The impact of macroeconomic variables on income distribution

Lyubomir Danov
Ana Lluberes
Björn Viergutz

Seminar on Statistical Programming Languages
Ladislaus von Bortkiewicz Chair of Statistics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Motivation and Outline

1. Model and Data
2. Stationarity
3. Testing for stationarity
4. Empirical analysis and discussion of the results
5. Conclusions

The Data

- Time series data from the United States
- Data taken from the Federal Reserve Bank of St. Louis
- Sample size: 48
- Yearly observations from 1967 - 2014

Variable	Explanation
GDP_t	Real gross domestic product
INF_t	Inflation rate
$UNEMPL_t$	Unemployment rate
$GINI_t$	Gini coefficient

Table 1: Variables in US time series data

The Model

- Suggested model: regression of the Gini coefficient on the inflation rate, GDP and unemployment rate.

$$GINI_t = \beta_0 + \beta_1 * INF_t + \beta_2 * GDP_t + \beta_3 * UNEMPL_t + u_t$$

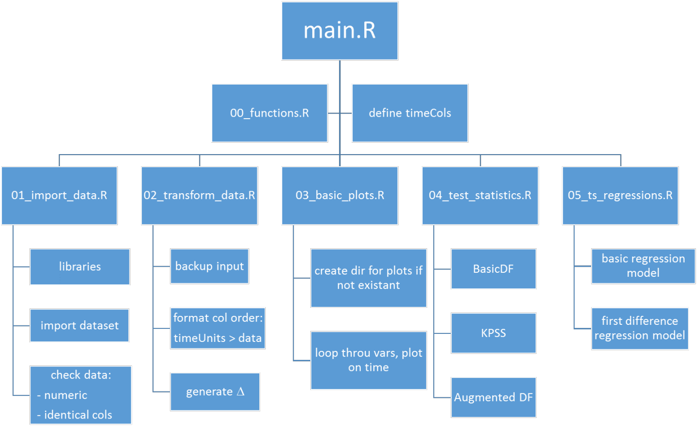
Stationarity

A *weak stationary* time series is define by:

$$E[X_t] = \mu \forall t$$
$$Var[X_t] = \sigma^2 \ \forall t$$
$$Cov[X_t; X_{t-j}] = \gamma(j) \ \forall t$$

- In other words: mean, variance and covariance are constant over time
    - Time series data does not follow any trends
- OLS estimates are only congruent when the variables are stationary

Code Structure



Importing the data

```
#### Import ####

### Import dataset
if (!exists("input")) {
input <- read.csv("SPL_data_v2.csv", header = TRUE, stringsAsFactors = FALSE)
print("input")
str(input)
input$timeIndex <- c(1:nrow(input))

} else {
print("input")
str(input)
}
```

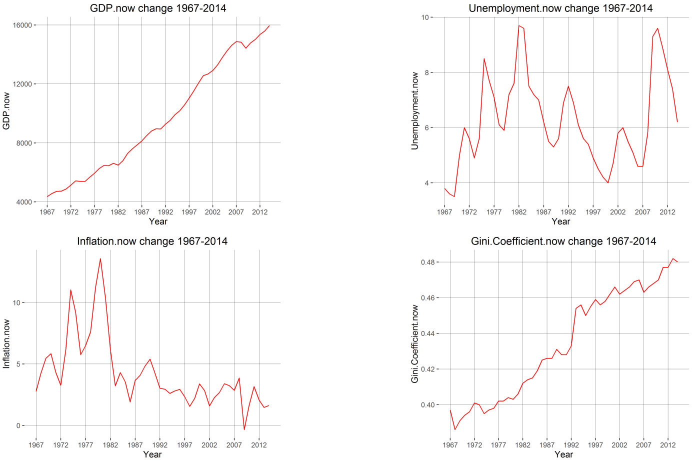
Plots

```
### Create plots in specified directory.
### If directory does not exist, create it.

if (!dir.exists(paste0(getwd(), "/plots"))) {
print("Creating folder '/plots'")
dir.create(paste0(getwd(), "/plots"))
} else {
print("Folder '/plots' exists. Careful to not
  overwrite existing plots")
}
```

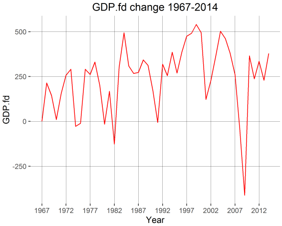
Plots

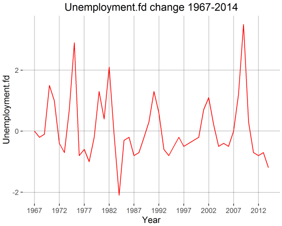
```r

foreach (i = 3:ncol(inputTest)) %do% {
if (!grepl(".prev", colnames(inputTest)[i])){
plotPath <- file.path(paste0(getwd(), "/plots/", colnames(inputTest)[i], ".png"))
png(file = plotPath, width = 640, height = 480)
plot(x = inputTest$Year,
y = inputTest[, i],
main=paste0(colnames(inputTest[i]), " change ", min(inputTest$Year), "-", max(inputTest$Year)),
xlab = "Year",
ylab = colnames(inputTest[i]))
dev.off()
}
}
```

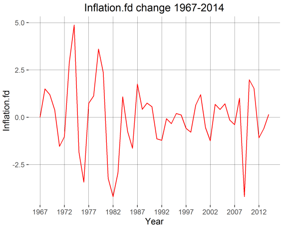
Plots: level data

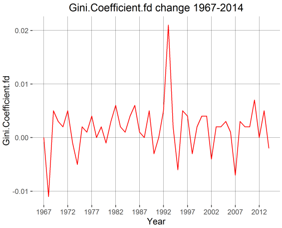


Plots: first differences









Dickey Fuller test

- Test for a unit root

$$\Delta y_t = \delta y_{t-1} + u_t$$

- Test for a unit root with drift

$$\Delta y_t = a_0 + \delta y_{t-1} + u_t$$

- Test for a unit root with drift and trend

$$\Delta y_t = a_0 + a_1 t + \delta y_{t-1} + u_t$$

Dickey Fuller test

	varName	test.level	test.stat	tval.test.stat	pval.regr
2	Year	UR.base	0.000502489215419904	995.000000000009	2.57819057362589e-101
3	Year	UR.drift	-2.02428340059621e-17	-1.73205080756865	2.5618547521896e-08
4	Year	UR.drift.tTrend	-2.02428340059621e-17	0.0901149418576962	2.5618547521896e-08
5	timeIndex	UR.base	0.0315789473684211	12	9.07085208886215e-16
6	timeIndex	UR.drift	-2.02428340059644e-17	-1.73205080756888	2.56185475218914e-08
7	timeIndex	UR.drift.tTrend	-2.02428340059644e-17	0.0901149418576563	2.56185475218914e-08
8	GDP.now	UR.base	0.023990553656017	8.64771656717408	3.34344369409078e-11
9	GDP.now	UR.drift	0.00957337078294869	1.28258321580246	0.206205611797258
10	GDP.now	UR.drift.tTrend	-0.097514195881221	-2.00068872131384	0.0438072708173052
11	GDP.fd	UR.base	-0.220842459526858	-2.2950849602844	0.0263390430210395
12	GDP.fd	UR.drift	-0.633114719835948	-4.63159311473371	3.10235419672735e-05
13	GDP.fd	UR.drift.tTrend	-0.669628600109888	-4.74730754635891	0.000110956461583952
14	Gini.Coefficient.now	UR.base	0.0040350007935066	2.58062307890578	0.0131184049594308
15	Gini.Coefficient.now	UR.drift	-0.00390932035790376	-0.170446203847744	0.865423281473382
16	Gini.Coefficient.now	UR.drift.tTrend	-0.326265092698418	-3.19420695667511	0.00928461688230808
17	Gini.Coefficient.fd	UR.base	-0.891132572431957	-6.0693180170345	2.27938645427238e-07
18	Gini.Coefficient.fd	UR.drift	-1.02649650168437	-6.84969428062005	1.70301747007719e-08
19	Gini.Coefficient.fd	UR.drift.tTrend	-1.03400317784933	-6.79716618159374	1.35498088809337e-07

Comparison with ADF test

H_0: time series has a unit root
H_1: time series is (trend-) stationary

	DF		ADF	
	Level	F.Diff	Level	F.Diff
GDP_t	0.044	0.003	0.590	< 0.010
INF_t	0.032	< 0.010	< 0.010	< 0.010
$UNEMPL_t$	0.014	< 0.010	0.034	< 0.010
$GINI_t$	0.009	< 0.010	0.590	< 0.010

Table 2: p-values from ADF test

KPSS test

```
KPSS_level <- data.frame(matrix(vector(), nrow = 0, ncol = 5))
colnames(KPSS_level) <- c("statistic.KPSS.Level", "parameter.Truncation.lag.parameter", "p.value", "method", "var")
for (i in 3:ncol(inputTest)) {
KPSS_level[i, ] <- unlist(kpss.test(inputTest[, i], null = "Level"))
KPSS_level[i, 5] <- colnames(inputTest)[i]
}
KPSS_level <- KPSS_level[-c(1:2), ]
KPSS_level <- within(KPSS_level, {
statistic.KPSS.Level                <- as.numeric(statistic.KPSS.Level)
parameter.Truncation.lag.parameter <- as.numeric(parameter.Truncation.lag.parameter)
p.value                             <- as.numeric(p.value)
}
)
```

KPSS test

H_0: time series is level-/trend-stationary
H_1: time series has a unit root

Variable	Levels	First Diff.
GDP_t	< 0.010	0.086
INF_t	< 0.010	0.100
$UNEMPL_t$	0.100	0.100
$GINI_t$	< 0.010	0.100

Table 3: p-values from KPSS test

The regression

```
reg1 <- lm(Gini.Coefficient.fDif ~ GDP.fDif + Inflation.fDif + Unemployment.fDif, inputTest)
summary(reg1)
str(summary(reg1))
plot(reg1$residuals)

reg2 <- lm(Gini.Coefficient.now ~ GDP.now + Inflation.now + Unemployment.now, inputTest)
summary(reg2)
str(summary(reg2))
plot(reg2$residuals)

kpss.test(reg1$residuals, null = "L")
kpss.test(reg2$residuals, null = "L")
```

Results

```
> summary(reg1)

Call:
lm(formula = Gini.Coefficient.fDif ~ GDP.fDif + Inflation.fDif +
    Unemployment.fDif, data = inputTest)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0119389 -0.0017641  0.0001196  0.0023929  0.0189450

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.753e-03  1.643e-03   1.067    0.292
GDP.fDif          -6.359e-08  6.018e-06  -0.011    0.992
Inflation.fDif    -5.938e-04  4.016e-04  -1.479    0.146
Unemployment.fDif -4.514e-04  1.149e-03  -0.393    0.696

Residual standard error: 0.00465 on 44 degrees of freedom
Multiple R-squared:  0.04805,	Adjusted R-squared:  -0.01686
F-statistic: 0.7402 on 3 and 44 DF,  p-value: 0.5338
```

Regression on level data

```
> summary(reg2)

Call:
lm(formula = Gini.Coefficient.now ~ GDP.now + Inflation.now +
    Unemployment.now, data = inputTest)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0126909 -0.0026484 -0.0006452  0.0016869  0.0183396

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.707e-01  5.193e-03  71.387  < 2e-16 ***
GDP.now          7.370e-06  3.066e-07  24.040  < 2e-16 ***
Inflation.now   -1.309e-03  4.052e-04  -3.231  0.00234 **
Unemployment.now -1.955e-04  5.664e-04  -0.345  0.73158
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.006243 on 44 degrees of freedom
Multiple R-squared:  0.9612,	Adjusted R-squared:  0.9585
F-statistic: 363.2 on 3 and 44 DF,  p-value: < 2.2e-16
```

Conclusions

- Introducing basic utility functions for unexperienced R users
- Test choice (DF, ADF, KPSS) based on visual detection of stationarity
- Our programmed DF test is less sensible to trend-stationarity compared to the ADF test and the KPSS test
- Regression model with first difference data
- Key findings: no significant impact of GDP, inflation rate and unemployment on income inequality (U.S. data from 1967 to 2014)
- Regression on level data yields different results (spurious regression)

Bibliography

Brockwell, Peter J. and Richard A. Davis, *Time series: theory and methods*, Springer Science & Business Media, 2013.
Granger, Clive W.J. and Paul Newbold, "Spurious regressions in econometrics," *Journal of econometrics*, 1974, *2* (2), 111-120.
Jäntti, Markus and Stephen P. Jenkins, "The impact of macroeconomic conditions on income inequality," *The Journal of Economic Inequality*, 2010, *8* (2), 221-240.
Moca, H. Naci, "Structural unemployment, cyclical unemployment, and income inequality," *Review of Economics and Statistics*, 1999, *81* (1), 122-134.
Parker, Simon C., "Opening a can of worms: the pitfalls of time-series regression analyses of income inequality," *Applied Economics*, 2000, *32* (2), 221-230.
Stock, James H. and Mark W. Watson, *Introduction to econometrics*, Vol. 104, Addison Wesley Boston, 2003.