R Package ArfimaMLM

An Effective Approach to the Repeated Cross-Sectional Design

Patrick Kraft Christopher Weber Matthew Lebo

Stony Brook University

University of Arizona

November 18, 2015

Introduction

- Implementation of ArfimaMLM approach presented by Lebo and Weber ("An Effective Approach to the Repeated Cross-Sectional Design", *AJPS* 2015) in R.
- Basic idea: Correcting for temporal autocorrelation in repeated cross-sectional data (as well as panel data).
- Key Aspects:
  - Individual observations are embedded within multiple, sequential time-points.
  - Retrieve estimates at the individual-level and at the aggregate level.
  - Allows use of variables that vary only within cross-sections and some that vary between cross-sections (e.g., unemployment rate)
  - Box-Jenkins and fractional differencing techniques can control for autocorrelation at level-2. (e.g. Box-Steffensmeier and Smith, 1996; Lebo et al., 2000; Clarke and Lebo, 2003)
  - Introduce double filtering to clean up two kinds of autocorrelation.

Description of Package

```
arfimaMLM(formula, data, timevar
        , d = "Hurst", arma = NULL
        , ecmformula = NULL, decm = "Hurst"
        , drop = 5, report.data = TRUE, ...)

arfimaMLM(y.ydif ~ x1.xdif + x1.fd + x2 + z1.fd + z2.fd
        + (1 | time)
        , data = data, timevar = "time", ...)

arfimaMLM(y.ydif ~ x1.xdif + x1.fd + x2 + z1.fd + z2.fd + ecm
        + (1 | time)
        , data = data, timevar = "time"
        , d = "Sperio"
        , ecmformula = y.mean ~ x1.mean
        , decm = "ML", ...)

arfimaMLM(y.ydif ~ x1.xdif + x1.fd + x2 + z1.fd + z2.fd + ecm
        + (1 + x1.dif | time)
        , data = data, timevar = "time"
        , d = list(y = "Hurst", x1 = "GPH", z1 = "Sperio", z2 = 0.25)
        , arma = list(y = c(1,0), z2 = c(0,1)), ...)
```

Simulational Scenario

- Scenario: repeated cross-sectional dataset with 100 timepoints and 500 units within each timepoint
- Independent variables:
  - $x_1 \sim \mathcal{N}(\mu = \bar{X}_{1t}, \sigma^2 = 2)$; $\bar{X}_{1t}$ follows a fractionally integrated series with $d = 0.3$ and a mean of 5
  - $x_2 \sim \mathcal{N}(\mu = 0, \sigma^2 = 40)$
  - $z_1 \sim \mathcal{N}(\mu = \bar{Z}_{1t}, \sigma^2 = 3)$; $\bar{Z}_{1t}$ follows a fractionally integrated series with $d = 0.1$ and a mean of 2
  - Z_{2t} follows a fractionally integrated series with $d = 0.25$ and a mean of 3 (Z_{2t} does not differ within timepoints)
- Dependent Variable:

$$y = \bar{Y}_t + \beta_{1t} * x_1 - 0.05 * x_2 + 0.3 * \bar{Z}_{1t} + 0 * Z_{2t} + \epsilon \quad \text{, where}$$

$$\beta_{1t} \sim \mathcal{N}(\mu = 0.2, \sigma^2 = 0.1)$$

$$\epsilon \sim \mathcal{N}(\mu = 0, \sigma^2 = 1),$$

where $\bar{Y}_t$ follows a fractionally integrated series with $d = 0.4$ and a mean of 10

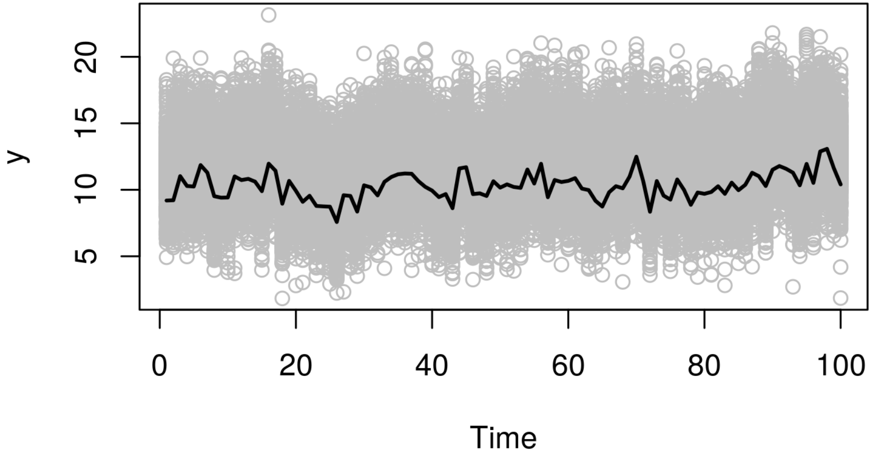
Data Overview I

```
data[496:505,]

    time       x1         x2          z1       z2         y
496    1 2.782894  99.649591  4.52390786 2.222936  6.556912
497    1 4.666993  15.769259  4.58101935 2.222936 10.326180
498    1 3.479995 -23.151975 -0.05688266 2.222936 13.371325
499    1 4.927197  -3.963808  6.34666239 2.222936 12.706198
500    1 4.671967  33.751661  0.60743779 2.222936  9.181135
501    2 2.888744  26.234304  2.96382521 1.969999 10.004992
502    2 3.334345 -22.937614  3.75093669 1.969999 13.311789
503    2 5.516160   5.706410  6.43928311 1.969999 11.906759
504    2 8.195076  31.802270  4.88271529 1.969999 14.047000
505    2 7.086206  95.629309  2.69268251 1.969999  6.587762
```

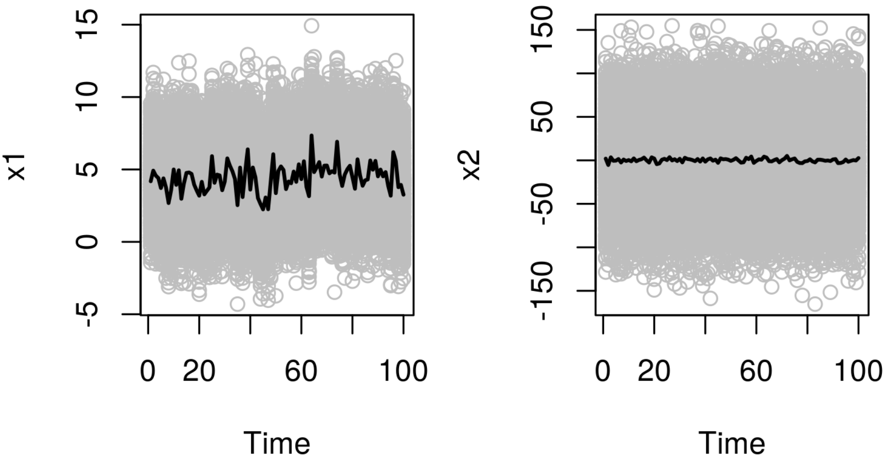
Data Overview II

Figure: Plot of Dependent Variable y Across Time



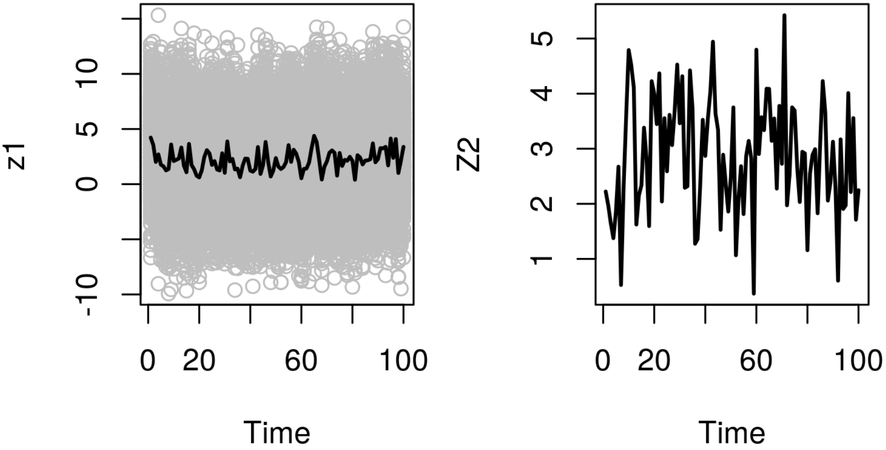
Data Overview III

Figure: Plot of Independent Variables Across Time



Data Overview IV

Figure: Plot of Independent Variables Across Time



Results without ArfimaMLM I

```
m1a <- lm(y ~ x1 + x2 + z1 + z2, data = data)

m1b <- lmer(y ~ x1 + x2 + z1 + z2
            + (1 | time), data = data)

m1c <- lmer(y ~ x1 + x2 + z1 + z2
            + (1 + x1 | time), data = data)
```

Results without ArfimaMLM II

Table: Results for Simple OLS and Multilevel Model

	Dependent variable:		
	y		
	OLS	*linear mixed-effects*	
	(1)	(2)	(3)
x1	0.200***	0.207***	0.207***
	(0.003)	(0.002)	(0.011)
x2	−0.050***	−0.050***	−0.050***
	(0.0002)	(0.0001)	(0.0001)
z1	0.025***	0.003*	0.003*
	(0.002)	(0.002)	(0.001)
z2	−0.223***	−0.224**	−0.181*
	(0.006)	(0.106)	(0.099)
Constant	11.546***	11.567***	11.455***
	(0.024)	(0.325)	(0.304)
Observations	50,000	50,000	50,000
R^2	0.655		
Log Likelihood		-72,501.370	-71,566.660
Bayesian Inf. Crit.		145,078.500	143,230.700
Note:			*p<0.1; **p<0.05; ***p<0.01

Results using ArfimaMLM I

```
m2a <- arfimaMLM(y.ydif ~ x1.xdif + x2 + z1.fd + z2.fd
                 + (1 | time)
                 , data = data, timevar = "time")

m2b <- arfimaMLM(y.ydif ~ x1.xdif + x2 + z1.fd + z2.fd
                  + (1 + x1.xdif | time)
                 , data = data, timevar = "time")
```

Results using ArfimaMLM II

Table: Results for ArfimaMLM

	Dependent variable:	
	y.ydif	
	(1)	(2)
x1.xdif	0.203***	0.204***
	(0.002)	(0.011)
x2	−0.050***	−0.050***
	(0.0001)	(0.0001)
z1.fd	0.204*	0.233**
	(0.118)	(0.105)
z2.fd	−0.071	−0.034
	(0.107)	(0.095)
Constant	0.055	0.055
	(0.110)	(0.110)
Observations	47,500	47,500
Log Likelihood	-69,004.340	-68,091.650
Akaike Inf. Crit.	138,022.700	136,201.300
Bayesian Inf. Crit.	138,084.100	136,280.200

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Getting the Package

- on CRAN:
  `http://cran.r-project.org/web/packages/ArfimaMLM/`
- on GitHub (development version):
  `https://github.com/pwkraft/ArfimaMLM/`

References

Box-Steffensmeier, Janet M and Renee M Smith. 1996. "The dynamics of aggregate partisanship." *American Political Science Review* 90(3):567–580.

Clarke, Harold D and Matthew Lebo. 2003. "Fractional (co) integration and governing party support in Britain." *British Journal of Political Science* 33(02):283–301.

Lebo, Matthew and Christopher Weber. 2015. "An Effective Approach to the Rolling Cross Sectional Design." *American Journal of Political Science* 59(1):242–258.

Lebo, Matthew J, Robert W Walker and Harold D Clarke. 2000. "You must remember this: dealing with long memory in political analyses." *Electoral Studies* 19(1):31–48.