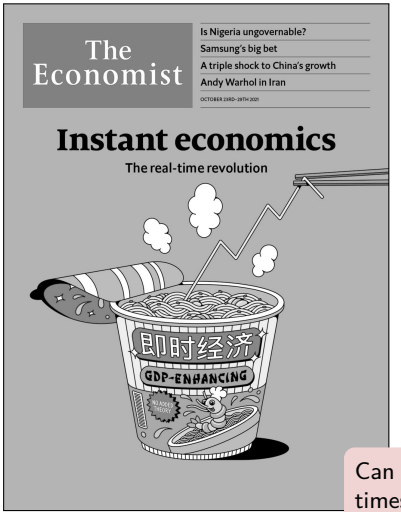
Combining small and big data: An integrated data framework for policy guidance in times of dynamic economic shocks

joint with Jan Kinne, David Lenz, Georg Licht, Peter Winker

submitted to PLoS ONE

Motivation - last week's Economist



Source: The Economist (2021a)

Does anyone really understand what is going on in the world economy? The pandemic has made plenty of observers look clueless.

Especially in times of rapid change, policymakers have operated in a fog.

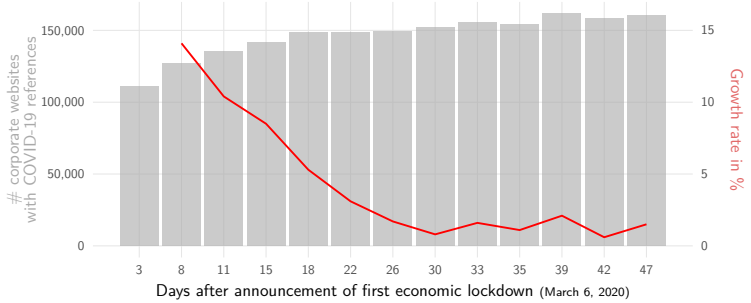
The gap between official data and what is happening in the real economy can still be glaring.

The Economist (2021a, 2021b)

Can we assist policy makers with **timely** and **insightful** firm-level data in times of dynamic economic shocks such as COVID-19?

Early firm communication and corporate websites

- accessed corporate websites of ~ 1.18M German companies from March 2020 - May 2020 twice a week searching for references related to the pandemic
- finding: companies used their websites intensively to report about the pandemic



Turn website references into knowledge

But: context of Corona references greatly differed across firms:

'The Corona pandemic is not only affecting ongoing [redacted] projects, but also the current selection rounds of the 13th and 14th funding seasons.'

* * *

'We have therefore decided to adapt our services to the current situation and to limit them until further notice. Although we want to continue to provide you with all indispensable services, we also want to meet the recommendations of the federal government on how to deal with the corona virus.'

* * *

'Your [redacted] advisor stands by your side - also in times of COVID-19.'

Turn website references into knowledge

But: context of Corona references greatly differed across firms:

'The Corona pandemic is not only affecting ongoing [redacted] projects, but also the current selection rounds of the 13th and 14th funding seasons.'

* * *

'We have therefore decided to adapt our services to the current situation and to limit them until further notice. Although we want to continue to provide you with all indispensable services, we also want to meet the recommendations of the federal government on how to deal with the corona virus.'

* * *

'Your [redacted] advisor stands by your side - also in times of COVID-19.'

Statistical learning approach:

1. introduced 5 meaningful & distinguishable classes

   (1) problem, (2) no problem, (3) adaption, (4) information, (5) unclear

Turn website references into knowledge

But: context of Corona references greatly differed across firms:

'The Corona pandemic is not only affecting ongoing [redacted] projects, but also the current selection rounds of the 13th and 14th funding seasons.'

* * *

'We have therefore decided to adapt our services to the current situation and to limit them until further notice. Although we want to continue to provide you with all indispensable services, we also want to meet the recommendations of the federal government on how to deal with the corona virus.'

* * *

'Your [redacted] advisor stands by your side - also in times of COVID-19.'

Statistical learning approach:

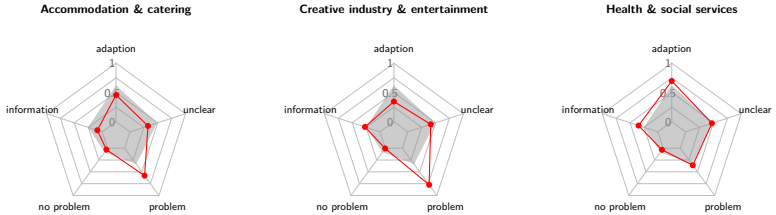
1. introduced 5 meaningful & distinguishable classes

   (1) problem, (2) no problem, (3) adaption, (4) information, (5) unclear
2. manually annotated ~ 4,000 references
3. fine-tuned multilingual language model (XLM-R by Conneau et al. (2019))
   - large neural network pre-trained on 2TB CommonCrawl data
   - model with general and multilingual language understanding
   - fine-tuning allows to adapt model to various NLP tasks (here: text classification)
   - classification performance: 89.5% accuracy

Insights from website analysis

- classified firm communication on websites revealed impact heterogeneity at sector level
- insights generated at near-real time (right after shutdown announcement in Mar 20)

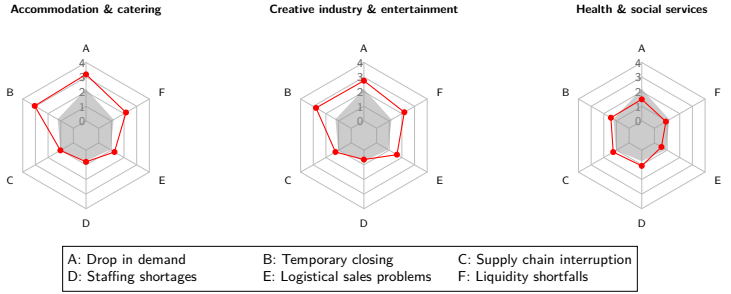
Figure: Early detection of impact heterogeneity at sector level



Follow up surveys

- based on the early findings construct targeted businesses surveys
- gain more detailed understanding about the sort of impact in order to design counter measures most effectively
- here: surveyed ~ 1,500 companies consecutively (Apr, Jun, Sep 2020) with targeted impact questions

Figure: Targeted impact questions at sector level



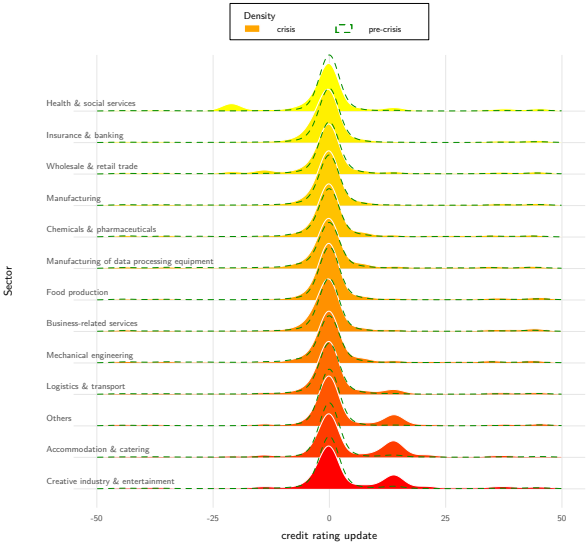
Retrospective analysis of firm outcomes

- after economic shock has materialized in economy, analyze firm outcomes
- understand possible long-term consequences and design stimulus programs
- here: examined credit rating updates between Jun 20 - Apr 21 for ~ 870,000 companies

Retrospective analysis of firm outcomes

- after economic shock has materialized in economy, analyze firm outcomes
- understand possible long-term consequences and design stimulus programs
- here: examined credit rating updates between Jun 20 - Apr 21 for ~ 870,000 companies

Figure: Credit rating movements at sector level



Classified website references as leading indicators

$$\Delta r_{i,\bar{t}+z} = \alpha + \beta_1 \text{Problem}_{i,\bar{t}} + \beta_2 \text{No problem}_{i,\bar{t}} + \beta_3 \text{Adaption}_{i,\bar{t}} + \beta_4 \text{Information}_{i,\bar{t}} + \beta_5 \text{Unclear}_{i,\bar{t}} + \gamma r_{i,\bar{t}-x} + \boldsymbol{\delta FE}_i + \epsilon_i$$

	(1) $\Delta r_{\bar{t}+z}$	(2) $\Delta r_{\bar{t}+z}$	(3) $\Delta r_{\bar{t}+z}$	(4) $\Delta r_{\bar{t}+z}$
$\text{Problem}_{\bar{t}}$	+1.66***	+1.68***	+1.62***	+0.42**
$\text{No problem}_{\bar{t}}$	-1.70***	-1.69***	-1.73***	-0.69
$\text{Adaption}_{\bar{t}}$	-0.46***	-0.47***	-0.33***	-0.13
$\text{Information}_{\bar{t}}$	-0.24***	-0.24***	-0.23***	-0.17***
$\text{Unclear}_{\bar{t}}$	-0.42***	-0.42***	-0.10	-0.08
$r_{\bar{t}-x}$	-0.09***	-0.10***	-0.11***	-0.13***
Age FE	No	Yes	Yes	Yes
Size FE	No	No	Yes	Yes
Sector FE	No	No	No	Yes
N	61,228	61,138	57,343	57,343

Δr_i: credit rating update (+ **down**grade, - **up**grade) of firm i
$\bar{t}$: 01/03/20 - 31/05/20, $\bar{t}+z$: z days after 01/06/20, $\bar{t}-x$: x days before 01/03/20
FE: fixed effects. Significance levels: *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$

Main contributions

- proposed a data framework for policy guidance in times of economic shocks
- to overcome information deficits policy makers are confronted with in highly dynamic situations
- possibly allowing more targeted liquidity injections to support affected companies instead of choosing the 'bazooka' as policy instrument