Dynamical models

for migration projections

Violeta Calian

Statistics Iceland

Joint Eurostat UNECE Work Session on Demographic Projections
Rome,October 29-31, 2013

Introduction

Goal

migration forecast during unstable economical conditions

Statistical problem

modelling with auto-correlated and non-stationary time series

uncertainty control

Our solution

dynamical models =
auto-regressive distributed lag (*ARDL*) models, where:

the dependent variable at time t is modelled as a function of its own values at different time lags and of the values of several simultaneous or lagged predictor variables.

Main points

1. why do we use dynamical models

2. what are the mathematical conditions for a reliable statistical inference and whether they are fulfilled by our data and models

3. how are our models for all migration components built and how do they perform

Data

- the number of Icelandic immigrants/emigrants, men
- the number of Icelandic immigrants/emigrants, women
- the number of immigrants/emigrants, women of foreign citizenship
- the number of foreign immigrants/emigrants, men

Data

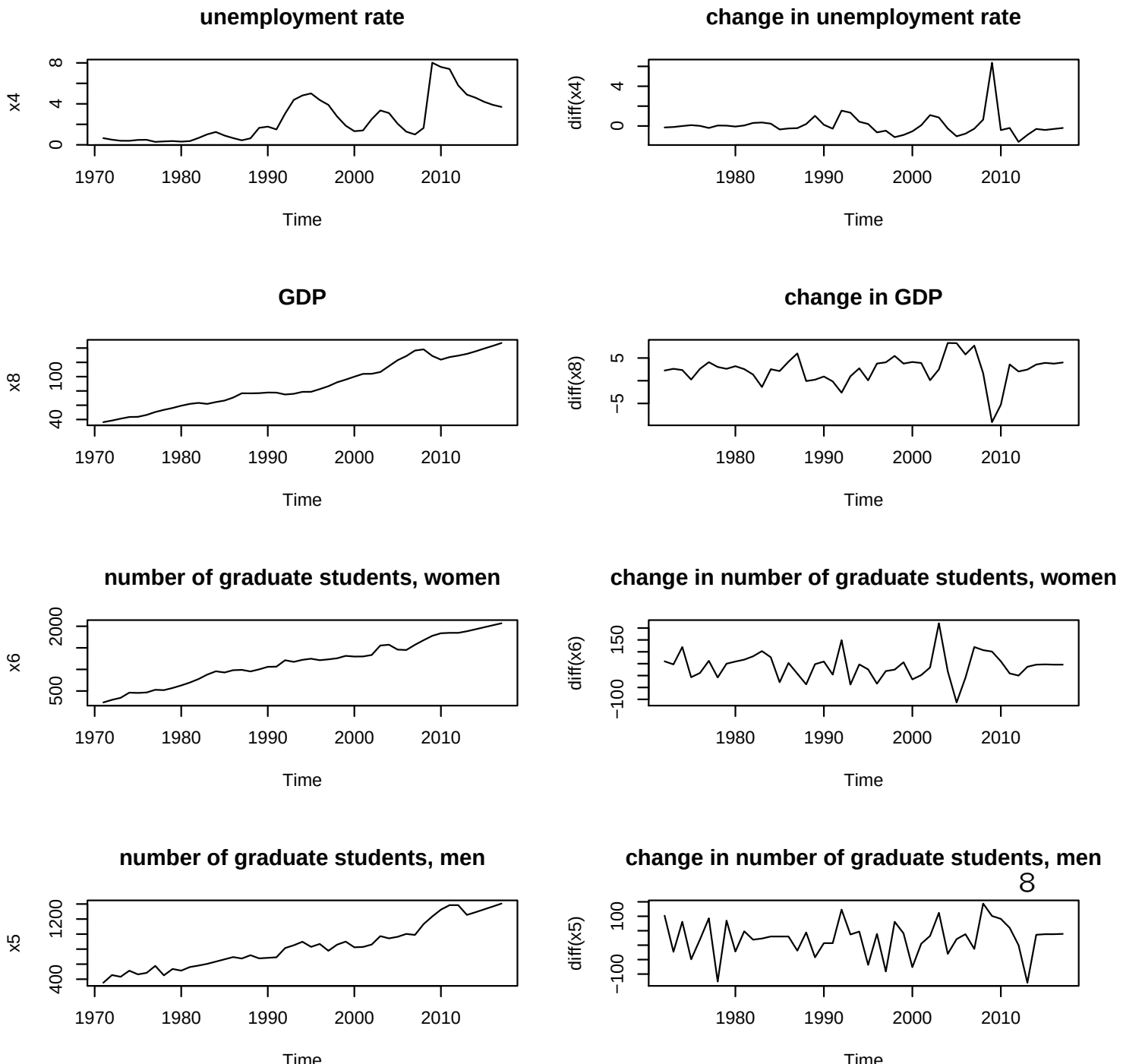
- the unemployment rate
- a measure of GDP
- the number of graduating students, men and women respectively
- a dummy variable coupled to the Icelandic economic boom
- a dummy variable which mirrors the re-sizing of the EEA

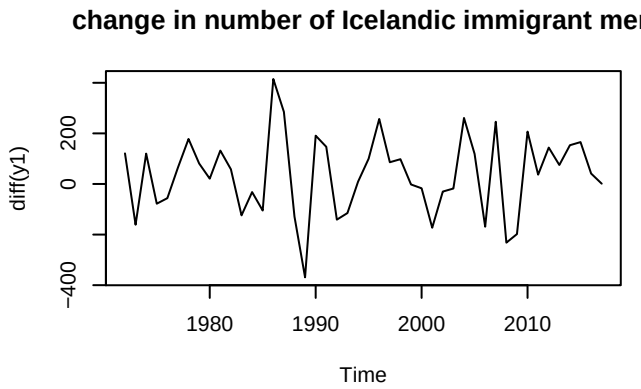
Data analysis

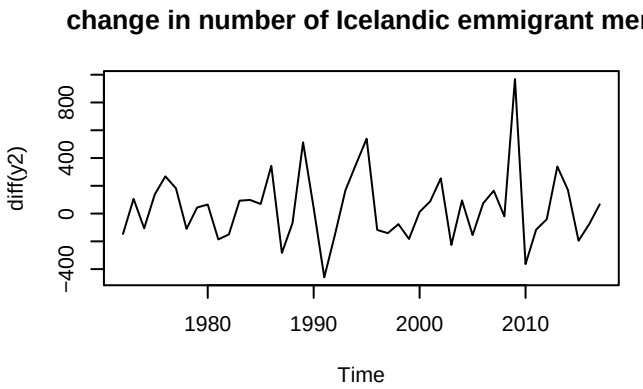
Statistical tests for all time series:

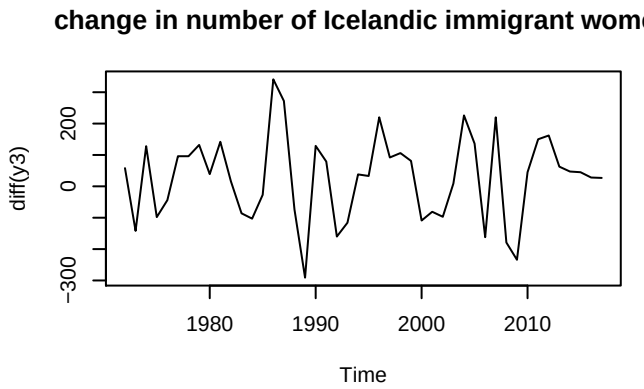
- stationarity: augumented Dickey-Fuller and Kwiatkowski-Philips-Schmidt-Shin (KPSS)

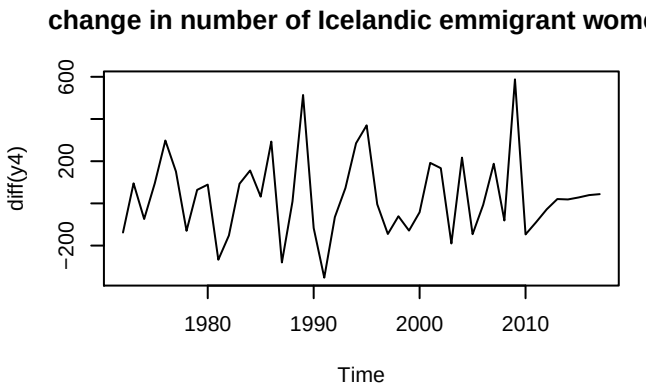
- auto-correlation of first and higher order, by using Durbin-Watson and Breusch-Gofrey tests

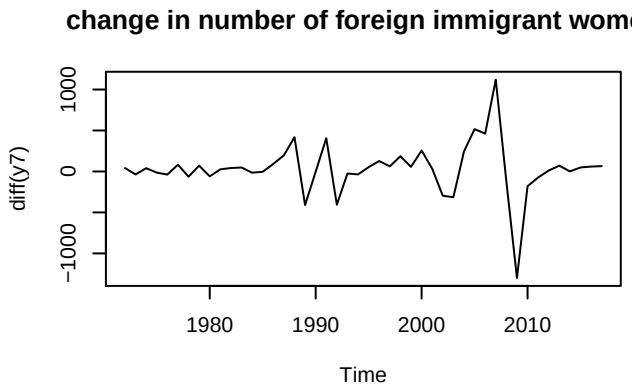

unemployment rate
x4
0
4
8
1970
1980
1990
2000
2010
Time
change in unemployment rate
diff(x4)
0
4
1980
1990
2000
2010
Time
GDP
x8
40
100
1970
1980
1990
2000
2010
Time
change in GDP
diff(x8)
−5
5
1980
1990
2000
2010
Time
number of graduate students, women
x6
500
2000
1970
1980
1990
2000
2010
Time
change in number of graduate students, women
diff(x6)
−100
150
1980
1990
2000
2010
Time
number of graduate students, men
x5
400
1200
1970
1980
1990
2000
2010
Time
change in number of graduate students, men
diff(x5)
−100
100
1980
1990
2000
2010
Time

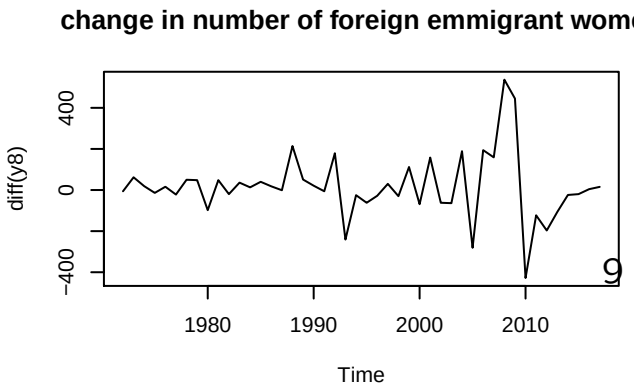

change in number of Icelandic immigrant men
diff(y1)
−400
0
200
1980
1990
2000
2010
Time


change in number of Icelandic emmigrant men
diff(y2)
−400
0
400
800
1980
1990
2000
2010
Time


change in number of Icelandic immigrant women
diff(y3)
−300
0
200
1980
1990
2000
2010
Time


change in number of Icelandic emmigrant women
diff(y4)
−200
200
600
1980
1990
2000
2010
Time


change in number of foreign immigrant women
diff(y7)
−1000
0
1000
1980
1990
2000
2010
Time


change in number of foreign emmigrant women
diff(y8)
−400
0
400
1980
1990
2000
2010
Time

Conditions for valid inference

- consistent model selection for the structure and the order of the model
- independent and identically distributed residuals
- non-biased and consistent point estimates
- correct and optimal calculation of confidence/prediction intervals

ARDL Models

$$y_\alpha(t) \sim \sum_{\beta \neq \alpha; i=0}^{n} y_\beta(t-i) + \sum_{i=1}^{p} y_\alpha(t-i)$$

$$+ \sum_{k; j=0}^{m} x_k(t-j)$$

$$y_1(t) \sim y_1(t-1) + x_4(t) + x_4(t-1) + y_2(t-1)$$

$$y_2(t) \sim y_2(t-1) + y_2(t-2) + x_5(t-2)$$

$$y_3(t) \sim y_3(t-1) + x_4(t) + x_8(t) + x_4(t-1) + x_8(t-1)$$

$$y_4(t) \sim y_4(t-1) + y_4(t-2) + x_6(t-2)$$

$$y_7(t) \sim y_7(t-1) + boom(t) + eea(t) + x_4(t) + x_8(t) + x_4(t-1) + x_8(t-1)$$

$$y_8(t) \sim y_8(t-1) + y_7(t) + y_7(t-1) + x_4(t) + x_8(t) + x_4(t-1) + x_8(t-1)$$

Interpretation

The interpretation and model diagnostics when using ARDL is very different from the classical so-called static models (see collinearity, short and long term effects)

One way to apply the classical notions is to transform a dynamical model into the equivalent error correction model (ECM):

$$\Delta y_\alpha(t) \sim \sum_\beta \Delta y_\beta(t) + \sum_k \Delta x_k(t)$$

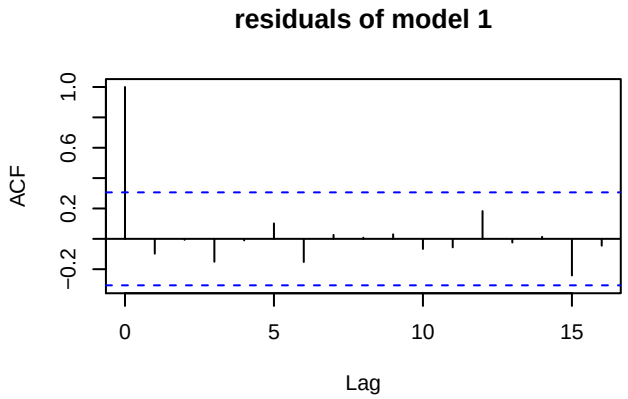
$$+ \left(y_\alpha(t-1) \sim \sum_\beta y_\beta(t-1) + \sum_k x_k(t-1) \right) + ...$$

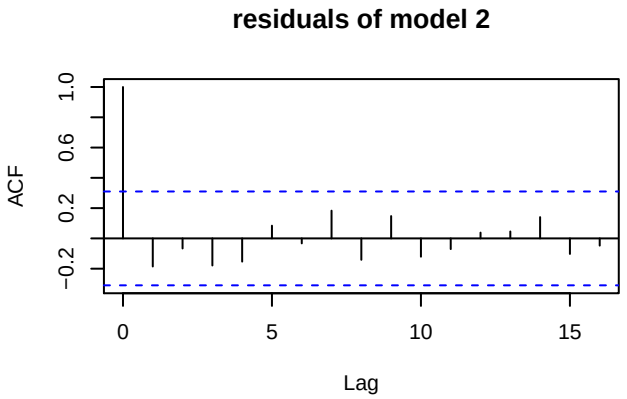
Behaviour of residuals and model fit

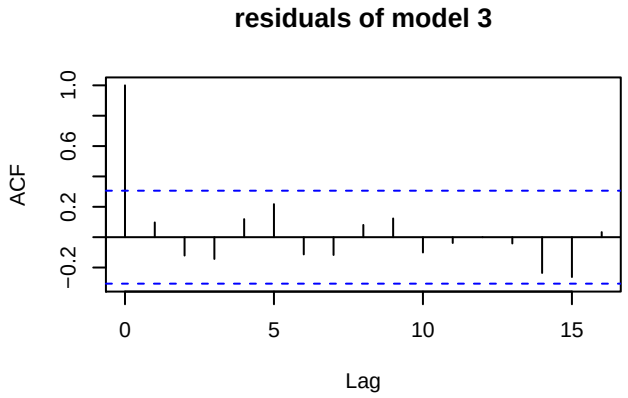
- Stationarity of residuals:
  - KPSS tests do not reject the hypothesis of stationarity (all p-values ≥ 0.1)
  - augumented Dickey-Fuller tests reject non-stationarity (all p-values ≤ 0.01)
- Normality of residuals: Jacques-Bera tests do not reject normality of model residuals - distributions. (all p values ≥ 0.1). Histograms.

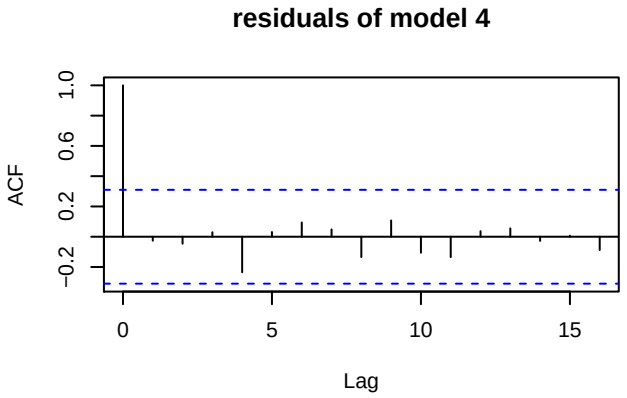
Behaviour of residuals and model fit

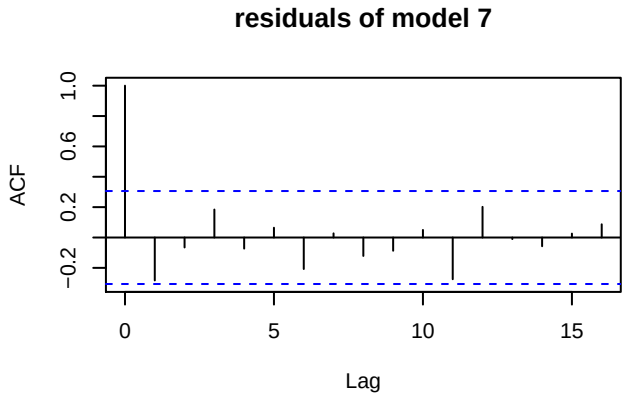
- Autocorrelation of residuals:
  - Box-Ljung tests do not reject the hypothesis of random residuals
  - direct calculation of autocorrelation for residuals (see Figure A1)
- Goodness of fit: rainbow tests

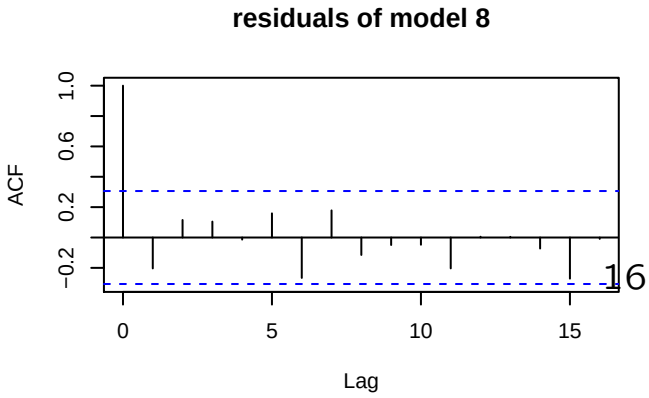

residuals of model 1
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

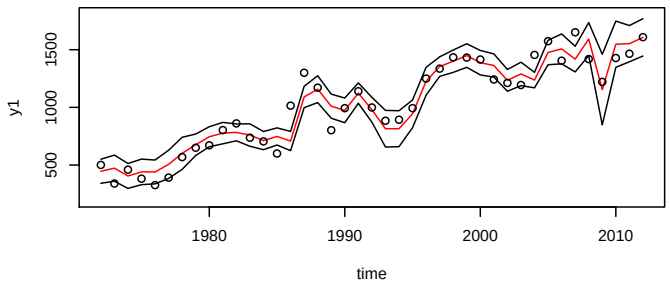

residuals of model 2
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

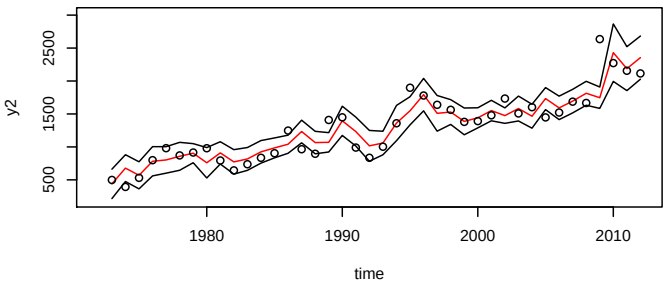

residuals of model 3
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

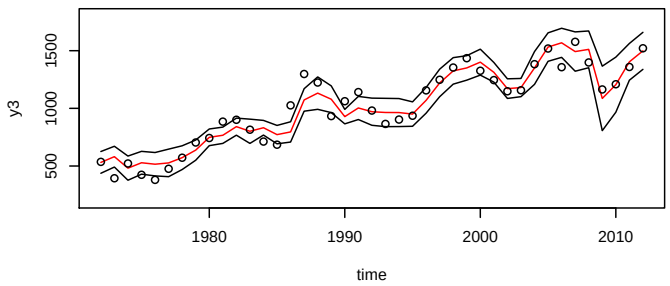

residuals of model 4
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

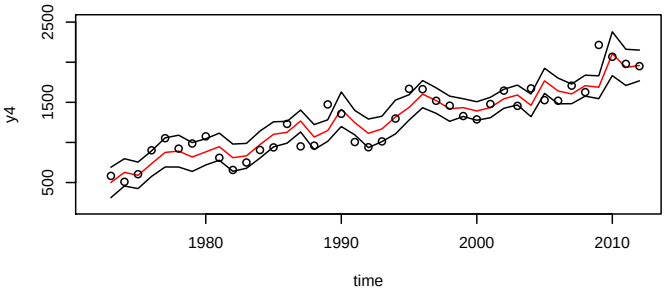

residuals of model 7
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

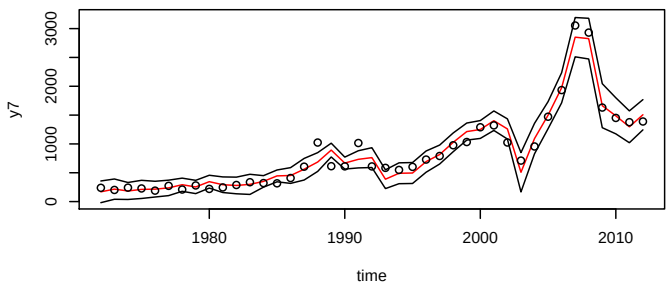

residuals of model 8
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
Lag

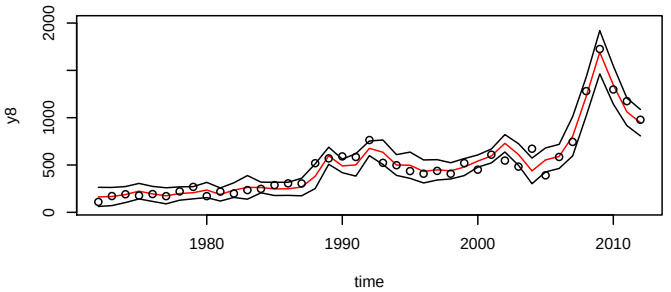

y1
500
1000
1500
1980
1990
2000
2010
time

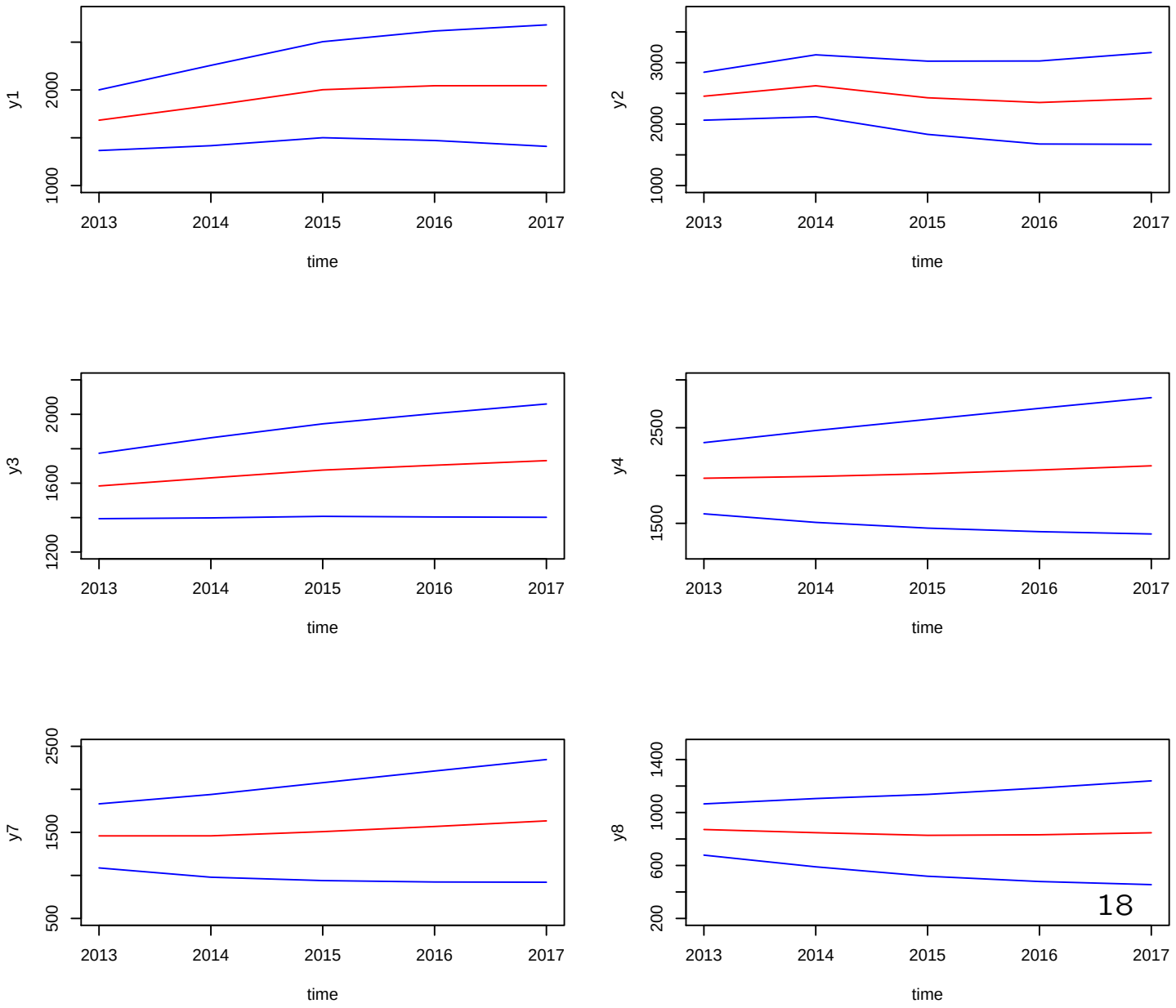

y2
500
1500
2500
1980
1990
2000
2010
time

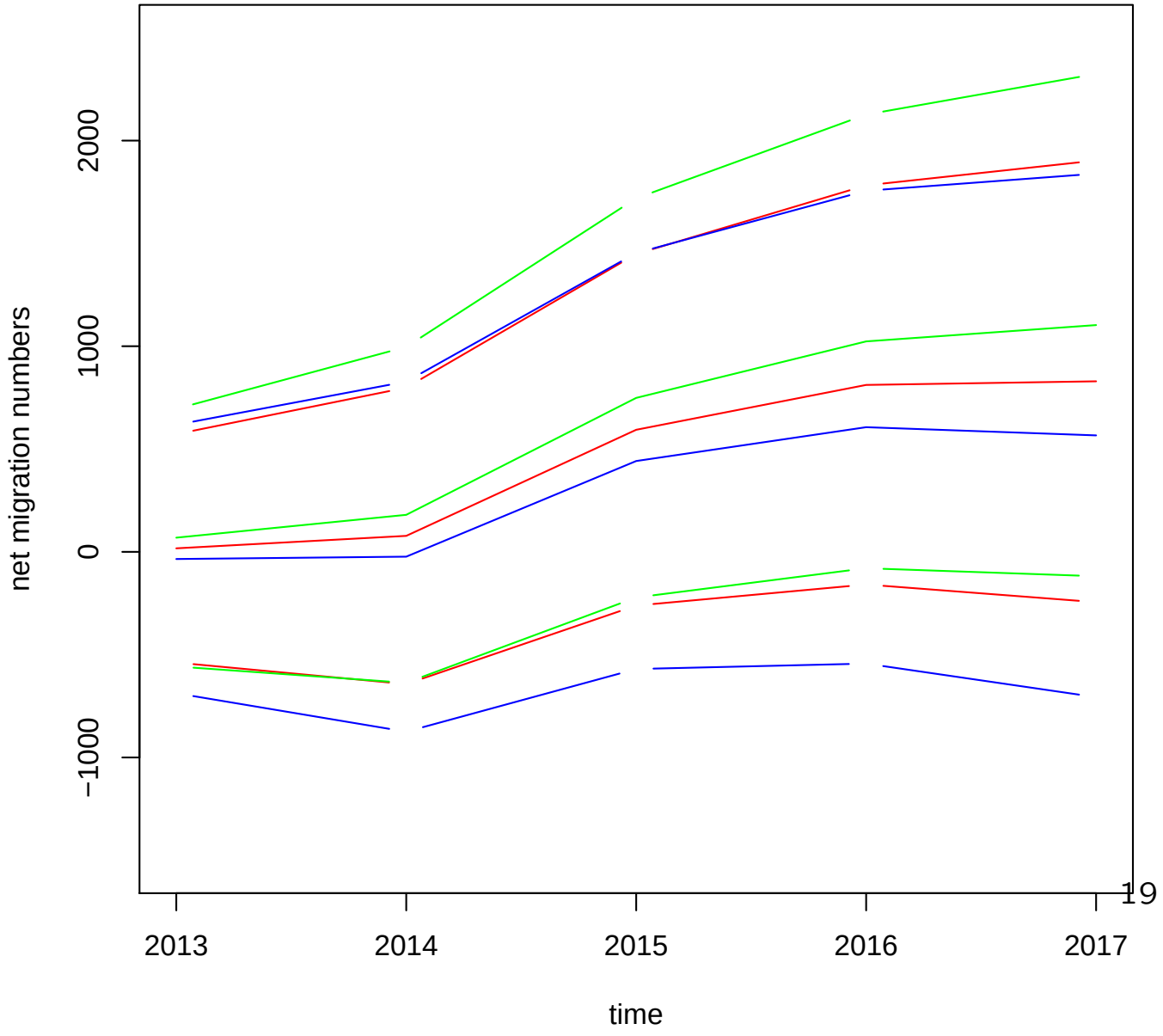

y3
500
1000
1500
1980
1990
2000
2010
time

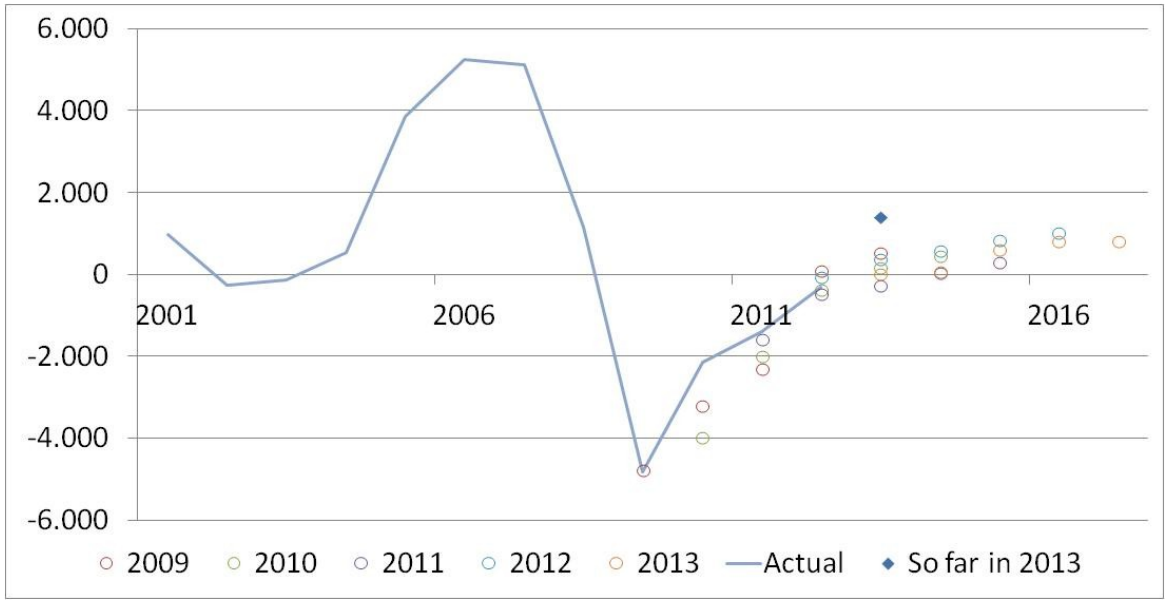

y4
500
1500
2500
1980
1990
2000
2010
time


y7
0
1000
2000
3000
1980
1990
2000
2010
time


y8
0
500
1000
2000
1980
1990
2000
2010
time


y1
1000
2000
2013
2014
2015
2016
2017
time
y2
1000
2000
3000
2013
2014
2015
2016
2017
time
y3
1200
1600
2000
2013
2014
2015
2016
2017
time
y4
1500
2500
2013
2014
2015
2016
2017
time
y7
500
1500
2500
2013
2014
2015
2016
2017
time
y8
200
600
1000
1400
2013
2014
2015
2016
2017
time


net migration numbers
−1000
0
1000
2000
2013
2014
2015
2016
2017
time
19



New and old predictions (with permission, from Omar Hardarson)

Conclusions and Discussion

- Good performance of ARDL models
- Limitations: regressor forecast, registration process

- Coverage and type II errors
- Vector ARDL models: how realistic
- External factors