Historical Analysis of Survey Data and Survey Results: *The Incredible Possibilities Afforded by Longitudinal Multilevel Analysis Using Time at a Higher level*

Claire Durand, Isabelle Valois and François Yale
Department of sociology,
Université de Montréal

Presented at the 6th ESRA Conference, July 13-17,
Reykjavik, Iceland

Outline

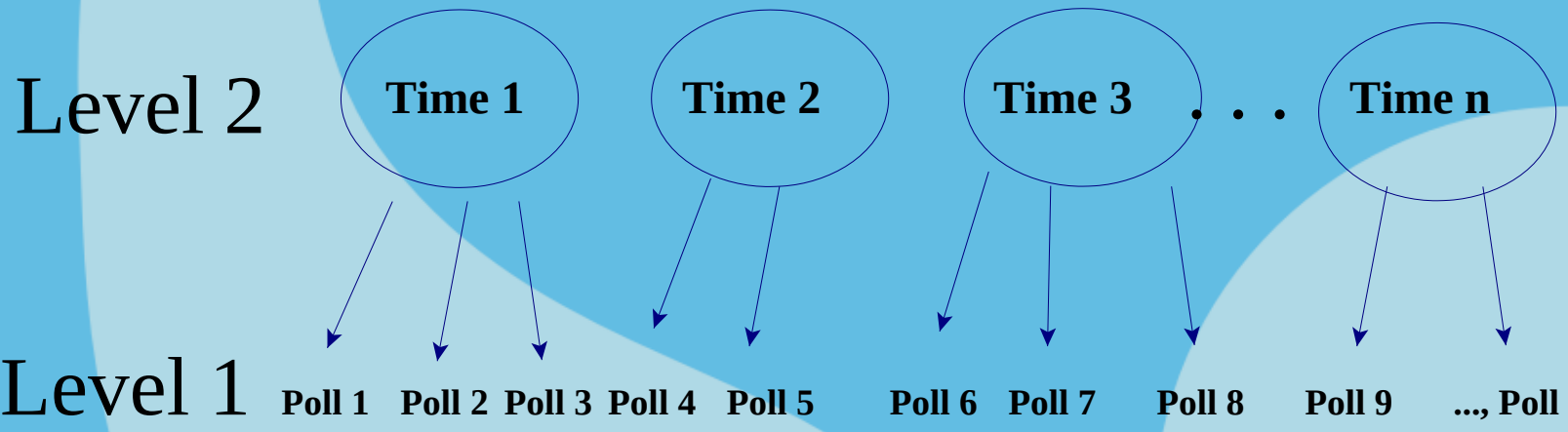
- The idea, the problem and the solution
- Example 1: Combining survey results: 2-level models: Evolution of support for sovereignty
  - Impact of methods, question wording and events
- Example 2: Combining survey results: 2-level model: Evolution of voting intention for Obama and Romney in 2012
  - Impact of methods
- Example 3: Combining survey data files: 3-level model
  - Evolution of trust in Canada
  - Impact of object of trust, sexe&age, question wording and time.
- Conclusion & future path

The idea, the problem & the solution

- We have access to huge quantities of data on similar topics from different sources.
- We would like to be able to use the full potential of all these data
  - To trace change over time of different attitudes and behaviors
  - To assess whether change over time is the same in different regions & contexts or for different age groups, etc.
- But we are hindered by the fact that
  - Measures of similar concepts are not always the same
  - The surveys that we use may have different methodological features that have to be taken into account.

Combining survey results using multilevel modeling: two levels

- At level 2: change over time and its predictors, i.e. events and time itself.



- At level 1: variation between polls and its predictors, ie., methods, sample size, question wording, etc.

Example 1: Change in support for sovereignty in Quebec (Yale & Durand, 2011)

- **Close to 700 polls between 1976 and 2008, 7 periods, 3 under study**
- **Poll results differ at level 1 in:**
  - Question wording, i.e. whether the question pertains to an opinion or to voting intention
  - Constitutional option, i.e. whether the question refers to sovereignty with an association/partnership with the rest of Canada, to sovereignty per se, to independence or to separation
  - Proportion of undecideds and sample size
- **The research questions are:**
  - Does support differ according to the question asked --voting intention or attitude, constitutional option?
  - What is the likely change in average support for sovereignty and which events, if any, influence change?
  - Is change similar for different constitutional options?

Example 1: Change in support for sovereignty in Quebec (Yale & Durand, 2011)

- At level 1:
  - Question wording (constitutional option):
    - Separation
    - Independence
    - Sovereignty
    - Sovereignty-association or sovereignty-partnership
  - Type of question: voter intent vs favorability, mandate
  - Proportion of undecideds
  - Sample size
- At level 2:
  - Time, time squared, time cubic;
  - Elections;
  - Events: Accords --Meech Lake, Charlottetown --and sponsorship scandal.

At level 1

Effects related to polls and questions (Yale & Durand, 2011)

Table 1 – Summary of Average Effects Linked to Measure

		1976–1979	1989–1995	1995–2008
Fixed effects				
	Intercept	**39.84***** **(1.51)**	**60.71***** **(1.05)**	**48.98***** **(1.57)**
	Voterint	***n.s.***	**-3.18***** **(0.76)**	**-4.04**** **(1.35)**
	Sovereignty	—	**-7.63***** **(1.09)**	**-6.75***** **(0.87)**
	Independence	—	**-13.46 (0.90)**	**-8.95***** **(1.31)**
	Separation	—	**-16.84***** **(1.08)**	**-11.38***** **(1.52)**
	Mandate	**14.01***** **(0.85)**	—	—
	Extreme	**-20.27***** **(1.67)**	—	—
	Size	***n.s.***	***n.s.***	***n.s.***
	Non-disclosers	***n.s.***	***n.s.***	**0.21*** **(0.0941)**
Variance component				
Level-1	*R*	**20.04**	**24.99**	**19.93**
	(%)	**66**	**49**	**61**
Level-2	*Intercept*	**10.10*****	**25.68*****	**12.53*****
	(%)	**34**	**51**	**39**
Deviation		**371.89**	**1762.09**	**1700.45**
	Parameters	**5**	**7**	**8**
	DL	**21**	**58**	**121**

* P< 0.05
** P<0.01
*** P<0.001
n.s. not significant. The variable was tested in one previous model and removed from the model.

Compared to attitudes re: Sovereignty-association.

Voter intent: -3 pts to -4 pts

Sovereignty: -7 pts to -8 pts

Independence: -9 pts to -14 pts

Separation: -12 pts to -17 pts

Mandate: + 14 pts

Extreme: -20 pts

+ item NR →+ support

49%-66% of variance btw polls, the rest, between time units

At level 2

Effects related to time and events (Yale & Durand, 2011)

Table 2 – Final Models of Change for 1989–1995 and 1995–2008.

	Sovereignty-partnership	Sovereignty	Independence	Separation
1989–1995				
intercept	45.07***	45.07	29.55***	37.74
month	1.79***	1.79	2.18*	0.73***
meech1	-2.30***	-2.51***	-2.79**	-1.10***
charlot1	0.48***	0.48	0.48	0.48
1995–2008				
intercept	55.30***	50.92***	47.10***	55.30
month	-0.31***	-0.31	-0.31	-0.87***
month²	0.002***	0.002	0.002	0.007***
sponsorship	4.76**	4.76	4.76	-9.46***
gomery1	-0.55***	-0.55	-0.55	-0.55

* P< 0.05
** P<0.01
*** P<0.001

Support for various options...

1989-1995

↑ with time

↓ after Meech failure

↑ after Charlottetown failure

1995-2008

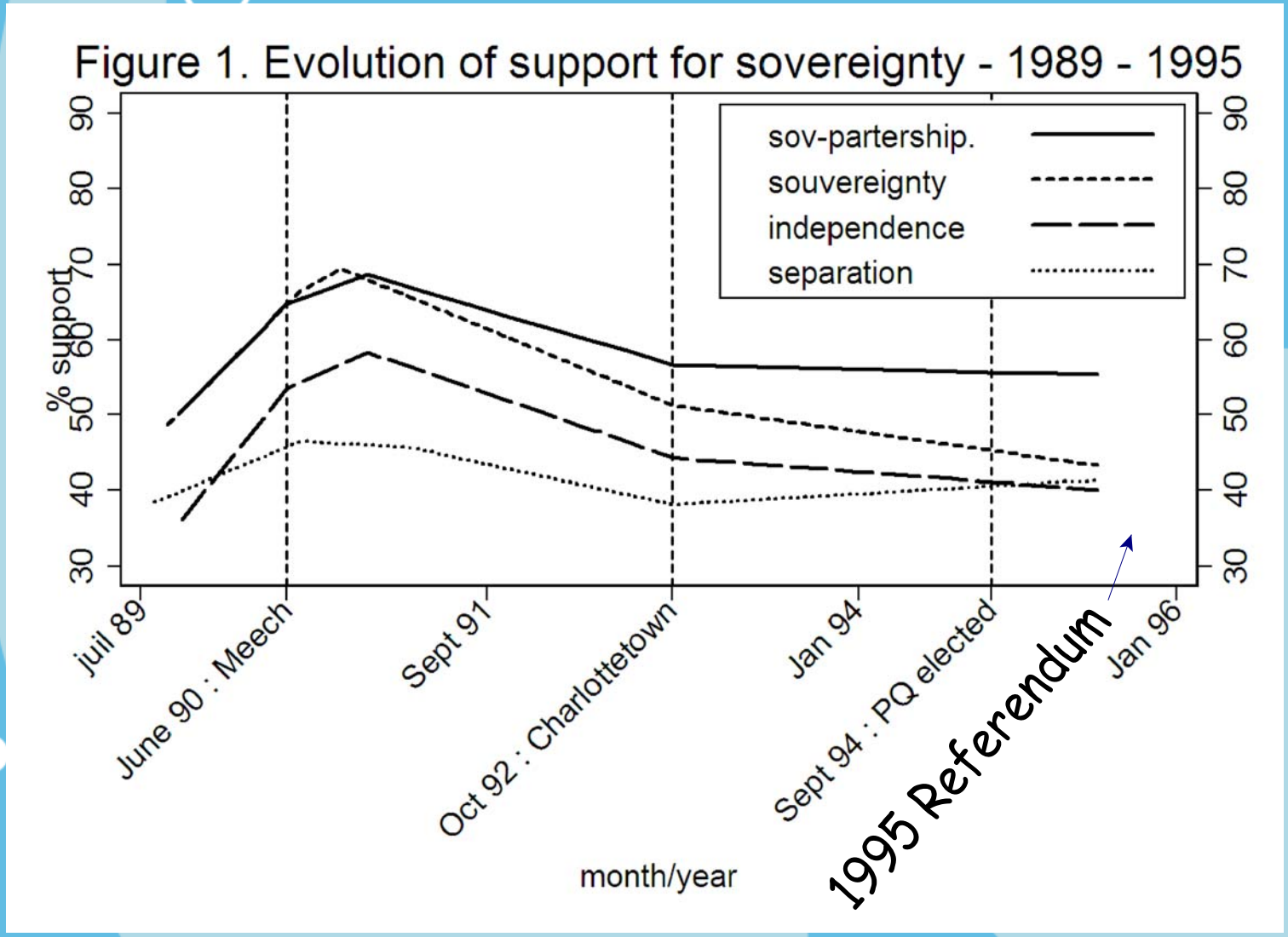
U shaped with time

↑ after spons. Scandal except for separation

↓ after Gomery report

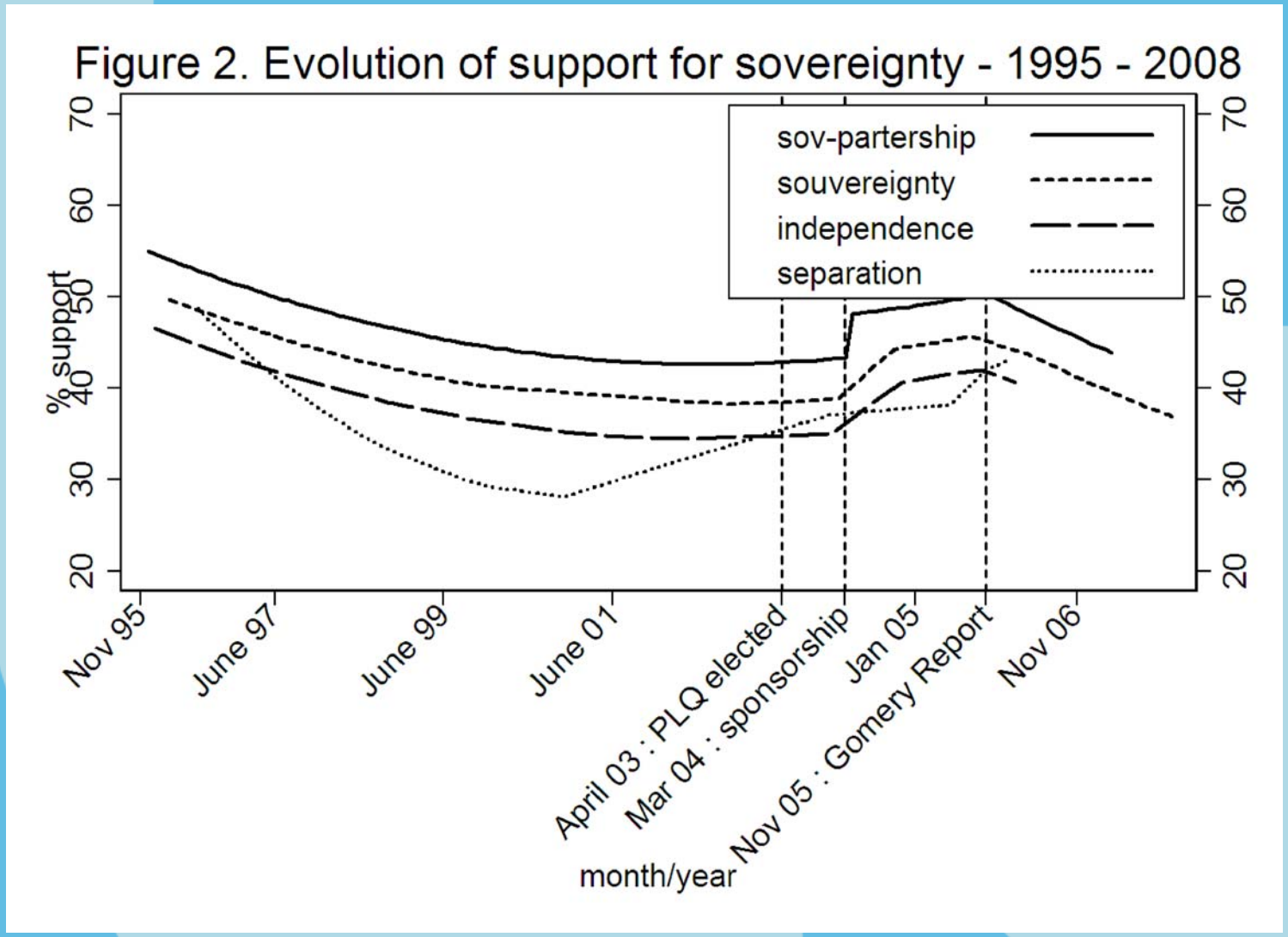
Combined model 1989-1995

Evolution according to constitutional option -voter intent (Yale & Durand, 2011)



Combined model 1995-2008

Evolution according to constitutional option -voter intent (Yale & Durand, 2011)



Example 2: Change in voting intentions for Obama and Romney, U.S. 2012

- The question:
  - What is the likely change in voting intentions for the 2012 US presidential election?
  - Is this change the same whatever the survey mode of administration?
  - What is the impact of using a likely voter model?

Example 2: Change in voting intentions for Obama and Romney, U.S. 2012

- At the poll level (Level 1):
  - Dependent variable:
    - Estimate of voting intention for Obama or Romney
  - Independent variables:
    - Mode of administration (not significant)
    - Number of days poll is in the field
    - Sample size
    - Proportion of non-disclosers
    - Use of a likely voter model
- At the time level (Level 2): Defining time: week (vs day)
  - Time, time squared, time cubic, power 4, power 5

Equations…Final model

LEVEL 1 MODEL (bold: group-mean centering; bold italic: grand-mean centering)

$$OBAMA = \beta_0 + \beta_1(UNDEC2) + \beta_2(NBJOURS) + \beta_3(LIKELY\_V) + \beta_4(SAMPLESQ) + r$$

LEVEL 2 MODEL (bold italic: grand-mean centering)

$$\beta_0 = \gamma_{00} + \gamma_{01}(TEMPS) + \gamma_{02}(TEMPS2) + \gamma_{03}(TEMPS3) + \gamma_{04}(TEMPS4) + u_0$$

$$\beta_1 = \gamma_{10}$$

$$\beta_2 = \gamma_{20}$$

$$\beta_3 = \gamma_{30} + \gamma_{31}(TEMPS) + \gamma_{32}(TEMPS2)$$

$$\beta_4 = \gamma_{40}$$

Only the intercept is allowed to vary per week.

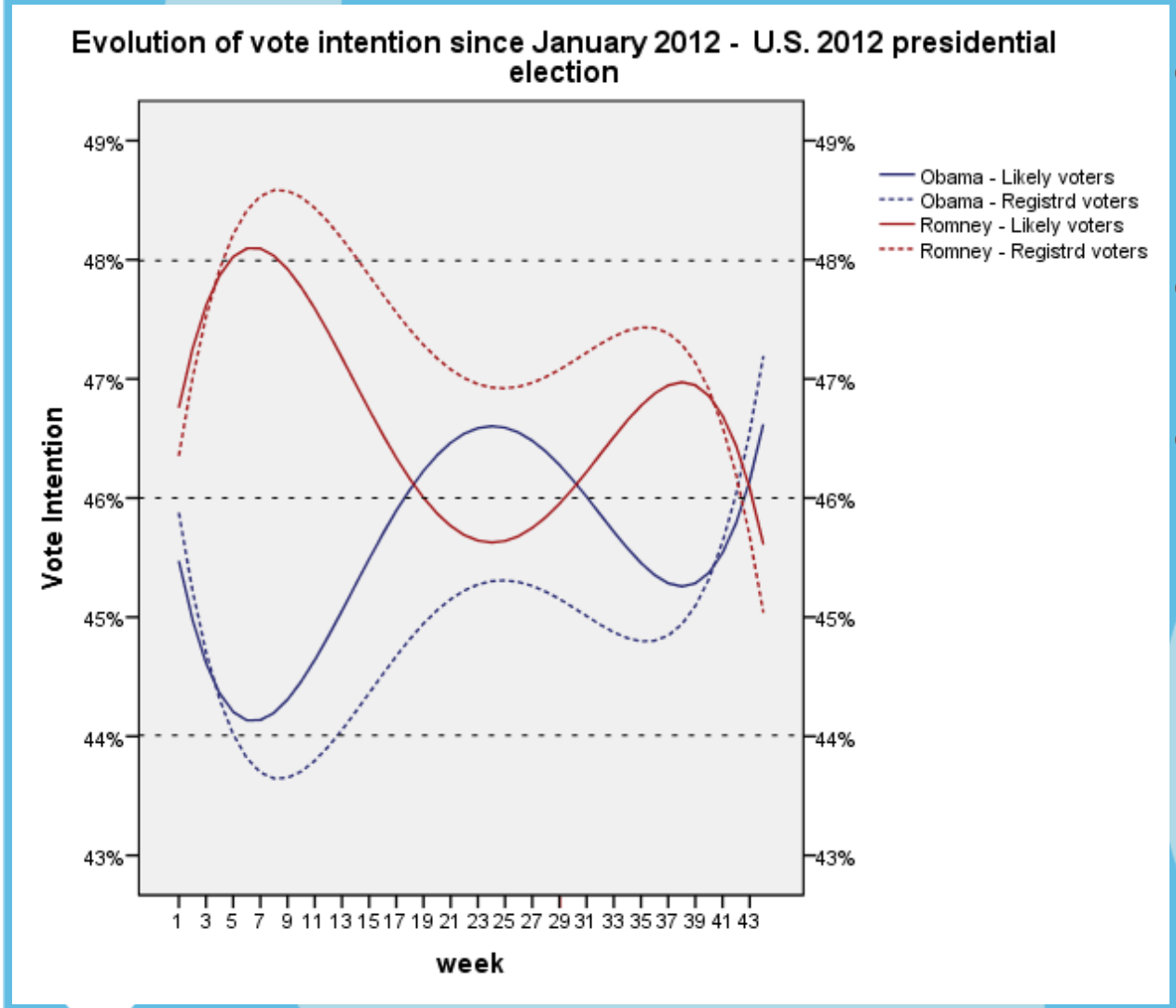
- At level 1: support for Obama is influenced by
  - The proportion of undecideds in the poll, the number of days the poll was in the field, the use of a Likely voter model and the sample size ($1/\sqrt{n}$)
- At level 2,
  - The intercept is influenced by time (linear, quadratic, cubic and power 4).
  - The influence of the likely voter model varies with time linear and quadratic.

Voting intentions for Obama

Fixed effects		Coefficient	Std error	T-ratio	d.f.	P-Value
Intercept1: B0						
Intercept2	G00	50.9356	0.7750	65.7220	39	0.000
Time	G01	0.0615	0.0214	2.8690	39	0.007
Time2	G02	-0.0110	0.0025	-4.4370	39	0.000
Time3	G03	-0.0001	0.0001	-1.4560	39	0.153
Time4	G04	0.0000	0.0000	5.2930	39	0.000
Prop. Undecided: B1						
Intercept2	G10	-0.5028	0.0280	-17.9280	381	0.000
Nb Days in field: B2						
Intercept2	G20	-0.1516	0.0390	-3.8900	381	0.000
Lilely Voter Model: B3						
Intercept2	G30	1.3170	0.4382	3.0060	381	0.003
Time	G31	0.0001	0.0148	0.0060	381	0.995
Time2	G32	-0.0039	0.0014	-2.8660	381	0.005
Sample Size: B4						
Intercept2	G40	-89.6405	14.5841	-6.1460	381	0.000

Note: Events could have been added but there was no cue that some important events had influenced voting intentions substantially.

Change in voting intention from January 2012 to Election Day



- Likely Voter Model: 59% of the polls.
- Registered voters or adults: 41%
- All the other variables have been put at the mean -- number of days (4.22), sample size (1268), proportion of non disclosers (7.77).

Final results: variance explained

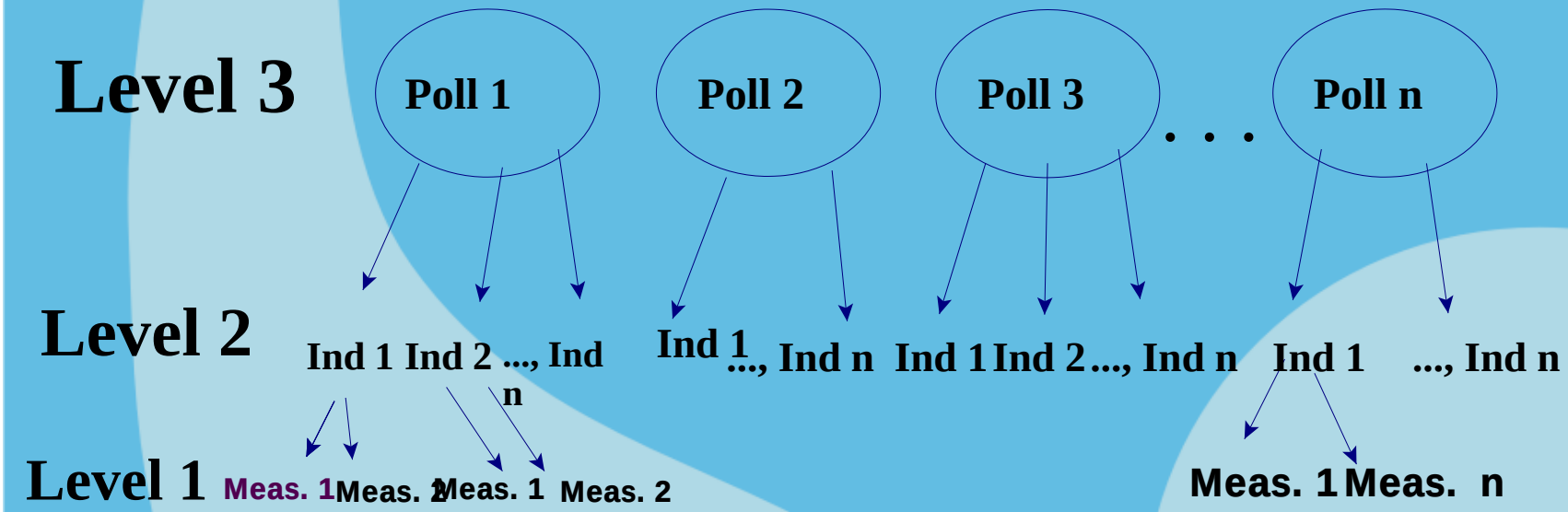
Prediction of voting intention for Obama

	Model 0	Model Niv1	Full model
Var. Niv. 2: weeks	.52	.52	.19
Var. Niv. 1: polls	4.19	1.82	1.76
Prop. var btw weeks	11.0%	22.2%	9.7%
Prop var. explained btw polls	-	56.6%	58.0%
Prop. var. explained btw we>		-	63.5%

- At the beginning, 11% of the variance is between weeks, 89% between polls.
- Variables at level one -- number of days in the field, sample size, proport. of undecideds and use of a likely voter model -- explain 57% of the variance between polls.
- Evolution with time -- including the effect of the varying impact of the likely voter model -- explains 63.5% of the variance over time.

Example 3: Combining data files i.e., individual records

- At Level 3: poll characteristics (incl. Question wording) and change over time and its predictors



- At level 2: Individuals and their characteristics (age, sex, education, region)
- At level 1: Trust and its objects and characteristics (police, religion, unions,...)

Example 3: Evolution of trust towards institutions in Canada

- 59 surveys with questions pertaining to trust in institutions from 1974 to 2012. Final model: 56.
- n=161,261 respondents. Final model: 134,802 with 606,540 measures.
- Measures vary according to:
  - The object of trust: religion, schools, unions, media, etc.
  - Whether the object is the institution itself or the people, i.e. religion or preasts, schools or teachers, unions or union leaders, media or journalists, etc.
  - The wording and the number of response categories (at the survey level).
- Each data base has to be put on a common basis, including socio-demographics: looking for the smallest common denominator.

How to proceed

- Data from all the surveys are combined into one file for each level.
- **At level 1, the measurement level:**
  - In each file, variable names changed to common names:
    - AnswerTrustReligion, objectTrustReligion (institutions or people),...
    - AnswerTrustSchools, objectTrustSchools (institutions or people),...
  - Then, restructure the file so that there are as many lines per respondent as the number of Trust questions asked.
    - Ident1, Line1: index (religion), AnswerTrust, objectTrust, GreatTrust,
    - Ident1, Line2: index (Schools), AnswerTrust, objectTrust, GreatTrust,
    - Ident1, Line3: index (Media), AnswerTrust, objectTrust, GreatTrust, etc.
- **At level 2:** For each respondent, age, sex, region, education, income, language, etc.
- **At level 3:** For each file, time of survey (month/year), wording of trust questions,...

Equations: Basic 3-level model

- At level 1: Trust at the question level (ref: religion)
  - GreatTrust= ψ_0+ ψ_1(Army) +ψ_2(Finance) +… + ψ_n(ObjectN) + ε
- At level 2: Trust at the individual level
  - ψ_0= π_{00}+ π_{01}(Man) +π_{02} (Young) + π_{03}(Old) +e_0
  - ψ_1= π_{10}
  - ψ_2= π_{20},…
  - ψ_n= π_{n0}
- At level 3: Trust at the survey level
  - π_{00}= β_{000}+β_{001}(Time) +β_{002}(Wording) + β_{000}(NbChoices)+r_{00}
  - π_{01}= β_{010}
  - π_{02}= β_{020}
  - π_{03}= β_{030}
  - π_{10}= β_{100}
  - π_{20}= β_{200},…

**Trust may evolve differently with time according to the object of confidence.*

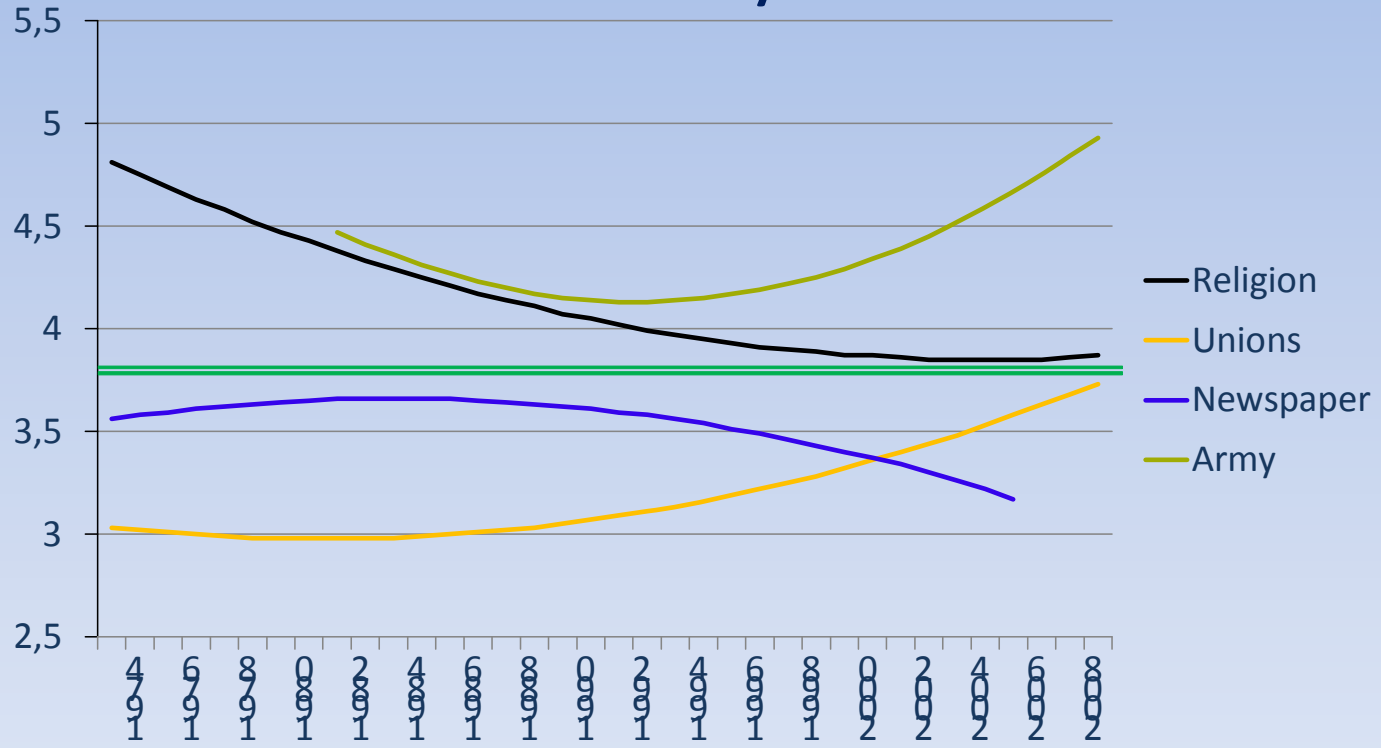
Results: Trust in Canada 1974-2012

Trust in institutions in Canada 1974-2012									
	Model 0		**model 1**		**Model 2**		**Model 3**		
Intercept	0.3788		0.1818		0.1775		0.3196		
Measure level ref: Religion)									
Army			0.4941		0.4942		0.4943		
Financial inst.			0.4465		0.4464		0.4464		
Private. Enterpr.			-0.0371	ns	-0.0372	ns	-0.0371	ns	
Justice			0.3714		0.3715		0.3716		
Media			0.2489		0.2489		0.2489		
Police			0.6078		0.6077		0.6077		
School			-0.0310	ns	-0.0307	ns	-0.0307	ns	
Union			-0.1268		-0.1267		-0.1267		
respondent level									
Man					-0.0136		-0.0136		
Young (18-29)					0.0150		0.0150		
Old (60+)					0.0369		0.0369		
Survey level									
Wording A							0.0535		
Nb choices							-0.0383	ns	
Time							-0.0001	ns	
Variance									exp. var
Measurem. leve	0.2268	58.1%	0.1538	84.5%	0.1533	84.1%	0.1538	85.2%	32.2%
Respond. Level	0.0073	1.9%	0.0210	11.6%	0.0220	12.1%	0.0208	11.5%	4.0%
Survey level	0.1566	40.1%	0.0071	3.9%	0.0069	3.8%	0.0060	3.3%	96.2%

Hypothesized evolution of trust in some institutions

Valois, Durand & Goyder, 2012

General trends in confidence: Religion, unions, the media and the army



Limits

- When combining results or data, we need to have enough information at all levels, for example,
  - Variation between question wording (example1) has to be spread on all time periods.
    - We had to perform analyses separately for the different periods in order to take this into account.
  - Use of likely voter model (example 2) has to be spread also throughout the period.
- When combining data files,
  - We cannot take into account all the subtilities of question wording
  - It may be difficult to find a common denominator for response categories.
    - Use mean, put on a 7 point scale, use proportion of high trust or of low trust as dependent variables, etc.
  - It may be very difficult to find a common denominator for variables like age, income, etc.

Conclusion

- There are incredible possibilities to combine data in order to get to the "big picture".
- Multilevel longitudinal analysis with time at the higher level allows for a thorough use of the data already collected in order to better understand different phenomena and their evolution over time and in different contexts.
- Next steps:
  - Refine the model.
  - Combine data files on other topics, for other countries.