SOC9052 Longitudinal Analysis

Session 2 (May 14th)

Eoin Flaherty
e.flaherty@qub.ac.uk

Mahoney and Thelen's (2010) model of institutional change processes

Conceptual Model	Characteristics
Displacement	Removal of existing rules and introduction of new (rapid breakdown of revolutions, gradual competition between old and new)
Layering	Introduction of new rules on top of or alongside old (amendment, revision, addition – challengers lack capacity to remove old)
Drift	Changed impact of existing rules due to environmental change (changing composition of electoral districts)
Conversion	Changed enactment of existing rules due to redeployment (actors exploiting ambiguities in existing structures)

The importance of theory

The comparative method (Rihoux and Ragin, 2009)

Macro-comparative logic: cases and theory

- Time-series work often deals with societies, economies, states, structural indicators
- Small N/Big T?
- To what should we limit our generalisations? OECD, advanced capitalist democracies, Europe, Antipodes?
- Single unit or pooled cross-sections?
- Upstream theory (input conditions) and downstream theory (assumptions and specification)

Modest generalisation

- To what should we limit our generalisations? OECD, advanced capitalist democracies, Europe, Antipodes?
- Can we generalise information on associations/mechanisms?
- Does discovering time-dependence enhance our explanatory power?

3

What is time series analysis?

Static model: $Y_t = \alpha_0 + \beta_o X_t + \varepsilon_t$

year	gini (Y_t)	union (X_t)
2005	34.9	28.6
2006	35.4	28.4
2007	35.7	28.3
2008	35.8	27.6
2009	35.7	27.5

Basic dynamic model: $Y_t = \alpha_0 + \beta_0 X_{t-1} + \varepsilon_t$

year	gini (Y_t)	union (X_{t-1})
2005	34.9	29.5
2006	35.4	28.6
2007	35.7	28.4
2008	35.8	28.3
2009	35.7	27.6

4

Three confounders: autocorrelation, non-stationarity, spuriousness

(1) Error autocorrelation / serial correlation

What is it?

Proximate error terms in cross-section regressions are uncorrelated ($Y_t = \alpha_0 + \beta_o X_t + \varepsilon_t$)

Errors are the unexplained part of the data-generating process, *residuals* are the unexplained part of our statistical model.

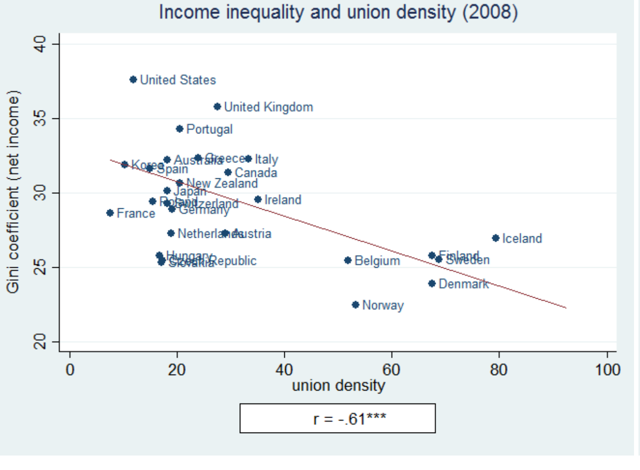
But…time-series data points have a natural sequence. Autocorrelation cannot arise in cross-section as all data are sampled at the same time (with exceptions, for example spatial autocorrelation).

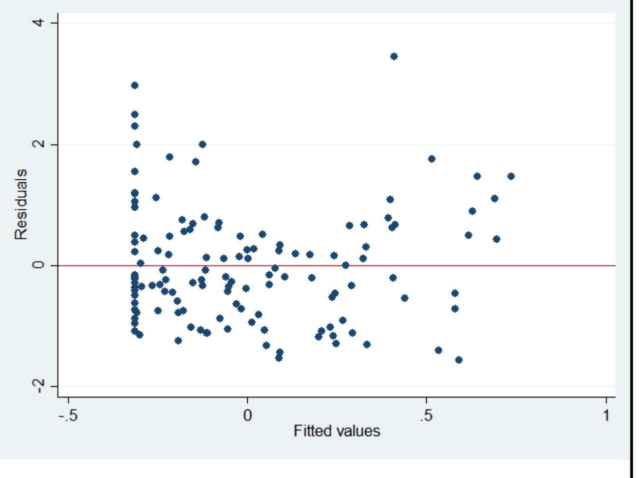
Why does it matter?

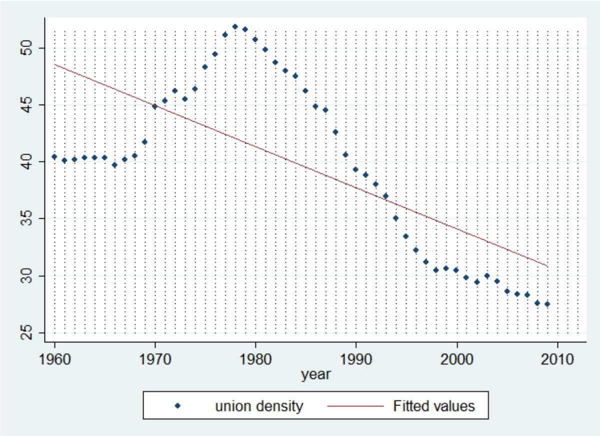
Error autocorrelation results in underestimated error variance ($se(\widehat{\beta})$), and inflated t: $t = \frac{\widehat{\beta}}{se(\widehat{\beta})}$

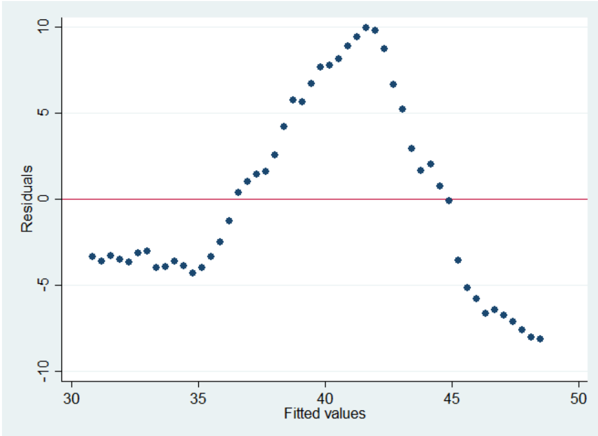
How do we deal with it?

Diagnose it, include dynamics, and model structural breaks.









7

Three confounders: autocorrelation, non-stationarity, spuriousness

(2) Non-stationarity

What is it?

Social science data series are often not mean-reverting, or strongly trended.

We can understand this quite intuitively, by thinking back to our discussion of dynamics and lags in social systems.

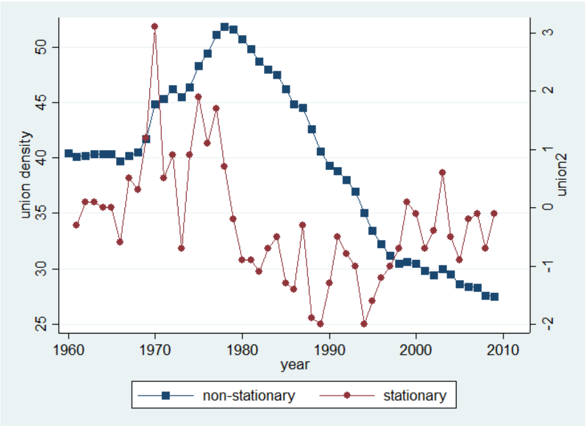
Why does it matter?

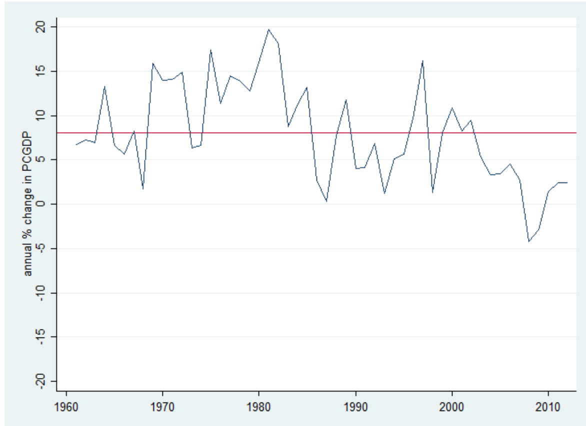
Regressions on non-stationary time series can inflate R-squared leading to incorrect conclusions on the model's explanatory power.

How do we deal with it?

Difference our variables, or try different model specifications able to cope with non-stationarity (cointegration).

8





9

Three confounders: autocorrelation, non-stationarity, spuriousness

(3) Spuriousness

What is it?

The end-product of autocorrelation and non-stationarity which produces biased inference.

The phenomenon of regressions on non-stationary time series producing results suggesting they are closely related.

Why does it matter?

Misleading diagnostics lead to inaccuracy model selection.

How do we deal with it?

Careful attention to pre-specification tests (for non-stationarity), post-estimation tests (for autocorrelation), and careful use of theory in our model selection.

10

Three confounders: autocorrelation, non-stationarity, spuriousness

(3) Spuriousness vs Super-Consistency

Spurious regression model:

A regression with two unrelated, non-stationary time series can reveal a close relationship, simply because the t-statistics are inflated due to non-stationarity. We will also find upward bias in our R^2, and F statistics.

Super-consistent cointegrated regression:

In some cases where the variables really are related, non-stationarity can improve overall model performance. A super-consistent regression produces more accurate estimates for larger values of T (i.e. longer series length). We can explore this possibility further in the context of Error-Correction Model (ECM) specifications.

Modelling dynamic systems

Modelling dynamic systems (deBoef and Keele, 2008)

> "Theories…typically tell us only generally how inputs relate to processes we care about. They are nearly always silent on which lags matter, whether levels or changes drive Y, what characterises equilibrium behaviour, or what effects are likely to be biggest in the long run"

Static model: $Y_t = \alpha_0 + \beta_o X_t + \varepsilon_t$

Finite distributed lag: $Y_t = \alpha_0 + \beta_o X_t + \beta_0 X_{t-1} + \varepsilon_t$

General model: $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_o X_t + \beta_0 X_{t-1} + \varepsilon_t$

A better (theoretical) question….? How does our model specification fit the underlying data-generation process?

Modelling dynamics systems – what specification?

Static model: $Y_t = \alpha_0 + \beta_o X_t + \varepsilon_t$

Impact of X on Y is instantaneous (assumes changes in X work through system instantaneously)

Finite distributed lag: $Y_t = \alpha_0 + \beta_o X_t + \beta_0 X_{t-1} + \varepsilon_t$

X impacts Y over a period of 1 (tax reform on net income)

General model: $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_o X_t + \beta_0 X_{t-1} + \varepsilon_t$

Past values of Y influence Yt, as well as current and past values of X (previous values of presidential support Y influence current, as well as level of confidence in economy)

Table 4. Non-agricultural production function estimates

Dependent variable: $\ln Q_N$	(1)	(2)	(3)
Estimation method:	**OLS**	**GLS**	**ML**
Variables:		**AR(1)**	**AR(1)**
$\ln L_N$	0.732** (12.82)	0.734** (9.61)	0.734** (10.31)
$\ln K_N$	0.208** (5.59)	0.212** (4.40)	0.212** (4.74)
T_{1984}	0.058** (25.41)	0.057** (20.13)	0.057** (20.67)
D_{1967}	−0.148** (−7.94)	−0.140** (−6.32)	−0.140** (−5.19)
D_{1968}	−0.268** (−15.06)	−0.238** (−5.47)	−0.238** (−4.37)
Constant	8.015** (19.80)	7.892** (15.59)	7.894** (14.60)
AR(1)		0.315* (2.18)	0.309 (1.26)
Observations	45	45	45
R^2	0.9993	0.9993	
Constant Returns to Scale test	F = 5.015 [0.031]	F = 2.332 [0.135]	$\chi^2 = 2.395$ [0.122]
Non-structural residual (e_t) diagnostics:			
Breusch-Godfrey autocorrelation LM χ^2 test	3.049 [0.081]	1.222 [0.269]	0.921 [0.337]
White heteroscedasticity χ^2 test	4.942 [0.085]	5.644 [0.059]	5.079 [0.079]
Jarque-Bera normality χ^2 test	2.635 [0.268]	2.609 [0.271]	2.473 [0.276]
Ramsey Reset F test	1.91 [0.146]	1.91 [0.145]	1.94 [0.140]