

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

David Rojas

9/8/2021

## Introduction

Se presenta una estrategia de investigación para investigar el efecto que han tenido las olas de desinformación en lengua rusa sobre temas de vacunación y salud en la tasa de aparición de nuevos casos de Covid-19 en Letonia para las localidades con mayor presencia de hablantes de ruso.

Se presenta un resumen de la literatura que se ha revisado hasta el momento y se plantea replicar una estrategia para detectar efectos de campañas de desinformación sobre vacunas utilizando una estrategia de diferencias en diferencias.

Se propone un argumento teórico y los canales de causalidad que podrían producir un efecto en la tasa de aparición de nuevos casos. Así mismo, se da una breve introducción al contexto social de Letonia, lo que nos ayudará a proponer algunas ideas de controles adicionales para la estrategia empírica.

En las siguientes secciones se describe la estrategia empírica así como las fuentes de datos utilizadas para desarrollar los resultados preliminares.

Ante la actual falta de una base de eventos de desinformación, se realizan cálculos preliminares para identificar los días y semanas en los que se presenta un efecto diferenciado y significativo durante la segunda mitad de 2020 y el 2021.

Se identifica una ventana de tiempo de la semana 48 del 2020 a la semana 2 del 2021 cuando los efectos observados en la aparición de casos es significativamente mayor en función de la proporción de población de habla rusa que contenga una unidad territorial dada.

Se agrega una estimación adicional tratando de estimar el efecto en la población de lengua bielorrusa.

Se proponen siguientes pasos y se expone la base utilizada para la estrategia empírica.

## 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 Pakistání 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” (Mölder, 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.<sup>1</sup>

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.<sup>2</sup>

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

---

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

<sup>2</sup>idem.

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.<sup>3</sup>

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.<sup>4</sup>

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.<sup>5</sup>

## 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  $rsp_m$  la proporción de 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  $\eta_m$  representa efectos fijos por unidad territorial y  $\delta_t$  representa efectos fijos por fecha.

La significancia del coeficiente  $\beta$  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

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

<sup>4</sup>Latvia 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>

<sup>5</sup>Which 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>

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  $\gamma$  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 `l0_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

### Indicadoras por día

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

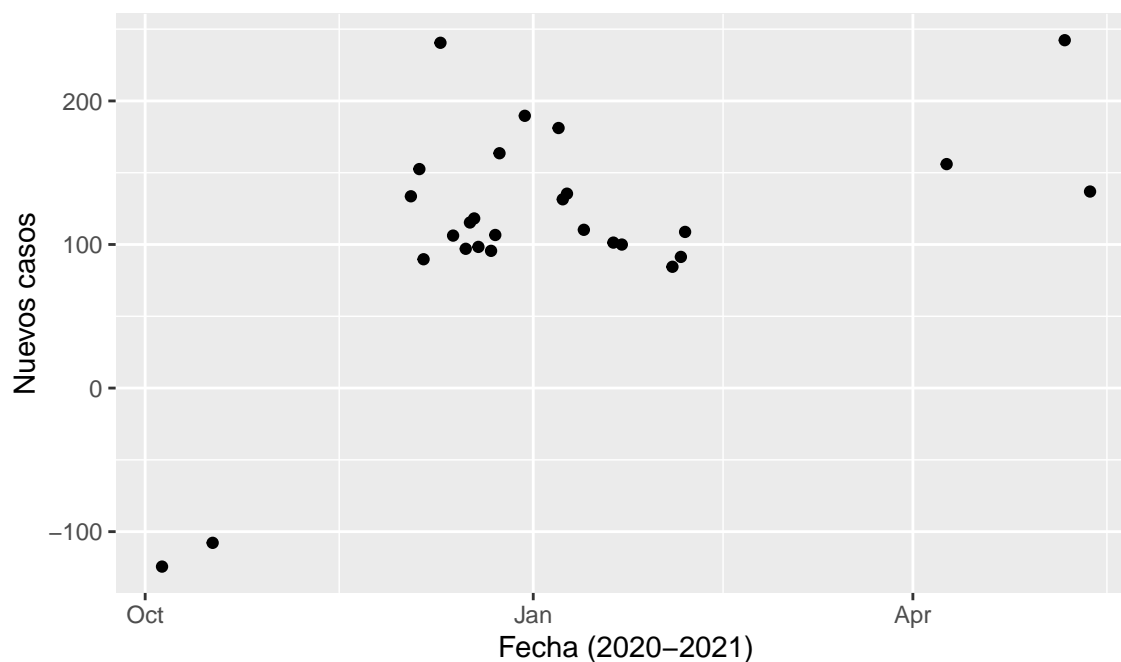
Fecha	Estimate	Std. Error	t value	Pr(> t )
2021-05-07	242.3061	32.41426	7.475293	0.00e+00
2020-12-10	240.5031	32.41449	7.419615	0.00e+00
2020-12-30	189.6444	32.42049	5.849522	0.00e+00
2021-01-07	181.1253	32.42135	5.586606	0.00e+00
2020-12-24	163.5553	32.42301	5.044421	5.00e-07
2021-04-09	155.9938	32.42368	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.

## Nuevos casos por cada 100 mil habitantes

Con respecto a la proporción de población rusa



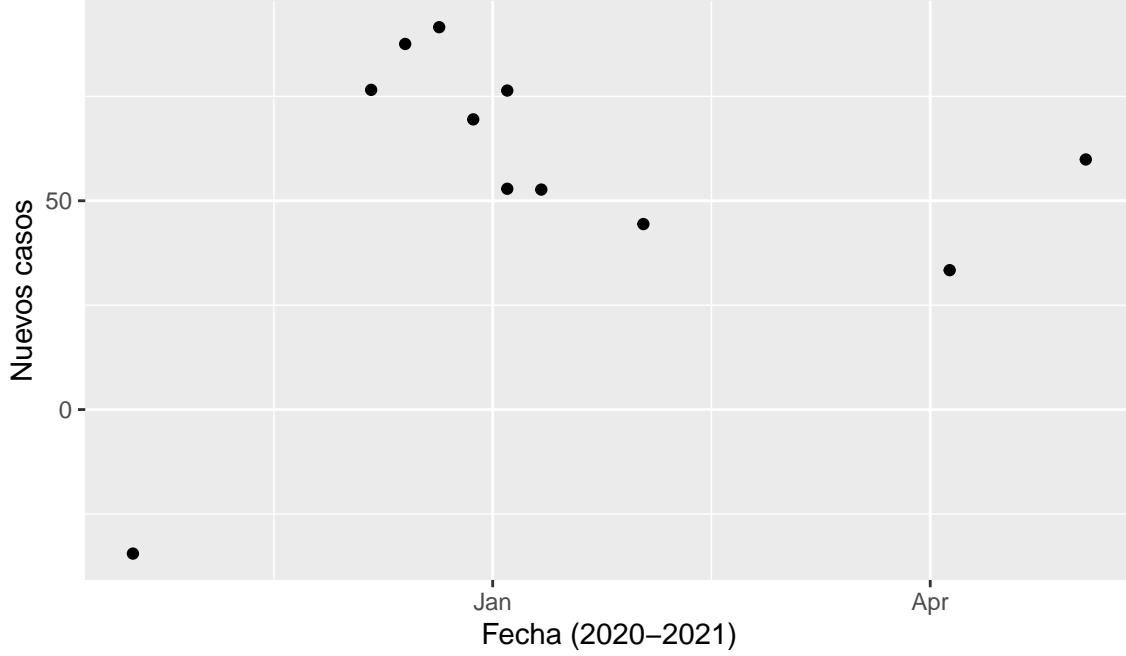
## Indicadoras por semana

Se repite la estrategia empírica pero esta vez usando indicadoras por semana. Coloquemos en una tabla las semanas con efectos más significativos. Las semanas se indican como `aaaass`, donde `ss` es el número de semana del año `aaaa`.

semana	Estimate	Std. Error	t value	Pr(> t )
202051	91.56465	12.32962	7.426395	0.0000000
202050	87.54928	12.33014	7.100428	0.0000000
202049	76.55663	12.33144	6.208247	0.0000000
202102	76.38915	12.33146	6.194656	0.0000000
202052	69.49346	12.33218	5.635130	0.0000000
202119	59.89112	12.33308	4.856137	0.0000012
202101	52.87281	12.33365	4.286875	0.0000182
202103	52.67673	12.33366	4.270972	0.0000195
202106	44.41907	12.33424	3.601281	0.0003169
202042	-34.49021	12.33480	-2.796170	0.0051730

Podemos confirmar que en las últimas semanas del 2020 y las primeras semanas del 2021 el efecto de aparición de casos de Covid-19 fue más fuerte con respecto a la proporción de hablantes de lengua rusa. Así mismo se confirma el efecto negativo en la semana 42 del 2020.

### Nuevos casos por cada 100 mil habitantes Con respecto a la proporción de población rusa



### Parallel trends assumption

Para comprobar el supuesto de tendencias paralelas se propondrá el cálculo de efectos lead/lag para detectar efectos de anticipación. La ecuación original para estimar efectos lead/lad es la siguiente.

$$y_{t,m} = \tau T_t + \gamma_m + \lambda_t + \sum_{t'=1}^f \delta_{t'} T_{t+t'} + \sum_{t'=1}^l \delta_{-t'} T_{t-t'} + \epsilon_{t,m} \quad (3)$$

Donde  $l$  es el número de rezagos,  $f$  es el número de adelantos y  $\gamma_m$  y  $\lambda_t$  son efectos fijos por unidad territorial y tiempo.

Sin embargo, debido a que en la especificación del modelo original considera una interacción entre el tratamiento y la variable  $rsp_m$ , se propone considerar el siguiente modelo de lead and lags.

$$y_{t,m} = \tau(rsp_m \times T_t) + \gamma_m + \lambda_t + \sum_{t'=1}^f \delta_{t'}(rsp_m \times T_{t+t'}) + \sum_{t'=1}^l \delta_{-t'}(rsp_m \times T_{t-t'}) + \epsilon_{t,m} \quad (4)$$

### First candidate to sociodemographic control

Se trabajará con el primer control sociodemográfico que se considera de especial relevancia, el nivel de los salarios por unidad territorial.



Sea  $x_m^1$  el salario mensual promedio en euros en la unidad territorial  $m$  durante el cuarto trimestre de 2020 (trimestre donde comienza la ventana de tiempo que va de la semana 48 del 2020 a la semana 2 del 2021). Se propone usar la siguiente especificación.

$$y_{t,m} = \beta(rsp_m \times T_t) + \eta_m + \delta_t + \sum_t \Gamma_{t,1} \delta_t \times x_m^1 + \epsilon_{t,m} \quad (5)$$

La variable  $x_m^1$  será depurada del archivo `Average monthly wages and salaries in cities and municipalities` de la carpeta `Census Data` que viene en el Dropbox del proyecto.

En caso de que se agreguen más controles, la especificación que calcularemos será la siguiente.

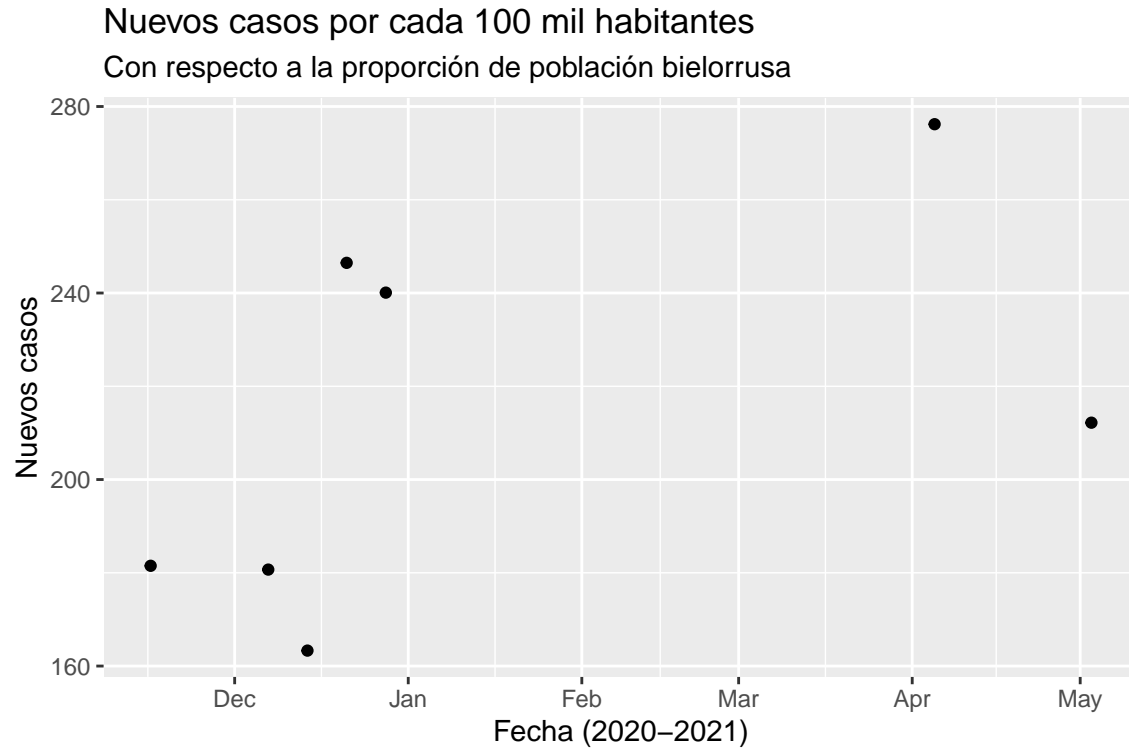
$$y_{t,m} = \beta(rsp_m \times T_t) + \eta_m + \delta_t + \sum_c^{nControls} \sum_t \Gamma_{t,1} \delta_t \times x_m^c + \epsilon_{t,m} \quad (6)$$

## Latvia also speaks Belarrussian

Se repite la estrategia empírica pero esta vez observando el efecto relacionado a la cantidad de hablantes de bielorruso. Coloquemos en una tabla las semanas con efectos más significativos. Las semanas se indican como `aaaass`, donde `ss` es el número de semana del año `aaaa`.

semana	Estimate	Std. Error	t value	Pr(> t )
202115	276.2093	55.33630	4.991467	0.0000006
202051	246.4728	55.33879	4.453888	0.0000084
202052	240.0824	55.33929	4.338371	0.0000144
202119	212.1934	55.34131	3.834267	0.0001261
202046	181.4995	55.34325	3.279523	0.0010405
202049	180.6868	55.34330	3.264837	0.0010959
202050	163.3122	55.34426	2.950842	0.0031704

Podemos observar que nos encontramos ante un patrón distinto a los hablantes de ruso en cuanto a la aparición de efectos significativos con respecto a la proporción de hablantes de lengua bielorrusa en una unidad territorial dada.



## Next steps

- Realizar regresión de lead and lags con el tratamiento interactuado con la variable  $rsp_m$ .
- Depurar la data de salarios mensuales promedio por unidad territorial.
- Continuar trabajando en la identificación de eventos de desinformación. E.g. identificar google trends olas de búsqueda de información de las vacunas (o de ciertas vacunas específicas).
- Comprender cómo realizar análisis de efectos placebo y pruebas de robustés.
- Seguir escribiendo.

## Example of data base

En el siguiente repositorio se pueden encontrar los avances del proyecto.

[github.com/rojasazules/research\\_project\\_1](https://github.com/rojasazules/research_project_1)

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

Contiene una base con las siguientes columnas.

- Fecha. En formato `aaaa-mm-dd`.

- ut. Código numérico para la unidad territorial.
- Nombre Unidad Territorial.
- tasa\_confirmados. Nuevos casos por día por unidad territorial por cada 100 mil habitantes.
- share\_Russian, share\_Latvian, share\_Belarusian, share\_Ukrainian, share\_Polish, share\_Lithuanian, share\_Other. Variables de composición demográfica de la unidad territorial.
- Total. Número total de habitantes por unidad territorial
- Semana. En formato aaaass. Indica el año y número de semana.

## Bibliography

- The Latvian Open Data Portal. <https://data.gov.lv/dati/eng/dataset/covid-19>
- (Kirch, 1992) Russians as a Minority in Contemporary Baltic States. Aksel Kirch. Bulletin of Peace Proposals , June 1992, Vol. 23, No. 2 (June 1992), pp. 205-212.
- (Martínez, 2021) In Vaccines We Trust? The Effects of the CIA's Vaccine Ruse on Immunization in Pakistan, Journal of the European Economic Association, Monica Martinez-Bravo, Andreas Stegmann, 2021;, jvab018, <https://doi.org/10.1093/jeea/jvab018>
- (Mölder, 2020) The Kremlin's strategic narratives on the Baltic states during the COVID-19 crisis. Holger Mölder and Vladimir Sazonov. December 2020. Kwartalnik Bellona.