How to Compare When Data Come from Diverse Sources: A 4 - Level Model of Change in Institutional Trust over Time

By Claire Durand,
Luis Patricio Peña Ibarra & Nadia Rezgui

Outline

- The problem: How to compare and keep the maximum information available
  - ▸ Various data sets and measures
  - ▸ Various political/electoral/economic situations

- The solution: Work with levels of analysis

- A multilevel approach and its problems

- Concretely, how does it work?

- Concretely, an example

- Discussion and conclusion

A different approach: Once the data have been collected, how to use it and keep as much information as possible?

- Since we are working with data that have been collected,
  - ‣ Similar concepts are often measured but different questions are used to measure them.

- We conceptualize the different measures as samples of all the measures of a similar concept that can be used.

- Therefore, we have samples at different levels, i.e., the levels of measures, of respondents, of countries and years.

Why use a multilevel approach?

- We can assess the different effects at the level where they operate, within individuals, at the individual level, over time and at the country level

- We do not have to deal with missing values and keep only the cases where the same information is available for all the cases, years, or countries.

- We can model cross-level effects,
  ‣ Like the possible effect of the countries' average GDP on trust in the government.
  ‣ Or the impact of age-group or of time on trust in religion.

A 4-level longitudinal model with repeated measures
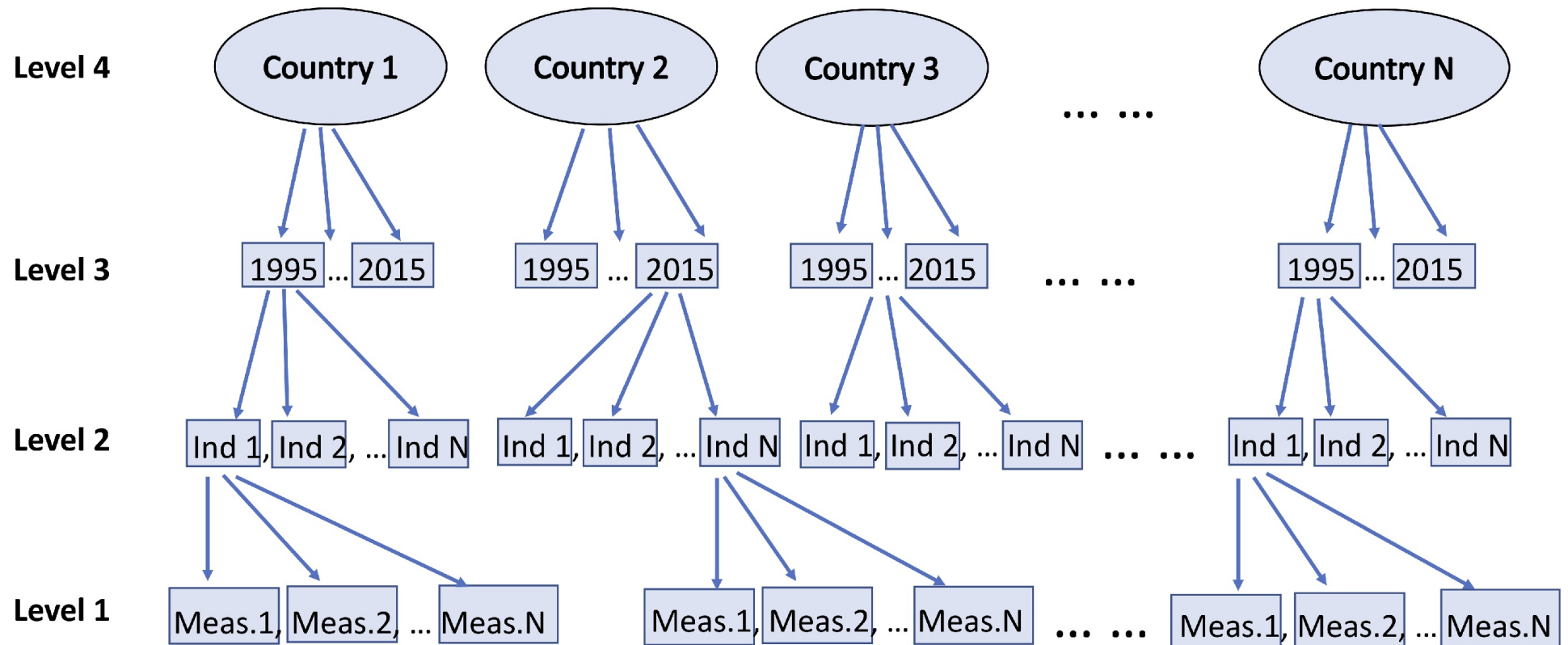
The Data

- The Barometers:
  - ▸ Latino Barometro (1995-2016)
  - ▸ East Asia Barometer (2001-2012), Asian Barometer (2003-2007)
  - ▸ Arab Barometer (2008-2014)
  - ▸ Africa Barometer (2001-2015)
- LAPOP (2004-2016)
- World Values surveys (WVS)
  - ▸ Surveys from 1995 to 2014 for Africa and West Asia.

A multilevel approach and its problems

At level 4: country or country-source?

- There are more than one survey conducted during the same year in a given country by different survey projects,
  - Which means that we need to be able to test whether, on average, there is a difference according to the source of data.
  - Solution: The highest level is a "country-source" level. For example, in 2013, the Arab Barometer and the WVS both conducted a survey in Algeria. Country-source codes are 120 & 122, depending on the source of data.

- Which allows for:
  - Adding a variable identifying the source of data in order to control for the different methodological features -- answer scale, question wording, etc. -- of the different projects.

A multilevel approach and its problems

At level 3: country-source-year = survey level

- The source of data is controlled at the country level.

- Multiple surveys are conducted over time in each country. The time level is intermediary, i.e., both
  - ▸ nested within country-source and
  - ▸ having respondents nested within each survey.

- Consequence: Add a variable identifying the year when the survey was conducted and, as a longer period of data becomes available, variables for a quadratic (or even cubic) effect of time.

- At this level, we can add methodological characteristics of specific surveys (within projects).

A multilevel approach and its problems

At level 2, the respondents' level: ex-post harmonization issues

- The respondent level is where most harmonization problems occur. The only non problematic variable – for now – is sex.

- Some projects/countries ask age in years, others in categories. Categories may not be the same in different projects.
  - ‣ Solution: The only common denominator: young (15-29), middle (30-59), old (60+)

- Level of education: the educational systems vary. May be difficult to place technical training. 3% not asked.
  - ‣ Harmonize in 5 categories:
    - – No formal education, primary, secondary, technical, university.

A multilevel approach and its problems

- The major problem is not harmonization but the fact that some questions were not asked in some surveys.

- Occupation: 23% not asked
  - ▸ Common categories: employed, out of work, homemaker, retired, student.

- Subjective Income: (36% not asked)
  - ▸ Four categories from "sufficient, can save" to "not sufficient, have big problems"

- Attitudes as independent variables.
  - – Satisfaction with democracy: 37% not asked
  - – Support for democracy: 22% not asked

A multilevel approach and its problems

At level 2, the respondents' level: methodological information

- We may ask whether the proportion of questions answered by a respondent, i.e., item non-response, is related to the level of trust.

- In order to control for item non-response, we recuperate the number of questions asked to a respondent (at the survey level) and the number of questions answered.

A multilevel approach and its problems

At level 1, the measurement level

- Measures asked in a survey are samples of all the measures that can be asked to measure a concept.

- The main concept for this study is institutional trust but we could use the method for other concepts like attitude towards the protection of environment, etc.

- There are 110 institutions to date if we consider each different institution for which trust is asked.
  - How to decide which institutions are similar enough to be grouped together and which ones should be kept separate?

A multilevel approach and its problems

At level 1, the measurement level

- **First problem**: The measures have to be on the same **scale**, which means
  - First, recode all scales so that the highest number corresponds to higher trust.
  - Second,
    - The current situation: Expanding the scale: from 1 to 4 (Barometers & WVS) to 1,3,5,7 (LAPOP).
    - To be solved: In new surveys, we have scales of 0 to 10, or 1 to 10 or scales of only 2 or 3 anchors.

- The **second problem**: Different question wordings:
  - Trust vs Confidence: In our case, not a problem because most if not all the surveys are not conducted in English. In other languages, there is only one word for trust.

A multilevel approach and its problems

At level 1, the measurement level

- The third problem: The institution on which the trust question is asked.
  - ▸ Example1: Four institutions related to elections: Elections, National Election Commission, Results of the next election and Secret Vote.
  - ▸ Example 2: Can we group together European Union, UN, International Monetary Fund, World Bank, World Trade Organization, UN Development Program in a category "International Organizations"?

- Solution: Check whether the institutions have similar averages and Std. Deviations in each region and for each survey project before grouping.
  - ▸ But some specificities: The "Russian" question, i.e., the state-governement vs the president.

Concretely, how does it work: Synthesis of the process

Inputs:

Barometers

WVS

LAPOP

MAIN FILE:
Individual level,
All variables related
to trust

**Level 1 file:
Measurement Level**
One line per question
on trust per respondent

Level 2 file: Respondent
Main file minus
trust questions

Level 3 file: Survey
Country-source-year
*Vars: Time, methodol.
features, mean trust*

**Level 4 file:
Country-source**
Vars: Region, mean trust

Time-Series file:
Mean trust per
institution-country-year
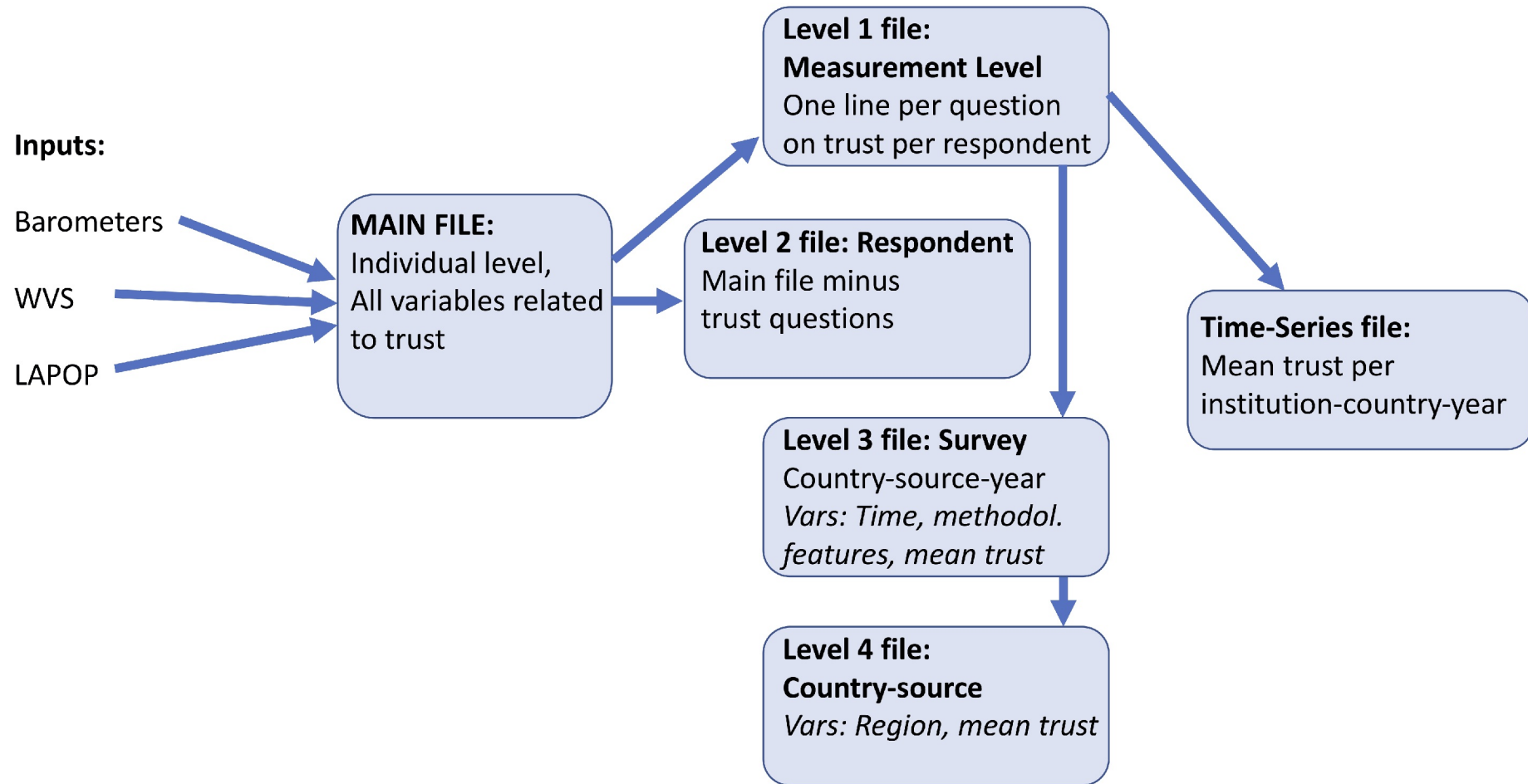
Concretely, how does it work?

- A Main database combines the data from all the survey projects at the individual level, including
  - ▸ Id for the country, year and source of data,
  - ▸ Harmonized indicators of socio-demographics,
  - ▸ Harmonized indicators of general attitudes of interest (perception of democracy, participation in political activities, etc.), if available in enough surveys.
  - ▸ One variable for each question on trust in an institution that has been asked in the survey.
    - – If institution already surveyed in a prior survey, use the same name for the variable.
    - – If not, introduce a new variable.

1st step: merging & harmonizing: The main file

Concretely, how does it work

- Restructure the file so that there is one line per question on trust in an institution asked to a given respondent and a variable identifiying the institution:
  - A line has five variables
    - 1. Country-source identifier
    - 2. Country-year-source identifier
    - 3. Respondent identifier
    - 4. Institution identifier
    - 5. Answer on a trust question
  - Recode the institutions into larger categories. And create dummy variables for each recoded institution.
  - Create a variable that indicates item non-response and send it back to the level 2 file.

Second step: the level 1 file

Concretely, how does it work?

Step 3: Create the level 3 file (country-year-source)

- It is necessary with HLM, the only software available now for 4-level models.

- Aggregate the level 1 file at the country-year-source (i.e., the survey) level:
  - ▸ Compute Time centered, Time centered at power 2.
  - ▸ Introduce variables related to survey methodology (nb questions asked on trust, for example)

The level 3 file:
Country_source_year = survey

Concretely, how does it work?

- Create the level 4 file (necessary with HLM), the only software available now for 4-level models.

- Aggregate at the Country-source level:
  - ▸ Polydichotomize the variable indicating the source of data in order to have one variable per source of data.
  - ▸ Compute a variable for region and polydichotomize to have one variable per region.

The level 4 file: Country-source level

The level 2 file is at the respondent level

Step 5: The level 2 file

- The original main harmonized file is a level 2 file.

- Clean it from all the unnecessary information, i.e. the information that is not at the respondent level: delete all the trust variables and save as level 2.

The time-series file

- From the level 1 file, aggregate by country, year and institution in order to have one line per country-year per institution.

- It becomes possible to perform local regressions that give an idea of the trends in trust for different institutions in different regions and overall.

Time-series file: country-year-institution category

Synthesis of the data

At the time being.

- Measures: 12,340,179

- Respondents: 1,023,681

- Countries-years: 659 (97 double)

- Countries-source-years: 756

- Countries: 102 (47 double)

- Countries-source:149

- Time-series: 5913

- In progress: Complete Africa and Mena with new waves, add WVS in Asia + new waves, add Eastern Europe and Russia.

An extra step: match external data

- Different sources of data can be matched at level 3 -
  - country-year -- or at level 4 -- country:
  - ‣ Quality of governement data
  - ‣ V-DEM project data
  - ‣ Word Governance indicators

- For example:
  - – Solt Gini, GDP, WGI, Polity index, proportion of urban population, ethnic and religious diversity, etc.

- The main problems:
  - ‣ Some indices do not vary enough over time: preferable to introduce them at the country level.
  - ‣ Lack of data outside the western world.
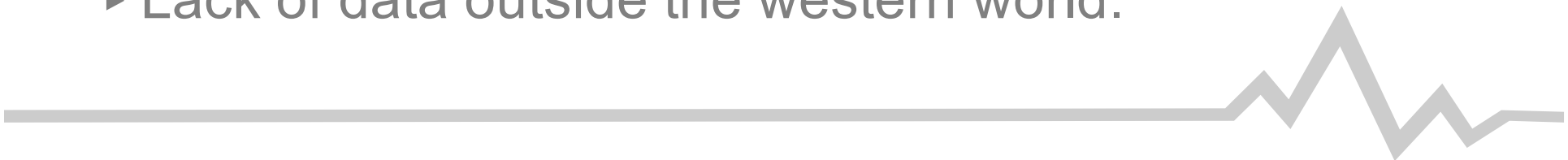
Concretely : an example of results

Trust in institutions - predictors at 3 & 4 levels						
	Model 0		Example 1		Example 2	
Intercept	4.145	***	4.239	***	3.598	***
Level Measure						
Media (REF)						
Church			0.817	***	0.823	***
Year			-0.090	***	-0.089	***
Trade Unions			-0.682	***	-0.682	***
ONG- Civil Society			-0.262	***	-0.262	***
Army			-0.005	n.s.	-0.005	n.s.
Police			-0.515	***	-0.515	***
Public Admin.			-0.442	***	-0.442	***
Judiciary			-0.500	***	-0.500	***
Finance			-0.204	***	-0.204	***
Enterprises			-0.372	***	-0.372	***
Governments			-0.295	***	-0.295	***
Polity2			-0.031	***	-0.031	***
Political Parties			-0.916	***	-0.916	***
Polity2			-0.038	***	-0.038	***
Elections- Elect. Commis.			-0.524	***	-0.524	***
Polity2			0.016	***	0.016	***
International Org.			-0.172	***	-0.172	***
Level Respondent						
woman			0.002	n.s.	0.002	n.s.
Young (Less than 30)			0.011	***	0.011	***
Old (60 plus)			0.099	***	0.099	***
Prop_Non-resp.			0.219	***	0.219	***
Level Country-Source-Year						
Time			0.023	***	0.021	**
Time2			0.001	n.s.	0.001	n.s.
Prop_urban population			-0.015	***	-0.010	**
LnGDP			0.128	*	0.122	*
Polity2			-0.002	n.s.	0.004	n.s.
Gini evolution			0.001	n.s.	-0.002	n.s.
Level Country-Source						
Central/South America (REF)						
Asia					0.339	*
Africa					0.574	***
West Asia N. Africa					0.444	*
LAPOP					0.267	*
WVS					0.417	**
Variance						
Measures	2.537	63.6%	2.294	61.5%	2.294	62.7%
Respondents	1.046	26.2%	1.064	28.6%	1.064	29.1%
Country-Source-Year	0.099	2.5%	0.168	4.5%	0.167	4.6%
Country-Source	0.304	7.6%	0.202	5.4%	0.132	3.6%
Total	*3.986*		*3.728*		*3.657*	
Deviance	38255971		32450232	*5805739*	32450199	*33*
dl	5		34	*29*	37	*3*

■ A complete analysis with 4 levels.

At level 1: Trust in institutions

Trust in institutions - predictors at 3 & 4 levels						
	Model 0		**Example 1**		**Example 2**	
Intercept	4.145389	***	4.239	***	3.598344	***
Level Measure						
Media (REF)						
Church			0.817	***	0.823	***
Year			-0.090	***	-0.089	***
Trade Unions			-0.682	***	-0.682	***
ONG- Civil Society			-0.262	***	-0.262	***
Army			-0.005	n.s.	-0.005	n.s.
Police			-0.515	***	-0.515	***
Public Admin.			-0.442	***	-0.442	***
Judiciary			-0.500	***	-0.500	***
Finance			-0.204	***	-0.204	***
Enterprises			-0.372	***	-0.372	***
Governments			-0.295	***	-0.295	***
Polity2			-0.031	***	-0.031	***
Political Parties			-0.916	***	-0.916	***
Polity2			-0.038	***	-0.038	***
Elections- Elect. Commis.			-0.524	***	-0.524	***
Polity2			0.016	***	0.016	***
International Org.			-0.172	***	-0.172	***

In blue: cross-level interactions

- Church highest, but decreasing.
- Political parties & trade unions lowest.
- More democratic: Hi Elections; Low GVT & pol. parties.

At level 2: Respondents

Trust in institutions - predictors at 3 & 4 levels						
	Model 0		**Example 1**		**Example 2**	
Intercept	4.145	***	4.239	***	3.598	***
Level Measure						
Media (REF)						
Church			0.817	***	0.823	***
Year			-0.090	***	-0.089	***
Trade Unions			-0.682	***	-0.682	***
ONG- Civil Society			-0.262	***	-0.262	***
Army			-0.005	n.s.	-0.005	n.s.
Police			-0.515	***	-0.515	***
Public Admin.			-0.442	***	-0.442	***
Judiciary			-0.500	***	-0.500	***
Finance			-0.204	***	-0.204	***
Enterprises			-0.372	***	-0.372	***
Governments			-0.295	***	-0.295	***
Polity2			-0.031	***	-0.031	***
Political Parties			-0.916	***	-0.916	***
Polity2			-0.038	***	-0.038	***
Elections- Elect. Commis.			-0.524	***	-0.524	***
Polity2			0.016	***	0.016	***
International Org.			-0.172	***	-0.172	***
Level Respondent						
woman			0.002	n.s.	0.002	n.s.
Young (Less than 30)			0.011	***	0.011	***
Old (60 plus)			0.099	***	0.099	***
Prop_Non-resp.			0.219	***	0.219	***

- No difference in average trust according to sex
- Both younger and older people are more trustful than middle-aged people.
- More Item non response = more trust.

At level 3: Country-year-source

Level Country-Source-Year						
Time			0.023	***	0.021	**
Time2			0.001	n.s.	0.001	n.s.
Prop_urban population			-0.015	***	-0.010	**
LnGDP			0.128	*	0.122	*
Polity2			-0.002	n.s.	0.004	n.s.
Gini evolution			0.001	n.s.	-0.002	n.s.
Level Country-Source						
Central/South America (REF)						
Asia					0.339	*
Africa					0.574	***
West Asia N. Africa					0.444	*
LAPOP					0.267	*
WVS					0.417	**
Variance						
Measures	2.537	63.6%	2.294	61.5%	2.294	62.7%
Respondents	1.046	26.2%	1.064	28.6%	1.064	29.1%
Country-Source-Year	0.099	2.5%	0.168	4.5%	0.167	4.6%
Country-Source	0.304	7.6%	0.202	5.4%	0.132	3.6%
Total	*3.986*		*3.728*		*3.657*	
Deviance	38255971		32450232	5805739	32450199	33
dl	5		34	29	37	3

- Trust increase with time.

- Hi Prop urban population= lower trust.

- Hi GDP= higher trust.

At level 4: Country-source

Level Country-Source-Year						
Time			0.023	***	0.021	**
Time2			0.001	n.s.	0.001	n.s.
Prop_urban population			-0.015	***	-0.010	**
LnGDP			0.128	*	0.122	*
Polity2			-0.002	n.s.	0.004	n.s.
Gini evolution			0.001	n.s.	-0.002	n.s.
Level Country-Source						
Central/South America (REF)						
Asia					0.339	*
Africa					0.574	***
West Asia N. Africa					0.444	*
LAPOP					0.267	*
WVS					0.417	**
Variance						
Measures	2.537	63.6%	2.294	61.5%	2.294	62.7%
Respondents	1.046	26.2%	1.064	28.6%	1.064	29.1%
Country-Source-Year	0.099	2.5%	0.168	4.5%	0.167	4.6%
Country-Source	0.304	7.6%	0.202	5.4%	0.132	3.6%
Total	*3.986*		*3.728*		*3.657*	
Deviance	38255971		32450232	*5805739*	32450199	33
dl	5		34	29	37	3

- Higher trust when source is LAPOP or WVS.

- Higher trust outside Latin America, even more in Sub Saharan Africa.

Distribution of Variance

Level Country-Source-Year						
Time			0.023	***	0.021	**
Time2			0.001	n.s.	0.001	n.s.
Prop_urban population			-0.015	***	-0.010	**
LnGDP			0.128	*	0.122	*
Polity2			-0.002	n.s.	0.004	n.s.
Gini evolution			0.001	n.s.	-0.002	n.s.
Level Country-Source						
Central/South America (REF)						
Asia					0.339	*
Africa					0.574	***
West Asia N. Africa					0.444	*
LAPOP					0.267	*
WVS					0.417	**
Variance						
Measures	2.537	63.6%	2.294	61.5%	2.294	62.7%
Respondents	1.046	26.2%	1.064	28.6%	1.064	29.1%
Country-Source-Year	0.099	2.5%	0.168	4.5%	0.167	4.6%
Country-Source	0.304	7.6%	0.202	5.4%	0.132	3.6%
Total	*3.986*		*3.728*		*3.657*	
Deviance	38255971		32450232	5805739	32450199	33
dl	5		34	29	37	3

- 63% of the variance is within individuals, between measures.
- Explained variance:
- At level 1: 9,6%
- Between countries: 56.6%

Questions to be resolved:

- What should we do about weighting?
  - At the individual level: not all files have equivalent weights, or even weights.
  - At the country-level: It would give a weight that is way too large to countries like Brasil in Latin America or China in Asia.

- Decide on the level at which external data should be matched.

- Find more relevant indicators of the context of each country.

Conclusion

- The method is now well developped, systematized and described.

- The distribution of variance between levels show how important it is to take into account the within individuals-between measures variance.

- Another advantage is the possibility of cross-level interactions.

- There is some more recent data to add in order to have more powerful analyses.
  - With the introduction of Eastern Europe this summer and new data for Asia and Africa, we will have covered all the countries outside of the "western world".