Crowd-Sourced Text Analysis: Reproducible and Agile Production of Political Data

Kenneth Benoit (LSE)
With: Drew Conway (NYU), Ben Lauderdale (LSE), Michael Laver (NYU), and Slava Mikhaylov (UCL)

American Political Science Review Vol. 110, No. 2 May 2016
doi:10.1017/S0003055416000058

Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data

KENNETH BENOIT *London School of Economics and Trinity College*
DREW CONWAY *New York University*
BENJAMIN E. LAUDERDALE *London School of Economics*
MICHAEL LAVER *New York University*
SLAVA MIKHAYLOV *University College London*

Empirical social science often relies on data that are not observed in the field, but are transformed into quantitative variables by expert researchers who analyze and interpret qualitative raw sources. While generally considered the most valid way to produce data, this expert-driven process is inherently difficult to replicate or to assess on grounds of reliability. Using crowd-sourcing to distribute text for reading and interpretation by massive numbers of nonexperts, we generate results comparable to those using experts to read and interpret the same texts, but do so far more quickly and flexibly. Crucially, the data we collect can be reproduced and extended transparently, making crowd-sourced datasets intrinsically reproducible. This focuses researchers' attention on the fundamental scientific objective of specifying reliable and replicable methods for collecting the data needed, rather than on the content of any particular dataset. We also show that our approach works straightforwardly with different types of political text, written in different languages. While findings reported here concern text analysis, they have far-reaching implications for expert-generated data in the social sciences.

what does it mean to be

agile?

reproducible?

crowd-sourcing for text analysis

proof-of-concept application(s)

agile data production

- interactive between a researcher and his or her needs
- flexible (rather than fixed)
- responsive
- rapid

most secondary data today is not agile

- general-purpose
- attempts to be comprehensive
- resource-intensive
- generates “lock-in”
- produced by a few (trained) experts:
  Example - expert-coded policy manifestos in political science

alternative: supplement automated approaches with human decision-making

- complex data generation projects are broken down into simple tasks, very targeted
- humans instead of algorithms ensures the validity of natural language processing in a wide range of applications
- retain the reproducibility of automated methods

our solution: draw on the crowd

- crowd-sourcing: outsourcing a task by distributing it in simplified parts to a large, unspecified group
- contrasts with expert coding:
  - less expertise, but in far greater numbers
  - no one sees a whole text: text analysis tasks are served partially and randomly
  - multiple coders per sentence, different coders treated as exchangeable
- we use a scaling model to aggregate judgments into quantities of interest
- very easy to reproduce

our problem: classifying text units

- the idea: to observe a political party's policy position by content analysis of its texts

- standard testing domain: party manifestos

- idea is to break them into sentences and use human judgment to apply pre-defined codes

- experts: all sentences, by one coder
  crowd: lots of coders, only some sentences

example: labelling immigration sentences

Coding immigration policy statements in political text (20140118)

Instructions ▲

Summary

This task involves reading sentences from political texts from the 2010 UK general election, and judging whether these statements deal with immigration policy. Each sentence may or may not be related to immigration policy. We tell you below what we mean by "immigration policy".</p>

First, you will read a short section from a party manifesto. For the sentence highlighted in red, enter your best judgment about whether it refers to some aspect of immigration policy, or not. **Most sentences will not relate to immigration policy** -- it is your job to find and rate those that do. **If the sentence does not refer to immigration policy**, you should select "Not immigration policy" and proceed directly to the next sentence. **If the sentence does refer to immigration policy**, you should indicate this by checking this option.

It's time for the truth. **A lot of lies have been told about the EU.** We're frequently told that we'll lose 3 million jobs if we leave _ a shameless lie.

Immigration Policy

Not immigration policy

findings

- non-experts produce valid results, just need a more of them
- experts are just another form of crowd: experts also have a variance
- crowd-sourced data production is reliable - and offers reproducibility
- flexibility, bespoke, low-cost
- can work for any data production job that is easily distributed into simple tasks

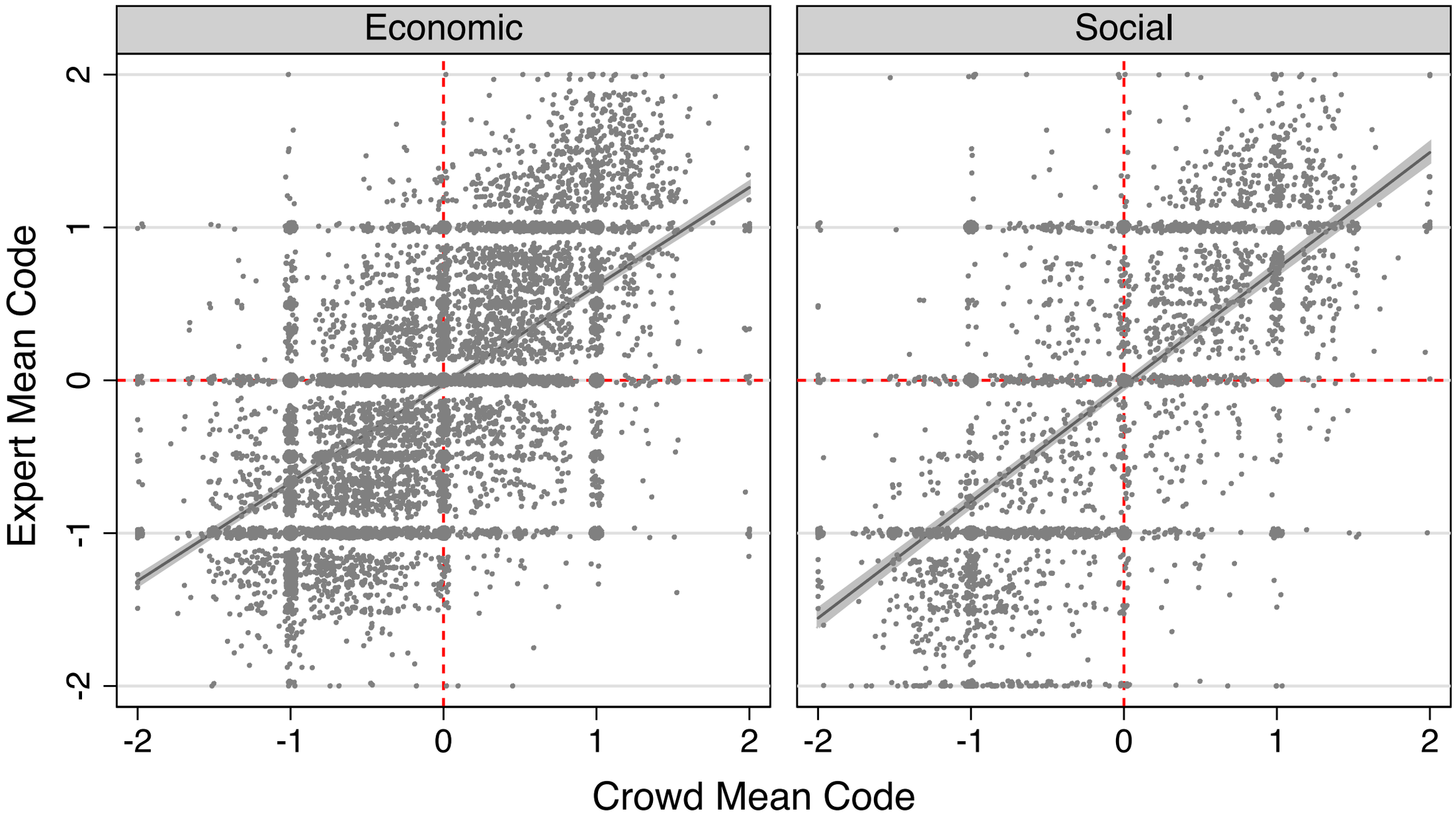
we used an IRT-type scaling model to estimate position

- also allows for coder effects and sentence difficulty effects
- can estimate uncertainty through posterior simulation (MCMC)

deployment on Crowdflower

- http://crowdflower.com
- a front-end to many crowd-sourcing platforms, not just Mechanical Turk
- uses a quality monitoring system so that you have to maintain an 80% “trust” score or be rejected
- trust maintained through “gold” questions carefully selected and agreed by experts

comparing experts to crowd-coders



overall correlations:

0.96 for economic, and 0.92 for social dimension

correlation of aggregate measures

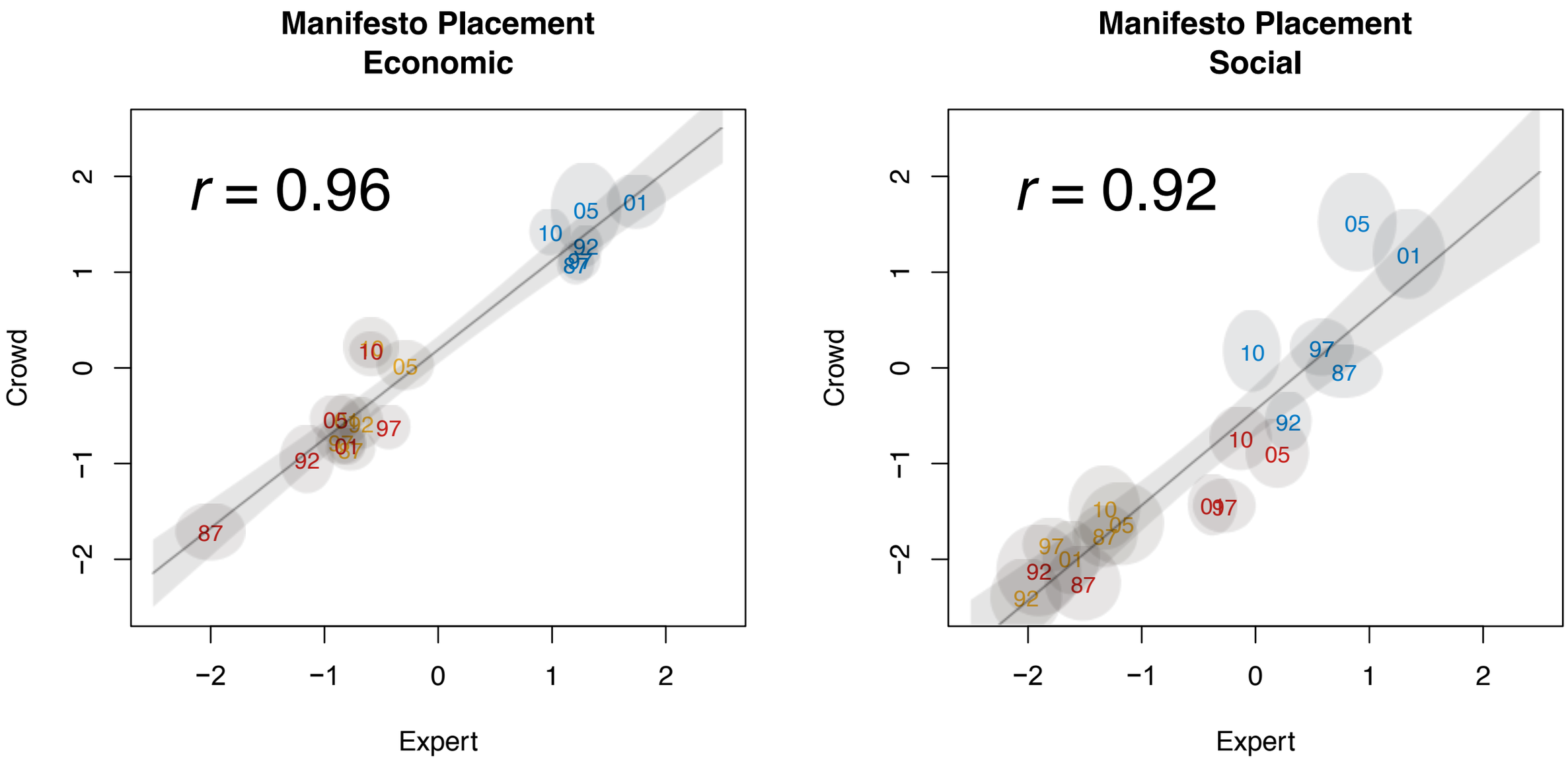
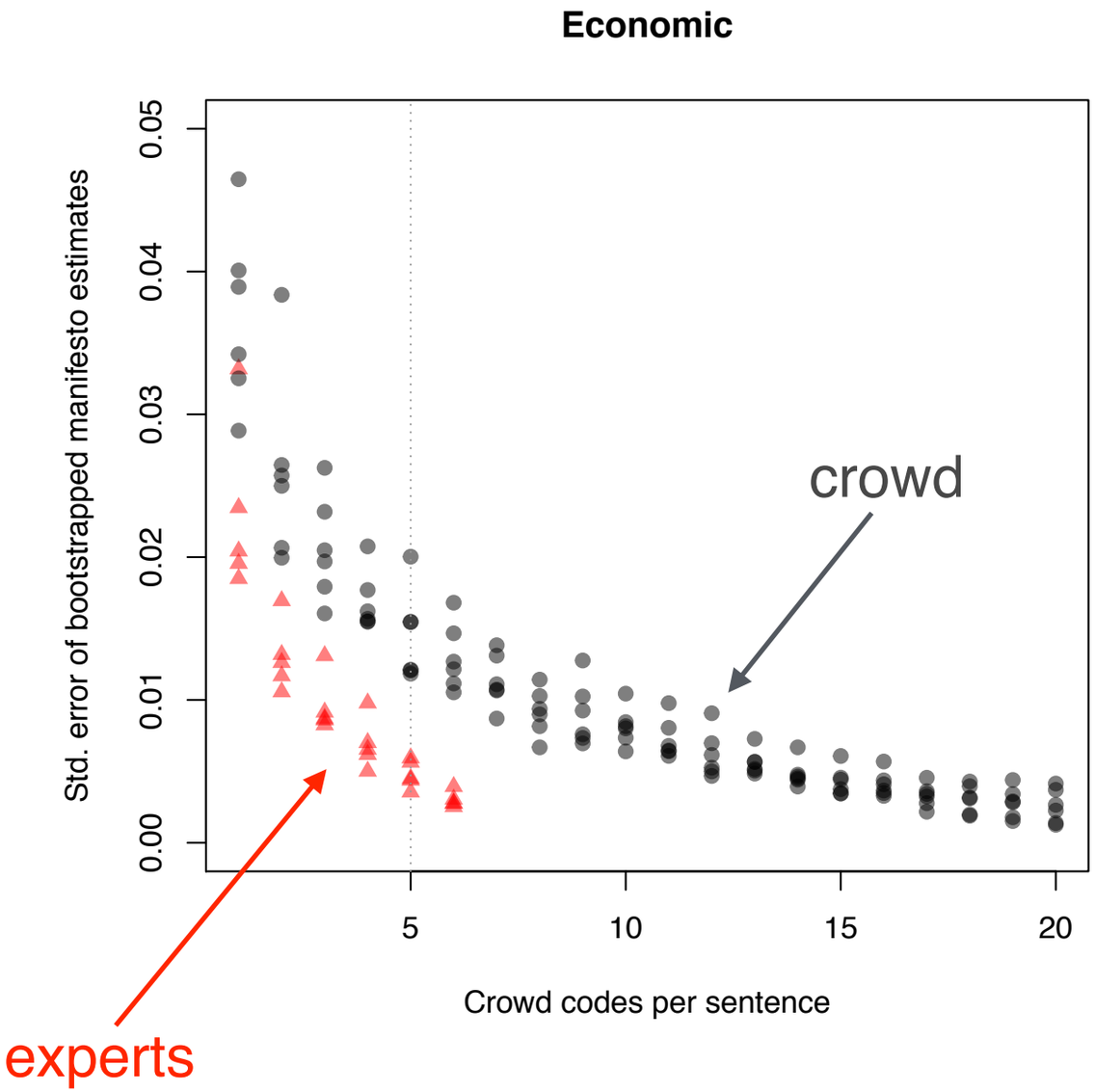
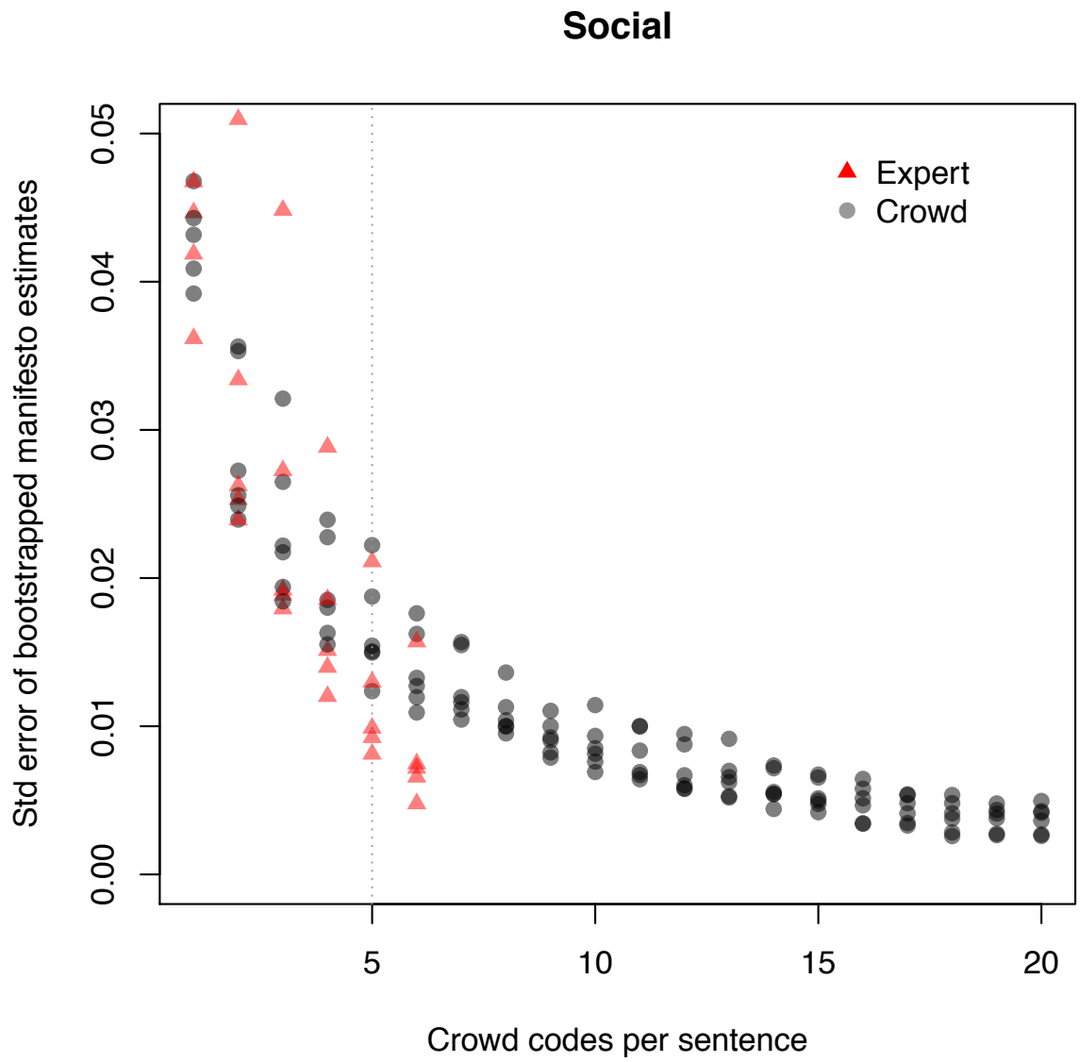


Figure 3. Expert and crowd-sourced estimates of economic and social policy positions.

how many judgments are needed?





we chose five per text unit

three other contexts, other languages

- new corpus: EP debate over state aid to uncompetitive coal mines
- 36 speakers, 11 countries, 10 original languages (only one English speech) translated into 22 languages
- we know the vote on the measure: can text-based measure predict it?

Procedure : **2010/0220(NLE)**

Document stages in plenary

Document selected : A7-0324/2010

Texts tabled :	Debates :	Votes :	Texts adopted :
A7-0324/2010	PV 23/11/2010 - 5 CRE 23/11/2010 - 5	PV 23/11/2010 - 6.22 Explanations of votes Explanations of votes Explanations of votes	P7_TA(2010)0424

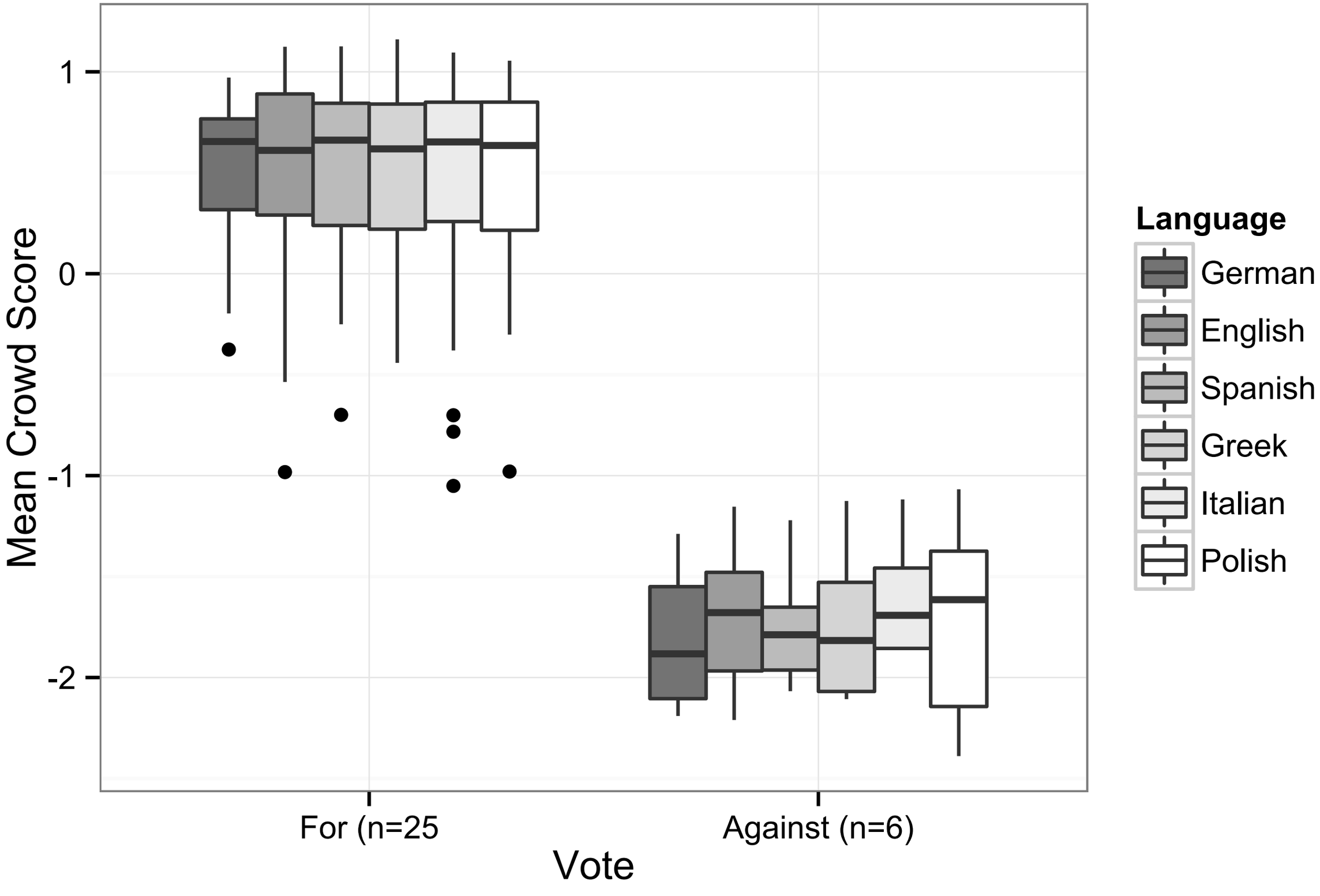
Debates

Tuesday, 23 November 2010 - Strasbourg — OJ edition

5. State aid to facilitate the closure of uncompetitive coal mines (debate)

Video of the speeches | PV

President. – The next item is the report by Mr Bernhard Rapkay, on behalf of the Committee on Economic and Monetary Affairs, on State aid to facilitate the closure of uncompetitive coal mines (COM(2010)0372 – C7-0296/2010 – 2010 /0220(NLE)) (A7-0324/2010).


Mean Crowd Score
1
0
-1
-2
For (n=25
Against (n=6)
Vote
Language
German
English
Spanish
Greek
Italian
Polish

Αξιολόγηση δηλώσεων από μια κοινοβουλευτική συζήτηση

Instructions ▲

Περίληψη

Το έργο αυτό ζητά έπειτα από την ανάγνωση δηλώσεων από μια συζήτηση πολιτικών στο Ευρωπαϊκό κοινοβούλιο να κριθεί κατά πόσο συγκεκριμένες δηλώσεις ήταν υπέρ ή κατά μιας προτεινόμενης πολιτικής.

Η εκτίμηση επιπτώσεων που παρουσιάσατε είναι πολύ καλή. Βασίζεται σε γεγονότα, είναι πειστική και εστιάζει στο προκείμενο. **Είναι κρίμα το γεγονός ότι τα άλλη μέλη του Σώματος των Επιτρόπων δεν διάβασαν αυτήν την εκτίμηση επιπτώσεων, διότι, αν το είχαν κάνει, δεν θα μπορούσαν να υποβάλουν την πρόταση αυτή, η οποία δεν έχει απολύτως καμία σχέση με την εκτίμηση επιπτώσεων.** Αναρωτιέμαι πώς το Σώμα των Επιτρόπων επέλεξε τη χρονιά, για παράδειγμα. Στην εκτίμηση επιπτώσεων δεν γίνεται απολύτως καμία αναφορά σε αυτό.

Όσον αφορά το ζήτημα της συνέχισης των επιδοτήσεων, αυτή η δήλωση είναι:

Κατά των επιδοτήσεων	Ουδέτερη ή άσχετη	Υπέρ των επιδότησεων
○	○	○