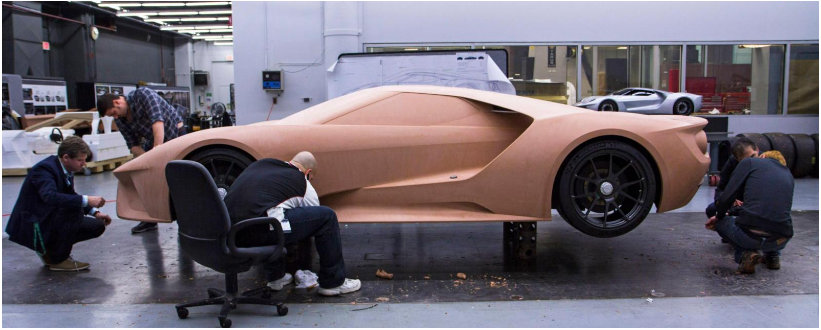
Building Empirical Configurational Typologies with Regularized Log-linear Models

Juraj Medzihorsky

ECPR General Conference 2018



Source: http://www.bbc.com/autos/story/20161111-why-car-designers-stick-with-clay

Motivation

Failing and Successful Sequences of Democratization (FASDEM)

"Failing and Successful Sequences of Democratization" (FASDEM) is a new cutting-edge research program led by Prof. Staffan I. Lindberg, Director of V-Dem Institute, which promises to revolutionize understanding of both the failing trajectories of democracy, and the successful pathways, and tell why this is the case.

FASDEM will capitalize on a set of novel analytical approaches and methods adapted from modeling in evolutionary biology; it will also take advantage of V-Dem dataset. With a large set of V-Dem indicators measured over many years, it would become possible for the first time to explore transition sequences. V-Dem data will make it possible to search for endogenous democratic sequences not necessarily contemplated by current theory – and do so with regards to long chains of sequential relationships between many factors.

Detailed information about the project can be found here FASDEM project.pdf (1.5 MB) .

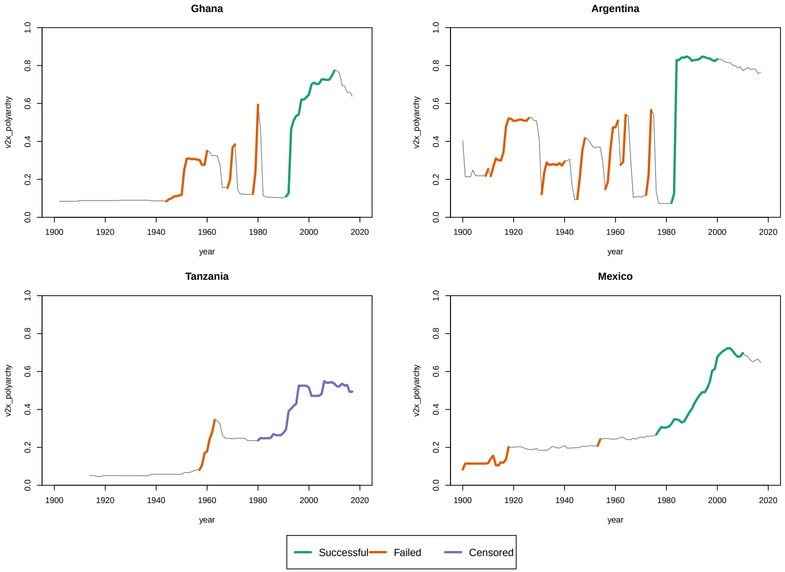
ERC Consolidator Grant 2017 - 2021

European Research Council

Research Questions and Methods

Sub-Project 1 Which are the failing & successful sequences of democratization?	**Sub-Project 2** What are the determining causal relationships in these sequences?
New Sequencing Methods Adapted from Evolutionary Biology (parasite-host systems) A. Graphical Investigation B. Frequency Tables C. Dependency Analysis D. Bayesian Dynamic Systems	Causal Identification Methods Genetics/Bayesian Statistics/Econometrics A. Sequencing Algorithms B. Dynamic Treatment Regimes C. Vector Auto-Regression

Democratization episodes



Source: Maxwell et al. (2018) “How democracy develops”

Empirical Configurational Types

Theory and empirics

Theory and empirics can drive typologies.

Theory *always* plays a role at least in attribute and unit definition and selection.

Empirical types summarize the observations.

Centrality and extremity

Empirical types can be *central* or *extreme*

Central types summarize a group of observations.

Extreme types define what is possible.

Emergence

How do empirical configurational types emerge?

Processes with *one* or *more* steps.

One step: units get the values and keep them.

More steps: units' attributes can change.

Multi-step processes

Some configurations *more* and others *less* persistent.

Sticky vs. *slippery* configurations.

Empirical types and their opposites

A configuration that is unusually common constitutes a *type*.

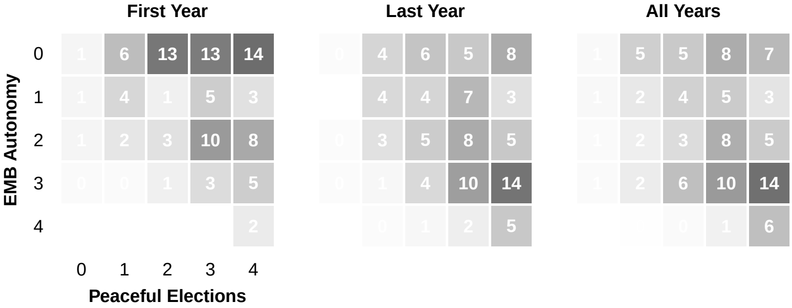
What about *rare* configurations?

What about configurations that are *neither*?

A simulation with sticky types

A simulation with a sticky and a slippery type

An empirical illustration with democratization episodes



Methods

Existing methods

Cluster Analysis

Finite Mixture Models

Qualitative Comparative Analysis

Configural Frequency Analysis

Hard partitioning into groups: everybody belongs somewhere and only there.

Not probabilistic.

Finite Mixture Models

Soft and *probabilistic* partitioning.

Conventional FMM: groups interpreted with central types.

Archetypal analysis: each observation as a mixture of realistic extreme types.

Used purely descriptively crisp-set QCA provides concise descriptions of conditional distributions.

Not probabilistic.

Configural Frequency Analysis

Statistical approach developed in psychometrics (Lienert and Krauth 1975).

Compares the observed frequencies to some baseline:

- A *type* is more common
- An *anti-type* is less common
- Some configurations can be neither

What is more and less common?

Both classical and Bayesian versions rely on arbitrary thresholds.

Extending CFA with regularized log-linear models

Conventional CFA baseline is

$$\hat{y}_{ab} = \lambda^{(0)} + \lambda_a^{(A)} + \lambda_b^{(B)}$$

The corresponding saturated model is

$$\hat{y}_{ab} = \lambda^{(0)} + \lambda_a^{(A)} + \lambda_b^{(B)} + \lambda_{ab}^{(AB)},$$

where $\{\lambda_{ab}^{(AB)}\}$ are the interactions and are $++$ for types and $--$ for anti-types.

Sparsity-inducing regularization shrink small estimates towards zero but keeps large ones retaining model fit.

Shrink the ‘type’ interactions $\{\lambda_{ab}^{(AB)}\}$ in

$$\hat{y}_{ab} = \lambda^{(0)} + \lambda_a^{(A)} + \lambda_b^{(B)} + \lambda_{ab}^{(AB)}$$

with a Bayesian lasso

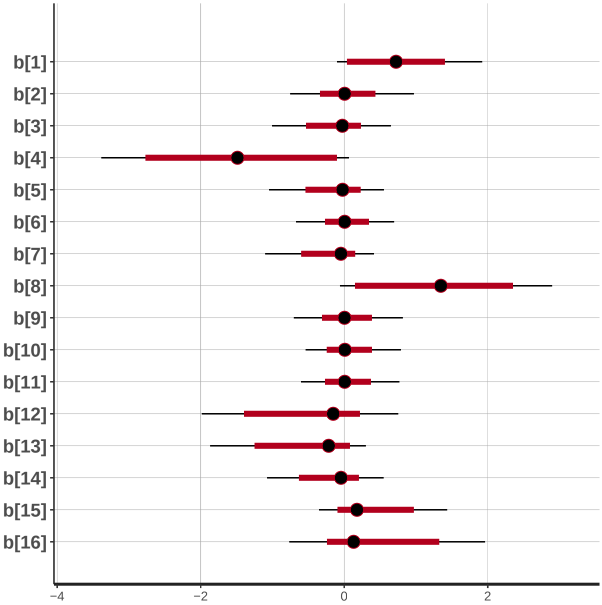
$$\lambda_{ab}^{(AB)} \sim Laplace(0, \tau\sigma_{ab}).$$

Source: Wikimedia Commons

Snee 1974, data

Hair \ Eye	Brown	Blue	Hazel	Green
Black	68	20	15	5
Brown	119	84	54	29
Red	26	17	14	14
Blond	7	94	10	16

Snee 1974, $\lambda_{ab}^{(AB)}$ shrinkage



Snee 1974, clear types & anti-types

Hair \ Eye	Brown	Blue	Hazel	Green
Black	68	20	15	5
Brown	119	84	54	29
Red	26	17	14	14
Blond	7	94	10	16

Discussion

Other sparsity-inducing regularization methods have different advantages

- smooth lasso
- horseshoe
- horseshoe+
- spike & slab
- ...

What about settings with > 2 right-hand side variables?

Hierarchical lasso.

Why not model the dynamical process more closely with a Hidden Markov Model?

Managing the bias-variance trade-off: large number of parameters.

Sometimes we have only a couple snapshots.

Thank you!