

The effect of Russian misinformation on COVID-19 cases in Latvia

David Rojas

9/8/2021

Introduction

The introduction should be a brief summary of the core things you will present in the rest of the document, and good practice to eventually develop an introduction for your paper.

Relevant literature review

El primer artículo relevante corresponde a “In Vaccines We Trust?” de Mónica Martínez y Andreas Stegmann (Martínez, 2021). Se expone el efecto que tuvo en los comportamientos de búsqueda de servicios de salud (y específicamente en las tasas de vacunación) una campaña de propaganda anti-vacunas en el territorio Paquistaní. La campaña estuvo dirigida a desacreditar tanto las vacunas como trabajadores de la salud.

Implementando una estrategia de diferencias en diferencias, el artículo concluye que se observaron efectos negativos y significativos en los comportamientos de búsqueda de servicios de salud en la población Pakistaní más expuesta a dicha campaña (esto adicional al efecto sobre la confianza en las vacunas mismas) tras la ola de propaganda.

La estrategia empírica de este artículo es relevante puesto que puede ser adaptada a nuestro caso de estudio.

Para contextualizar las condiciones de las poblaciones rusas en los territorios bálticos, se utilizará “Russians as a Minority in Contemporary Baltic States” (Kirch, 1992).

Para contextualizar la naturaleza y el desarrollo de las narrativas pro-Rusas diseminadas en los territorios Bálticos durante la pandemia Covid-19 se utilizará “The Kremlin’s strategic narratives on the Baltic states during the COVID-19 crisis” (Mfilder, 2020).

Theoretical argument and predictions

La pandemia Covid-19 ha provocado una coyuntura en temas de salud, económicos, políticos y de organización social en todo el mundo.

Este contexto ha sido aprovechado por algunos grupos internacionales de poder para diseminar narrativas que ponen en tela de juicio las narrativas oficiales de algunos Estados, en un afán de mermar la confianza de los ciudadanos y hacer avanzar su agenda política.

Letonia, junto los demás países Bálticos, se ha visto sumergida en una oleada de propaganda, principalmente en lengua rusa, que pretende generar un efecto de desconfianza y desinformación en el tema de las vacunas (dirigiendo las narrativas tanto a los preparados biológicos como a la capacidad del gobierno letón de administrar la emergencia sanitaria).

Las unidades territoriales donde la presencia de hablantes de lengua rusa es mayor, podrían verse especialmente afectadas en caso de que las personas eligieran diferenciar su comportamiento del de otras unidades territoriales con otra composición demográfica. Esto a partir de la desinformación que las personas en estas unidades territoriales consumen.

Se puede presentar un efecto diferenciado en la tasa de aparición de nuevos casos de Covid-19 en dichas unidades territoriales por alguna de las siguientes razones.

1. Una mayor exposición a campañas anti-vacunas puede provocar una disminución en la confianza en las vacunas y, por ende, una disminución en la tasa de vacunación. Al tener un bajo porcentaje de población vacunada, la unidad territorial es más sensible a la aparición de nuevos casos.
2. Una mayor exposición a campañas de desinformación (aunado a una menor disponibilidad de información oficial en lengua rusa) puede generar desconfianza en los servicios de salud. Personas con poca confianza en los servicios de salud podrían optar por no buscar tratamiento médico ante un caso de contagio, lo que puede incidir en la prevalencia de la enfermedad en estas unidades territoriales.
3. Una falta de acceso a información gubernamental de calidad en lengua rusa puede provocar una falta de coordinación en la estrategia de contención de la enfermedad.

Las causas anteriores pueden generar un aumento en el número de casos por Covid-19 en las unidades territoriales letonas con mayor presencia de rusoparlantes.

Context of the project

El contexto geopolítico y demográfico con el que Letonia entra a la pandemia Covid-19 es sumamente complejo. Con un pasado de república soviética, Letonia forma parte de la OTAN y la Unión Europea desde 2004.

A diferencia del exitoso proceso lituano de integración de la población rusófona, en Letonia este proceso ha sido más complejo, en parte por haber una mayor cantidad de hablantes de ruso. Además, las personas que hablan ruso son más proclives a identificarse como rusas que como letonas.¹

La región de Latgale es densa en rusoparlantes. Es, además, una de las regiones con menor ingreso per cápita. El Partido de la Armonía (Saskara) es popular en esta y otras regiones con importante presencia rusófila y su (ahora auto-censurada) afinidad con Rusia Unida, el partido de Vladimir Putin, no es menos que controversial.²

Podría decirse que la población rusoparlante se encuentra agrupada geográficamente, marginada económicamente y unida políticamente en la protección de sus intereses.

En Letonia, una de cada cuatro personas tienen ascendencia rusa, pero además, de aquellas personas que hablan letón, nueve de cada diez comprende la lengua rusa. Es decir, “todos” en Letonia entienden ruso. Debido a que aprender letón es un requisito para la ciudadanía, aproximadamente trescientas mil personas, i.e. el 15% de la población que se encuentra en el territorio letón, permanece sin ciudadanía, por lo que no puede votar ni trabajar para el gobierno.³

En mayo de 2020 más de la mitad de eventos virales de desinformación sobre Covid-19 en internet en los países bálticos, ocurrieron en Letonia. Además, siete de cada diez fake news se diseminaron a través de Facebook.⁴

Es importante mencionar que Letonia se adhiere plenamente al portafolio de vacunación de la Unión Europea, el cual es aprobado por la Agencia Europea de Medicamentos con sede en Ámsterdam.⁵

Empirical strategy

La estrategia empírica consiste en un método de diferencias en diferencias.

Sea t la variable que indexa las fechas. Sea m la variable que indexa las unidades territoriales de Letonia. Sea $y_{t,m}$ la tasa de nuevos casos por cada cien mil habitantes. Sea r_{sp_m} la proporción de

¹Russian Minority Populations in the Baltics. <https://storymaps.arcgis.com/stories/6cb4f0a7dcd64278b52840a7dc364127>

²idem.

³Latvia: Battling Russian Propaganda, Moving slowly to the West. <https://nationalsecurityzone.medill.northwestern.edu/blog/snowball/latvia/>

⁴Latvia had the widest spread of COVID-19 related disinformation in May, <https://www.debunkeu.org/post/debunk-eu-latvia-had-the-widest-spread-of-covid-19-related-disinformation-in-may>

⁵Which piece of the puzzle is Latvia in the Covid-19 vaccination campaign of European Union? <https://www.debunkeu.org/post/which-piece-of-the-puzzle-is-latvia-in-the-covid-19-vaccination-campaign-of-european-union>

población que habla ruso en la unidad territorial m .

Consideremos la variable T_t para indicar la presencia de un evento viral de desinformación sobre Covid-19 en internet al tiempo t .

Se propone el siguiente modelo.

$$y_{t,m} = \beta(rsp_m \times T_t) + \eta_m + \delta_t + \epsilon_{t,m} \quad (1)$$

Donde η_m representa efectos fijos por unidad territorial y δ_t representa efectos fijos por fecha.

La significancia del coeficiente β nos indicará la relevancia de los efectos de las campañas rusas de desinformación sobre Covid-19 en la población de habla rusa en Letonia.

Possible control

Una de las variables que podría ser utilizada como control es el ingreso per cápita de la unidad territorial m . Una población con más o menos ingreso podría ser más o menos proclive a los efectos de la desinformación con respecto a las vacunas (y la confianza en los servicios de salud) y por ende a los efectos de la aparición de nuevos casos.

¿Subestimación de casos?

Surge aquí, sin embargo, la primera consideración que es posible hacer. La población de habla rusa, al ser una minoría en circunstancia de marginación, podría tener un acceso diferenciado a los servicios de salud, específicamente podría tener acceso diferenciado a pruebas de detección de Covid-19, lo que podría generar una subestimación de los nuevos casos.

Sea $\mu_{t,m}$ el número de pruebas realizadas por cada cien mil habitantes en la unidad territorial m a tiempo t . Consideremos que $rsp_{t,m} = rsp_m$ para toda t . Podemos usar el siguiente modelo OLS para identificar el efecto de la cantidad de rusoparlantes en una unidad territorial con respecto al número de pruebas realizadas.

$$\mu_{t,m} = \gamma rsp_{t,m} + \epsilon_{t,m} \quad (2)$$

El coeficiente γ nos ayudaría a identificar si existe un efecto de sub-testeo en la población de habla rusa.

Data

Las bases de datos de nuevos casos por unidad territorial se pueden descargar del Latvian Open Data Portal. Desde este portal se puede descargar los archivos `covid_19_pa_adm_terit.csv` y `covid_19_pa_adm_terit_new.csv` que contienen el detalle de casos.

La base de datos demográficos corresponde al archivo `Ethnic_nationality_2020_Alone.dta` que a su vez se deriva del archivo `v17-2020-01-01Polygon.xlsx`.

La base de datos actual se obtiene de juntar estas dos bases. El criterio para unir los datos es usar como llave la variable del Nombre de la Unidad Territorial, que en el archivo de casos se llama `Administrativi Teritorialas Vienibas Nosaukums` y en el de etnicidad `10_name`.

Variable description and sources

Llamaremos $confirmados_{t,m}$ a la variable `ApstiprinataCOVID19infekcija` (Covid 19 infections confirmed) de las bases de nuevos casos del Latvian Open Data Portal. Esta variable nos indica el número de nuevos confirmados en la unidad territorial m a tiempo t .

Definimos $casosNuevos_{t,m} = confirmados_{t,m} - confirmados_{t-1,m}$.

La variable $nHabitantes_m$ corresponde a la variable `Total` de la base de datos demográficos `Ethnic_nationality_2020_Alone.dta`.

La variable rsp_m corresponde a la variable `share_Russian` de la base de datos demográficos `Ethnic_nationality_2020_Alone.dta`.

La variable $y_{t,m}$, tasa de nuevos casos por cada cien mil habitantes, se calcula como cien mil veces el cociente entre $casosNuevos_{t,m}$ y $nHabitantes_m$.

La variable T_t será obtenida de la base de eventos de desinformación. Por el momento será generada con indicadoras de día y semana.

La variable $\mu_{t,m}$ para el número de pruebas realizadas por cada cien mil habitantes en la unidad territorial m a tiempo t no se ha localizado al momento de la realización de este avance.

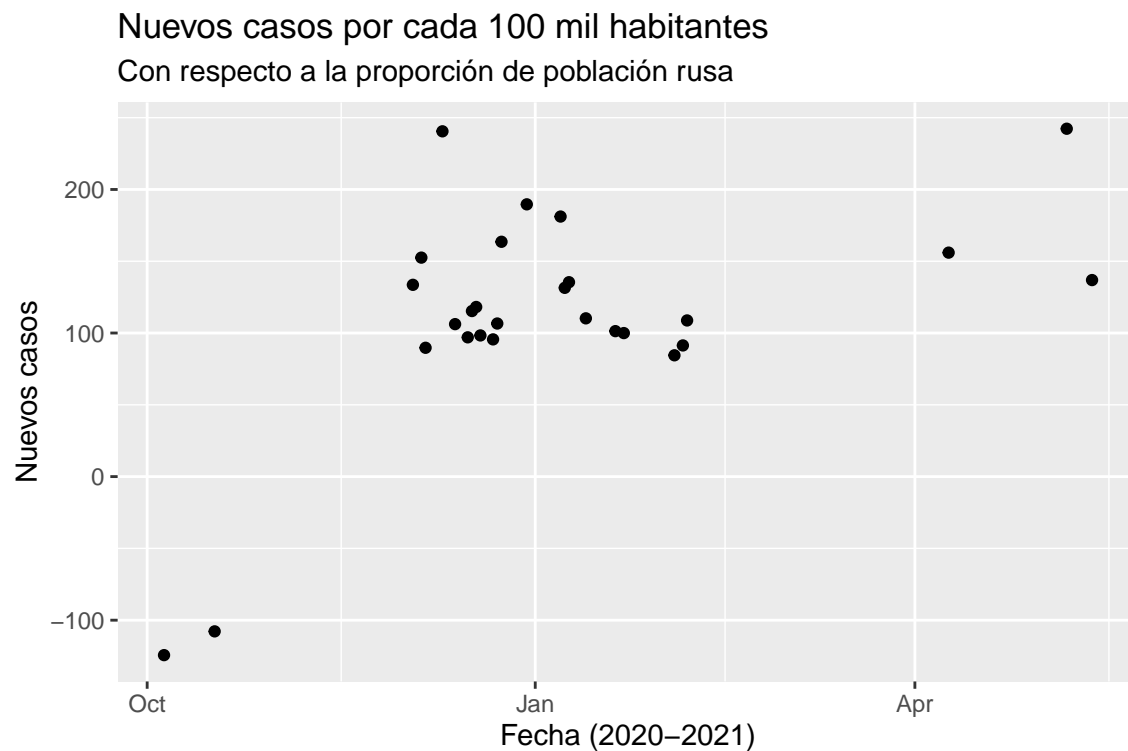
Completed and next steps

Se calculan las regresiones para la estrategia empírica propuesta usando variables indicadoras por día. Se muestran en la siguiente tabla los días días en los que se observan los efectos más significativos. Cabe señalar que todos son altamente significativos.

Fecha	Estimate	Std. Error	t value	Pr(> t)
2021-05-07	242.3061	32.41426	7.475293	0.00e+00
2020-12-10	240.5031	32.41449	7.419615	0.00e+00
2020-12-30	189.6444	32.42049	5.849522	0.00e+00
2021-01-07	181.1253	32.42135	5.586606	0.00e+00
2020-12-24	163.5553	32.42301	5.044421	5.00e-07
2021-04-09	155.9938	32.42368	4.811107	1.50e-06
2020-12-05	152.5301	32.42397	4.704238	2.60e-06
2021-05-13	136.8900	32.42521	4.221716	2.43e-05
2021-01-09	135.4062	32.42532	4.175940	2.97e-05
2020-12-03	133.5798	32.42545	4.119598	3.80e-05

Esto quiere decir por ejemplo que, los días 7 de mayo del 2021 y 10 de diciembre de 2020, se observaron 2.4 casos de Covid 19 adicionales por cada punto porcentual de población rusoparlante en una unidad territorial dada por cada cien mil habitantes.

Se grafican a continuación los días en los que se observaron los efectos significativos al 1%. Se puede observar un cluster de días con efectos positivos y significativos durante diciembre y enero. Se pueden apreciar además dos días en octubre de 2020 donde el efecto fue negativo.



Example of the data base

En el siguiente repositorio se pueden encontrar los avances del proyecto. www.github.com/rojasazules/proyecto.

En la carpeta `/data/processed` se encuentra el archivo `casos.dta`.

Contiene una base con las siguientes columnas.

-
-
-

The example of your final database should be how the data should look (and ideally is looking) right before you can implement the empirical strategy. So, I do not want you to show me photos of the original data sets and say that somehow magically those will end up being the database you need to run the analysis. So, there are two options. First, you actually have that database, and thus I would like to put a print screen and explain to me broadly what you did. Second, you do not have the data, and thus I would like you to put a print screen of an Excel of the data set you need to arrive to so that you can run the analysis. You will also need to explain every single detail you need to do with each data set (not vaguely, but very precisely), including aggregation (if applicable), how you will merge them (exactly using which variables in each data set), etc to arrive to how the data should look like right before you can implement the empirical strategy

Bibliography

- The Latvian Open Data Portal. <https://data.gov.lv/dati/eng/dataset/covid-19>
-