(S)AR(F)IMA Modelling

Roman Horváth

Lecture 3

Applied Econometrics

Contents

- Summary of last lecture
- Autocorrelation Function (Correlogram)
- Partial Autocorrelation Function
- Ljung-Box test
- Examining ARMA (p,q) by ACF and PACF
- Information criteria
- Evaluating forecasting performance
- Fractional integration
- Spatial autoregressive models

ARMA Model:

- **AR – autoregressive models**

  AR(p) process: $Y_t = \alpha + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + e_t$

- **MA – moving average models**

MA(q) process: $Y_t = \beta + e_t + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_q e_{t-q}$,
$e \sim IID(0,\sigma^2)$

combination $\Rightarrow$

- **ARMA – autoregressive moving average models**

ARMA(p,q):

$Y_t = c + a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_p Y_{t-p} + e_t + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_q e_{t-q}$

Stationarity

- Stationarity (independence on time)
    - compute mean, variance and covariance
    - check time dependency
- MA process always stationary
- AR proces: stationary only, if $|\alpha| < 1$ (in case of AR(1), or sum of $\alpha < 1$ in abs. value in case of AR(k))
- A series, which becomes stationary after first differencing, is said to be integrated of order one, denoted I(1).
- If Δy_t is described by a stationary ARMA(p,q) model, we say that y_t is described by an autoregressive *integrated* moving average (ARIMA (p,v,q)) model of order (p,1,q).

Autocorrelation Function

- Autocorrelation ϱ_k – correlation between $\boldsymbol{X_t}$ and $\boldsymbol{X_{t-k}}$
- **$\boldsymbol{\varrho_k}$ as a function of *k*** (k = 1,2,…) is called sample **autocorrelation function (ACF)** or **correlogram**
- You can do a joint test of the first *m* autocorrelations,
- Null hypothesis (H0): all correlations are 0, alternative (H1): at least one is significantly different from 0
- Use Box-Ljung test - test for significance of autocorrelations

Box-Ljung test

- Test statistics:

$$Q = T(T+2)\sum_{k=1}^{m}\frac{\rho_k^2}{T-k}$$

- T – # of observations,k – # of lags,correlation ρ_k – autocorrelation, m - # of autocorrelation terms
- If Q > critical value (p-value < 0.05) at least one of the first *m* autocorrelations is significant, autocorrelation is significant.

Partial Autocorrelation Function (PACF)

- Note that autocorrelation at lag *k* is given by the coefficient in

$$X_t = c + \varrho_k X_{t-k} + u_t$$

- PACF, denoted by Φ_{kk}, gives the correlation between X_t and X_{t-k} allowing for the effects of intermediate lags $X_{t-1},\ldots,X_{t-k-1}$
- Φ_{kk} is given by the last coefficient in
- $X_t = c + \Phi_{1k} * X_{t-1} + \ldots + \Phi_{kk} * X_{t-k} + u_t$

Application:
Does PX-50 follows random walk?

- If $X_t = X_{t-1} + u_t$, then $\Delta X_t \sim IID(0, \sigma^2)$

- To test random walk hypothesis, look for significant autocorrelation in $\Delta \boldsymbol{X_t}$,
  where $\boldsymbol{X_t}$ is PX-50 index

  (PX-50 is the Czech stock market index, daily data from 1998-2002, roughly 2200 observations)

Application:

Does PX-50 follows random walk?

(so changes in PX-50 is white noise)

Results from Ljung-Box statistic

	AC	Q-Stat	P-value
1	**0.128**	36.002	0
2	**0.091**	54.296	0
3	**0.059**	62.046	0
4	0.019	62.836	0
5	-0.036	65.674	0
6	-0.008	65.829	0
7	0.014	66.251	0
8	0.012	66.549	0
9	0.014	67.011	0
10	0.046	71.689	0

MA(1) process

- $\mathbf{X_t = \beta + u_t + \alpha_1 * u_{t-1}}$

- $\mathbf{E(X_t)} = \beta$
- $\mathbf{Var(X_t)} = E(\beta + u_t + \alpha_1 * u_{t-1} - \beta)^2 =$

  $= E(u_t^2 + 2\alpha_1 u_t u_{t-1} + \alpha_1^2 u_{t-1}^2) = u_t^2(1 + \alpha_1^2)$
- $\mathbf{Cov(X_t, X_{t-k})} = E[(X_t - E(X_t))(X_{t-k} - E(X_{t-k}))] = u_t^{2*}\alpha_1,$

  if $t = 1$ and 0, if $t > 1$.
- $\mathbf{Corr(X_t, X_{t-k})} = Cov(X_t, X_{t-k}) / Var(X_t)^{1/2} * Var(X_{t-k})^{1/2} =$

$\alpha_1 / (1 + \alpha_1^2)$ if $t = 1$ and 0, if $t > 1$.

AR(1) process

- Similarly as for MA(1), first invert AR(1) to MA process
  this is simple difference equation, for variance you have to sum up infinite series
- $\boldsymbol{X_t = \alpha X_{t\text{-}1} + u_t}$

- **E(*X*$_t$)**$=0$
- **Var(*X*$_t$)**$= u_t^2/(1- \alpha^2)$
- **Cov($\mathbf{X_t,X_{t\text{-}k}}$)**$= \alpha^{k*} u_t^2/(1- \alpha^2)$
- **Corr($\mathbf{X_t,X_{t\text{-}k}}$)**$=\alpha^k$ - correlogram declines exponentially with *k*

ACF, PACF

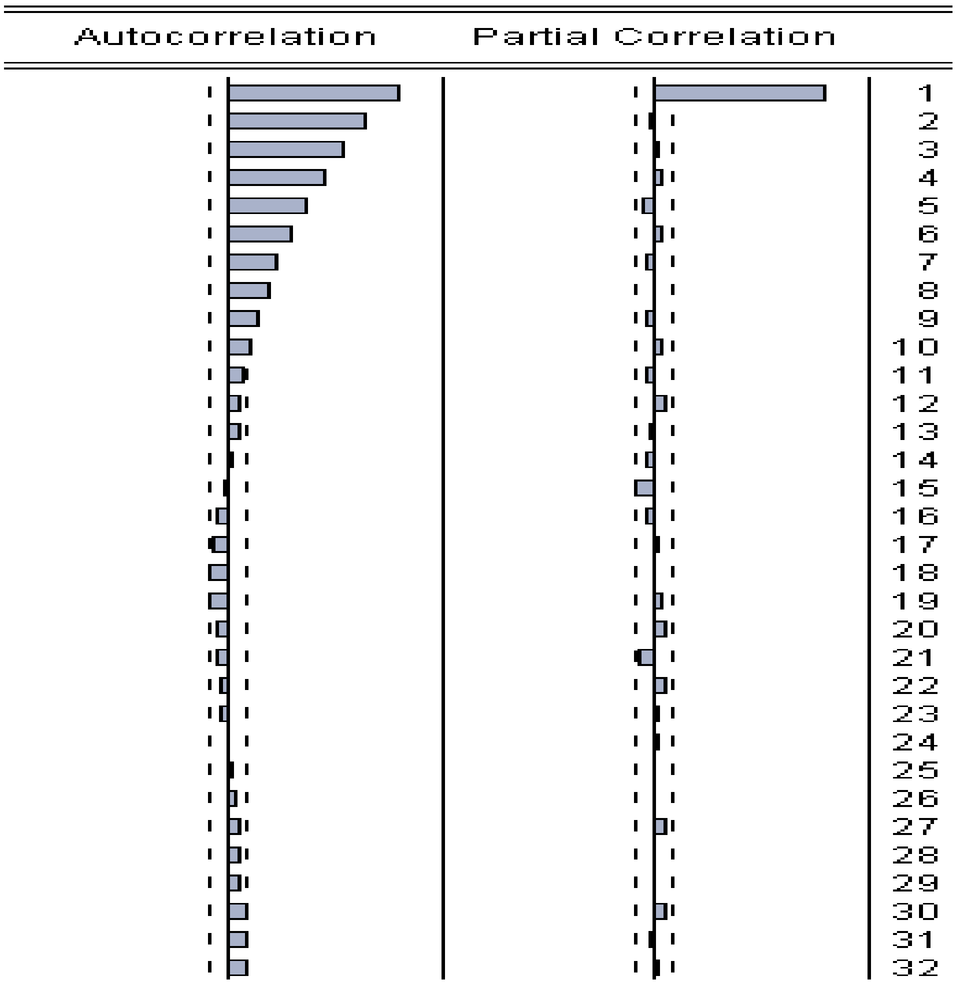
to determine a number of AR and MA terms

- **AR(p)** – ACF declines, PACF = 0 if k > p
- **MA(q)** – ACF = 0 if k > q, PACF declines
- **ARMA(p,q)** – ACF, PACF declines

- For ACF see the previous slides, where *corr* of MA and AR processes were derived
- To realize that PACF for AR(p) is PACF = 0 if k > p, note that there is no direct connection between $\boldsymbol{Y_t}$ and $\boldsymbol{Y_{t-k}}$ as AR(p) is: $\boldsymbol{Y_t = \alpha + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} + e_t}$
- To realize that PACF declines for MA(q), need to invert MA(q) to AR(∞)

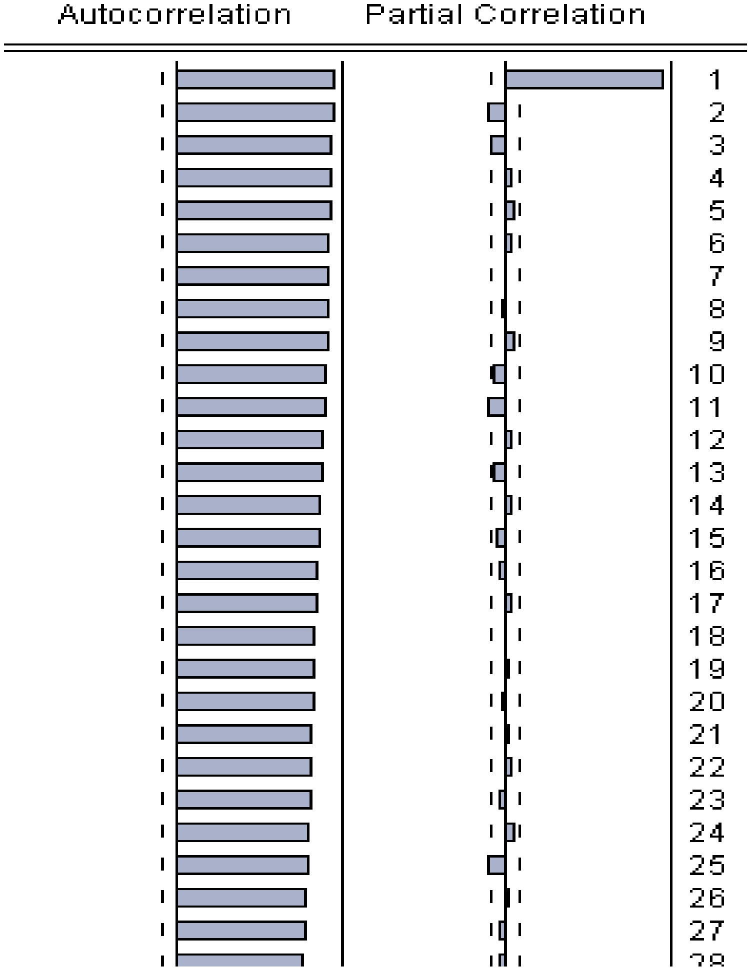
Examples – AR(1)

- Simulated data
- $X_t = 0.8*X_{t-1} + u_t$
- ACF decays
- PACF drops to 0 for 2nd lag and higher



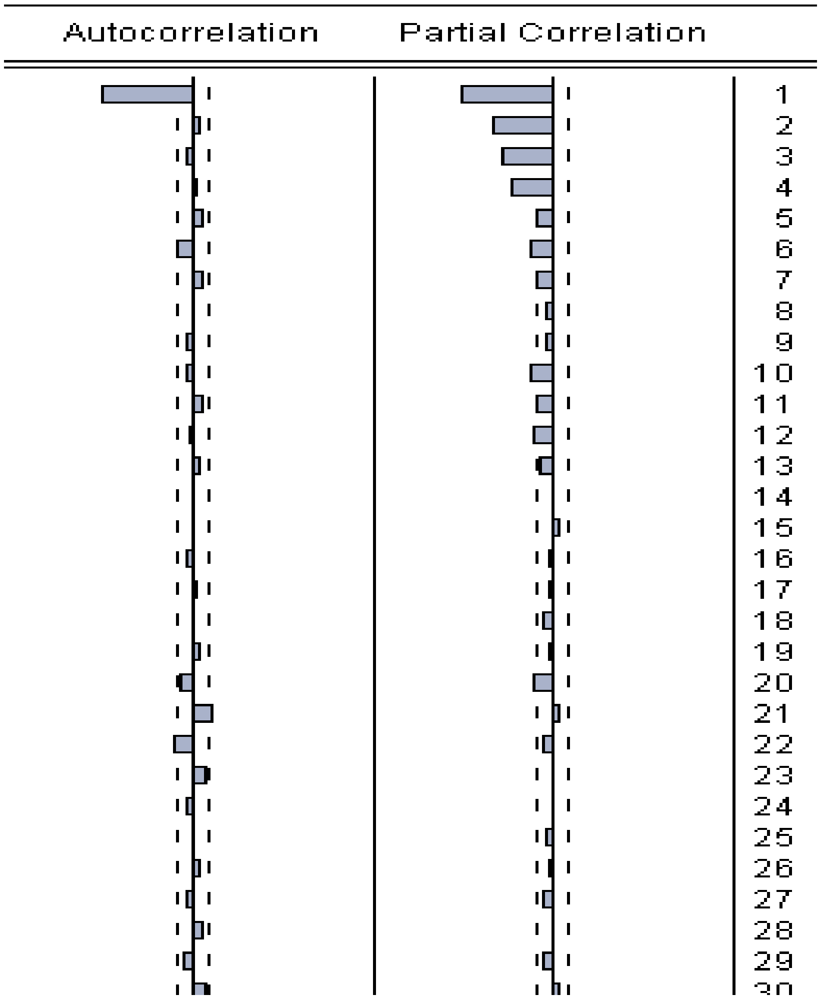
Examples – Random walk

- Simulated data
- $X_t = 1{*}X_{t-1} + u_t$
- ACF decays extremely slowly
- PACF drops to 0 for 2nd lag and higher



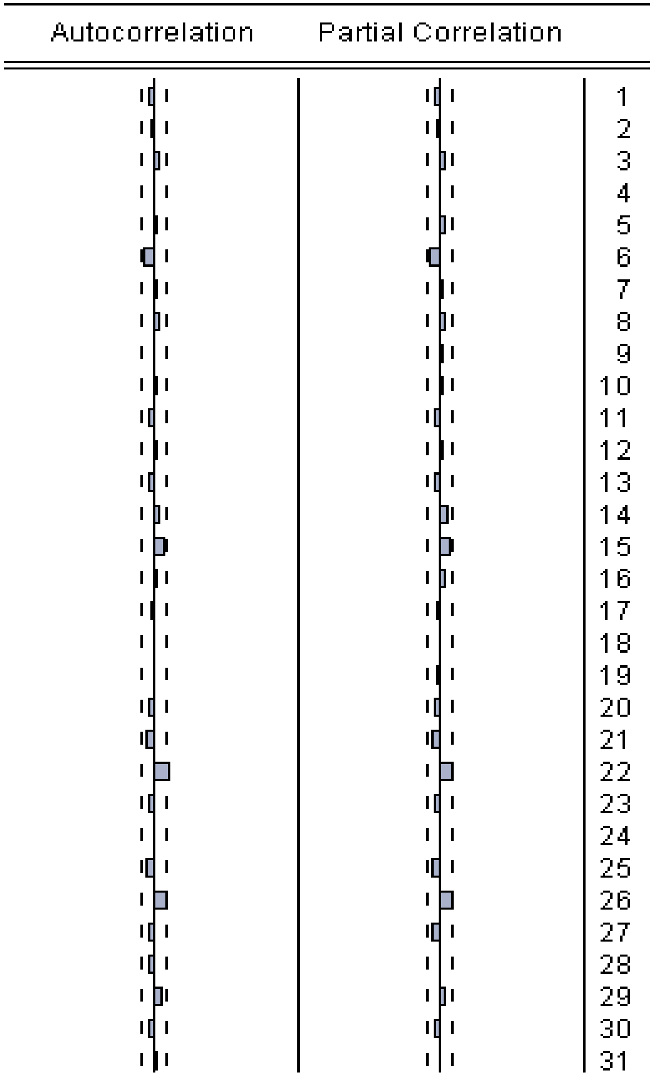
Examples – MA(1)

- Simulated data
- $X_t = u_t - 0.9^* u_{t-1}$
- ACF drops to 0 for 2nd lag and higher
- PACF decays slowly



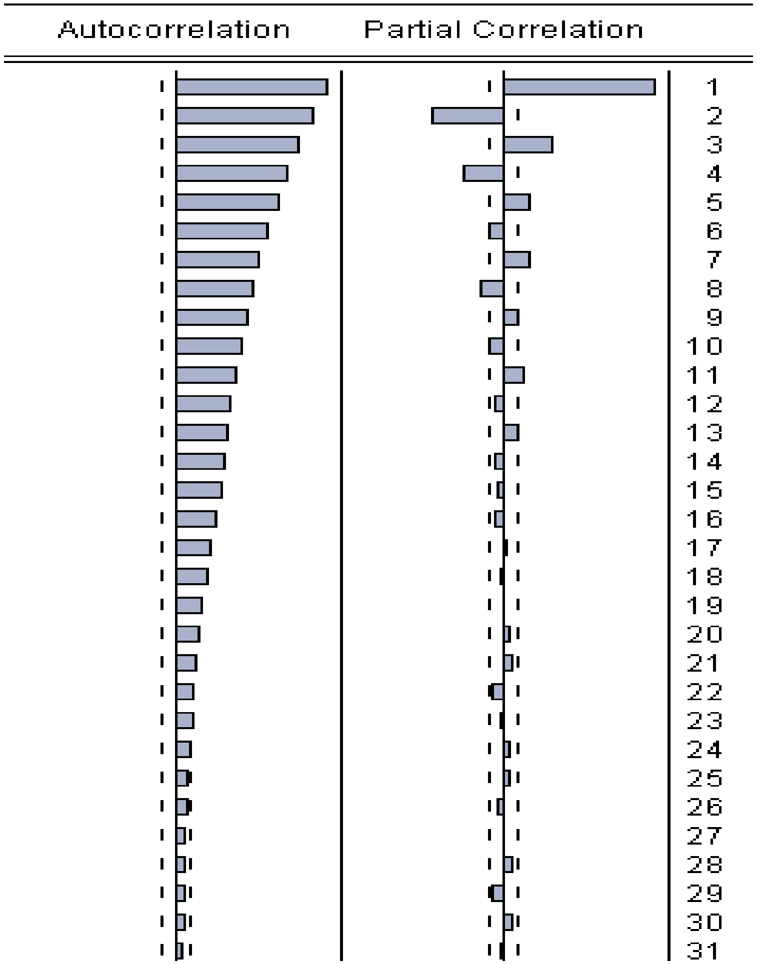
Examples – White noise

- Simulated data
- $X_t = u_t$
- ACF close to 0
- PACF close to 0



Examples – ARMA(1,1)

- Simulated data
- $X_t = 0.9{*}X_{t-1} + u_t - 0.9{*}\ u_{t-1}$
- ACF decays slowly
- PACF decays slowly



Fitting ARMA Models: ***Box-Jenkins Methodology***

1. Visually examine the time plot of series, ACF and PACF
2. Estimate the degree of integration and difference the data to achieve stationarity
3. Estimate ARIMA model and select the 'best' model based on following criteria:

   - parsimony (do not overfit),

   - goodness of fit (based on information criteria)

Information Criteria

- Information Criteria often use model selection for non-nested alternatives (non-nested: different set of explanatory variables among the regression models)
- Schwarz criterion = $-2L/T + (k*\log(T))/T$
- Akaike information criterion = $-2L/T + 2k/T$
  - where $L = -T/2(1+\log(2\pi)+\log(RSS/T))$...log likelihood function, T...number of obs., k...number of explanatory variables+1, RSS … residual sum of squares
- AIC and SC trade off parsimony and goodness of fit
- So AIC or SC should be as low as possible

Forecasting

- In the long run the forecast converges to a constant (for stationary model)
- In the short run, estimate the model and simply take the expected value in order to get the next period forecast

$X_t = aX_{t-1} + u_t$

$E(X_{t+1}) = E(aX_t + u_{t+1}) = aX_t$

$\mathrm{E}(X_{t+2}) = aX_{t+1} = aaX_t$

$\mathrm{E}(X_{t+k}) = a^{\mathrm{k}}X_t$

- As k becomes large, the forecast converges to a constant, which is 0 in our case.
- ARMA modelling is for short-term forecasting, not long-term!

Evaluating Forecasting Performance

- Forecast error, e: **error = fitted – actual value**
- **Root square mean error** (RMSE)

$$RMSE = \sqrt{\frac{1}{k}\sum_{j=1}^{k} e_{T+j}^{2}}$$

- **Mean Absolute Error** (MAE)

$$MAE = \frac{1}{k}\sum_{j=1}^{k}\left|e_{T+j}\right|$$

- **Mean absolute percentage error** (MAPE)

$$MAPE = \frac{1}{k}\sum_{j=1}^{k}\left|\frac{e_{T+j}}{y_{T+j}}\right| * 100\%$$

Evaluating Forecasting Performance (cont.)

% correct sign predictions = $$\frac{1}{N}\sum_{t=1}^{N} z_{t+s}$$

where z_{t+s} = 1 if $(x_{t+s} \, . \, f_{t,s}) > 0$

z_{t+s} = 0 otherwise

$f_{t,s}$ forecast s periods ahead

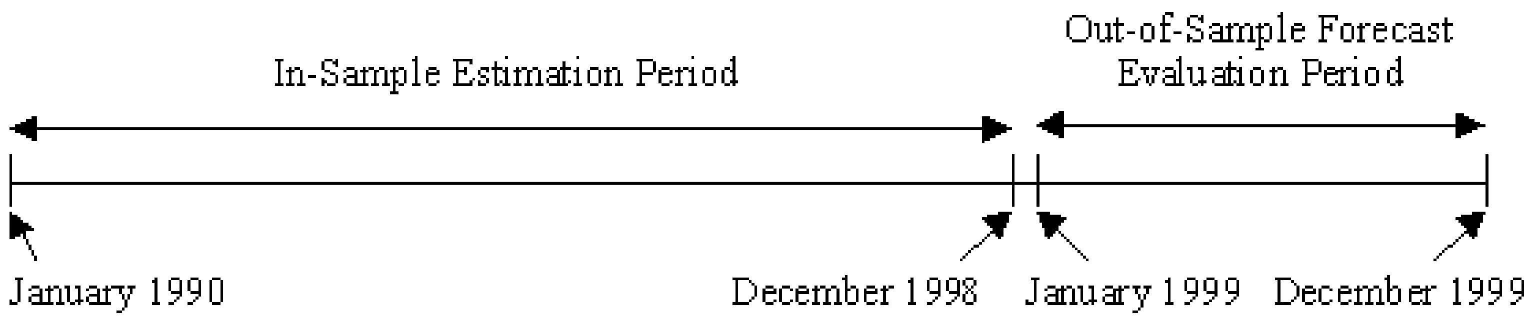
Analogously, one can form *% correct direction change predictions* (these two statistics proved to be more profitable for market trading strategies)

Evaluating Forecasting Performance (cont.)

- Root square mean error penalizes disproportionally more larger than smaller errors
- Mean absolute error penalizes proportionally equally large errors as small errors
- Sign prediction error does not penalize large errors any more than small errors

In-Sample Versus Out-of-Sample Forecasting

- Expect the "forecast" of the model to be good in-sample.
- Say we have some data - e.g. monthly FTSE returns for 120 months: 1990M1 – 1999M12. We could use all of it to build the model, or keep some observations back:



- A good test of the model since we have not used the information from 1999M1 onwards when we estimated the model parameters.
- Pseudo-out-of-sample forecasting (pretending to have the data up to Jan 1999 but knowing the future)

Fitting ARIMA models

- Same as for ARMA
  - Use augmented ADF test to evaluate the order of integration
  - Difference the data to achieve stationarity

Fractional integration

- Some series are autocorrelated up to very high lag
- Long-memory processes
- So far, we assumed that series is I(0) or I(1)
- But series can be I(d), where $0 < d < 1$
- Fractional integration
- ARFIMA – a generalization of ARIMA

Fractional integration (II)

- AR(1) proces can be written as $(1- \phi B)Y_t = e_t$
- B – backshift operator (L is a lag operator, B or L is alternative notation but identical meaning)
- Note that $BY_t = Y_{t-1}$, $BBY_t = Y_{t-2}$, $B^2Y_t = Y_{t-2}$,
- $(1- \phi B)^d Y_t = e_t$, fractional integration of order d
- $d \geq 0.5$ – non-stationary process
- $d < 0.5$ – stationary process
- ARFIMA: $\phi_p(B)\ (1 - B)^d\ Y_t = \theta_q(B)\ \varepsilon_t$

ARFIMA Model Estimation

- ARFIMA(0; d; 0). For example, ARFIMA(0; 0,55; 0)
- Two-step method: 1) d is estimated, 2) ARIMA is estimated
- One-step method: Joint estimation with maximum likelihood
- Two-step method needs a lot of observations
- One-step method: local maxima, computational complexity, initial values issues

Short vs. Long Memory

- No memory – ACF is zero
- Short memory – stationary
  - exponential decay in the ACF
- Long memory – stationary
  - The "differencing" exponent (d) is between
  - $0 < |d| < ½$.
  - Hyperbolic (slower) decay in the ACF
- Non-stationary – random walk

Spatial econometrics

- Space matters
  - Two geographically close countries more likely to adopt similar reforms because of knowledge spillovers
  - Two central banks with identical monetary policy regimes more likely to cooperate
  - House prices depend on location

  If there are spatial effects, OLS can be either inefficient with wrong standard errors, or biased and inconsistent

Weighting

- Assumption: structure of spatial dependence is *a priori* known, not estimated
- The specification of the weighting matrix "is a mater of considerable arbitrariness and a wide range of suggestions in the literature"
- "Connectivity matrix" specifies the degree of interdependence among observations
  - Based either on contiguity, Euclidean distance, or even non-geographical distance-based measure
- Strong assumption, but not as strong as assuming it is zero and all observations are spatially independent

Weighting (II)

- Denote the weighting matrix, W
- Contiguity matrix
  - NxN symetric matrix where $w_{ij} = 1$ when i and j are neighbors and 0 when they are not
  - W matrix is usually standardized so all rows sum to 1
    - $w^s_{ij} = w_{ij} / \Sigma_j w_{ij}$
    - Makes operations with the W matrix as an average of neighboring values

Weighting (III)

- W matrix used to generate spatial lag operator, Wy
  - $\Sigma_j w_{ij} y_j$, Weighted average of the y values based on neighbors

W matrix:

0	1	0	1
1	0	0	0
0	0	0	1
1	0	1	0

Type of models

- Spatial lag model
  - Spatial lag model corresponds to the time-series lagged dependent variable model
- Spatial error model
  - Spatial error model analogous to time-series serially correlated errors

Spatial lag model

- Model
  - $y = \varphi w_i y_i + \beta x_i + e_i$
    - Can also include $w_i x_i$ term
  - OLS in this case is biased and inconsistent

- Can be extended to spatial Durbin model
- $y = \varphi w_i y_i + \beta x_i + \beta w_i x_i + e_i$

Spatial error model

- Model
  - Start with basic OLS model
    - $y= \beta x+e \quad e \sim N(0,\sigma^2)$
  - $y= \beta x+e+\lambda we$
    - If $\lambda=0$, reduces to OLS, if $\lambda \neq 0$, OLS is unbiased and consistent, but SE will be wrong and the betas will be inefficient if spatial errors ignored

Testing for spatial effects

- Morans' I test
  - Indicates general spatial misspecification
  - I = e'We'/e'e (for row standard weights)
    - e = vector of OLS residuals
    - W = spatial weights
  - Similar to the Durbin-Watson test
  - Does not provide insight into suggesting which alternative regression specification to use