# 27/05/2020

Aggregating on country and phase (short)

1. Code Phase=0 if country is in restriction (1 if in relaxation).
   1. Restriction: when index is going up or staying constant
   2. Relaxation: from when it’s reduced onwards until it goes up again. Only Turkey I think was one such instance but I think it’s an anomaly
   3. Challenge: Every country might have as many restriction and relaxation periods as Oxford categories
      1. Solution 1: if we are to make our entire analysis only on one aspect of policy (combined with one dimension of Google Movement), then no worries.
      2. Solution 2: maybe we define lockdown when certain conditions are met and we register the date when the last of these conditions was met and relaxation when any of the 3 conditions was lifted.
2. Elastic net regressions to identify the right behavioural variable to be used. Risktaking very promising.
   1. Note that we might need to run elastic nets for 15 dependent variables (8-9 Oxford policies and 6-7 Google sub-categories).
   2. Include household residents
   3. Include good weather variable
3. The duration of each phase might be telling as well.
   1. Might correlate itself with behavioural measures (e.g. patience).

Aggregating on Country (long)

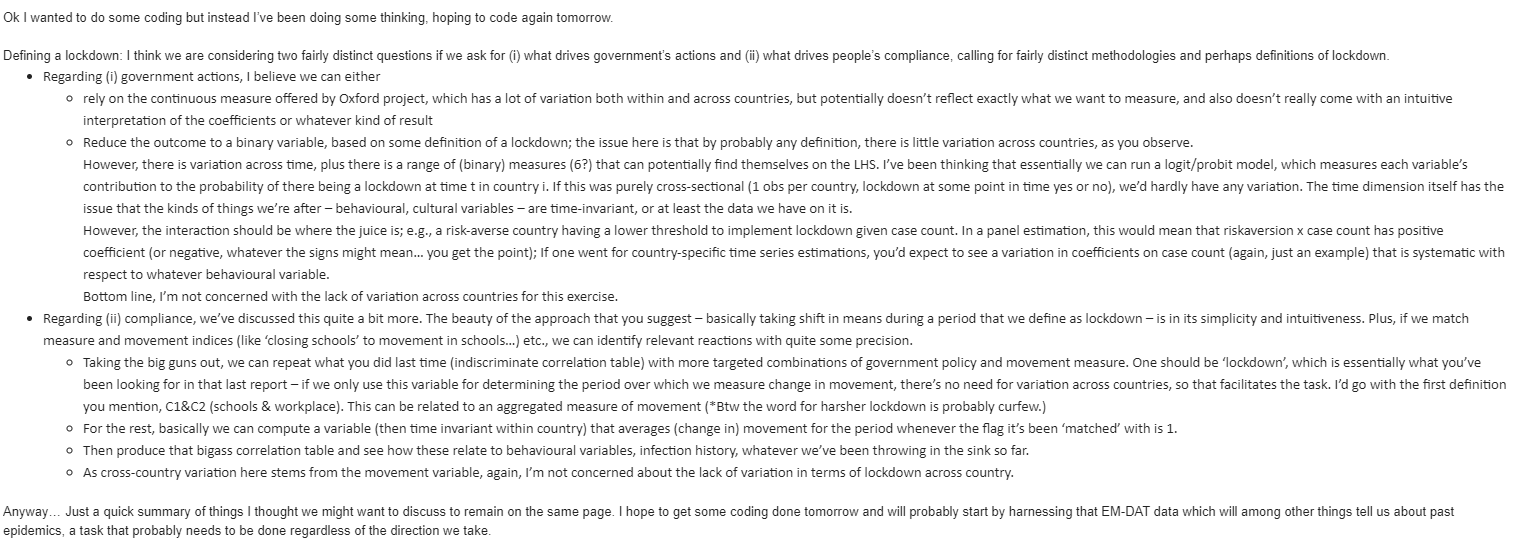
1. After we identify behavioural measures, sub-policies and sub-movements we can consider fancier measures (like ARDA) to calculate adherence.
2. The duration

# 28/05/2020

* The more I think about it, the more I come to realise that finding a universal criterion upon which we can binarize lockdown is really hard if not impossible.
* Perhaps a way to go about it is to go on a very micro level. Here is a strategy for looking at more disaggregated data and try to reverse engineer the external shocks we should be looking for (if this is possible). Here are the steps:
  + Let’s pick a country and quantify the daily differences in movement by taking first differences.
  + Then, let’s normalise these differences and test the days when the drop in movement was more than 2 standard deviations from the mean drop over a certain period.
  + Is there something special about these days? Was there a policy implementation? If so, what sort of policy was it?
    - Disaggregate further: do the same exercise for this country for a every Google variable.
      * Disaggregate even further: do the same exercise for every city of this country. For the UK we have daily data for 153 cities!

# 12/06/2020

## Country level

* Redo the “Kitchen and Sink” correlation table (KSC-table) but this time, correctly.
  + In fact, we should probably do three of those tables.
    - One where we focus on a country’s decision for **policies**. Factors:
      * Foreign influence
        + Recent paper: <https://www.aeaweb.org/articles?id=10.1257/jel.20201481&from=f>
      * State of the economy (cyclicality included). For example, Greece went hard during the early phase and plans to reopen everything during the peak-touristic period.
      * Health care capacity
      * Household occupants
    - One where we focus on people’s **compliance**. Factors:
      * Behavioural measures
      * Weather
    - **Efficiency**
      * This is more composite. Takes into account policy and compliance.
      * High efficiency is measured in total cases/deaths as a function of intensity of measures and movement reduction.
  + Challenges:
    - How continuous/ discrete do we want our policy index to be?
      * This point is discussed in the email Lio sent on June 1st:
        + 
      * Let’s try many possible indexes, throw them in the KSC table and take it from there

## Within country level

Early thoughts here:

* “I was thinking that it might be interesting to look within certain countries for which we have a lot of reliable info. For example, Google records the movement in 160 cities in the UK. The cool thing about this exercise is that within country analysis (like the UK in this example) offers us a singular and unambiguous time stamp for the introduction of a policy. Yet, I am fairly sure we will observe significant variation across different cities. What predicts this variation? This exercise can help us sharpen the tools for the country level  analysis or make for an interesting study on its own, provided the results are juicy enough. Sure, we don't have access to behavioural measures at a city level. But we do have access (or can potentially find) for avg income, population density, household residents (?), temperature/weather...etc.
* A challenge: we start dealing with big data, even when we look at a single country: 160 cities\*120 days \*10 (long format data that R handles best)
* An interesting advantage: look at specific events like the George Floyd death and how they impacted the trend.

## Main analysis

How will the main analysis look? We have variables that change by the day like Movement (and derivatives), Policy Index (and derivatives), Weather and the rest that are country specific.

|  |  |  |
| --- | --- | --- |
|  | Daily | SubCountry |
| Movement | 1 | 1 |
| Policy Oxford | 1 | 0 |
| Policy Corona | 1 | 1 |
| Behavuoural | 0 | 0 |
| Economy | 0 | ? |
| Covid | 1 | ? |

“?” means that for some we have info in sub-regions of a country too.

The amount of information will influence the type of analysis. If the primary regression is regarding movement (and derivatives) we could run a time series with country-fixed effects captured by the variables that don’t change…

If we want to run a more cross-section analysis, then we might want to find an aggregation rule for the variables that change daily.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **COL** | **ROL** | **patience** | **risktaking** | **altruism** | **trust** | **MAS** | **UAI** | **LTO** | **total\_cases** | **new\_cases** | **total\_deaths** | **new\_deaths** | **Gini** | **GDP.capita** | **polity2** | **cor\_simple** | **coef\_simple** |
| **COL** | 1.000 | -0.645 | -0.736 | -0.021 | 0.034 | -0.198 | -0.103 | 0.276 | 0.269 | 0.231 | 0.214 | -0.003 | -0.146 | 0.507 | -0.705 | -0.274 | -0.109 | -0.147 |
| **ROL** | -0.645 | 1.000 | 0.716 | -0.093 | -0.062 | 0.193 | -0.157 | -0.226 | 0.174 | 0.263 | 0.160 | 0.054 | 0.110 | -0.433 | 0.813 | 0.309 | 0.196 | 0.081 |
| **patience** | -0.736 | 0.716 | 1.000 | 0.260 | -0.039 | 0.182 | -0.123 | -0.429 | -0.125 | 0.153 | -0.040 | 0.027 | 0.082 | -0.328 | 0.833 | 0.335 | 0.283 | 0.273 |
| **risktaking** | -0.021 | -0.093 | 0.260 | 1.000 | -0.063 | -0.047 | -0.063 | -0.413 | -0.827 | -0.112 | -0.259 | 0.007 | 0.004 | 0.389 | -0.022 | -0.078 | -0.148 | 0.038 |
| **altruism** | 0.034 | -0.062 | -0.039 | -0.063 | 1.000 | 0.199 | -0.071 | -0.098 | 0.073 | 0.085 | -0.051 | 0.012 | -0.071 | 0.155 | -0.037 | -0.215 | -0.059 | -0.123 |
| **trust** | -0.198 | 0.193 | 0.182 | -0.047 | 0.199 | 1.000 | -0.085 | -0.349 | -0.263 | -0.054 | -0.286 | -0.167 | -0.129 | -0.173 | 0.247 | -0.191 | 0.012 | -0.003 |
| **MAS** | -0.103 | -0.157 | -0.123 | -0.063 | -0.071 | -0.085 | 1.000 | 0.072 | 0.177 | 0.111 | 0.230 | 0.260 | -0.070 | 0.140 | -0.065 | -0.047 | -0.299 | -0.330 |
| **UAI** | 0.276 | -0.226 | -0.429 | -0.413 | -0.098 | -0.349 | 0.072 | 1.000 | 0.521 | -0.214 | -0.182 | 0.169 | 0.187 | 0.025 | -0.281 | 0.054 | -0.092 | -0.190 |
| **LTO** | 0.269 | 0.174 | -0.125 | -0.827 | 0.073 | -0.263 | 0.177 | 0.521 | 1.000 | 0.425 | 0.327 | 0.089 | *NA* | -0.064 | 0.016 | -0.015 | 0.514 | 0.239 |
| **total\_cases** | 0.231 | 0.263 | 0.153 | -0.112 | 0.085 | -0.054 | 0.111 | -0.214 | 0.425 | 1.000 | 0.790 | 0.227 | 0.058 | 0.185 | 0.135 | -0.149 | 0.210 | 0.068 |
| **new\_cases** | 0.214 | 0.160 | -0.040 | -0.259 | -0.051 | -0.286 | 0.230 | -0.182 | 0.327 | 0.790 | 1.000 | 0.329 | -0.037 | 0.213 | 0.118 | -0.086 | 0.072 | -0.077 |
| **total\_deaths** | -0.003 | 0.054 | 0.027 | 0.007 | 0.012 | -0.167 | 0.260 | 0.169 | 0.089 | 0.227 | 0.329 | 1.000 | 0.567 | -0.026 | -0.011 | 0.132 | 0.059 | -0.123 |
| **new\_deaths** | -0.146 | 0.110 | 0.082 | 0.004 | -0.071 | -0.129 | -0.070 | 0.187 | *NA* | 0.058 | -0.037 | 0.567 | 1.000 | -0.080 | 0.063 | 0.075 | -0.062 | -0.143 |
| **Gini** | 0.507 | -0.433 | -0.328 | 0.389 | 0.155 | -0.173 | 0.140 | 0.025 | -0.064 | 0.185 | 0.213 | -0.026 | -0.080 | 1.000 | -0.357 | -0.218 | -0.285 | -0.295 |
| **GDP.capita** | -0.705 | 0.813 | 0.833 | -0.022 | -0.037 | 0.247 | -0.065 | -0.281 | 0.016 | 0.135 | 0.118 | -0.011 | 0.063 | -0.357 | 1.000 | 0.187 | 0.130 | 0.093 |
| **polity2** | -0.274 | 0.309 | 0.335 | -0.078 | -0.215 | -0.191 | -0.047 | 0.054 | -0.015 | -0.149 | -0.086 | 0.132 | 0.075 | -0.218 | 0.187 | 1.000 | 0.220 | -0.049 |
| **cor\_simple** | -0.109 | 0.196 | 0.283 | -0.148 | -0.059 | 0.012 | -0.299 | -0.092 | 0.514 | 0.210 | 0.072 | 0.059 | -0.062 | -0.285 | 0.130 | 0.220 | 1.000 | 0.588 |
| **coef\_simple** | -0.147 | 0.081 | 0.273 | 0.038 | -0.123 | -0.003 | -0.330 | -0.190 | 0.239 | 0.068 | -0.077 | -0.123 | -0.143 | -0.295 | 0.093 | -0.049 | 0.588 | 1.000 |

* Cor\_simple: the Pearson correlation coefficient between Movement and StringencyIndex for every country.
* Coef\_simple: The coefficient of a linear model (so including a constant).
* I think that the correlation simple is good enough for this first step. I would perhaps just break the observation period in two: Phase 1 (closing down) Phase 2: reopening.
* The LTO (long term orientation) is very interesting. Correlates intuitively with cor\_simple and risktaking (to a spectacular degree!), even though they are completely independent measurements (LTO by Hofstede, risktaking by Briq). It also seems really independent to GDP, etc. Unfortunately, we have information for only 21 countries. Is this a bad thing? Should we perhaps restrict our analysis to just a few countries?

**Dimensions of national cultures**

* **Power distance index** (**PDI**): The power distance index is defined as "the extent to which the less powerful members of organizations and institutions (like the family) accept and expect that power is distributed unequally". In this dimension, inequality and power is perceived from the followers, or the lower strata. A higher degree of the Index indicates that hierarchy is clearly established and executed in society, without doubt or reason. A lower degree of the Index signifies that people question authority and attempt to distribute power.[[7]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-IACCP-7)
* **Individualism vs. collectivism** (**IDV**): This index explores the "degree to which people in a society are integrated into groups". Individualistic societies have loose ties that often only relate an individual to his/her immediate family. They emphasize the "I" versus the "we". Its counterpart, collectivism, describes a society in which tightly-integrated relationships tie extended families and others into [in-groups](https://en.wikipedia.org/wiki/Ingroups_and_outgroups). These in-groups are laced with undoubted loyalty and support each other when a conflict arises with another in-group.[[7]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-IACCP-7)[[8]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-OffSite-8)
* **Uncertainty avoidance** (**UAI**): The uncertainty avoidance index is defined as "a society's tolerance for ambiguity", in which people embrace or avert an event of something unexpected, unknown, or away from the status quo. Societies that score a high degree in this index opt for stiff codes of behavior, guidelines, laws, and generally rely on absolute truth, or the belief that one lone truth dictates everything and people know what it is. A lower degree in this index shows more acceptance of differing thoughts or ideas. Society tends to impose fewer regulations, ambiguity is more accustomed to, and the environment is more free-flowing.[[7]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-IACCP-7)[[8]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-OffSite-8)
* [**Masculinity**](https://en.wikipedia.org/wiki/Masculinity) **vs.** [**femininity**](https://en.wikipedia.org/wiki/Femininity) (**MAS**): In this dimension, masculinity is defined as "a preference in society for achievement, heroism, assertiveness and material rewards for success". Its counterpart represents "a preference for cooperation, modesty, caring for the weak and quality of life". Women in the respective societies tend to display different values. In feminine societies, they share modest and caring views equally with men. In more masculine societies, women are somewhat assertive and competitive, but notably less than men. In other words, they still recognize a gap between male and female values. This dimension is frequently viewed as taboo in highly masculine societies.[[7]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-IACCP-7)[[8]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-OffSite-8)
* **Long-term orientation vs. short-term orientation** (**LTO**): This dimension associates the connection of the past with the current and future actions/challenges. A higher degree of this index (long-term) indicates that [traditions](https://en.wikipedia.org/wiki/Tradition) are honored and kept, while steadfastness is valued. Societies with a low degree in this index (short-term) view adaptation and circumstantial, pragmatic problem-solving as a necessity. A poor country that is short-term oriented usually has little to no economic development, while long-term oriented countries continue to develop to a point.[[7]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-IACCP-7)[[8]](https://en.wikipedia.org/wiki/Hofstede%27s_cultural_dimensions_theory#cite_note-OffSite-8)

# 25/06/2020

1. The **simulation** verified that **ARDL** coefficients are doing what they are supposed to. Nonetheless, the coefficients do not seem to correlate with any of the key variables we are examining.
   1. To do: automate the ARDL estimation. Probably already done already, but if not, having a script that upon “source” generates the ARDL coefficient, would be neat. Perhaps a function would be even better, that takes as arguments the columns and returns the ARDL of appropriate lag.
2. Merits of **forecasting** exercise were discussed.
   1. Create a forecasting folder
   2. Identify the period of restrictive measures for each country. Let’s call this period “wave 1”.
   3. Split “wave 1” it in two: the **training** set and the **prediction** set.
   4. Fit a model explaining Movement in the training period and use the parameters to measure MSE in the prediction set.
   5. Machine learning model. Can ML do better than “theory”? Let’s use ML’s usual suspects (in a fast and frugal first step) as benchmarks.
   6. I realise that forecasting time series is very different than the type of prediction exercises I have used to. First of all, linear models are often substituted for exponential ones.
3. Why do we care about explaining (predicting) movement? Does it correlate with fewer cases and deaths? This should be a straight-forward point to make but early analysis seemed to not necessarily corroborate this assertion.
   1. Look more closely into what predicts death and cases.
      1. Perhaps there are countries that Movement reduction helps them reduce death while others do it with less reduction in Movement. Why is that? Efficiency? Weather? Etc…
4. The weather comes up very often as a latent variable that we don’t account for.
   1. Let’s find daily weather data. I have tried in the past; surprised how hard it is…
5. Literature review:
   1. Keep in developing the literature review document. So far focusing on compliance towards Covid policy. But also:
   2. Identify papers that have tackled compliance towards new policies (how it is measured, what are the scores of different countries) outside of Covid.