks. An example of such an extension is adding MA(2) components below:

$$y_t = \mu + \boldsymbol{\varepsilon_t} + \theta_1(\boldsymbol{\varepsilon_{t-1}}) + \theta_2(\boldsymbol{\varepsilon_{t-2}})$$

Thus, $y_t =$ described by a **linear combination** of **past disturbances:** $(\boldsymbol{\varepsilon_{t-1}}$ & $\boldsymbol{\varepsilon_{t-2}})$, and $\boldsymbol{\varepsilon_t}$ **only impacts** $\boldsymbol{Y_t}$ **for two periods…**

If we then set: $\boldsymbol{\varepsilon_t} + \theta_1(\boldsymbol{\varepsilon_{t-1}}) + \theta_2(\boldsymbol{\varepsilon_{t-2}}) = \boldsymbol{\lambda_t}$ **And if** $\boldsymbol{\lambda_t \sim WN}$:

our model $(y_t = \mu + \boldsymbol{\lambda_t})$ has been made a **Stationary** process with **WN** residuals $\boldsymbol{\lambda_t}$

- **<u>This is known as a second order Moving Average, or: MA(2)-model.</u>**

[Check for yourself that the MA(Q) model will be stationary now if all past residuals are stationary]

Combining AR and MA: ARIMA(pdq)

- Box & Jenkins then suggested fitting an ARIMA(p,d,q) model in order to control for serial dependence on the past after fitting an explanatory mean equation
- This way, we should ideally be left only with White Noise residuals.

- **NOTE::**
  - ARIMA models can be used with other regressors (or fundamental) models that aim to **explain** the time-series behaviour.
  - In such a model, ARIMA models would be applied to “misbehaving” **residual series** of fundamental models that show discernible patterns remaining in model residuals that cannot be explained by the included regressors.

ARiMAX

- When we have a structural model, with explanatory variables included in the mean equation (μ): Then we call such a structure an ARIMAX model.
  - For our purposes, we will just refer to all such models as being **ARIMA (i.e. drop the X)**
- NOTE: After doing certain basic transformations, high frequency data especially tends to represent naïve mean models – i.e. series where the dlog vary almost randomly around a constant mean… making fundamental analysis extremely difficult.
  - Of course, the literature on financial asset prices and currencies display such inherently unpredictable behaviour at higher frequencies… With the returns series behaving similar to Random Walk processes.

ARIMAX

- Conceptually, it should be intuitively clear that the impact on a macroeconomic variable of events which occurred in ***earlier periods*** (e.g., changes in various economic factors that make up the index of a leading indicator) is often most clearly represented in the prior history of **that variable itself**.
- Hence, lagged or contemporaneous values of other macroeconomic time series **may have little to add** to a univariate model which has already fully internalised its history– thus accounting for X in its own past – by taking the difference.
  - Thus **X would be most easily picked up** if it has a strong contemporaneous impact on Y not reflected in its past.

ARiMAX

- So does that mean we should throw away X and only focus on the past of Y?
- Not necessarily…
- If we are interested in establishing **causal** relationships over time – we use VARs and VECMs to establish such relationships.
- If we want to build a model to forecast Y – including a forecast of X might be very helpful: especially if X can be forecasted more easily.

Confucious says…

Confusion corner…

- When first confronted with ARIMA theory, it seems plausible that the following two equations are equal:

$$Y_t = \alpha + AR(1) + e_t$$

$$Y_t = \alpha + L(Y_t) + e_t = \alpha + \beta . y_{t-1} + e_t$$

Note – they are not the same…

Although the AR(1) term effectively also studies past behaviour as explaining contemporary levels (similar to a lagged variable)… It does so for the model's structure as a whole

The crucial difference is that with an AR-model you are NOT measuring whether <u>Y</u> follows an AR process, but rather that the <u>**residuals of the model**</u> are following an AR process

Re-written

- If we re-write the AR(1) system this may make more sense:
- Remember that any AR model can be rewritten in terms of auto-regressed residuals:

$$Y_t = \alpha + \mu_t$$

So that: $$\mu_t = \rho\mu_{t-1} + e_t$$

And substituting:

$$Y_t = \alpha + \rho(y_{t-1} - \alpha) + e_t$$

So that: $$Y_t = (1 - \rho)\alpha + \rho y_{t-1} + e_t$$

Which differs from the ordinary Y regressed on Y(-1) ito **only the intercept** coefficient (check this in Eviews). They are though, equivalent for only the AR(1) model…

Including regressors

- When including X, the difference between including an AR term (to control for residual persistence) and a lag (persistence in the series) becomes more visible. Consider the ARIMAX model:

AR(1) model: $Y_t = \alpha + \beta X_t + \mu_t$

$$\mu_t = \alpha + \rho \mu_{t-1} + e_t$$

This then becomes:

$$Y_t = (1 - \rho).\alpha + \rho Y_{t-1} + \beta(X_t - \rho X_{t-1}) + e_t$$

Which clearly is not the same as including a lagged variable of Y. Looking at the parameters estimated – it is clear that the impact of the past is accounted for in the **entire process** (including the lagged variables)

Including MA terms…

MA(1) model: $Y_t = \alpha + \beta X_t + \mu_t$

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

Thus: $Y_t = \alpha + \beta X_t + e_t + \theta e_{t-1}$

This estimation requires iterative non-linear fitting procedures (i.s.o. OLS), and as such is more complicated than fitting AR's.

Here shock-impact differs from AR models, in that they are propagated directly (appearing on the RHS of the equation), as opposed to indirectly through its influence on Y_{t-1}.

Thus an MA-model implies a direct "news" effect on Y_t, which dies away after q-lags.

In contrast, in an AR model the shocks have an infinite impact on Y_t (although it dies away).

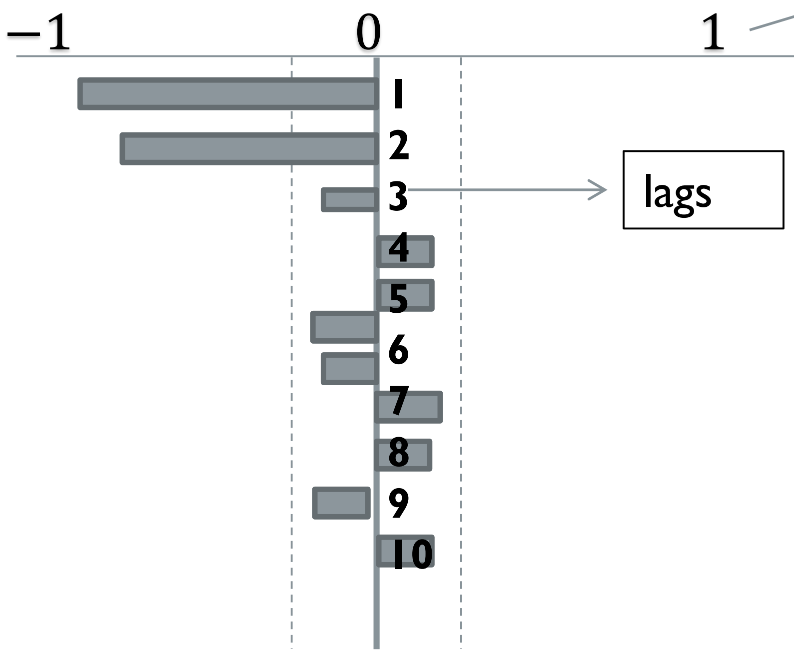
THAT IMPLIES: an AR(P) model can always be rewritten as an MA(∞) model.

Autocorrelation and the ACF

- The simplest way of describing temporal dependence left in the residuals is by viewing the sample Autocorrelation function (ACF) : ρ_k
- Plotting the ACF tells us, on average, *how much of* $\boldsymbol{e_t}$ *is explained by its* k^{th} *lag :* $\boldsymbol{e_{t-k}}$

- A strongly stationary process should **<u>have zero dependence</u>** on **all** past residuals, as it does not exhibit a trend and merely oscillates around a constant mean. <u>Its ACF should therefore be close to zero **at all lags** –</u> which would imply e_t is not correlated with any of its past errors on aggregate (which is the definition of stationarity).

Typical ACF Plot



- The ACF plots the autocorrelation coefficients of time lags of the **residuals** of the model explaining y_t (up to 10 in this illustration, normally about 28).
- The dotted lines ($\pm 2/\sqrt{T}$) provide critical values for testing independence at the 5% level for the estimated sample autocorrelation coefficients → so that if it lies within the lines, the relevant lag may be regarded as **statistically insignificant** in explaining e_t (i.e. uncorrelated with y_t)

- <u>In this example</u>, **the residuals of y_t is strongly dependent on its immediate past** (2 lagged parameter values are significantly different from 0). This implies the residual series of the model is **not** approximately WN.

Partial Autocorrelation Function (PACF)

- The partial autocorrelation function (PACF) of the residuals is the **conditional autocorrelation between** $\boldsymbol{e_t}$ **and** $\boldsymbol{e_{t-h}}$ $\boldsymbol{given\ e_{t-1}, \ldots, e_{t-h+1}}$, or written formally:

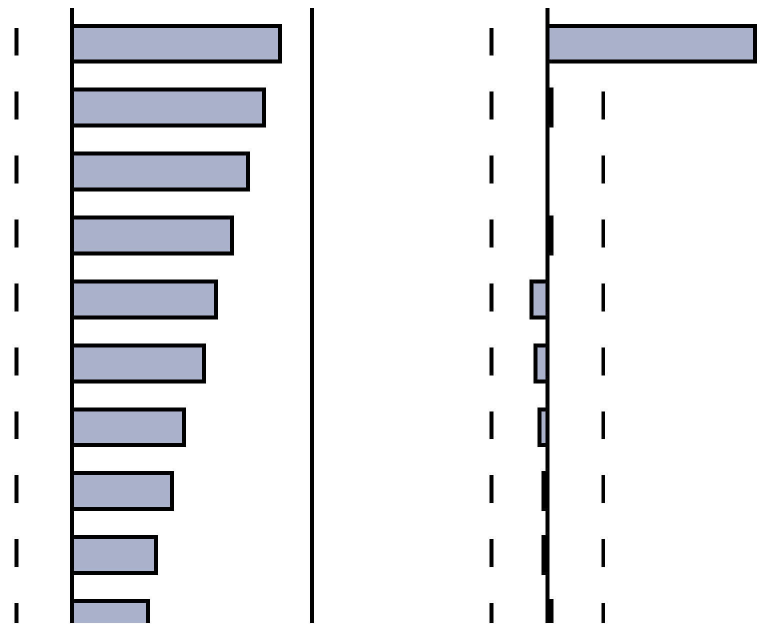
$$\boldsymbol{a_h = corr(e_t, e_{t-h} | e_t, \ldots, e_{t-h+1})}.$$

What this implies is very simply that: $\boldsymbol{a_h} \rightarrow$ measures the correlation between the **contemporary residuals** (e_t) and the model's **residual** $\mathbf{h-}$ **periods ago**: (e_{t-h}), given all information up to that lag (i.e. all info on y for all lags $< h$).

Thus it effectively **isolates** the **correlation** between e_t & e_{t-h}, after removing the effects of [e_{t-1} , … , e_{t-h+1}] … Of course, if $\mathrm{h} = 1 \rightarrow$ $\boldsymbol{ACF = PACF}$

ACF & PACF : How the Eviews output looks

What is this saying?



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.912	0.912	59.144	0.000
		2	0.836	0.021	109.55	0.000
		3	0.766	0.003	152.57	0.000
		4	0.704	0.010	189.46	0.000
		5	0.636	-0.069	219.98	0.000
		6	0.566	-0.049	244.55	0.000
		7	0.499	-0.029	263.96	0.000
		8	0.436	-0.024	279.01	0.000
		9	0.375	-0.024	290.36	0.000
		10	0.324	0.021	299.00	0.000

Ljung-Box Q-Stat

- Remember: what we are trying to do is verify **whether our model is any good at explaining the past of y_t** by fitting the model $F(Y_t)$
- We do this by looking at whether there remains any ***persistence*** in the residual terms. Remember that we need to be sure that the residuals are **white noise**, else our estimates will be biased!!
  - If the residual process is **White Noise**, both Autocorrelations and Partial Autocorrelations should $\rightarrow 0$ (as per definition).

Ljung-Box Q-Stat

- For both these ACF and PACF statistics, the **Ljung-Box Q-statistics** are given as a way of testing **formally** whether a series displays significant remaining autocorrelation.
  - This is often referred to as the Ljung Box **Q-stats**, and have χ^2 distributions.
  - It is normally given next to the correlogram, together with its $\boldsymbol{p-values}$
- The LBQ-stat is calculated as:

$$T(T+2)\sum_{k=1}^{n}\frac{\rho_k^2}{T-k} \sim \chi_n^2$$

Ljung-Box Q-Stat

<u>The $p - value$ is the probability that the null hypothesis of stationarity holds</u>:

H_0: the first k-lags of the cumulative autocorrelation coefficients $= 0$

<u>This implies</u>: H_0: ε_t *is* ***independently distributed with the past*** *(i.e.* ***no serial correlation*** *exists with the past)* → ***<u>implying White Noise Stationarity</u>***

Thus if $\boldsymbol{p < 0.05}$ → we **reject** (at the 95%-level) the $\boldsymbol{H_0}$ of **no significant autocorrelation**, and hence the series is **non-stationary**

Note then, that what we seek is low $\boldsymbol{Q}$ and High $\boldsymbol{p - values}$ for WN.

What's the previous ACF and PACF saying?

- Consider again the Eviews outputs of the ACF and PACF plots a few slides back.
- It is suggesting that there is **significant** serial autocorrelation with the past, but that the Partial Autocorrelation **dies out (or cuts-off)** after the first lag.
- Notice that the ACF dies off slower, but gradually…
  - What does this imply about the persistence of the residuals?

What's the previous ACF and PACF saying?

- This probably implies **very strong first order persistence** (i.e. a unit root, or close to it, being present), after which the relative strength of longer lags' persistence is smaller, as suggested by the partial autocorrelation dying out after the first lag.

- THUS: The significant ACF that is picked up at, say, the fifth lag – is likely picking up first order correlation with the subsequent lags that only dies out slowly over time (as a shock today impacts all the future processes, but with less and less impact), and not specific ***partial*** correlation with the fifth lag itself…

Fitting ARIMA models

- In the next subsection we will thus be discussing how to use the ACF and PACF figures to guide our fitting of models that seek to control for such serial residual dependence.

- The example just discussed (cutting off PACF and dying out ACF): probably suggests an AR process… as AR models have an infinite (yet shrinking) impact of past residuals – something that is not reflected in the PACF

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Fitting

ARIMA-models

ARMA models

- If we combine the AR(P) and MA(Q) processes discussed earlier, we get an ARMA(P,Q) process describing the evolution of a time-series' non-stationary (persistent) residual process.
- <u>This can be represented as:</u>

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_1 \, y_{t-2} + \cdots + \beta_1 y_{t-p} + \epsilon_t + \alpha_1 \epsilon_{t-1} + \cdots + \alpha_q \epsilon_{t-q}$$

Or:

$$A(L) Y_t = B(L) \varepsilon_t$$

With $A(L) = 1 - \alpha_1 L - \cdots \alpha_p L^p$ and $L^p(Y_t) = Y_{t-p}$ and $A \& B \rightarrow param's$

Again, stationarity conditions require the roots of $A(L) \; and \; B(L)$ to lie **outside** the unit circle $(|z| > 1)$

Box and Jenkins approach

- Box and Jenkins suggest a three step approach to fitting an ARMA model:

- 1) Identification
- 2) Estimation
- 3) Diagnostic checking

Example of an AR(1) process…

Residual auto-correlation measures:

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.387	0.387	56.063	0.000
		2	0.196	0.054	70.477	0.000
		3	0.120	0.032	75.921	0.000
		4	0.031	-0.040	76.275	0.000
		5	-0.078	-0.104	78.554	0.000
		6	-0.028	0.039	78.860	0.000
		7	0.071	0.110	80.774	0.000
		8	0.077	0.036	83.045	0.000
		9	0.057	-0.003	84.301	0.000
		10	0.071	0.017	86.214	0.000
		11	0.017	-0.038	86.321	0.000
		12	-0.076	-0.078	88.543	0.000
		13	-0.029	0.044	88.879	0.000
		14	-0.084	-0.079	91.586	0.000
		15	-0.011	0.065	91.634	0.000
		16	-0.011	-0.019	91.679	0.000
		17	-0.009	-0.029	91.709	0.000
		18	0.019	0.028	91.851	0.000
		19	-0.045	-0.073	92.647	0.000

Geometric decay

Cut-off at lag 1

Thus it probably requires an AR(1) type model

AR selection

- The previous correlogram we say has an **AR-signature**, in that it suggests we include **AR** terms (PACF cuts off, ACF dies out slowly).
- Most often this would be associated with a **positive ACF value at lag 1** – implying the series is slightly *underdifferenced* (thus the AR term almost acts like a **partial** difference).
- The fact that ACF doesn't die out – is signature to AR as well, as an AR model has an infinite (yet shrinking) impact of past residuals – something that is not reflected in the PACF!

MA Selection

- Note that in theory we could just keep adding AR terms until the residuals behave well. **<u>This is bad econometrics though…</u>**
- A better alternative might be to add MA terms: here we look at the PACF, which plays the same role as ACF for AR's.

- **MA-terms** imply the **direct impact of past residuals on the contemporaneous model**: and note the **<u>PACF</u> suggests** the direct impact of a residual (or shock) at a lag k, directly attributed to the residual.
  - Typically, MA signature data have a **<u>negative ACF</u>** at the first lag… implying MA's are most needed in slightly ***<u>over</u>*differenced** series.
  - And / or it has an ACF cut-off point – suggesting the amount of MA terms to be included

1) Identification: Visual indication of the lags

- How do go about deciding how many lags (for P and Q) to include in order to fit the most parsimonious ARMA model?
- We can employ a visual analysis of the ACF and PACF plots – noting the following:

Series	ACF Plot	PACF Plot
AR(P)	Decays slowly to zero	Cuts-off after lag P
MA(Q)	Cuts-off after lag Q	Decays slowly to zero
ARMA(P,Q)	Decays to zero	Decays to zero.

- Note: If ACF cuts-off: it implies that the residuals do not have a slowly decaying nature as would be suggested by the design of an AR model – where residual impacts die over time.

2) Estimation

- This step involves estimating the values for the parameters fitted by the model. Once fitted, the modeller should check for stationarity in the roots of the model and for serial correlation in the remaining residuals.
- As ARMA models employ Maximum Likelihood Estimation (**MLE**) techniques if it has a MA component, it can be a laboriously arbitrary exercise of choosing the lags which maximize the MLE function.
- For this reason, we use (in addition to the visual techniques which provide an indication of which lags to use as a starting point) certain **selection criteria** to evaluate the goodness of fit and compare different models' explanatory power…

The constant?

- The constant represents the ***mean* of the series if no differencing is performed**, it represents the ***average trend*** in the series **if one order of differencing is used**, and it represents that ***average trend-in-the-trend*** (i.e., curvature) if there are two orders of differencing.
- Thus in a model with one order of differencing (e.g. Dlog'd data), the constant may or may not be included, depending on whether we do or do not want to allow for an *average trend* over time.

- As a rule of thumb: **In a model with <u>one</u> order of total differencing, a constant term should be included if the series has a <u>non-zero</u> average trend.**

HOW DO WE PROCEED in selection?!

- Care must be taken when fitting ARIMA models.
- In particular, the coefficients AR & MA **can cancel out** each-others' effects – even if they are significant {this can be shown by e.g. comparing an ARIMA(1,1,2) and ARIMA(0,1,1)}.
- Also NB: **ARIMA models cannot be identified in backward-stepwise fashion** as OLS models can. Instead, it requires "**forward"estimation** – **adding** terms and comparing fits.
- Remember too: $\sum AR\ terms < 1$ → else it shows a unit root.
- This also applies to the MA parameters…

3) Diagnostic Checking

- The information criteria discussed below provides an objective indication of the best model fitted – as opposed to the subjective visual indication of the ACF and PACFs.
- Three most popular information criteria used to compare models with different lags include:

$$Akaike\ info\ criterion - AIC(p,q) = \ln\left(\hat{\sigma}^2(p,q)\right) + \frac{2}{T}(p+q+1)$$

$$Schwarz\ criterion - SBIC(p,q) = \ln\left(\hat{\sigma}^2(p,q)\right) + \frac{\ln(T)}{T}(p+q+1)$$

$$Hana\ Quin\ - HQIC(p,q) = \ln\left(\hat{\sigma}^2(p,q)\right) + \frac{\ln(\ln(T))}{T}2.(p+q+1)$$

Using the diagnostic checks

- Ideally, we want the criteria (AIC, SBIC, HQIC) to be **as small as possible**

  i.e. AIC, SBIC, HQIC $\rightarrow -\infty$

- Comparing two models, A and B, with A's AIC = -3.2 and B's AIC = -2.9 → we should choose model A on the basis of this specification of the goodness-of-fit.
- These measures punish the modeller for adding regressors with no significant explanatory power, with the SBC giving a higher cost (i.t.o. higher SBC value) for adding redundant variables to the model.
- Although used together and mostly providing the same conclusions → **SBC** is **asymptotically consistent** (with AIC more biased toward over-parameterized), while **AIC** performs **better** on **smaller samples**.

ARIMA modelling: Steps summary

- If a series is integrated to the first order (i.e. has a unit root) – then remember from the last session that we need to take a **FD** to remove the near perfect persistence before we can do anything else.
- We **then** fit an explanatory model $\boldsymbol{\mu_t}$
- **Then** we should check the ACFs and PACFs for any **remaining** serial correlation in the residual series…
- If it is found, fit an ARMA(P, Q) model in **forward estimation** to control for such remaining persistence, until we are left with residuals that are approximately White Noise Stationary.
- Such a model would then be called an $\boldsymbol{ARIMAX}(P, 1, Q)$

Accuracy of ARMA models

- For many time series models, fitting an ARMA process results in accurate fits of past values and provides relatively good future forecasts (especially for high frequency data and short forecast periods).
- However, **only** using the past values of a time series would be an incredibly inaccurate exercise, void of explanatory power to the modeller (as occurrences outside the historic framework cannot be predicted).
  - **Black Swan** events (especially in finance) are not picked up by models based solely on historic data.
- Hence the need to include other variables in the model to better explain our target time series' behaviour over time.
- This will be returned to in our VAR analyses…

Example in Eviews of ARIMA(2,1,1)

- **Type in Eviews:** Ls dlg c ar(1) ar(2) ma(1) :

Dependent Variable: DLG
Method: Least Squares
Date: 02/12/13 Time: 19:59
Sample (adjusted): 1995Q1 2013Q1
Included observations: 73 after adjustments
Convergence achieved after 24 iterations
MA Backcast: 1994Q4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009699	0.002229	4.351191	0.0000
AR(1)	0.429072	0.176039	2.437370	0.0174
AR(2)	0.273751	0.115469	2.370779	0.0205
MA(1)	-0.496177	0.172267	-2.880285	0.0053

R-squared	0.251575	Mean dependent var	0.008090
Adjusted R-squared	0.219035	S.D. dependent var	0.011983
S.E. of regression	0.010590	Akaike info criterion	-6.204660
Sum squared resid	0.007738	Schwarz criterion	-6.079155
Log likelihood	230.4701	Hannan-Quinn criter.	-6.154644
F-statistic	7.731203	Durbin-Watson stat	1.878352
Prob(F-statistic)	0.000159		

Inverted AR Roots	.78	-.35
Inverted MA Roots	.50	

Checking whether our model is stable

- Remember from the last session that in order for an autoregressive process (and in this case, having potentially both AR and MA terms) to be regarded as stationary (note that it does not imply that ***the series*** y_t is necessarily made stationary, rather this is referring to the ***model*** describing it being a stationary process…)– the roots of the system must lie **outside** the unit circle.

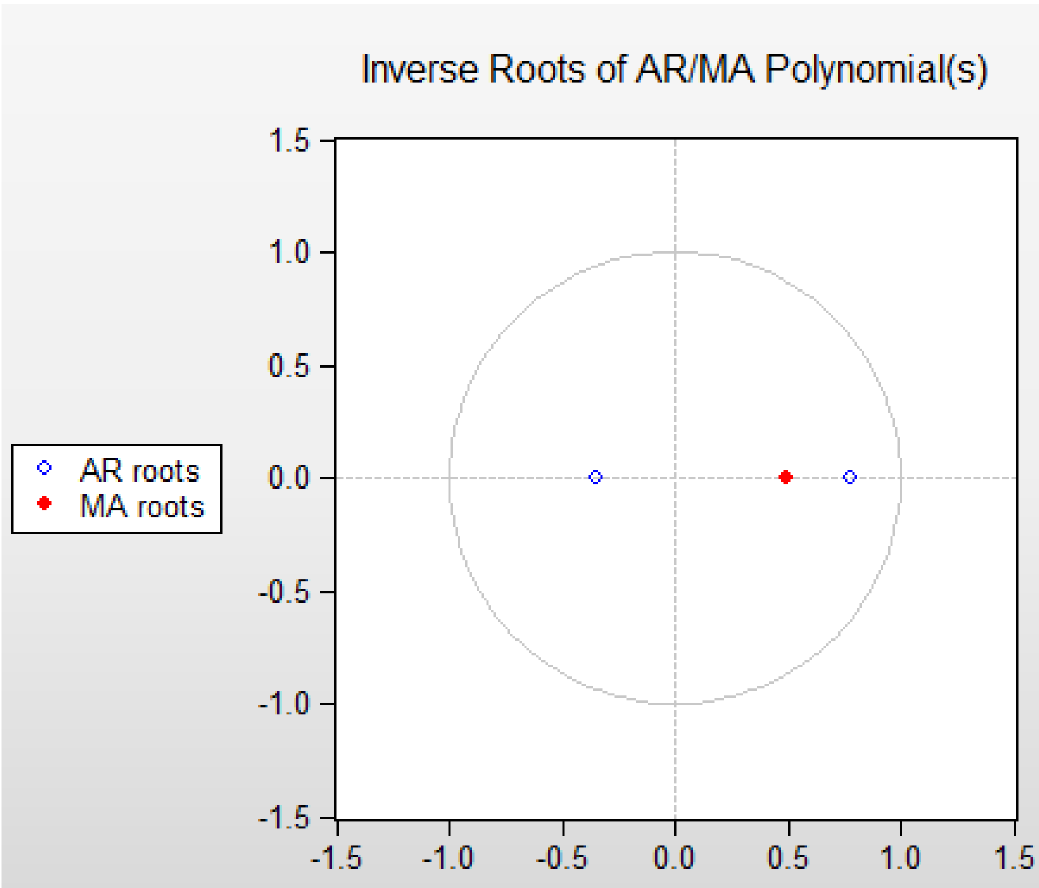
  - Eviews allows us to easily check whether the roots of the system lie outside the unit circle… But it tests for the **inverse roots** lying **inside** the unit circle (easier to assess visually).

Note that this implies the same thing mathematically:

$$X > 1 \quad \leftrightarrow \quad X^{-1} < 1$$

Checking stationarity of ARIMA(2,1,2) model…

- Select: View – ARMA structure to get the following:



REMEMBER:

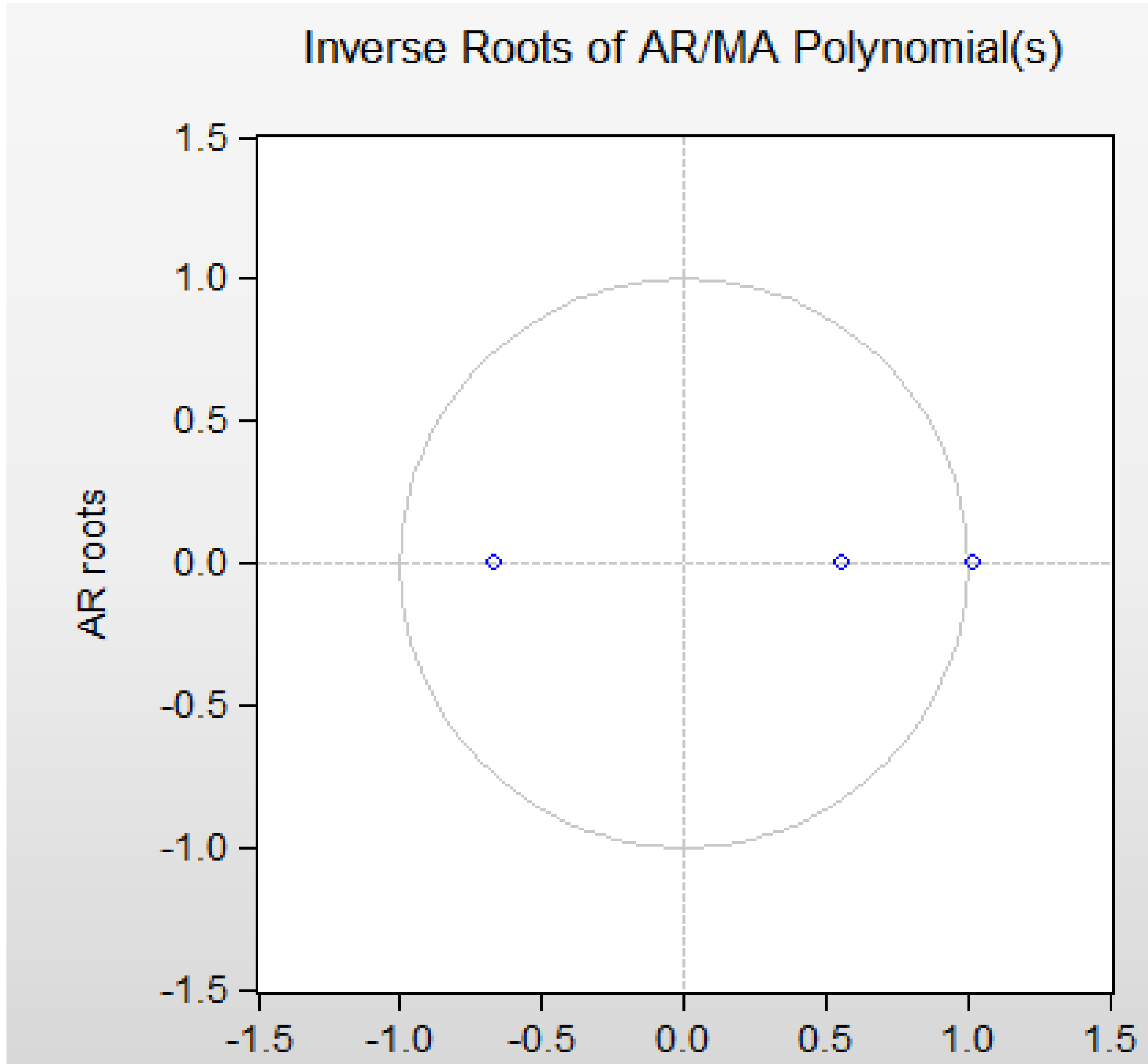
The roots of the solution must lie **outside** the unit circle for stationarity…

Which implies the **INVERSE** of the roots **must lie within** the unit circle…

From the figure on the left, the **model** is stationary

Non-stationary roots

- Notice that one of the roots lies on the unit circle (all should lie **within** for stationarity). This implies that the process probably has a unit root and perhaps requires a FD.
- Go do an ADF to test for this formally…



- Notice that we could also choose to view the result in table form, providing a straight forward interpretation from it:

Inverse Roots of AR/MA Polynomial(s)
Specification: D(Y1) C AR(1) AR(2) MA(1)
Date: 04/05/13 Time: 10:10
Sample: 1990M04 2013M01
Included observations: 274

AR Root(s)	Modulus	Cycle
-0.012758 ± 0.536398i	0.536549	3.940346

No root lies outside the unit circle.
ARMA model is stationary.

MA Root(s)	Modulus	Cycle
0.367410	0.367410	

No root lies outside the unit circle.
ARMA model is invertible.

SARIMA

- Another add-on to our model could be a seasonal dummy SAR(p) – which implies controlling for a p^{th} order systematic seasonal correlation – which could e.g. be controlling for a twelve month correlation (year on year factors if data = monthly), or $p = 5$ could control for weekly repetitions.

Dependent Variable: DLY1
Method: Least Squares
Date: 02/12/13 Time: 21:07
Sample (adjusted): 1995M09 2012M01
Included observations: 197 after adjustments
Convergence achieved after 16 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004032	0.002556	1.577327	0.1164
AR(1)	-0.359278	0.073729	-4.872973	0.0000
AR(2)	0.156026	0.091692	1.701631	0.0904
SAR(2)	-0.659539	0.070626	-9.338504	0.0000

R-squared	0.307275	Mean dependent var	0.004054
Adjusted R-squared	0.296507	S.D. dependent var	0.085411
S.E. of regression	0.071638	Akaike info criterion	-2.414276
Sum squared resid	0.990487	Schwarz criterion	-2.347612
Log likelihood	241.8062	Hannan-Quinn criter.	-2.387290
F-statistic	28.53656	Durbin-Watson stat	1.984543
Prob(F-statistic)	0.000000		

Inverted AR Roots	.25

SARIMA

- A seasonal factor could also be included for the MA term – SMA.
- In terms of the lag operators this can be written as:
- SARIMA(2,1,0,12,0)

$$(1-\rho_1 L-\rho_2 L^2)(1-\theta L^{12})\ u_t=\varepsilon_t$$

For a SARIMA(2,1,0,12,4):

$$(1-\rho_1 L-\rho_2 L^2)(1-\theta L^{12})\ u_t=(1-\theta L^4)\varepsilon_t$$

Why the low R^2 with ARIMA models?

- Why in time-series is the R^2 so low? And why do we care less about it than with cross-sectional regressions?
- This follows, as we often model **differenced** (e.g. dlog) dependent variables in our time-series regressions. By definition, this implies that *much of what is explainable has been removed*, and a large component of what remains is noise.
- Thus, although AR and MA terms might be strongly significant – much of the period to period variations are unpredictable, hence the low R^2 (and hence it not being that important).
- Explaining levels with auto-lags of the same series would then, mostly, yield a higher R^2 - as explaining levels as opposed to incremental changes is typically more achievable with macroeconomic data.

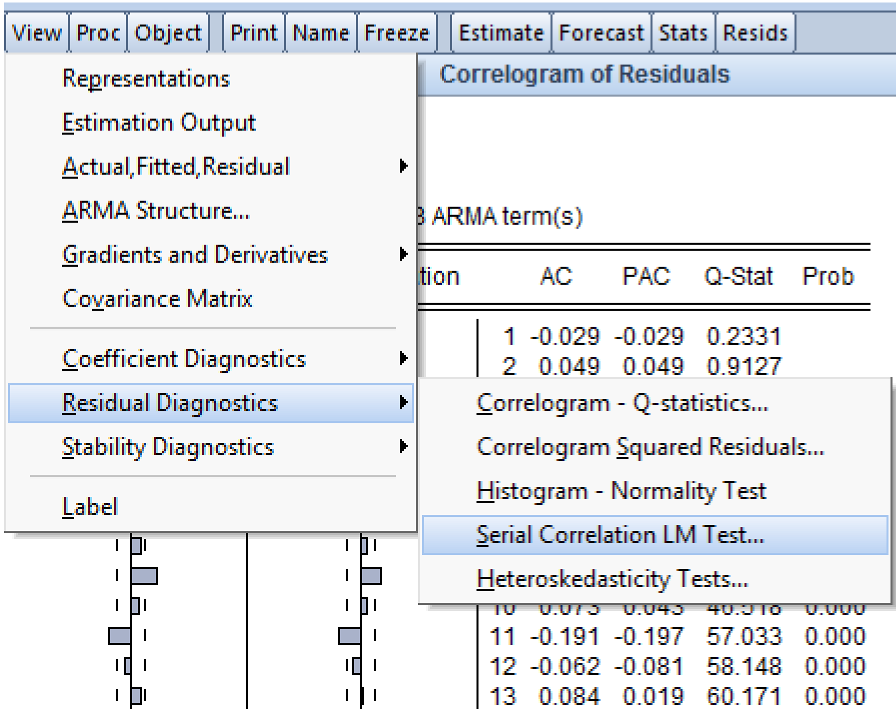
Testing for remaining serial correlation

- How do we now know that all possible serial correlation with the past is effectively controlled for by our model?

- Graphically: View the correlogram and check that the residuals all lie within the dotted lines and that the Ljung Box $Q - stat's\ p - values$ suggest that there remains no significant correlation.

- A more formal way would be to conduct a **Breusch-Godfrey LM-test** to check whether there remains any remaining serial correlation in the residual series of our estimated model…

LM test



Breusch-Godfrey Serial Correlation LM Test:

F-statistic	16.01750	Prob. F(2,268)	0.0000
Obs*R-squared	29.25521	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 04/05/13 Time: 10:18
Sample: 1990M04 2013M01
Included observations: 274
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.001300	0.263005	-0.004941	0.9961
AR(1)	2.128319	0.443137	4.802848	0.0000
AR(2)	0.205318	0.259577	0.790969	0.4297
MA(1)	4.420189	0.998117	4.428529	0.0000
RESID(-1)	-6.627854	1.426712	-4.645545	0.0000
RESID(-2)	-1.729025	0.567422	-3.047158	0.0025

R-squared	0.106771	Mean dependent var	0.001843
Adjusted R-squared	0.090106	S.D. dependent var	9.453299
S.E. of regression	9.017347	Akaike info criterion	7.257832
Sum squared resid	21791.76	Schwarz criterion	7.336952
Log likelihood	-988.3230	Hannan-Quinn criter.	7.289588
F-statistic	6.406997	Durbin-Watson stat	1.970979
Prob(F-statistic)	0.000012		

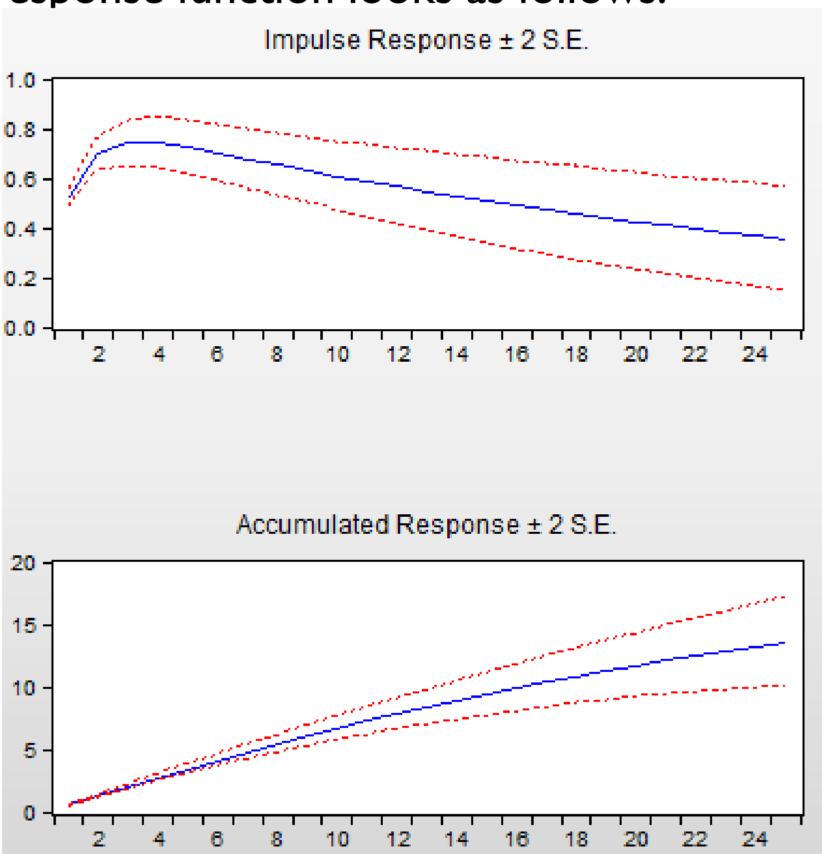
- The Breusch Godfrey **LM test** in this example suggests that there remains significant serial autocorrelation in the series ($p < 0.05$)
- Specifically this suggests that we need to either include another AR / MA term, or another variable to help explain past innovations of y_t.
- **If, however,** the LM test shows a joint F distr with $p > 0.05$ – It implies that there are **no serial correlation remaining**.

Impulse responses of the series

- Eviews also allows us to view the predicted impulse responses of the model to shocks.
- Thus we can estimate, e.g., the accumulated response of the model to a one standard-deviation (σ) shock to the system.
  - Seeing how it impacts future values of the series is useful in deducing how a large shock would impact the system.

Impulse responses of the series

- We can check the impulse response function of the ARIMA model by selecting: ***View – ARMA structure – Impulse Response.***
- Suppose we test for the impact of a one standard deviation shock. The impulse response function looks as follows:



The blue line is the expected impulse response for 24 lags, while the red line is the confidence interval (+/− 2 S.E.).

The impulse response is the impact of the shock on future values, while the accumulated response is the total response to the system following a $\sigma - shock$ at time t

Forecasting with our ARIMA model…

- Remember from our stated objectives for time-series analysis that we intend to use our model to **describe the past** (and uncover useful relationships), and then [very important for practical purpose] use the model to **forecast** the **future**.

Forecasting

- Broadly, **forecasting** can be **based** either **on quantitative** or **qualitative** features.
- Qualitative approaches include looking at typical movements (e.g. subjective measures like trough and peak predictions, typical or similar product life cycles, the **Delphi method**, experience, intuition, etc.).
- Our main interest for this course lies in the quantitative approach – which seeks to find a **statistical means of forecasting** (used in motivating our qualitative premise for forecast evaluation).

Forecasting

- Qualitative forecasting can be split between ***univariate*** *forecasting models*, which predict based solely on past values of the time series of interest (i.e. using an autoregressive approach) and ***causal*** *forecasting models,* which include other explanatory series to explain the relationship of the variable of interest.
- There are no clear consensus as to which of the methods report the most accurate results *ex-ante*. A combination of quantitative and qualitative methods are of course ideal, but where should the emphasis lie?
- It also depends on the frequency and forecasting time frame. In particular, for shorter forecasting periods univariate models can be regarded as more accurate (most of the time at least).

Forecasting

- After deciding on the model(s) used to forecast a time series, we can proceed to make both point estimates and prediction interval forecasts.

- At times, due to the inherently unpredictable nature of a series being forecast, making **point estimate** forecasts are of little value – hence the **interval** that the variable's values might likely be in is of most interest (as an example –are we more interested in the inflation band [say 5-6.5% for next year] that the SARB forecasts, than in their point forecast of 6.32% inflation for next year)?

Forecasting using Iteration

- There are various methods used by time series econometricians in order to forecast time series data. The most intuitive method is using the process of iteration.
- This involves fitting a model describing y_t using past data, and then forecasting K-periods ahead, each time using the previous forecasted data point.
- E.g., forecasting 2 periods ahead using iteration, if a function f(.) explains y_t:

$$\widehat{y_{t+1}} \approx f(y_t)$$

$$\widehat{y_{t+2}} \approx f(\widehat{y_{t+1}})$$

$$\widehat{y_{t+3}} \approx f(\widehat{y_{t+2}})$$

It is clear then that if y_t is known at time (t), forecasting ahead becomes less accurate each period, as the subsequent forecasts all use estimated (not true) values of y_t. [And thus includes the sum of all previous error terms]

Also, if f(.) is a function which includes **other variables**, it adds to the uncertainty as their values **ALSO** need to be approximated!

Evaluating forecasting performance

- Since we cannot observe forecasted values in the future at the current period (we only know whether our forecasts were accurate *ex post* which is hardly ever looked back on!?*)*, how can we evaluate the accuracy of our model in forecasting future values at the current date to motivate the use of our model in forecasting to the future?

- The easiest way we can evaluate a model's forecasting performance is by comparing our model forecast to **historic data**, fitting our models in previous time periods *as if we were at such a historic period*, and forecasting ahead as normal. This allows us to check the robustness of our model's forecast ability, by comparing the estimate to the true past outcome that we know.

Evaluating forecasting performance

- The method of using historic data to test the predictability of our model is called **in-sample** testing.

- Of course, we would hope that there is no clear pattern in the errors, as this would contradict our assumption of only an irregular component remaining in our transformed data series (or stationary process with WN errors)
- Note that there is also no way to know whether a good in-sample forecast performance will lead to a good out-of-sample performance…

Evaluating forecasting performance

- Our in-sample test would then provide us with an error:

$$\varepsilon_{t+h} = Y_{t+h} - \widehat{Y_{t+h}} = forecast\ error$$

- Graphing the error terms should give us an indication of whether there is still a trend / seasonal component present in the series, which we should then take into account before making any formal forecasts on it!
  - **NB to note** → even if all the measurable components are accounted for and only the irregular pattern remain, the irregular component might still be incredibly large so that ***accurate*** forecasting is not at all possible.
  - ***Are financial asset pricing such an example?***

Forecasting with our ARIMA model

- Let's consider two ways of conducting a forecast:
  - **Dynamic Forecasts**: Starts from the first date in the forecast period (2010) and forecasts to the last (2014). Here we use previously forecasted values of y_t and forward estimates of x_t as inputs to the model (the iteration procedure).
  - **Static Forecasts**: Calculates a sequence of one-step ahead forecasts using the actual (not forecasted) values for the lagged dependent variables (i.e. the true previous period values: y_{t-1} and x_{t-1} as inputs to the model describing y_t for each period)– thus no iteration. This process checks the immediate forecasts of the series.
- It is therefore glaringly obvious that our **static forecasts** will be "more accurate" (as the model corrects its past mistakes and using updated data, while a dynamic forecast works with its errors).
- Depending on what is forecasted, what is tested (forecastability or one-period ahead precision) and for how far into the future, the choice between Dynamic and Static is made.

Let’s consider the difference

- Suppose we study South Africa’s imports, and decide to fit an ARIMAX model to explain the past.
- We thus take the log of Imports (we should of course really include several explanatory factors as well) and then decide to include interest rates as explanatory factors, and then fit an ARIMA(2,1,0) to control for remaining serial dependence on past residuals.
- We then seek to evaluate its forecasting ability by testing **it in-sample** (for the unfortunate period of 2008q1 – 2013q1) by assuming that interest rates stay the same for the forecast period when using the dynamic forecasts.

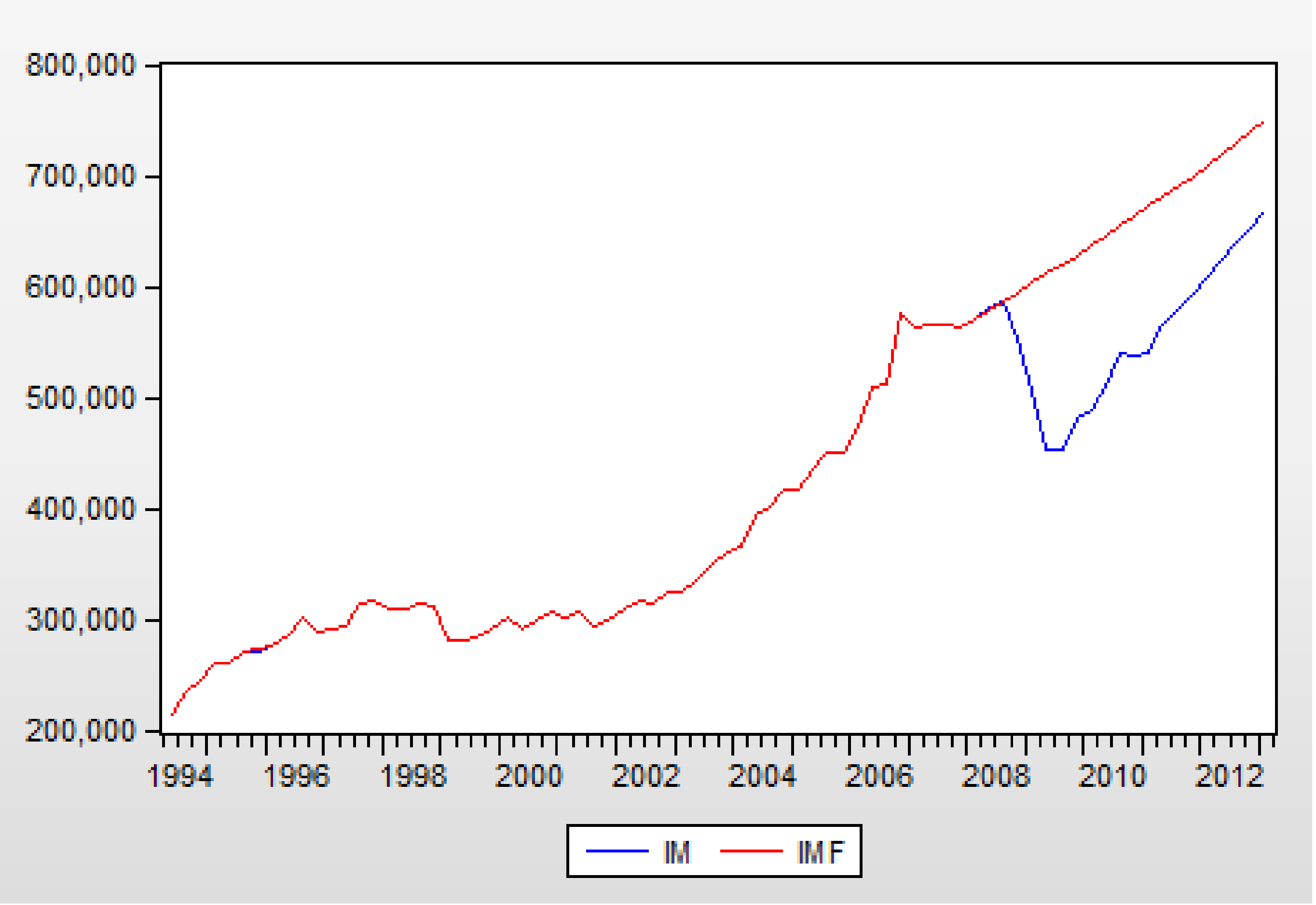
Let’s look at the results…

Dynamic

- Consider a **<u>Dynamic Forecast</u>** performance of an in-sample test of the Imports in SA for the period: 2008q1 – 2013q1.

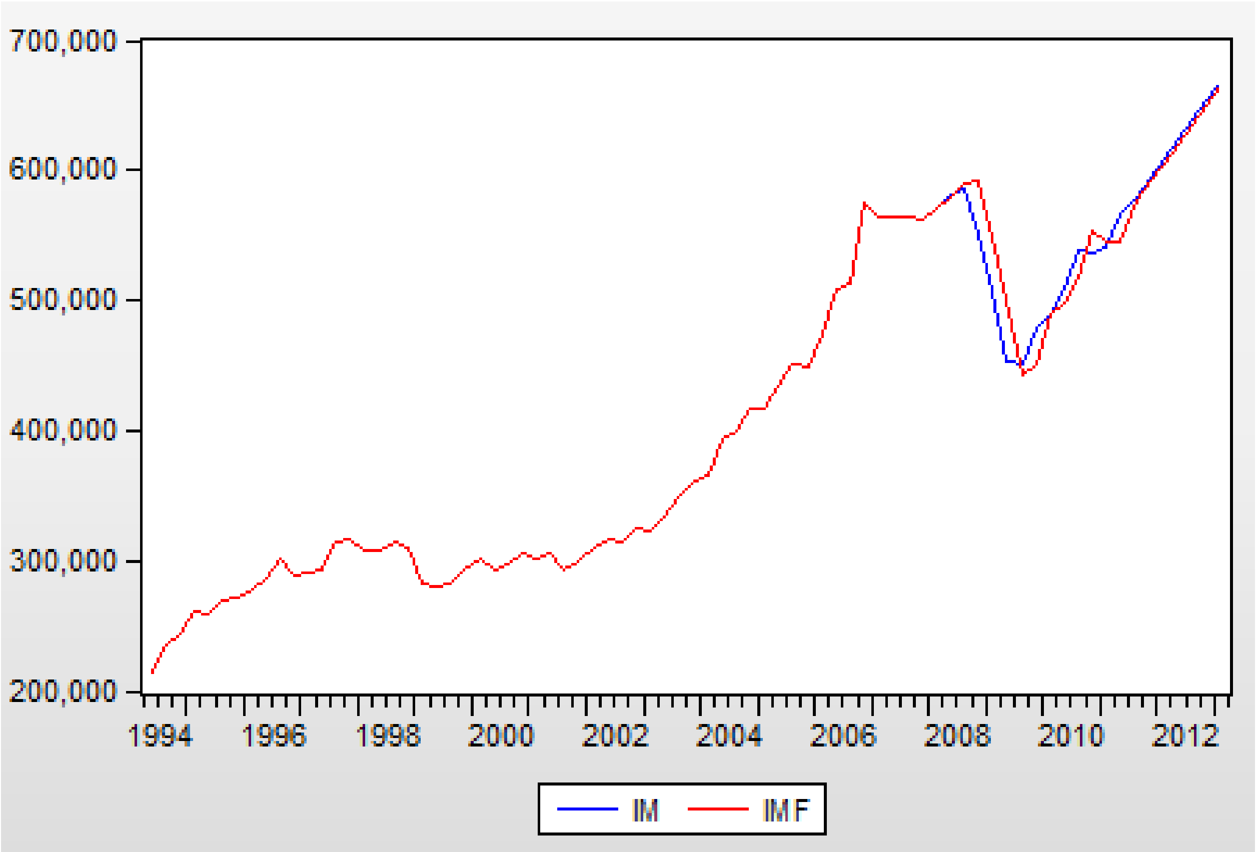
- **Forecast:** Red-line
- **True outcome:** Blue Line
- (Notice that the model did not pick up the recession in

2008 – 2009 for obvious reasons)

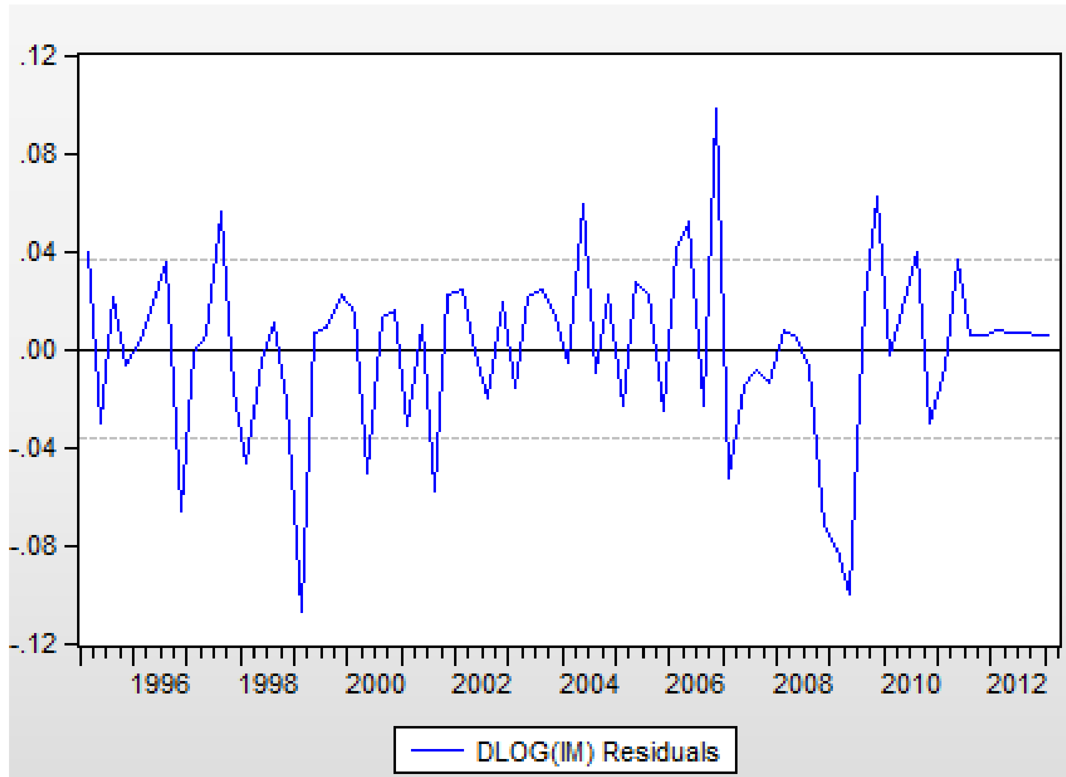


Static

- The **static forecast** picked it up (with a lag), because the inputs of the series (the lags of IM and i) are updated each period – whereas the inputs to a dynamic forecast are not.



- We can also view the residuals of the in-sample forecast to see whether there were any biases in our estimates…



A more formal evaluation required…

- Now, the graphs on the previous slides are good at giving us a first impression of the model's forecasting abilities.
- However we need a more concrete statistical indication of the predictive ability of our model in order to assess whether it would be useful in forecasting.
- We use several measures including…

Prediction accuracy measures

$$\varepsilon_{t+h} = Y_{t+h} - \widehat{Y_{t+h}} = forecast\ error$$

- **We can use the MSE:**

$$MSE = \frac{1}{T-(T_1-1)}\sum_{T_1}^{T}\left(y_{t+h} - f_{t,h}\right)^2$$

- With: $t = T_1 \rightarrow T$ (with T=total sample=in+out of sample, T_1=first obs in in-sample)
- In-sample: $T - (T_1 - 1)$.
- A series with: $h - step$ ahead forecasts.
- $f_{t,s} \rightarrow$ Forecast of y_{t+s} at time t.

Prediction accuracy measures

$$MAE = \frac{1}{T-(T_1-1)}\sum_{T_1}^{T}|y_{t+h}-f_{t,h}|$$

$$MAPE = \frac{100}{T-(T_1-1)}\sum_{T_1}^{T}\frac{|y_{t+h}-f_{t,h}|}{|y_{t+h}|}$$

MAPE:
Interpreted as % error

A more formal evaluation given…

Forecast: DLGF
Actual: DLG
Forecast sample: 2010Q2 2013Q1
Included observations: 12

Root Mean Squared Error	0.005926
Mean Absolute Error	0.003469
Mean Absolute Percentage Error	76.55108
Theil Inequality Coefficient	0.272561
Bias Proportion	0.003348
Variance Proportion	0.723276
Covariance Proportion	0.273376

- MSE is normally used in this regard, as it penalizes for large errors disproportionally more heavily than small errors.
- MAE, however, penalizes large errors proportionally equally to smaller errors
- We want both to be as small as possible…

A more formal evaluation given…

Forecast: DLGF
Actual: DLG
Forecast sample: 2010Q2 2013Q1
Included observations: 12

Root Mean Squared Error	0.005926
Mean Absolute Error	0.003469
Mean Absolute Percentage Error	76.55108
Theil Inequality Coefficient	0.272561
Bias Proportion	0.003348
Variance Proportion	0.723276
Covariance Proportion	0.273376

- The ***Bias Proportion*** measures the extent to which the mean o/t forecast differs from that of the actual series.
- ***Variance Proportion:***
  Δ btwn Forecast variance & actual variance
- ***Covariance Proportion***:
  Whether there are any remaining irregular part in the forecast errors (which we hope to be larger than variance proportion – so that most of the variance is attributable to irregular movements and not a bias)

Diebold-Mariano test

- In the tutorial solution, we will discuss in more depth how to use this test in evaluating the forecast performance of a pair of models.

Density forecast testing.

- In order for us to compare the true accuracy of forecasting models, we need to consider more than just the point estimate.
  - This follows as we are essentially more interested in the distribution **around** the mean forecast (i.e. the confidence bands around our forecasts).
- This requires, however, Bayesian techniques that allow us to calculate the **density** forecast estimates – which gives distributional forecasts from which we can construct likelihood intervals of forecasts.

- Although it remains outside the scope of this course, interested readers can check out our paper on using density forecasts: https://www.sciencedirect.com/science/article/abs/pii/S1062940817300414

END