Ukrainization and the effect of Russian language on the Web: the Google Trends case study

Hui Yao
School of Mathematics and Physics
The University of Queensland
St Lucia QLD, 4072, Australia: hui.yao@uq.net.au

Andrew Crowden
School of Historical and Philosophical Inquiry
The University of Queensland
St Lucia QLD, 4072, Australia; a.crowden@uq.edu.au

Radislav Vaisman
School of Mathematics and Physics
The University of Queensland
St Lucia QLD, 4072, Australia; r.vaisman@uq.edu.au

Corresponding author: Dr. Radislav Vaisman

Academic Position: Lecturer (equivalent to US Assistant Professor), at the School of Mathematics and Physics, The University of Queensland, QLD, Australia 4072.

Web: `https://researchers.uq.edu.au/researcher/11107`

Email: `r.vaisman@uq.edu.au`

Phone: `+61-733653264`

Conflicts of interest

The authors have no conflicts of interest to report. The authors contributed equally to this paper.

Ukrainization and the effect of Russian language on the Web: the Google Trends case study

April 3, 2023

Abstract

In this paper we consider the question of language diffusion. Specifically, we examine the latest attempt to popularise the Ukrainian language. This attempt has been undertaken by the Ukrainian government since 1991 and is a major Ukrainian objective. Despite encouraging reports from the Ukrainian side, in practice the language adoption progress might be slower than reported. However, sociological studies report positive dynamics in Ukrainian language dissemination. We test this hypothesis by focusing on google trends web search data. We apply a Bayesian beta regression model where we explore the effect on language use in all Ukrainian regions and in regions bordering the Russian Federation. Our analysis shows that the proposed model is appropriate. Overall, the results suggest that the Ukrainian language popularisation policy is successful. Interestingly, our model suggests that the 2022 Russian invasion has considerably intensified the usage of the Ukrainian language.

Keywords— methodology, culture, nationalism, identity, soft power

1 Introduction

The Russian Federation has enormous military power. According to the *global fire power* ranking,[1] its army is placed *second* on our planet. Moscow also possesses a powerful cultural weapon which is frequently overlooked — the Russian language (Decker 2021). The Russian language has an enormous uniting power for 193 ethnic groups that live in Russia, with more than 95% speaking Russian (Gladkova 2015). Despite the dissolution of the Soviet Union which preceded the creation of many new independent states, Moscow is more than willing to use both cultural and linguistic ties with the view to promoting its *geopolitical interests* (Feklyunina 2016, Decker 2021, Forsberg & Smith 2016). The usage of this (soft) force is especially applicable (but is not limited), to a range of former Soviet Union republics such as Ukraine, Belarus, and Kazakhstan. In this study, we examine the case of Ukraine and in particular, the most recent Ukrainization attempt that followed the 2014 *Euromaidan* revolution. Starting from 1991, the Ukrainian government has opposed both the dissemination of Russian cultural and linguistic establishment in the Ukrainian territory. These attempts have become more comprehensive since 2014. The extensive Ukrainization policy (see Kiss (2022), was perceived by Donetsk and Luhansk regions as an effort to forbid the Russian language. This could be an important element that contributed to the development of separatists movements in eastern Ukraine. In addition, this situation was a perfect opportunity for Vladimir Putin, President of Russia, to pursue his geopolitical aims. The Ukrainian case gives us a chance to *rigorously* examine the effect of *de-Russification* policies in the post Soviet space; the latter being the main objective of this study.

Several studies suggest that language is an important aspect of society's self-identification (Ushchyna 2020, Poses & Revilla 2021, Kulyk 2011, Arel 2002, Reznik 2018). An implementation of Ukrainization policy implies the increased usage of the Ukrainian language, promoting the local culture in various domains, and discouraging both the Russian cultural influence and the Russian language. For a comprehensive historical overview of the Ukrainian language development, we refer to Flier & Graziosi (2017-2018). While such policy can be beneficial for Ukrainian identity and selfhood, it is important to remember that for some regions, especially for those that share a border with the Russian Federation, the corresponding ties to Russia are extremely strong. In addition, many people who live in these regions consider Russian as their native language and have (historically) held pro-Russian views (Bureiko & Moga 2018). Such a regional diversity effect has been often exploited by politicians. Toivanen (2007), states that “Language boundaries, real or imagined, can easily become exploited politically.” Not surprisingly, Ukrainian politicians have also capitalised on linguistic and cultural differences to accumulate electoral support (Kulyk 2011).

Language planning development as part of Ukrainization policy has always been an important concern in Ukraine (Kiss 2022). In 1989, Ukrainian Soviet Socialist Republic adopted the *Law of Languages*; this law declared the Ukrainian language as the *only* official language in the state. It is convenient to divide the 1991–2021 time frame into three periods. Namely, the early independence years (1991–2004), the *Orange revolution*, which took place in 2004, and the post Euromaidan era which started in 2014. Under President Kravchuk (1991-1994), a de-Russification of schools have begun. President Yushchenko (2005-2010) started to de-Russify the media landscape. However, only mild Ukrainization attempts were made by president Yushchenko's government. One example of such actions, is the requirement that television and radio will have a *quota* of 75% minimum Ukrainian-language programs, and that there should be an audio-dubbing in Ukrainian for programs that are originally broadcasted using other languages. In 2010, when the election was won by Yanukovich, Ukrainization policy implementation experienced a significant slowdown. A law of Kivalov and Kolesnichenko (both members of the Verkhovna Rada), called the *Principles of State language policy* from 2012, granted the Russian language the status of a *regional language*. Effectively, the law allowed the usage of minority languages (and Russian in particular) in government institutions such as schools and courts within the regions where the national minorities exceed 10% of the population (Elder 2012). The most important event that happened in 2014 at the linguistic front was when parliament repealed (February 23, 2014) the “Principles of State language policy” law. While President Oleksandr Turchynov (23 February 2014 – 7 June 2014) and President Petro Poroshenko (2014-2019) refused to sign the removal of the law (so the law remained in force until February 2018), the parliament decision provided a pretext for Moscow to militarily annex Crimea and promote separatist movements in the east (Reznik 2018).

After the Euromaidan events, Ukrainian society has experienced major changes. The loss of Crimea Peninsula and the war in the eastern Ukraine led to a dramatic increase in national identity and influenced the growing usage of the Ukrainian language. For example, in 2018, the Kivalov-Kolesnichenko Language Law was declared unconstitutional by the Constitutional Court of Ukraine. According to the Ukrainian centre for economic & political studies, (Centr Razumkova 2016), 69%, 27%, and 2% of Ukrainians, consider the Ukrainian language, the Russian language and other language, as their native languages, respectively. Nevertheless, certain care should be taken when discussing the "native language" term (Zeller 2021, Hentschel & Palinska 2022). Specifically, it does not necessarily mean that this language is used in practice but instead, it might correspond to nationality, heritage, or the country of residence.

The de-Russification efforts of Ukrainian government included Ukrainian language quotas for television and radio broadcasting (Ogarkova 2018) and a *de-communization* law under which the majority of geographical names with reference to Soviet era were changed. To further support de-Russification laws the Ukrainian government banned the commercial importation of books from Russia in February 2017. According to Ukrainian sources the Ukrainization policy is very fruitful. The 2021 sociological service poll of Ilko Kucheriv Democratic Initiatives Foundation and Razumkov Centre (2021), which considered 2,019 respondents aged 18 and older, showed that 78% of Ukrainians identify the Ukrainian language as their native language as opposed to 18% who stated that their native language is Russian. As expected, Russia-oriented media sources like *Russia Today* (RT), report that the majority of the population in Ukraine is actively using the Russian language and in fact prefers to use Russian. For example, based on polls from *Social Monitoring Centre*, RT reports that more than 50% of Ukrainian citizens are willing to consume books and media that are delivered in Russian. Moreover, they also report that less than a third of the population supports the usage of the Ukrainian language only. RT thus arrives at the conclusion that the "forceful" Ukrainization of the population, which started in 2014, is not effective overall; for details, please see (Latyshev et al. 2021). We would like to note that we could not independently verify the *Social Monitoring Centre* poll, since this resource is no longer available.

In this work, we do not intend to make any political claims in favour of Ukrainization nor Russification of Ukraine. We do want to note that due to different polls, there is a considerable lack of clarity and uncertainty about the actual usage of Ukrainian and Russian languages in Ukraine. In addition, some responders might be reluctant about stating their true preferences in both the controlled and the uncontrolled territories. With this in mind, we aim to rigorously investigate the hypothesis that the usage of the Ukrainian language is actually growing, by examining an *independent* and *self-sufficient* data source. In order to accomplish this, we utilise the *google trend data* from 2011 to 2021 in order to study the dynamic of change in percentage of Ukrainian language usage. In order to understand the effect of Ukrainization on different regions, namely, regions that are geographically (historically and culturally), closer to Russia or having such a proximity to the "West",[2] we apply a *Bayesian beta* *regression model* which can take into account the effect of regions (significant parts of two regions were annexed by Russian Federation in 2014). We verify that the model fits the data well and that the Ukrainization policy is successful in all regions (except of Crimea and Sevastopol). However, we also show that in practice, the Ukrainian language adoption by the Ukrainian population might be slower than reported.

In addition, we provide several conceptual and methodological contributions. First, we show that the annexed regions of Crimea and Sevastopol should be have experienced a severe deterioration in the Ukrainian language usage. A possible reason for this decline is the fact that the usage of the Ukrainian language is discouraged and that many Ukrainians, including Crimean Tatars, who fled from Crimea since 2014. For additional details about linguistic conflicts, we refer to Müller & Wingender (2021) (specifically, see the *Characterisation of the Language Situation in the Republic of Crimea from the Perspective of Geolinguistics*, by Yuri Dorofeev, Part III in Müller & Wingender (2021)), where a situation in Crimea is discussed. Other regions, that are now under control of the Ukrainian government, show a *statistically significant* increase of Ukrainian language usage. Nevertheless, our model suggests that some regions (Donetsk and Luhansk), in Ukraine show a very slow adoption of the Ukrainian language. This is not very surprising since Donetsk and Luhansk regions have a strong pro-Russian agenda and are partly controlled by separatists. Moreover, with respect to these regions, the question of multiethnicity of Ukraine is raised. The question of Ukraine being a multiethnic country, was recently examined by Kulyk (2022*b*), where the author focuses on the disappearing differentiation between the two largest groups of Soviet times, Ukrainians and Russians.

The rest of the paper is organised as follows. In Section 2, we formally define the methods used in this paper. We show that one can use the Bayesian beta regression models to fit a language usage data extracted from google trends, and also validate the model and perform efficient prior selection. The results are discussed in Section 3. Section 4 is dedicated to the full-scale Russian invasion in 2022, where we show that the proposed model can be easily adjusted to account for large-scale changes. Finally, in Section 5, we summarise our findings and discuss both the limitations of the proposed method and the possible directions for future research.

2 Methods

2.1 Data collection

Google and other media companies collect useful data about their customers preferences. Here, we take advantage of the publicly available *Google trends* data, and in particular, of the *interest by sub-region* option which is attainable for specific search terms. Under our setting, a search term is translated into two languages: Ukrainian and Russian, while making sure that these terms are *written differently*, in order to distinguish between them. Effectively, we now have *two* search terms that have the *same meaning* but, are treated as *different* search labels by *Google*. Then, the *sub-region* data provides the percentage of queries recorded for these two terms for a specific region and for a given time frame. Using the *Leipzig Corpora Collection* (Quasthoff et al. 2014),[3] a list of 50 popular terms was created (for the full list, please see Appendix A). The recorded terms are associated with the *News*, *Web*, and *Wikipedia* domains. To ensure a fair comparison, both frequent Russian and Ukrainian words were recorded. Finally, for each term, region, and time frame (year), the data from google trends website was extracted. Eventually, for each region and for years from 2011 to 2021, we calculated the average proportion of searches in Ukrainian language. Formally, we are working with a quantity:

$$\text{proportion} \stackrel{\text{def}}{=} \frac{\text{percentage of searches in Ukrainian}}{\text{percentage of searches in Ukrainian} + \text{percentage of searches in Russian}}. \tag{1}$$

For the rest of the paper we refer to (1) as the *proportion*. In order to ensure that the sample of size 50 is indeed representative, we took several random samples (without replacement), of sizes 30 and 40 out of 50. Using these reduced data-sets, we performed a statistical analysis (see Sections 2.3, 2.4, and 3). The obtained results were similar to the ones presented in this manuscript.

2.2 Exploratory analysis

Figure 1 depicts the map of Ukraine divided to administrative districts (regions); the detailed region mapping is given in Table 1.

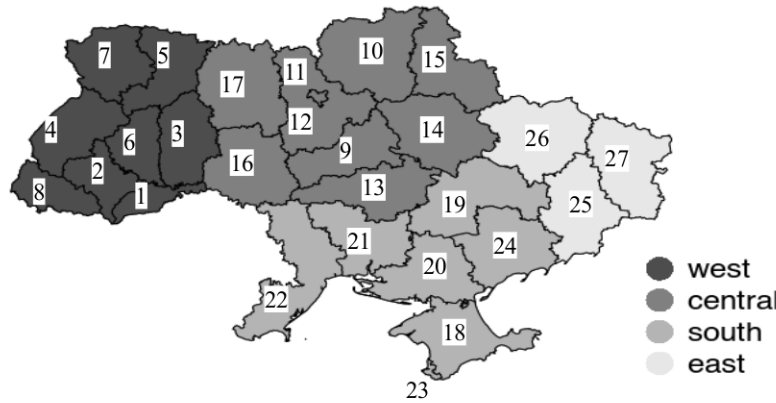


Figure 1: Ukraine map with regions divided into western, central, southern, and eastern regions. The west part of Ukraine consists of regions that are geographically close to “Western” counties, namely, to Poland, Slovakia, Hungary, and Roumania. The eastern part of Ukraine contains regions that are geographically close to Russian Federation.

Table 1: 27 Ukrainian administrative divisions; the Crimea* peninsula and the City of Sevastopol* were annexed in 2014 by Russian Federation; the regions of Donetsk△ and Luhansk△ are partially controlled by separatists.

region id	region name	admin. division	region id	region name	admin. division
1	Chernivtsi	West	15	Sumy	Centre
2	Ivano-Frankivsk	West	16	Vinnytsia	Centre
3	Khmelnytskyi	West	17	Zhytomyr	Centre
4	Lviv	West	18	Crimea*	South
5	Rivne	West	19	Dnipropetrovsk	South
6	Ternopil	West	20	Kherson	South
7	Volyn	West	21	Mykolaiv	South
8	Zakarpattia	West	22	Odessa	South
9	Cherkasy	Centre	23	Sevastopol*	South
10	Chernihiv	Centre	24	Zaporizhzhia	South
11	Kyiv (city)	Centre	25	Donetsk△	East
12	Kyiv (region)	Centre	26	Kharkiv	East
13	Kirovohrad	Centre	27	Luhansk△	East
14	Poltava	Centre			

The average proportion of Ukrainian language usage for each region between 2011 to 2021 is depicted in Figure 2. While Figure 2 shows a positive dynamics for the Ukrainian language, it is important to rigorously investigate the phenomena using an appropriate regression analysis. With the view to modelling proportions and in order to allow a natural interpretation of the obtained results, we propose to utilise the *beta regression* model (Ferrari & Cribari-Neto 2004, Figueroa-Zúñiga et al. 2013), which is discussed in Section 2.3.

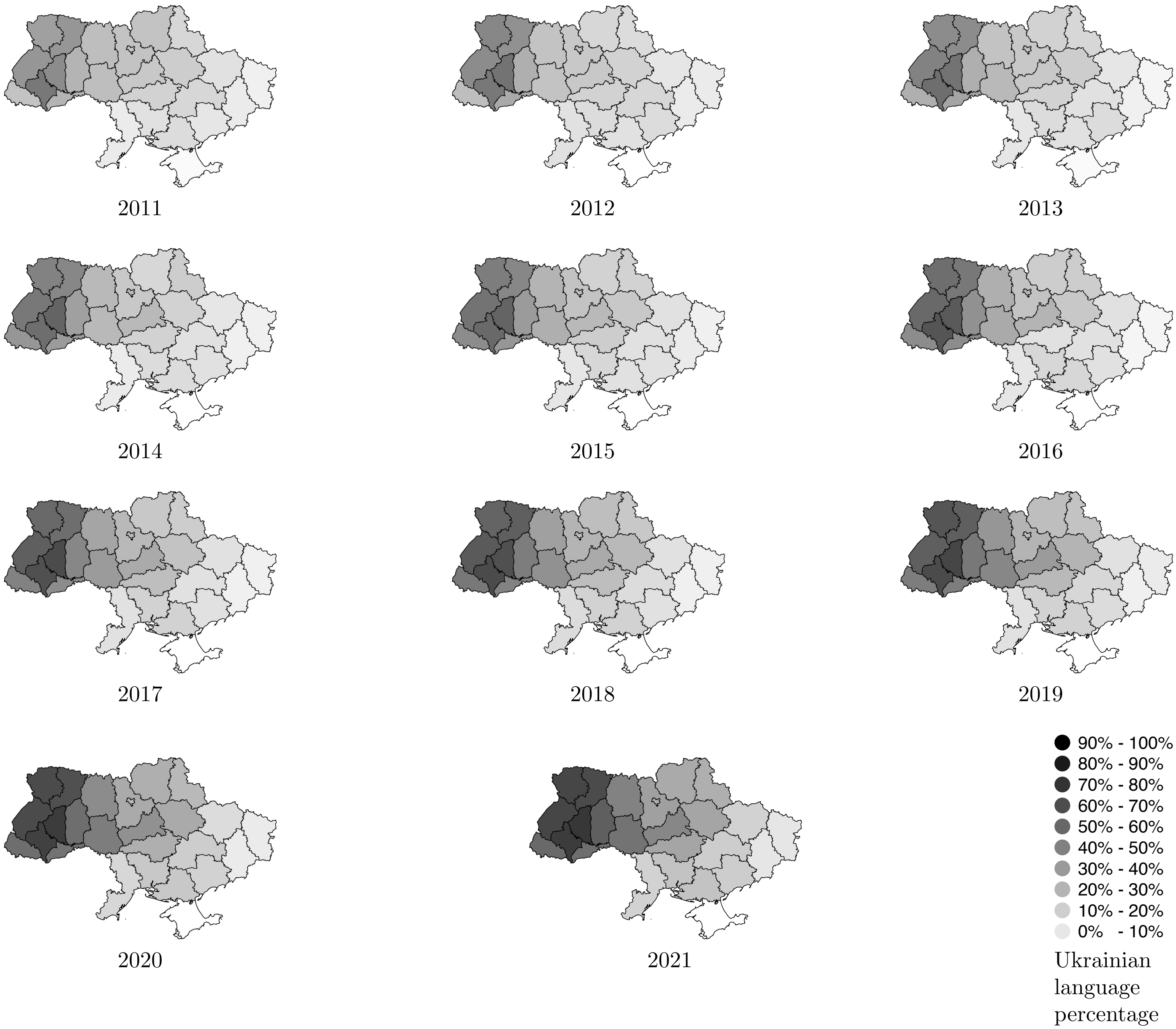


Figure 2: The dynamic of the proportion of the Ukrainian language used from 2011 to 2021.

2.3 The proposed beta regression model

A continuous random variable Y is said to have a Beta distribution if its probability density function is given by

$$f(y|\alpha,\beta) = \begin{cases} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{\mathcal{B}(\alpha,\beta)} & 0 \le y \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha > 0$, $\beta > 0$, and $\mathcal{B}(\alpha,\beta)$ is the *beta function* (Grimmett & Stirzaker 2001). For $Y \sim \mathsf{Beta}(\alpha,\beta)$, it holds that $\mathbb{E}Y = \frac{\alpha}{\alpha+\beta}$, and therefore, for convenience, we consider the *reparametrisation*: $\alpha = \mu\phi,$ and $\beta = (1-\mu)\phi$.

Under this reparametrisation, we write $Y \sim \mathsf{Beta}(\mu, \phi)$, and arrive at $\mathbb{E}Y = \mu$. We further assume that the Ukrainian language proportion in region $i \in \{1, \ldots, 27\}$ and year $j \in \{2011, \ldots, 2021\}$ is y_{ij}, and that $y_{ij} \sim \mathsf{Beta}(\mu_{ij}, \phi)$. In order to ensure that $\mu_{ij} \in [0, 1]$ and with the view to providing a natural interpretation via *odds ratios*, we utilise the *logit link* function and define:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \left(\beta_0 + \beta_0^{(i)}\right) + \left(\beta_1 + \beta_1^{(i)}\right)\ x_j \stackrel{\text{def}}{=} \eta_{ij},$$

where the effect β_0 is the intercept that characterises baseline state of proportion, and β_1 is the baseline rate of proportion growth. The regional effects $\beta_0^{(i)}$ and $\beta_1^{(i)}$ are associated with region i, and x_j is the covariate, specifically, x_j is a function of a year $j \in \{2011, \ldots, 2021\}$. Under the proposed model, the slopes $\left(\beta_1 + \beta_1^{(i)}\right)$, have an appealing interpretation as the change of log-odds that corresponds to a one unit increase in x_j, namely

$$\begin{aligned}\log\left(\frac{\mu_{ij+1}}{1-\mu_{ij+1}}\right) - \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) &= \eta_{ij+1} - \eta_{ij} = \\ &= \left(\beta_0 + \beta_0^{(i)}\right) + \left(\beta_1 + \beta_1^{(i)}\right)\ (x_j+1) - \left(\beta_0 + \beta_0^{(i)}\right) + \left(\beta_1 + \beta_1^{(i)}\right)\ x_j = \left(\beta_1 + \beta_1^{(i)}\right).\end{aligned}$$

For computational efficiency, we centre the year covariate and define $x_j = j - 2016$, where 2016 is the mean of the $\{2011, \ldots, 2021\}$ set. The available data size is not large, and thus we believe that the latter justifies the usage of the Bayesian approach. In addition, we aim to explore a general machinery for future research which might include a good prior knowledge about the model parameters. Specifically, we propose to use a Bayesian model which is defined via:

$$\begin{aligned}& y_{ij}|\mu_{ij}, \phi \sim \mathsf{Beta}(\mu_{ij}, \phi), \quad i \in \{1, \ldots, 27\},\ j \in \{2011, \ldots, 2021\}, \\ & \mu_{ij} = \frac{\mathrm{e}^{\eta_{ij}}}{1+\mathrm{e}^{\eta_{ij}}}, \quad i \in \{1, \ldots, 27\},\ j \in \{2011, \ldots, 2021\}, \\ & \phi \sim \mathsf{U}(0, 10^4), \beta_0, \beta_1, \beta_0^{(i)}, \beta_1^{(i)} \sim \mathsf{N}\left(0, \sigma^2\right), \quad i \in \{1, \ldots, 27\}.\end{aligned} \tag{2}$$

The proposed model can be potentially extended to include change-points (regime switching), with the view to specifying different behaviours of the proportion time series, and by specifying distinct parameters ϕ_{ij}, instead of the single parameter ϕ. Our experiments imply that the model in (2) (with $\sigma = 1$ parameter for the $\beta_0, \beta_1, \beta_0^{(i)}, \beta_1^{(i)}$ for $i \in \{1, \ldots, 27\}$ coefficients prior), fits the data well. A more detailed discussion regarding the choice of the σ parameter (prior sensitivity) and the goodness of fit, is provided in Sections 2.4.2 and 2.4.3, respectively. We proceed with the computational aspects.

2.4 Computation and Validation

2.4.1 Computation

Using *No U-Turn Sampling* scheme (Carpenter et al. 2017), we generated three chains with 10,000 iterations per chain, where the first 5,000 iterations were used as a warm-up. The 5,000 remaining samples for each chain were thinned by a factor of 10. Therefore, we had the total of 1,500 samples to perform inference. The Gelman-Rubin diagnostics (Brooks & Gelman 1998, Gelman & Rubin 1992) shows good convergence characteristics with the corresponding statistic values around 1.0 for all model parameters.

Convergence diagnostics

The MCMC sampler shows good convergence results. Figure 3 shows the convergence of the Gelman-Rubin statistic (Brooks & Gelman 1998, Gelman & Rubin 1992) of three independent MCMC runs of the No U-Turn sampler and Figure 4 depicts a graphical summary. The first, the second, and the third column of Figure 3, correspond to trace, sample auto correlation function, and density plots, respectively. The first, the second, and the third row of Figure 3, correspond to the first, the second and the third independent MCMC run, respectively. For additional typical convergence results, please see Appendix B.

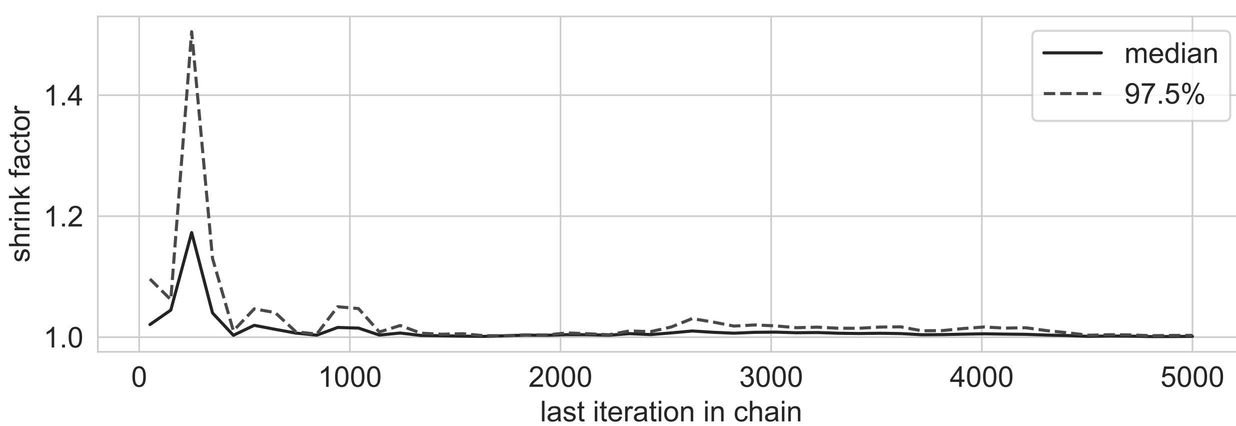


Figure 3: Gelman-Rubin diagnostic for parameter ϕ.

2.4.2 Prior sensitivity analysis for models with same parameter vector

Note that since our attention is restricted to models with priors $\phi \sim \mathsf{U}(0, 10^4)$, and $\beta_0, \beta_1, \beta_0^{(i)}, \beta_1^{(i)} \sim \mathsf{N}(0, \sigma^2)$ for $i \in \{1, \ldots, 27\}$, for different values of σ, all competitive models have the same parameter vector $\boldsymbol{\theta} = \left(\phi, \beta_0, \beta_1, \beta_0^{(1)}, \ldots, \beta_0^{(27)}, \beta_1^{(1)}, \ldots, \beta_1^{(27)}\right)$. Let M_i and M_j be two models that have the same parameter vector.

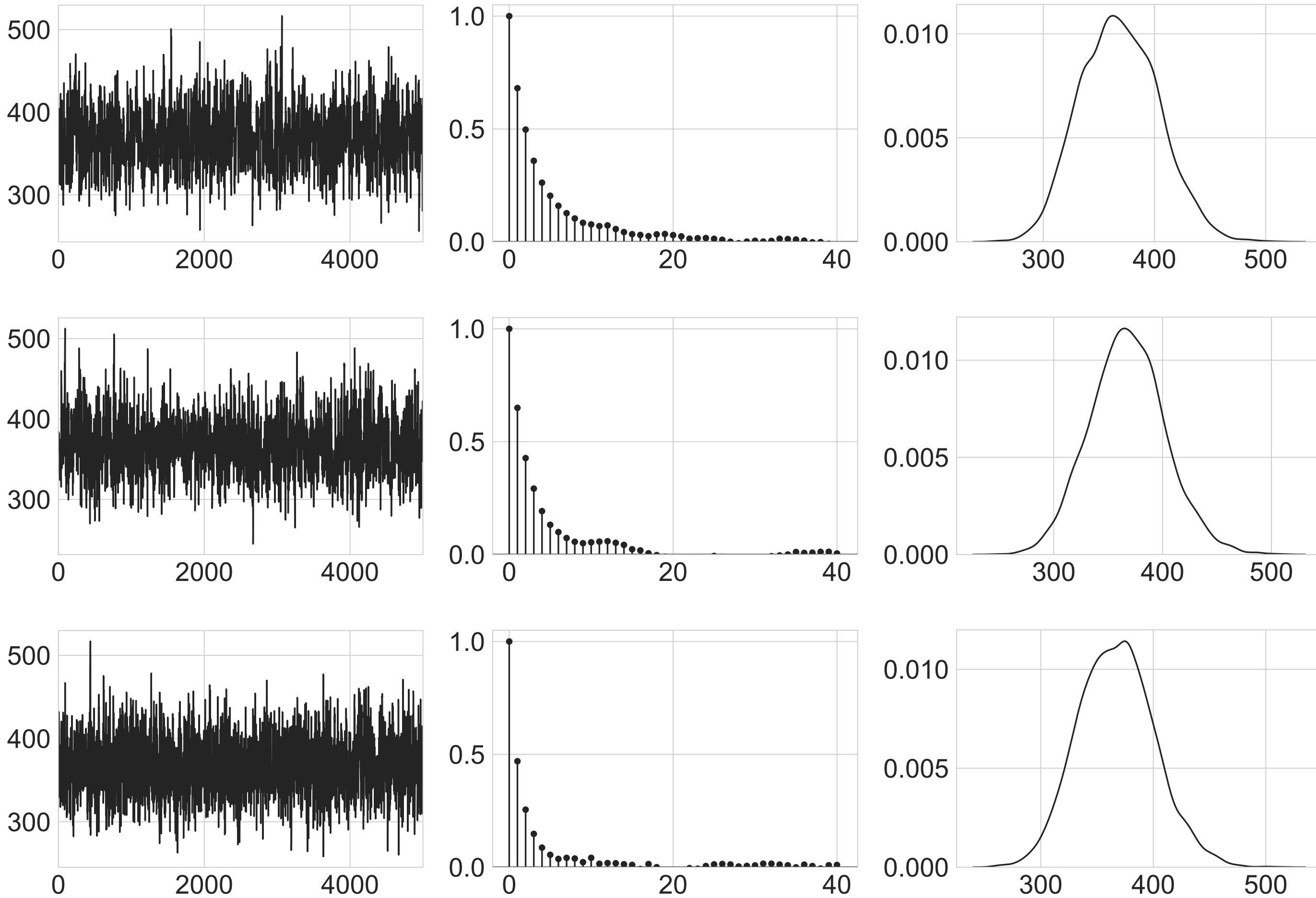


Figure 4: Summary of three Markov Chain Monte Carlo runs for the parameter ϕ.

Then, the Bayesian Factors (BF) for models M_i and M_j, is given by:

$$\mathrm{BF}_{ij} \stackrel{\text{def}}{=} \frac{p(\mathbf{y}|M_i)}{p(\mathbf{y}|M_j)}.$$

Under the *same parameter vector assumption*, it holds that

$$\begin{aligned} p(\mathbf{y}|M_i) &= \int_\Theta p(\mathbf{y}|\boldsymbol{\theta}, M_i)p(\boldsymbol{\theta}|M_i)d\boldsymbol{\theta} = \int_\Theta p(\mathbf{y}|\boldsymbol{\theta}, M_i)p(\boldsymbol{\theta}|M_i)\frac{p(\boldsymbol{\theta}|\mathbf{y}, M_j)}{p(\boldsymbol{\theta}|\mathbf{y}, M_j)}d\boldsymbol{\theta} \\ &= \int_\Theta \frac{p(\mathbf{y}|\boldsymbol{\theta}, M_i)p(\boldsymbol{\theta}|M_i)}{p(\boldsymbol{\theta}|\mathbf{y}, M_j)}p(\boldsymbol{\theta}|\mathbf{y}, M_j)d\boldsymbol{\theta} = \int_\Theta \frac{p(\mathbf{y}|\boldsymbol{\theta}, M_i)p(\boldsymbol{\theta}|M_i)}{p(\mathbf{y}|\boldsymbol{\theta}, M_j)p(\boldsymbol{\theta}|M_j)p(\mathbf{y}|M_j)^{-1}}p(\boldsymbol{\theta}|\mathbf{y}, M_j)d\boldsymbol{\theta} \\ &= p(\mathbf{y}|M_j)\int_\Theta \frac{p(\mathbf{y}|\boldsymbol{\theta}, M_i)p(\boldsymbol{\theta}|M_i)}{p(\mathbf{y}|\boldsymbol{\theta}, M_j)p(\boldsymbol{\theta}|M_j)}p(\boldsymbol{\theta}|\mathbf{y}, M_j)d\boldsymbol{\theta}. \end{aligned}$$

Therefore,

$$\mathrm{BF}_{ij} = \frac{p(\mathbf{y}|M_i)}{p(\mathbf{y}|M_j)} = \int_\Theta \frac{p(\mathbf{y}|\boldsymbol{\theta}, M_i)p(\boldsymbol{\theta}|M_i)}{p(\mathbf{y}|\boldsymbol{\theta}, M_j)p(\boldsymbol{\theta}|M_j)}p(\boldsymbol{\theta}|\mathbf{y}, M_j)d\boldsymbol{\theta}. \tag{3}$$

Suppose further that for any two models, M_i and M_j and for any $\boldsymbol{\theta}$, it holds that

$$p(\mathbf{y}|\boldsymbol{\theta}, M_i) = p(\mathbf{y}|\boldsymbol{\theta}, M_j).$$

Note that this condition corresponds to the beta regression model in the manuscript, since the only difference between two models is the prior parameter σ. In this case, (3) simplifies to

$$\mathrm{BF}_{ij} = \int_{\Theta} \frac{p(\boldsymbol{\theta}|M_i)}{p(\boldsymbol{\theta}|M_j)} p(\boldsymbol{\theta}|\mathbf{y}, M_j) d\boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{y},M_j} \frac{p(\boldsymbol{\theta}|M_i)}{p(\boldsymbol{\theta}|M_j)}.$$

If one have an access to samples from the posterior distribution that corresponds to the M_j model, it is possible to compare M_j to any model M_i without even fitting the M_i model. Specifically, it holds that

$$\widehat{\mathrm{BF}}_{ij} = \frac{1}{N} \sum_{i=1}^{N} \frac{p(\boldsymbol{\theta}_i|M_i)}{p(\boldsymbol{\theta}_i|M_j)},$$

where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ are samples from the posterior distribution that corresponds to M_j.

For the proposed model in the manuscript, let M_{σ_1} and M_{σ_2} be two competitive models. Then, an estimator $\widehat{\mathrm{BF}}_{M_{\sigma_1}, M_{\sigma_2}}$ for $\mathrm{BF}_{M_{\sigma_1}, M_{\sigma_2}} \overset{\text{def}}{=} p(\mathbf{y}|M_{\sigma_1}) p(\mathbf{y}|M_{\sigma_2})^{-1}$, can be obtained via

$$\widehat{\mathrm{BF}}_{M_{\sigma_1}, M_{\sigma_2}} = \frac{1}{N} \sum_{i=1}^{N} \frac{p(\boldsymbol{\theta}_i|M_{\sigma_1})}{p(\boldsymbol{\theta}_i|M_{\sigma_2})},$$

where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N$ are samples from the posterior distribution that corresponds to M_{σ_2}, and $p(\boldsymbol{\theta}_i|M)$ is the joint *prior* distribution that corresponds to model M (Chan & Eisenstat 2015). We used posterior samples that correspond to two models, M_1 and M_5 to produce Figure 5. The left panel of Figure 5, shows the logarithm of Bayes factor ($\widehat{\mathrm{BF}}_{M_\sigma, M_5}$) as a function of σ and one can observe that the largest $\widehat{\mathrm{BF}}_{M_\sigma, M_5}$ is located around $\sigma = 1$. The right panel of Figure 5 depicts the logarithm of Bayes factor ($\widehat{\mathrm{BF}}_{M_\sigma, M_1}$) as a function of σ. Combining the observations from the left and the right panel with BF interpretation (see (Kass & Raftery 1995) for details), we conclude that M_1, namely $\sigma = 1$, constitutes an appropriate prior.

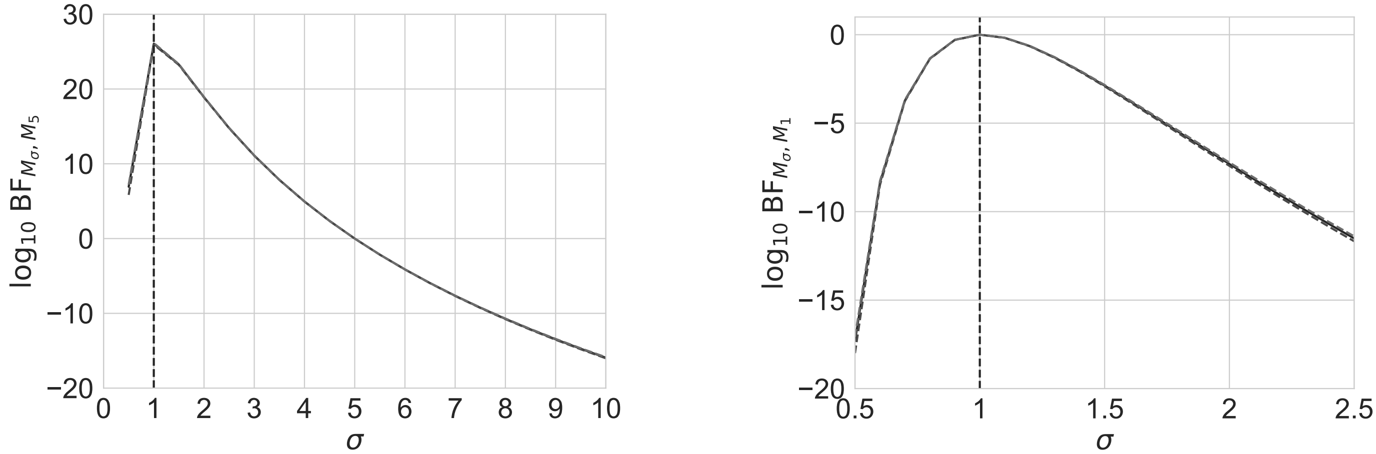


Figure 5: Logarithms of Bayes factors as a function of σ.

Remark 1 (A hierarchical model). As an alternative, it is possible to consider a *hierarchical model*:

$$
\begin{aligned}
& y_{ij}|\mu_{ij}, \phi \sim \mathsf{Beta}(\mu_{ij}, \phi), \quad i \in \{1, \ldots, 27\},\ j \in \{2011, \ldots, 2021\}, \\
& \mu_{ij} = \frac{\mathrm{e}^{\eta_{ij}}}{1 + \mathrm{e}^{\eta_{ij}}}, \quad i \in \{1, \ldots, 27\},\ j \in \{2011, \ldots, 2021\}, \\
& \beta_0, \beta_1, \beta_0^{(i)}, \beta_1^{(i)} \sim \mathsf{N}\left(0, \sigma^2\right), \quad i \in \{1, \ldots, 27\}. \\
& \phi \sim \mathsf{U}(0, 10^4), \\
& \sigma \sim \mathsf{U}(0, 100),
\end{aligned} \tag{4}
$$

The obtained results are similar to the ones reported in Section 3; please see Appendix C for additional details. For example, the posterior summary for ϕ, β_0, β_1, and σ parameters that are associated with the hierarchical model 7 provided in Table 6. Indeed, the σ parameter is around 1.0 as expected. Figure 23 shows the slopes and the intercept of the hierarchical model.

2.4.3 Goodness of fit

In order to validate that the proposed model fits the data, we use the extension of the classical χ^2 test proposed by Johnson (Johnson 2004). Johnson showed that if one draws parameters samples from the posterior distribution and evaluate the Parson's goodness-of-fit statistic at these values, then, regardless of the dimension of the parameter vector, the Parson's goodness-of-fit statistic is asymptotically distributed as a χ^2 random variable with $K - 1$ degrees of freedom.

Let $y_1, \ldots, y_n$ be a scalar-valued, continuous, identically distributed, conditionally independent observations drawn from probability density function $f(y|\boldsymbol{\theta})$ ($\boldsymbol{\theta}$ is a multidimensional parameter vector). Let $p(\boldsymbol{\theta}|\mathbf{y})$ be the posterior density of $\boldsymbol{\theta}$ based on the data $\mathbf{y}$ and let $\widetilde{\boldsymbol{\theta}}$ be a sample from the posterior distribution, namely, from $p(\boldsymbol{\theta}|\mathbf{y})$. The procedure of constructing the Bayesian χ^2 test for goodness of fit is as follows.

1. Let $0 = a_0 < a_1 < \cdots < a_{K-1} < a_K = 1$, be (user-predefined) quantiles from a uniform distribution. In addition, let $p_j \stackrel{\text{def}}{=} a_j - a_{j-1}$.

2. Define the vector $\mathbf{z}_j(\widetilde{\boldsymbol{\theta}})$ for $1 \le j \le K$ to be a K-length vector such that its jth element is 1 and all other elements are zero if

$$F(y_i|\widetilde{\boldsymbol{\theta}}) \in (a_{j-1}, a_j], \tag{5}$$

where F is the cumulative distribution function which corresponds to $f(y|\boldsymbol{\theta})$.

3. Using the definition in (5), let: $m(\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \mathbf{z}_i(\widetilde{\boldsymbol{\theta}})$. Essentially, the jth component of $m(\widetilde{\boldsymbol{\theta}})$ (let us call it $m_j(\widetilde{\boldsymbol{\theta}})$), is the number of observations that fell into the jth bin. Note that the bins are determined by the quantiles of the inverse distribution function evaluated at $\widetilde{\boldsymbol{\theta}}$.

4. Finally, we define

$$R^B(\widetilde{\boldsymbol{\theta}}) = \sum_{j=1}^{K} \frac{\left(m_j(\widetilde{\boldsymbol{\theta}}) - np_j\right)^2}{np_j}, \tag{6}$$

where np_j is the expected number of points that should lend in bin j.

Theorem 2.1. *Under some regularity conditions (Johnson 2004), R^B converges to the χ^2 distribution with $K-1$ degrees of freedom as $n \to \infty$.*

Under this setting, the *null hypothesis* is that there is no significant difference between the observed and the expected values.

Practical considerations: It was shown that the number of bins $K = \lceil n^{0.4} \rceil$ works well in practice. In principle, it is preferred to base the goodness-of-fit statistic on more than a single sampled value from the posterior distribution. This means that in practice we should aim to calculate an average with respect to samples from the posterior distribution, namely, we use

$$\widehat{R}^B = \frac{1}{N} \sum_{i=1}^{N} R^B(\widetilde{\boldsymbol{\theta}}_i),$$

where $\widetilde{\boldsymbol{\theta}}_i$ for $i \in \{1, \ldots, N\}$ are N samples from the posterior distribution.

In order to test the adequacy of the proposed model, we perform the Bayesian χ^2 goodness of fit test (Johnson 2004). Following the recommendation of Johnson (Johnson 2004), we define the number of bins $K = \lceil n^{0.4} \rceil$, where n is the sample size. In our case, there are 27 regions with 11 observations for each district and thus $K = 10$. The corresponding χ^2 test statistic estimator $\widehat{R}^B$ for goodness of fit, was calculated based on 1,500 posterior samples. For the $\sigma = 1$ model, the point estimator $\widehat{R}^B$ is 12.649 and the corresponding 95% confidence interval is $(12.734, 13.305)$. Since it holds that $\chi^2_{K-1,0.95} \approx 16.919$, we conclude that $\widehat{R}^B < \chi^2_{K-1,0.95}$, so this suggests that the proposed model indeed provides an adequate fit to the data.

3 Results

As mentioned in Section 2.4.1, the inference is based on 1,500 posterior samples. In this section, all the results are reported with respect to $\sigma = 1$ prior parameter. For the posterior distribution summary tables with respect to $\sigma = 5$ prior parameters, please refer to Appendix D. Tables 2, 3, and 4, show the full posterior distribution summaries associated with model (2) and with respect to $\sigma = 1$ prior parameter.

Table 2: Posterior distribution summaries for ϕ and baseline effects β_0 and β_1; the summary is with respect to $\sigma = 1$ prior parameter.

			Quantiles				
Parameter	Mean	Std. Dev.	0.025	0.25	0.5	0.75	0.975
ϕ	369.0	34.6	305.7	344.5	368.1	391.9	437.9
β_0	-1.218	0.182	-1.569	-1.342	-1.224	-1.097	-0.854
β_1	0.066	0.181	-0.289	-0.063	0.069	0.190	0.414

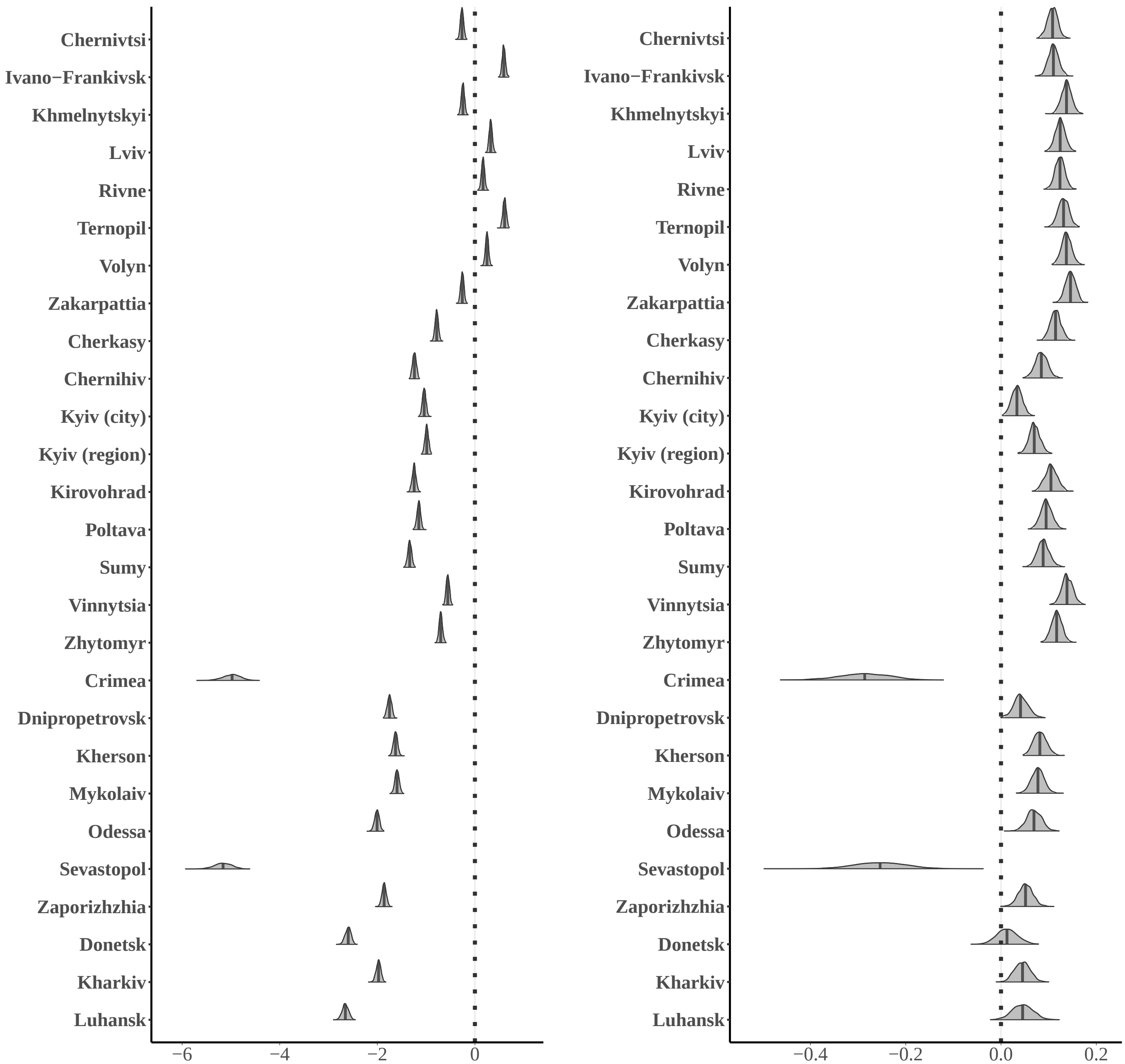
Table 3: Posterior distribution summaries for intercepts; the summary is with respect to $\sigma = 1$ prior parameter.

				Quantiles				
Region	Parameter	Mean	Std. Dev.	0.025	0.25	0.5	0.75	0.975
Chernivtsi	$\beta_0 + \beta_0^{(1)}$	-0.265	0.033	-0.328	-0.287	-0.265	-0.243	-0.2
Ivano-Frankivsk	$\beta_0 + \beta_0^{(2)}$	0.592	0.033	0.528	0.571	0.591	0.614	0.66
Khmelnytskyi	$\beta_0 + \beta_0^{(3)}$	-0.246	0.033	-0.31	-0.268	-0.245	-0.225	-0.183
Lviv	$\beta_0 + \beta_0^{(4)}$	0.324	0.032	0.263	0.302	0.323	0.345	0.387
Rivne	$\beta_0 + \beta_0^{(5)}$	0.166	0.032	0.103	0.146	0.167	0.188	0.23
Ternopil	$\beta_0 + \beta_0^{(6)}$	0.608	0.034	0.543	0.585	0.609	0.631	0.674
Volyn	$\beta_0 + \beta_0^{(7)}$	0.248	0.031	0.187	0.228	0.248	0.269	0.309
Zakarpattia	$\beta_0 + \beta_0^{(8)}$	-0.259	0.033	-0.322	-0.28	-0.258	-0.236	-0.196
Cherkasy	$\beta_0 + \beta_0^{(9)}$	-0.783	0.034	-0.85	-0.806	-0.783	-0.76	-0.716
Chernihiv	$\beta_0 + \beta_0^{(10)}$	-1.239	0.037	-1.31	-1.265	-1.239	-1.214	-1.168
Kyiv (city)	$\beta_0 + \beta_0^{(11)}$	-1.039	0.036	-1.107	-1.064	-1.039	-1.016	-0.972
Kyiv (region)	$\beta_0 + \beta_0^{(12)}$	-0.987	0.036	-1.056	-1.01	-0.987	-0.963	-0.917
Kirovohrad	$\beta_0 + \beta_0^{(13)}$	-1.246	0.039	-1.324	-1.271	-1.245	-1.221	-1.168
Poltava	$\beta_0 + \beta_0^{(14)}$	-1.15	0.037	-1.225	-1.174	-1.15	-1.126	-1.08
Sumy	$\beta_0 + \beta_0^{(15)}$	-1.336	0.039	-1.413	-1.361	-1.336	-1.309	-1.259
Vinnytsia	$\beta_0 + \beta_0^{(16)}$	-0.556	0.033	-0.618	-0.579	-0.556	-0.532	-0.492
Zhytomyr	$\beta_0 + \beta_0^{(17)}$	-0.7	0.034	-0.766	-0.722	-0.701	-0.679	-0.629
Crimea	$\beta_0 + \beta_0^{(18)}$	-4.986	0.18	-5.363	-5.102	-4.976	-4.865	-4.653
Dnipropetrovsk	$\beta_0 + \beta_0^{(19)}$	-1.751	0.044	-1.837	-1.78	-1.75	-1.72	-1.668
Kherson	$\beta_0 + \beta_0^{(20)}$	-1.627	0.043	-1.709	-1.655	-1.627	-1.599	-1.546
Mykolaiv	$\beta_0 + \beta_0^{(21)}$	-1.595	0.043	-1.68	-1.624	-1.596	-1.567	-1.511
Odessa	$\beta_0 + \beta_0^{(22)}$	-2.007	0.049	-2.104	-2.039	-2.005	-1.973	-1.912
Sevastopol	$\beta_0 + \beta_0^{(23)}$	-5.162	0.179	-5.519	-5.275	-5.16	-5.036	-4.825
Zaporizhzhia	$\beta_0 + \beta_0^{(24)}$	-1.861	0.045	-1.951	-1.89	-1.86	-1.831	-1.774
Donetsk	$\beta_0 + \beta_0^{(25)}$	-2.599	0.06	-2.719	-2.639	-2.596	-2.559	-2.489
Kharkiv	$\beta_0 + \beta_0^{(26)}$	-1.976	0.048	-2.069	-2.007	-1.974	-1.943	-1.887
Luhansk	$\beta_0 + \beta_0^{(27)}$	-2.654	0.065	-2.78	-2.695	-2.654	-2.61	-2.53

Table 4: Posterior distribution summaries for slopes; the summary is with respect to $\sigma = 1$ prior parameter.

				Quantiles				
Region	Parameter	Mean	Std. Dev.	0.025	0.25	0.5	0.75	0.975
Chernivtsi	$\beta_1 + \beta_1^{(1)}$	0.108	0.011	0.087	0.101	0.108	0.115	0.128
Ivano-Frankivsk	$\beta_1 + \beta_1^{(2)}$	0.11	0.011	0.089	0.103	0.11	0.117	0.132
Khmelnytskyi	$\beta_1 + \beta_1^{(3)}$	0.138	0.011	0.117	0.13	0.137	0.145	0.159
Lviv	$\beta_1 + \beta_1^{(4)}$	0.124	0.011	0.103	0.118	0.124	0.131	0.145
Rivne	$\beta_1 + \beta_1^{(5)}$	0.124	0.01	0.104	0.117	0.124	0.131	0.144
Ternopil	$\beta_1 + \beta_1^{(6)}$	0.131	0.011	0.11	0.124	0.131	0.139	0.154
Volyn	$\beta_1 + \beta_1^{(7)}$	0.137	0.011	0.117	0.13	0.137	0.145	0.159
Zakarpattia	$\beta_1 + \beta_1^{(8)}$	0.146	0.01	0.125	0.139	0.146	0.153	0.165
Cherkasy	$\beta_1 + \beta_1^{(9)}$	0.115	0.011	0.094	0.107	0.114	0.122	0.136
Chernihiv	$\beta_1 + \beta_1^{(10)}$	0.085	0.013	0.06	0.076	0.085	0.094	0.11
Kyiv (city)	$\beta_1 + \beta_1^{(11)}$	0.033	0.011	0.012	0.026	0.033	0.041	0.055
Kyiv (region)	$\beta_1 + \beta_1^{(12)}$	0.07	0.011	0.048	0.063	0.07	0.077	0.094
Kirovohrad	$\beta_1 + \beta_1^{(13)}$	0.105	0.013	0.08	0.097	0.105	0.114	0.131
Poltava	$\beta_1 + \beta_1^{(14)}$	0.095	0.012	0.072	0.087	0.095	0.103	0.119
Sumy	$\beta_1 + \beta_1^{(15)}$	0.089	0.013	0.065	0.08	0.088	0.097	0.114
Vinnytsia	$\beta_1 + \beta_1^{(16)}$	0.139	0.011	0.117	0.131	0.138	0.147	0.162
Zhytomyr	$\beta_1 + \beta_1^{(17)}$	0.117	0.011	0.096	0.11	0.117	0.124	0.138
Crimea	$\beta_1 + \beta_1^{(18)}$	-0.287	0.054	-0.396	-0.321	-0.286	-0.249	-0.186
Dnipropetrovsk	$\beta_1 + \beta_1^{(19)}$	0.042	0.015	0.011	0.032	0.041	0.052	0.073
Kherson	$\beta_1 + \beta_1^{(20)}$	0.082	0.014	0.057	0.072	0.082	0.091	0.11
Mykolaiv	$\beta_1 + \beta_1^{(21)}$	0.077	0.013	0.051	0.068	0.078	0.086	0.104
Odessa	$\beta_1 + \beta_1^{(22)}$	0.07	0.016	0.039	0.059	0.069	0.081	0.1
Sevastopol	$\beta_1 + \beta_1^{(23)}$	-0.254	0.054	-0.361	-0.291	-0.254	-0.218	-0.146
Zaporizhzhia	$\beta_1 + \beta_1^{(24)}$	0.052	0.015	0.022	0.042	0.051	0.062	0.081
Donetsk	$\beta_1 + \beta_1^{(25)}$	0.013	0.022	-0.028	-0.002	0.013	0.027	0.056
Kharkiv	$\beta_1 + \beta_1^{(26)}$	0.045	0.016	0.015	0.034	0.045	0.056	0.076
Luhansk	$\beta_1 + \beta_1^{(27)}$	0.046	0.022	0.003	0.031	0.045	0.06	0.088

Figure 6 depicts interval estimates from posterior draws associated with model (2).



(a) intercepts (b) slopes

Figure 6: Left panel shows the regional intercepts and right panel shows the regional slopes.

The left plot shows the intercepts. The intercept of the Western part of Ukraine are generally larger as compared to the eastern part thus indicating that the usage of Ukrainian language was always stronger in the Western part as compared to other regions. On the other hand, the intercepts of Crimea, Sevastopol, and eastern regions such as Luhansk and Donetsk, that have geographic (historical and cultural), proximity to Russian Federation, indicate a lower usage of Ukrainian language as compared to the Russian language.

While the left panel of Figure 6 aligns well with our intuition, it is even more instructive to consider the right panel of Figure 6. The latter corresponds to the regression slopes and thus shows the dynamics of the proportion, namely, the growth of the proportion of Ukrainian language usage. By observing the slopes, we can see that all the government controlled territories have positive slopes (the results are statistically significant, please see Table 4 for details). This indicates that the Ukrainization policy is indeed quite successful. Of course, we can also see that the slopes in Crimea and Sevastopol (two regions that were annexed by Russia in 2014), have negative slopes (this result is also statistically significant).

The mean and the corresponding 95% credible intervals (CI) for slopes in Crimea are -0.287 and (-0.396, -0.186), respectively. The situation in Sevastopol is similar, the mean and the corresponding 95% credible intervals (CI) for slopes are -0.254 and $(-0.361, -0.146)$, respectively. Namely, we see a strong Russification of the annexed regions. On the other hand, there are regions in Ukraine that show a rapid growth of the Ukrainian language usage. Two such regions are Zakarpattia and Khmelnytskyi, that introduce the slopes of 0.146, $(0.125, 0.165)$ and 0.138, $(0.117, 0.159)$, respectively. The results suggest that the effect of the Russification of Crimea is stronger than the corresponding Ukrainization effects in other Ukrainian regions.

The regional effect slopes in the Ukrainian territory, indicate that the western regions in Ukraine show the biggest increase in the proportion, which is not surprising. We acknowledge that Donetsk and Luhansk are partially controlled by separatists, so negative slopes are expected there. Surprisingly enough, the Luhansk region shows a slow growth with mean slope of 0.046 and 95% CI of $(0.003, 0.088)$. However, for the Donetsk region, we observe the mean slope of 0.013 and the corresponding 95% CI of $(-0.028, 0.056)$. The later might indicate that the situation in the Donetsk region is more radical as compared to the Luhansk region.

Overall, the obtained regression slopes indicate that the Ukrainization policy is working, although, the progress might be slower then reported by the Ukrainian sources. Using the posterior samples and the 2021 population estimates,[4] we were able to approximate the overall population percentage that perform Google searches in Ukrainian language. Our data shows that about 35% (the 95% credible interval is (32.56%, 37.48%), see Table 5), perform their search in Ukrainian.

The posterior summary of the number of people who perform search in Ukrainian language for each region is given in Table 5. From Table 5, we arrive to the conclusion that the percentage of the Ukrainian population that performs Ukrainian language searches is about 35%; this might not align well with the reported 78% who reported that they consider Ukrainian as their native language.

Nevertheless, our model predicts that the number of regions that will exceed 50% usage threshold of the Ukrainian language grows. Specifically, in 2021, only 9 regions (out of 27), exceeded the 50% threshold. However,

Table 5: Posterior summary of the number of people who perform Google search using Ukrainian language. The *mean UA search* column corresponds to the estimator of the number of people that search using the Ukrainian language.

Region	population estimate	mean UA search	95 Cred Interval
Chernivtsi	8.92×10^5	5.07×10^5	(4.81×10^5 , 5.33×10^5)
Ivano-Frankivsk	1.35×10^6	1.03×10^6	(9.95×10^5 , 1.06×10^6)
Khmelnytskyi	1.23×10^6	7.50×10^5	(7.14×10^5 , 7.84×10^5)
Lviv	2.48×10^6	1.79×10^6	(1.72×10^6 , 1.85×10^6)
Rivne	1.14×10^6	7.85×10^5	(7.57×10^5 , 8.12×10^5)
Ternopil	1.02×10^6	7.98×10^5	(7.74×10^5 , 8.23×10^5)
Volyn	1.02×10^6	7.35×10^5	(7.09×10^5 , 7.60×10^5)
Zakarpattia	1.25×10^6	7.67×10^5	(7.32×10^5 , 8.02×10^5)
Cherkasy	1.16×10^6	5.22×10^5	(4.88×10^5 , 5.57×10^5)
Chernihiv	9.63×10^5	2.96×10^5	(2.67×10^5 , 3.25×10^5)
Kyiv (city)	2.95×10^6	8.72×10^5	(7.94×10^5 , 9.51×10^5)
Kyiv (region)	1.80×10^6	6.22×10^5	(5.73×10^5 , 6.75×10^5)
Kirovohrad	9.08×10^5	2.97×10^5	(2.70×10^5 , 3.23×10^5)
Poltava	1.36×10^6	4.58×10^5	(4.21×10^5 , 4.99×10^5)
Sumy	1.04×10^6	3.02×10^5	(2.75×10^5 , 3.33×10^5)
Vinnytsia	1.51×10^6	8.10×10^5	(7.65×10^5 , 8.54×10^5)
Zhytomyr	1.18×10^6	5.57×10^5	(5.24×10^5 , 5.93×10^5)
Crimea			
Dnipropetrovsk	3.11×10^6	5.49×10^5	(4.76×10^5 , 6.30×10^5)
Kherson	1.01×10^6	2.30×10^5	(2.04×10^5 , 2.57×10^5)
Mykolaiv	1.10×10^6	2.52×10^5	(2.24×10^5 , 2.80×10^5)
Odessa	2.36×10^6	3.77×10^5	(3.27×10^5 , 4.31×10^5)
Sevastopol			
Zaporizhzhia	1.65×10^6	2.76×10^5	(2.39×10^5 , 3.14×10^5)
Donetsk	4.07×10^6	3.01×10^5	(2.36×10^5 , 3.74×10^5)
Kharkiv	2.61×10^6	3.86×10^5	(3.31×10^5 , 4.43×10^5)
Luhansk	2.11×10^6	1.72×10^5	(1.36×10^5 , 2.11×10^5)
Summary	4.13×10^7	1.44×10^7	(1.34×10^7 , 1.55×10^7)
Percentage	100.00 %	34.98 %	(32.56 % , 37.48 %)

in 2026 and 2031, we predict that 11 and 15 regions will exceed this threshold. The later is an indication of the overall success of the Ukrainization policy, at least with respect to the Ukrainian language usage.

The full-scale Russian invasion in 2022, requires an additional validation of the proposed model. In Section 4 we examine the model and the consequences of the recent events on the Ukrainian language popularisation. We show that the proposed model is still valid subject to an introduction of a change point which occurs in 2022.

4 The 2022 Russian invasion

First, we examine how well the current 2011-2021 model fits the data when including the results from 2022. In this case, there are again 27 regions with 12 observations for each district and thus $K = 11$. The corresponding χ^2 test statistic estimator $\widehat{R}^B$ for goodness of fit, was calculated based on 1,500 posterior samples. For the $\sigma = 1$ model, the point estimator $\widehat{R}^B$ is 19.921 and the corresponding 95% confidence interval is $(19.570, 20.271)$. Since it holds that $\chi^2_{K-1,0.95} \approx 18.307$, we conclude that $\widehat{R}^B > \chi^2_{K-1,0.95}$, so this suggests that the proposed model is not adequate for the 2022 data. While the χ^2 test statistic is important from the mathematical point of view, Figure 7, which depicts the 2022 data and the 2011-2021 model prediction intervals for every region is quite instructive. It is interesting to note that the 2011-2021 model is adequate for the Western regions. Nevertheless, the majority of data points for the central, south, and eastern regions, are above the prediction intervals. This might be due to the fact that the Western regions were less exposed to military actions.

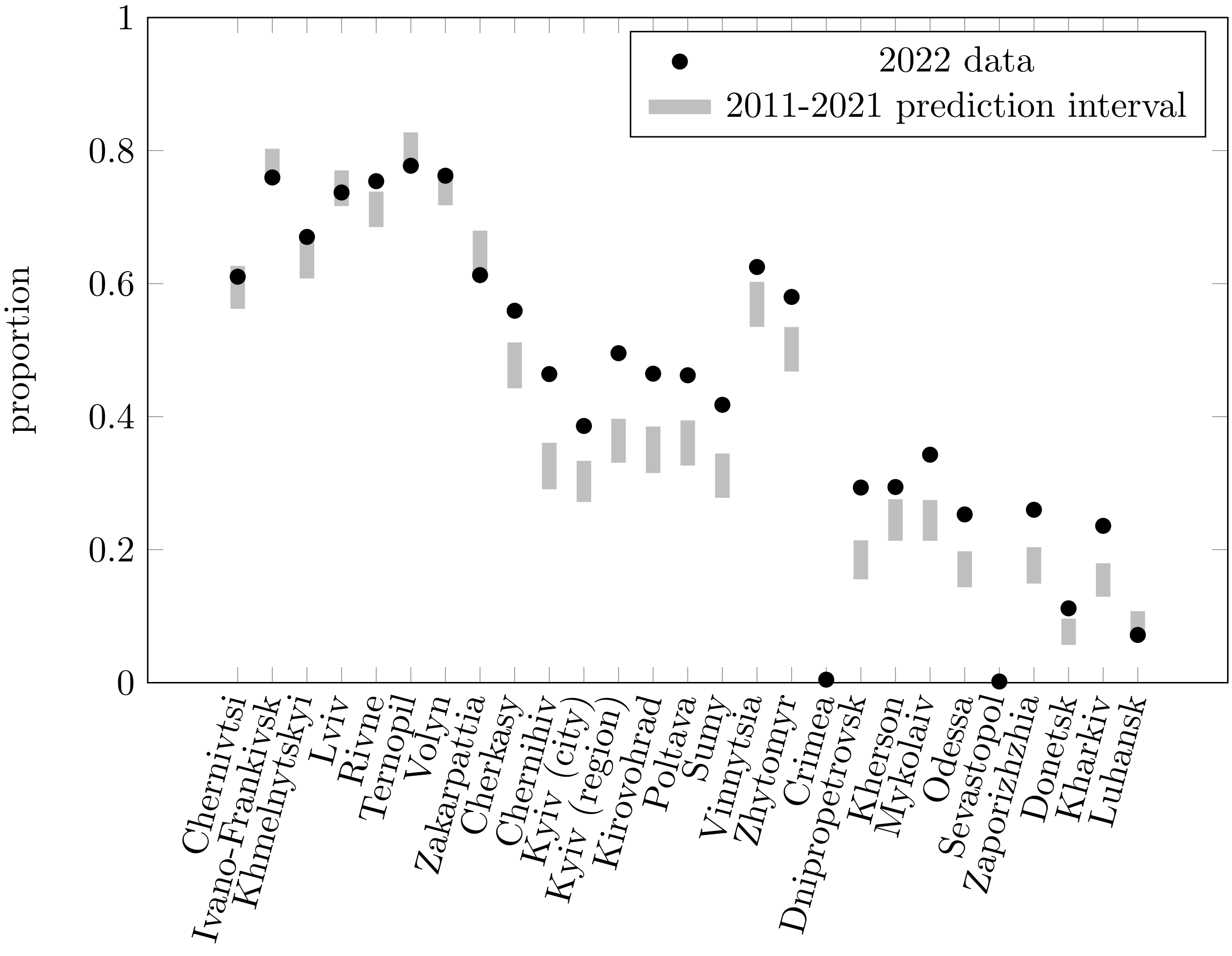


Figure 7: The 2022 data and the 2011-2021 model 95% prediction intervals for all regions.

The above finding are not very surprising and are basically supported by recent studies of Rating (2022) and Kulyk (2022*a*). As noted by Kulyk (2022*a*), many Ukrainians tend to blame the Russian population for the war and the associated crimes of the occupying forces. Specifically, according to Kulyk's study, Russian is now considered to be *the language of the enemy* and thus, many Ukrainians refuse to utilise Russian and instead try to use Ukrainian as it is now considered to be *the language of the resistance*. Our findings support this claim.

The obtained results and in particular, a careful observation of Figure 7, indicates that the 2011-2021 model should be adjusted. We propose to extend the original model by introducing a *change point* (Rizzo 2019, Chapter 11). The extension of the (2) is as follows.

$$
\begin{aligned}
y_{ij} &\sim \begin{cases} \mathsf{Beta}(\mu_{ij}, \phi_1) & i \in \{1, \dots, 27\},\ j \in \{2011, \dots, 2021\} \\ \mathsf{Beta}(\mu'_{ij}, \phi_2) & i \in \{1, \dots, 27\},\ j = 2022 \end{cases} \qquad (7) \\
\mu_{ij} &= \frac{\mathrm{e}^{\eta_{ij}}}{1 + \mathrm{e}^{\eta_{ij}}}, \quad i \in \{1, \dots, 27\},\ j \in \{2011, \dots, 2021, 2022\}, \\
\mu'_{ij} &= \frac{\mathrm{e}^{\eta'_{ij}}}{1 + \mathrm{e}^{\eta'_{ij}}}, \quad i \in \{1, \dots, 27\},\ j = 2022, \\
&\phi_1, \phi_2 \sim \mathsf{U}(0, 10^4), \\
&\beta_0, \beta_1, \beta_0^{(i)}, \beta_1^{(i)}, \beta_0^{\prime(i)}, \beta_1^{\prime(i)} \sim \mathsf{N}\left(0, \sigma^2\right), \quad i \in \{1, \dots, 27\}.
\end{aligned}
$$

Here, similarly to the original model, we define:

$$
\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \eta_{ij} = \beta_0 + \beta_0^{(i)} + (\beta_1 + \beta_1^{(i)})\, x_j \text{ and } \log\left(\frac{\mu'_{ij}}{1 - \mu_{ij}}\right) = \eta'_{ij} = \beta_0 + \beta_0^{\prime(i)} + (\beta_1 + \beta_1^{\prime(i)})\, x_j.
$$

Now, for the new 2011-2022 ($\sigma = 1$) model, the point estimator $\widehat{R}^B$ is 15.348 and the corresponding 95% confidence interval is $(14.180, 16.254)$. It holds that $\chi^2_{K-1,0.95} \approx 18.307$, and we conclude that $\widehat{R}^B < \chi^2_{K-1,0.95}$, so this suggests that the proposed model fits the data well. Figure 8 shows a comparison of slopes for the original 2011-2021 model and the new 2011-2022 model.

Figure 8 is very instructive. The majority of slopes in the 2011-2022 model are much higher when the corresponding slopes in the 2011-2021 model. This further indicates that the Russian invasion contributes to the development and the acceptance of the Ukrainian language and supports the findings of Rating (2022) and Kulyk (2022*a*). In Crimea and Sevastopol, the change in the slope is actually negative, namely, from -0.2871 to -0.7409 in Crimea and from -0.2537 to -0.9045 in Sevastopol.

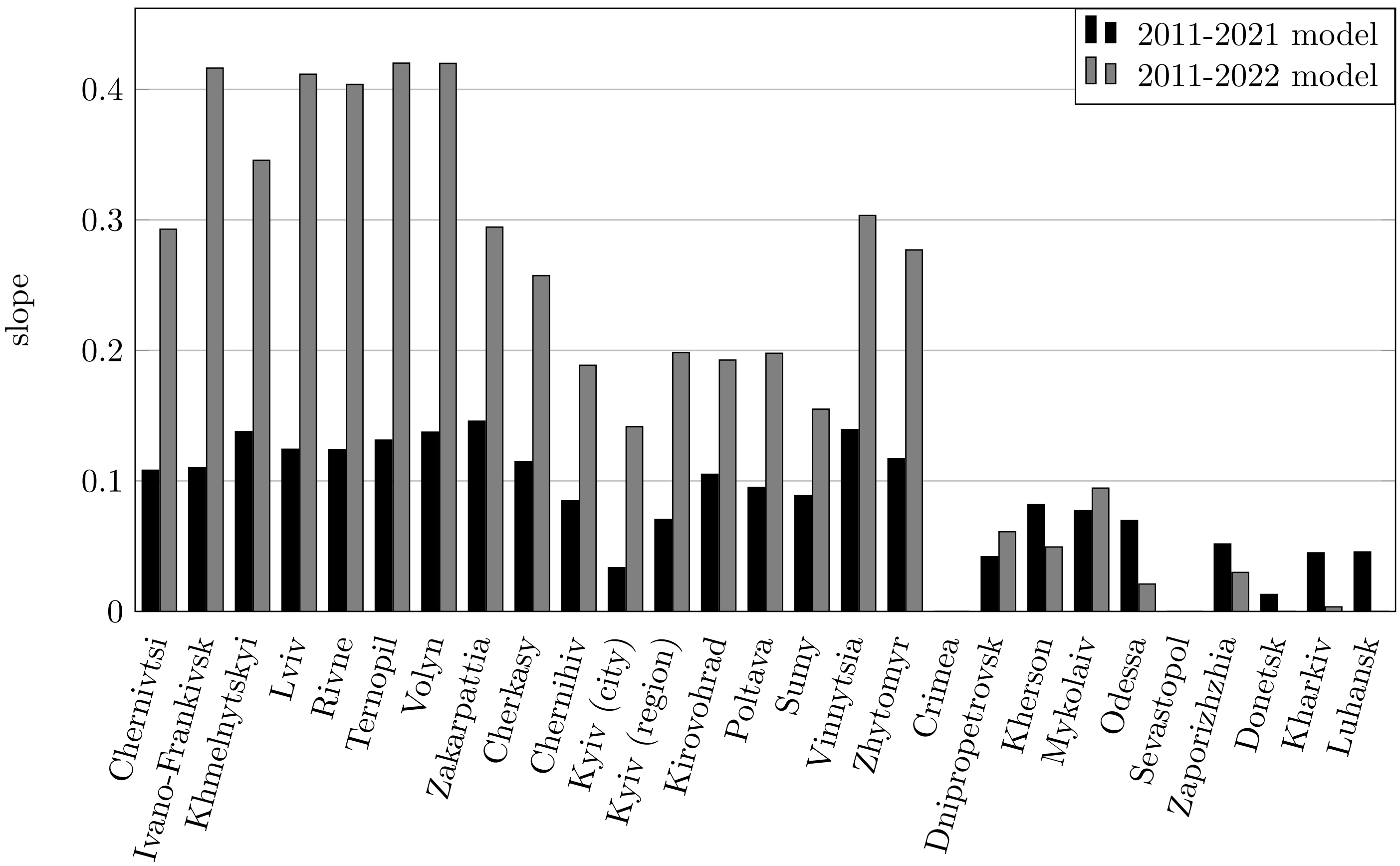
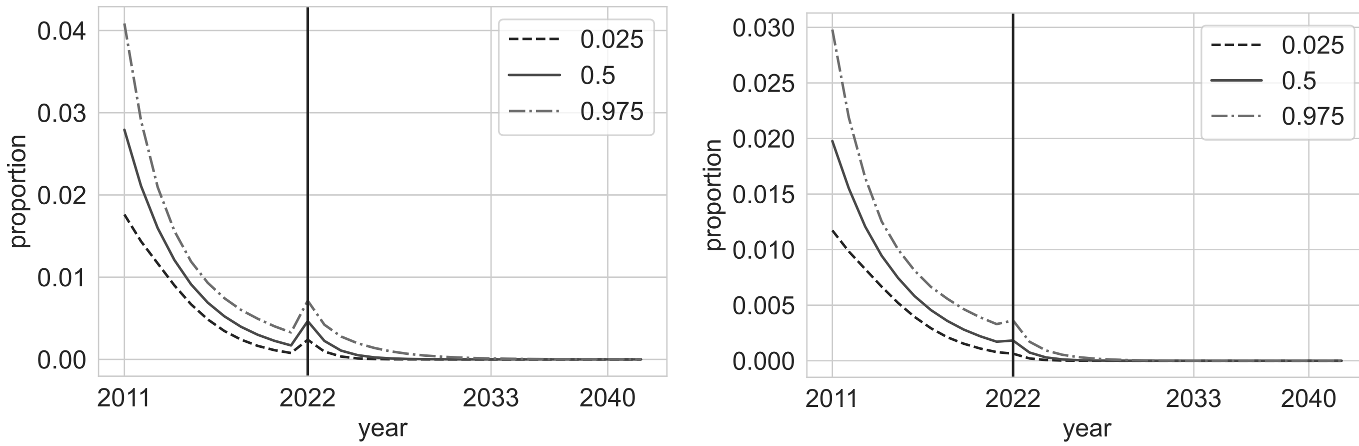


Figure 8: The 2011-2021 vs 2011-2022 models slope comparison.

5 Discussion and conclusion

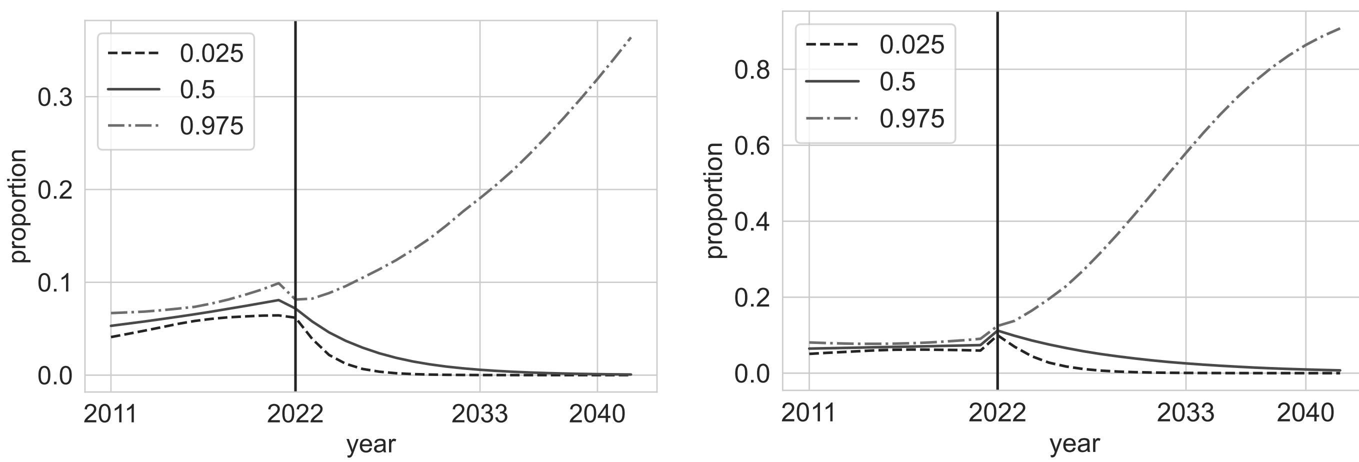
In this paper, we have demonstrated that google trend data combined with spatial information, can provide important insights into language dissemination trends. The proposed beta regression model fits the data well and is able to explain spatial variations. It is important to note that the data is open and no costly experiments are required. The proposed model combined with google trends data, can potentially serve as a verification mechanism to language poll experiments.

Under our model, the prediction of the proportion for the forthcoming years is straightforward. Using the posterior samples associated with the model, the prediction and confidence intervals of the proportion for region i and year $j \geq 2023$ can be derived from (7). From Figure 9, which shows the prediction of the Ukrainian language usage proportion in the annexed Crimea and Sevastopol, we conclude that the situation is quite distressing for the Ukrainian language.



(a) Crimea (b) Sevastopol

Figure 9: Crimea and Sevastopol; proportion of Ukrainian language usage prediction until 2040. The graph shows the 0.025, 0.5, and the 0.975 quantiles of the proportion.



(a) Luhansk (b) Donetsk

Figure 10: Donetsk and Luhansk; proportion of Ukrainian language usage prediction until 2040. The graph shows the 0.025, 0.5, and the 0.975 quantiles of the proportion.

This is also very important to consider the situation in partly controlled territories of Donetsk and Luhansk. The prediction of the proportion for Donetsk and Luhansk is depicted in Figure 10. Our analysis indicates that the situation in both regions looks dreadful for the Ukrainian language. The prediction of the Ukrainian language proportion usage for all regions is depicted in Figure 11 and Figure 12.

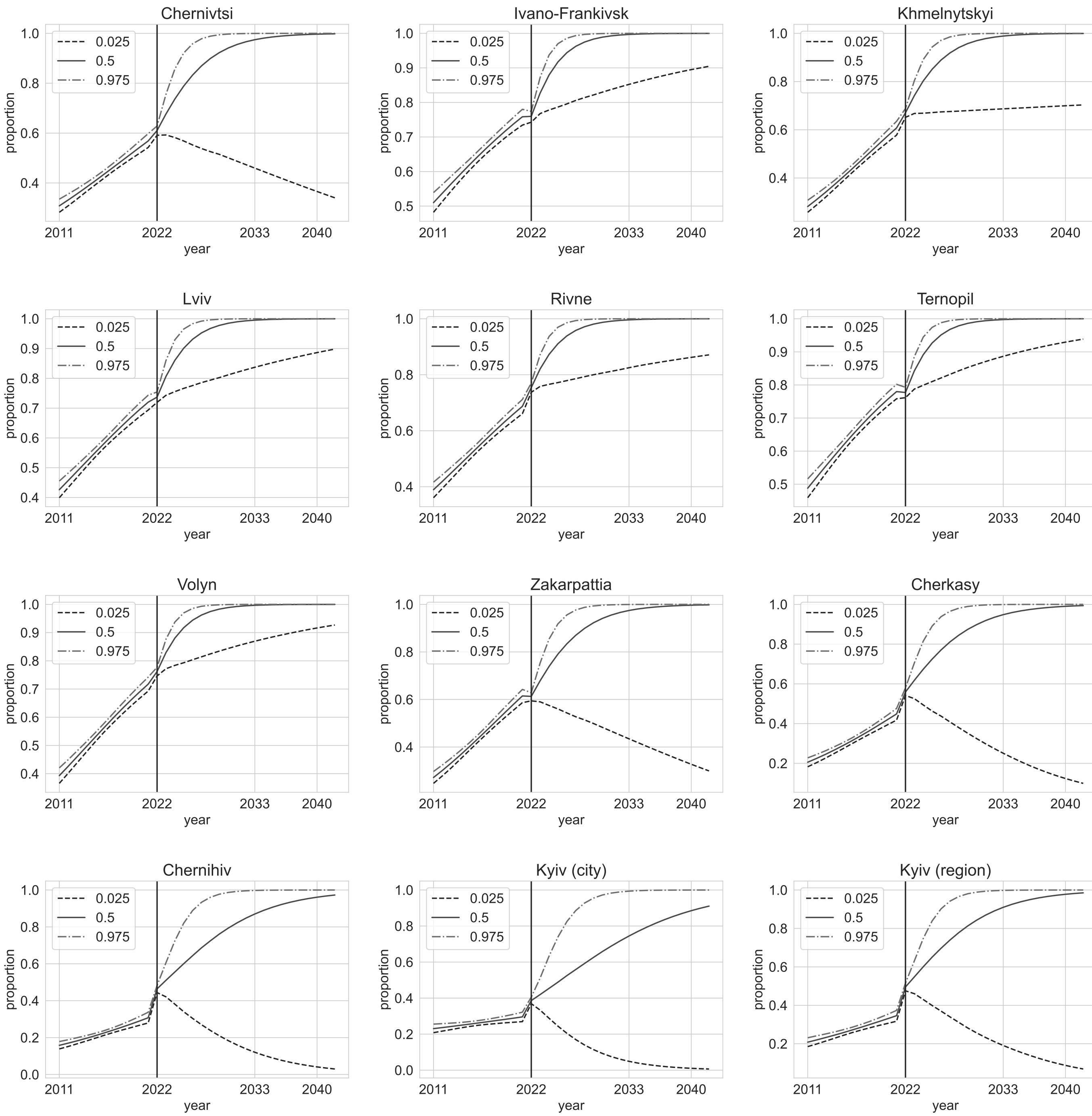


Figure 11: Prediction of Ukrainian language usage proportion until year 2040 (part 1)

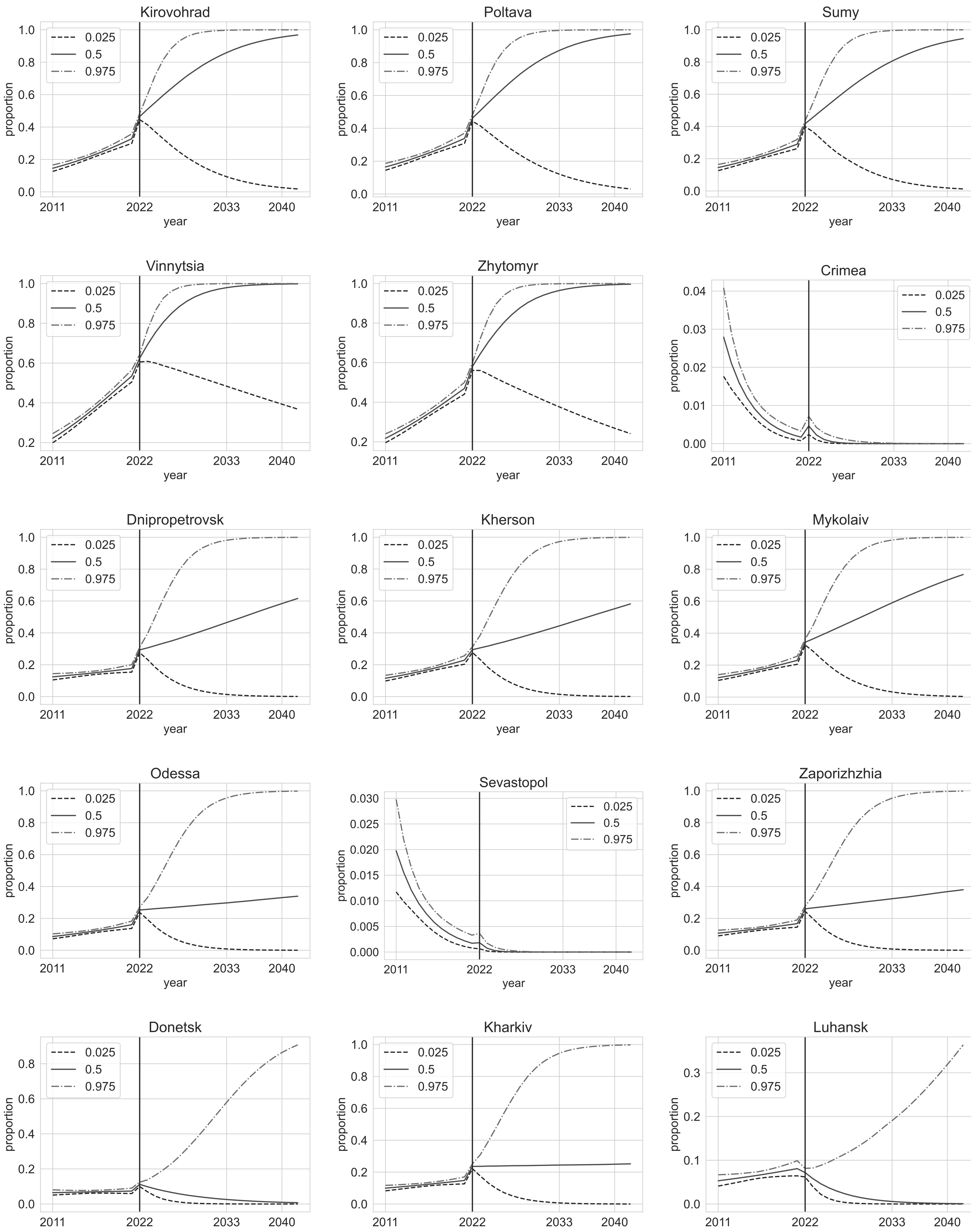


Figure 12: Prediction of Ukrainian language usage proportion until year 2040 (part 2)

There are several limitations of this study that should be discussed. First, due to the geopolitical situation in Donetsk and Luhansk regions, we cannot distinguish between the territories that are under control of the Ukrainian government or under control of separatists. That is, the data is combined for these two regions and therefore we cannot observe the corresponding effect directly. This is also important to consider the effect of the Russian invasion of Ukraine which started on 24 February 2022. Specifically, one should keep in mind that many people died or fled these regions and a considerable number of cities and villages were devastated. Moreover, if we consider the annexed regions of Crimea and the city of Sevastopol, there might be an additional effect which is related to service availability. Namely, an individual located in Crimea and seeking say government assistance, will need to use the Russian language. However, this logic also applies to the territories that are under the control of the Ukrainian government. An additional limitation of this study is the web content availability. Specifically, there exists more content in Russian language. Furthermore, by using the open google trend data, we cannot distinguish between individuals with respect to say age, education, etc. However, this work can be potentially extended by designing appropriate experiments with human subjects; the statistical machinery will remain almost identical. Finally, it is important to note that many Ukrainian citizens use Surzhyk (Hentschel & Palinska 2022, Hentschel & Taranenko 2021), a mixed language that contains both Russian and Ukrainian words. The proposed method can not distinguish between pure Ukrainian, pure Russian, and Surzhyk speakers. However, the method can provide an evidence regarding the proportion of Ukrainian and Russian words used.

Despite the above limitations, this work demonstrates the value of the spatial google trend language data availability. Moreover, it lays a foundation for various extensions and future work. For example, starting from February 2022, additional queries and keywords might become popular. It will be of interest to develop a model that both considers the available 2011-2022 data, and, takes into account the new set of war-related queries. One possible direction is to consider the Beta rectangular distribution link function as suggested by Bayes et al. (Bayes et al. 2012), since it can provide a more robust modelling of proportions with respect to outliers. Applying such a model will be increasingly important as time passes, and additional data from google trends becomes available. To conclude, we conjecture that the application of the proposed model is of great value since it can help to examine the effectiveness of government policies with respect to language dissemination.

Acknowledgments

We are thoroughly grateful to the editor and to anonymous reviewers for their valuable and constructive remarks and suggestions.

Notes

[1]https://www.globalfirepower.com

[2]Poland, Slovakia, Hungary, and Roumania

[3]https://wortschatz.uni-leipzig.de/en

[4]http://database.ukrcensus.gov.ua/PXWEB2007/eng/news/op_popul_e.asp

References

Arel, D. (2002), 'Interpreting "Nationality" and "Language" in the 2001 Ukrainian Census', *Post-Soviet affairs* **18**(3), 213–249.

Bayes, C. L., Bazán, J. L. & García, C. (2012), 'A new robust regression model for proportions', *Bayesian analysis* **7**(4), 841–866.

Brooks, S. P. & Gelman, A. (1998), 'General methods for monitoring convergence of iterative simulations', *Journal of computational and graphical statistics* **7**(4), 434–455.

Bureiko, N. & Moga, T. L. (2018), Bounded Europeanisation: the case of Ukraine, *in* 'The European Union and its eastern neighbourhood', Manchester University Press, pp. 71–85.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017), 'Stan: A probabilistic programming language', *Journal of statistical software* **76**(1), 1–32.

Centr Razumkova (2016), 'Consolidation of Ukrainian society: challenges, opportunities, pathways'.
URL: *https://razumkov.org.ua/uploads/journal/eng/NSD165-166_2016_eng.pdf*

Chan, J. C. C. & Eisenstat, E. (2015), 'Marginal likelihood estimation with the cross-entropy method', *Econometric reviews* **34**(3), 256–285.

Decker, P. K. (2021), '"We Show What Is Concealed": Russian Soft Power in Germany', *Problems of post-communism* **68**(3), 216–230.

Elder, M. (2012), 'Ukrainians protest against Russian language law', *The Guardian* .
URL: *https://www.theguardian.com/world/2012/jul/04/ukrainians-protest-russian-language-law*

Feklyunina, V. (2016), 'Soft power and identity: Russia, Ukraine and the 'Russian world(s)'', *European journal of international relations* **22**(4), 773–796.

Ferrari, S. & Cribari-Neto, F. (2004), 'Beta regression for modelling rates and proportions', *Journal of applied statistics* **31**(7), 799–815.

Figueroa-Zúñiga, J. I., Arellano-Valle, R. B. & Ferrari, S. (2013), 'Mixed beta regression: A Bayesian perspective', *Computational statistics & data analysis* **61**, 137–147.

Flier, M. S. & Graziosi, A. (2017-2018), 'The Battle for Ukrainian: a Comparative Perspective', *Harvard Ukrainian Studies* **35**(1/4).

Forsberg, T. & Smith, H. (2016), 'Russian cultural statecraft in the eurasian space', *Problems of post-communism* **63**(3), 129–134.

Gelman, A. & Rubin, D. B. (1992), 'Inference from iterative simulation using multiple sequences', *Statistical science* **7**(4), 457–472.

Gladkova, A. (2015), 'Linguistic and cultural diversity in Russian cyberspace: examining four ethnic groups online', *Journal of multicultural discourses* **10**(1), 49–66.

Grimmett, G. & Stirzaker, D. (2001), *Probability and random processes*, 3rd edn, Oxford university press, New York.

Hentschel, G. & Palinska, O. (2022), 'The linguistic situation on the Ukrainian Black Sea coast – Ukrainian, Russian and Suržyk as "native language", "primary code", frequently used codes and codes of linguistic socialization during childhood', **46**(3), 259–290.

Hentschel, G. & Taranenko, O. (2021), 'Bilingualism or tricodalism: Ukrainian, Russian and "Suržyk" in Ukraine. Analysis and linguistic-geographical mapping', **66**(2), 268–299.

Ilko Kucheriv Democratic Initiatives Foundation and Razumkov Centre (2021), 'The day of Ukrainian writing and language-2021: Is the use of the state language in the public sphere increasing?'.
URL: *https://dif.org.ua/article/the-day-of-ukrainian-writing-and-language-2021-is-the-use-of-the-state-language-in-the-public-sphere-increasing*

Johnson, V. E. (2004), 'A Bayesian chi(2) test for goodness-of-fit', *The Annals of statistics* **32**(6), 2361–2384.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Kiss, N. (2022), Key actors in the organized language management of ukraine: On the materials of language legislation development and adoption, *in* 'Interests and Power in Language Management', Peter Lang International Academic Publishing Group, Bern.

Kulyk, V. (2011), 'Language identity, linguistic diversity and political cleavages: evidence from Ukraine', *Nations and nationalism* **17**(3), 627–648.

Kulyk, V. (2022*a*), 'Die Sprache der Widerstands. Der Krieg und der Aufschwung des Ukrainischen', *Osteuropa* **6–8**(72), 237–248.

Kulyk, V. (2022*b*), 'Is Ukraine a Multiethnic Country?', *Slavic review* **81**(2), 299–323.

Latyshev, A., Onischuk, E. & Medvedeva, A. (2021), '"Too strong": why more than half of the population of Ukraine continues to actively use the Russian language', *Russia Today* pp. 1–1.
URL: *https://russian.rt.com/ussr/article/896925-ukraina-opros-yazyk*

Müller, D. & Wingender, M. (2021), *Language Politics, Language Situations and Conflicts in Multilingual Societies Case Studies from Contemporary Russia, Ukraine and Belarus*, Harrassowitz Verlag, Wiesbaden.

Ogarkova, T. (2018), 'The truth behind ukraine's language policy', *Atlantic Council* .

Poses, C. & Revilla, M. (2021), 'Measuring satisfaction with democracy: How good are different scales across countries and languages?', *European political science review* pp. 1–18.

Quasthoff, U., Goldhahn, D. & Eckart, T. (2014), Building large resources for text mining: The leipzig corpora collection, *in* 'Text Mining', Theory and Applications of Natural Language Processing, Springer International Publishing, Cham, pp. 3–24.

Rating (2022), 'Seventeenth national survey: Identity. Patriotism. Values'.
URL: *https://ratinggroup.ua/research/ukraine/s_mnadcyate_zagalnonac_onalne_opituvannya_dentichn_st_patr_otizm_c_nnost_17-18_serpnya_2022.html*

Reznik, V. (2018), Language Policy in Independent Ukraine: A Battle for National and Linguistic Empowerment, *in* 'Language Planning in the Post-Communist Era', Springer International Publishing, Cham, pp. 169–192.

Rizzo, M. L. (2019), *Statistical computing with R Maria L. Rizzo.*, Chapman & Hall/CRC The R Series, second edition. edn, CRC Press, Taylor & Francis Group, Boca Raton.

Toivanen, R. (2007), Linguistic diversity and the paradox of rights discourse, *in* C. L. Dario Castiglione, ed., 'The Language Question in Europe and Diverse Societies: Political, Legal and Social Perspectives', 1st edn, Oñati international series in law and society, Hart Publishing Ltd, London, chapter 6.

Ushchyna, V. (2020), 'Corinne A. Seals Choosing a mother tongue: The politics of language and identity in Ukraine. Bristol: Multilingual Matters, 2019. Pp. 213.', *Language in Society* **49**(3), 491–492.

Zeller, J. (2021), ‘The geographical and social distribution of native languages in central ukraine’, *Linguistica Copernicana* **18**, 105–136.

Appendices

A Google trends search terms

English	Ukrainian	Russian	English	Ukrainian	Russian
activity	діяльність	деятельность	news	новини	новости
after	після	после	now	зараз	теперь
as	як	как	number	кількість	количество
attention	увага	внимание	only	тільки	только
baby	дитячий	детский	others	інших	других
buy	купити	купить	parents	батьки	родители
child	дитина	ребенок	part	частина	часть
children	діти	дети	person	людина	человек
choice	вибір	выбор	place	місце	место
city	місто	город	present	подарунок	подарок
different	різних	разных	question	питання	вопрос
doctor	лікар	врач	Russia	Росія	Россия
even	навіть	даже	seeking	шукаю	ищу
has/have	має	имеет	several	кілька	несколько
help	допомога	помощь	together	разом	вместе
if	якщо	если	Ukraine	Україна	Украина
In this way	чином	образом	very	дуже	очень
Kyiv	Київ	Киев	virus	вірус	вирус
life	життя	жизнь	war	війна	война
make	зробити	сделать	what	що	что
many	багато	много	when	коли	когда
medicine	ліки	лекарство	which	яка	какая
more	більше	больше	who	хто	кто
near	біля	рядом	work	роботу	работу
necessary	необхідно	необходимо	years	років	лет

B Some typical convergence results

For each model parameter, we present a graphical summary (top figure) and the convergence of the Gelman-Rubin statistic (bottom figure) (Brooks & Gelman 1998, Gelman & Rubin 1992) of three independent MCMC runs of the No U-Turn sampler. The first, the second, and the third row of the top figure, correspond to the first, the second and the third independent MCMC run, respectively. The first, the second, and the third column of the top figure, correspond to trace, sample auto correlation function, and density plots, respectively.

MCMC graphical summary for the parameter ϕ

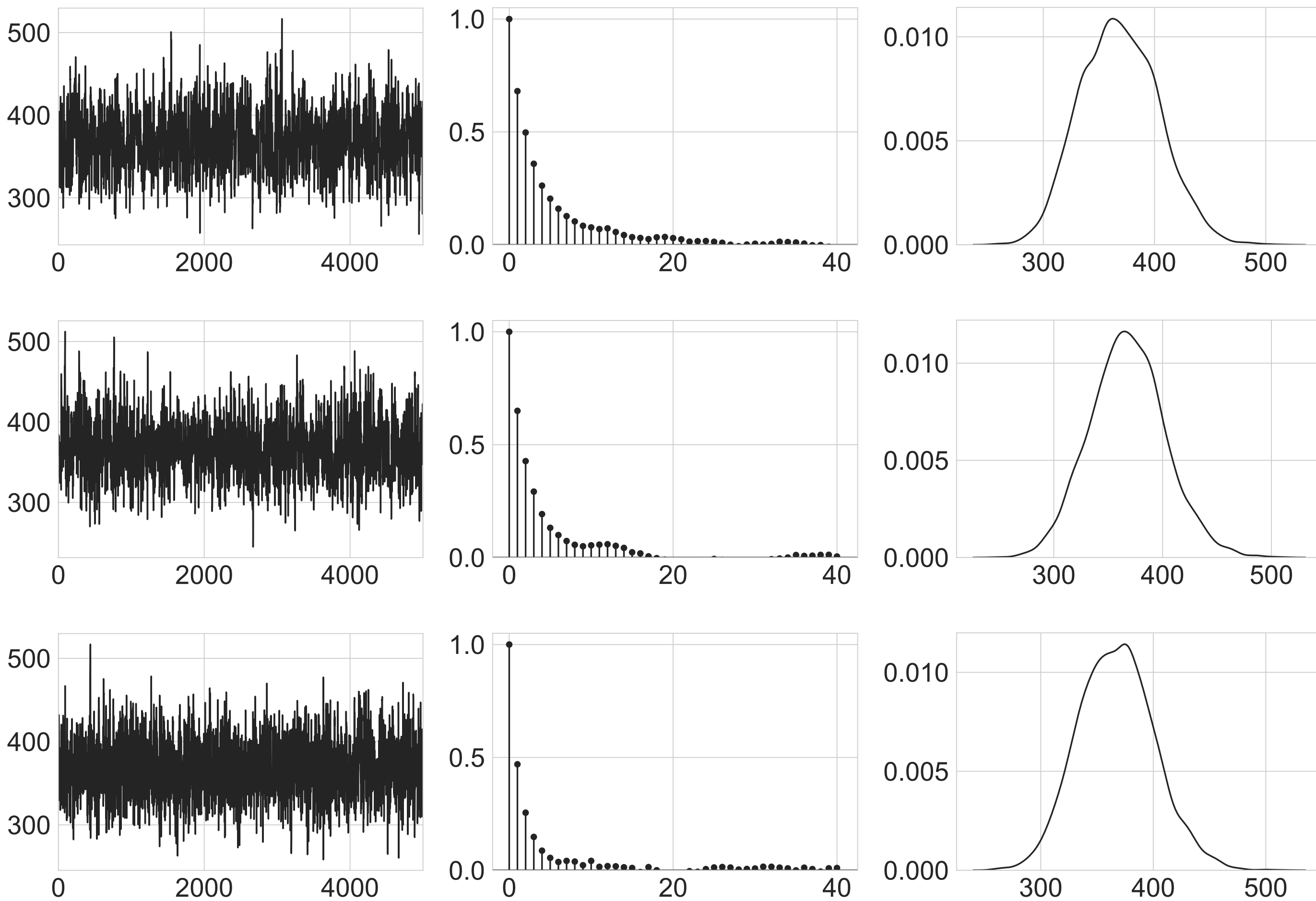


Figure 13: Summary of three Markov Chain Monte Carlo runs for the parameter ϕ.

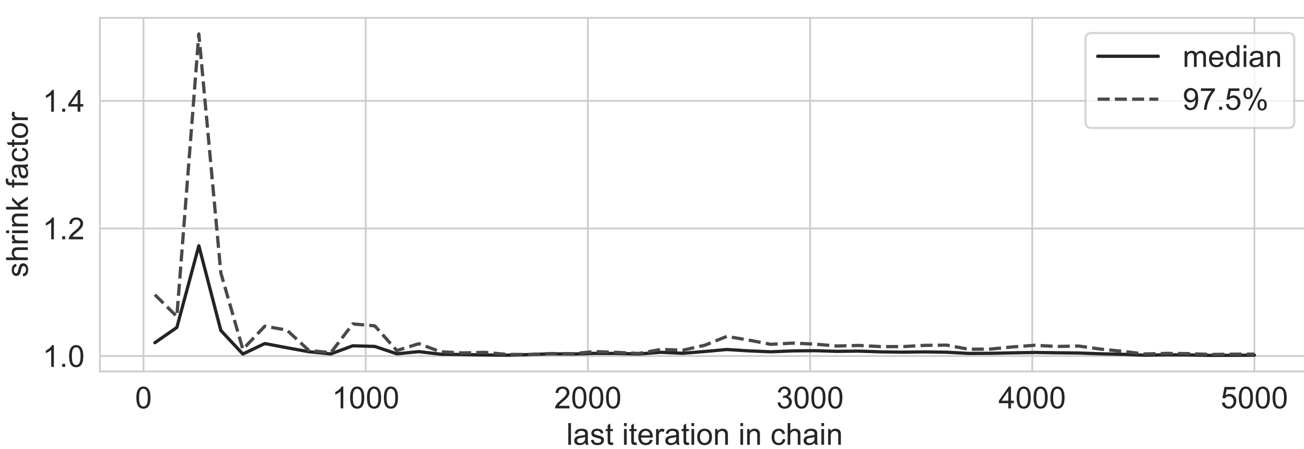


Figure 14: Gelman-Rubin diagnostic for parameter ϕ.

MCMC graphical summary for the parameter β_0

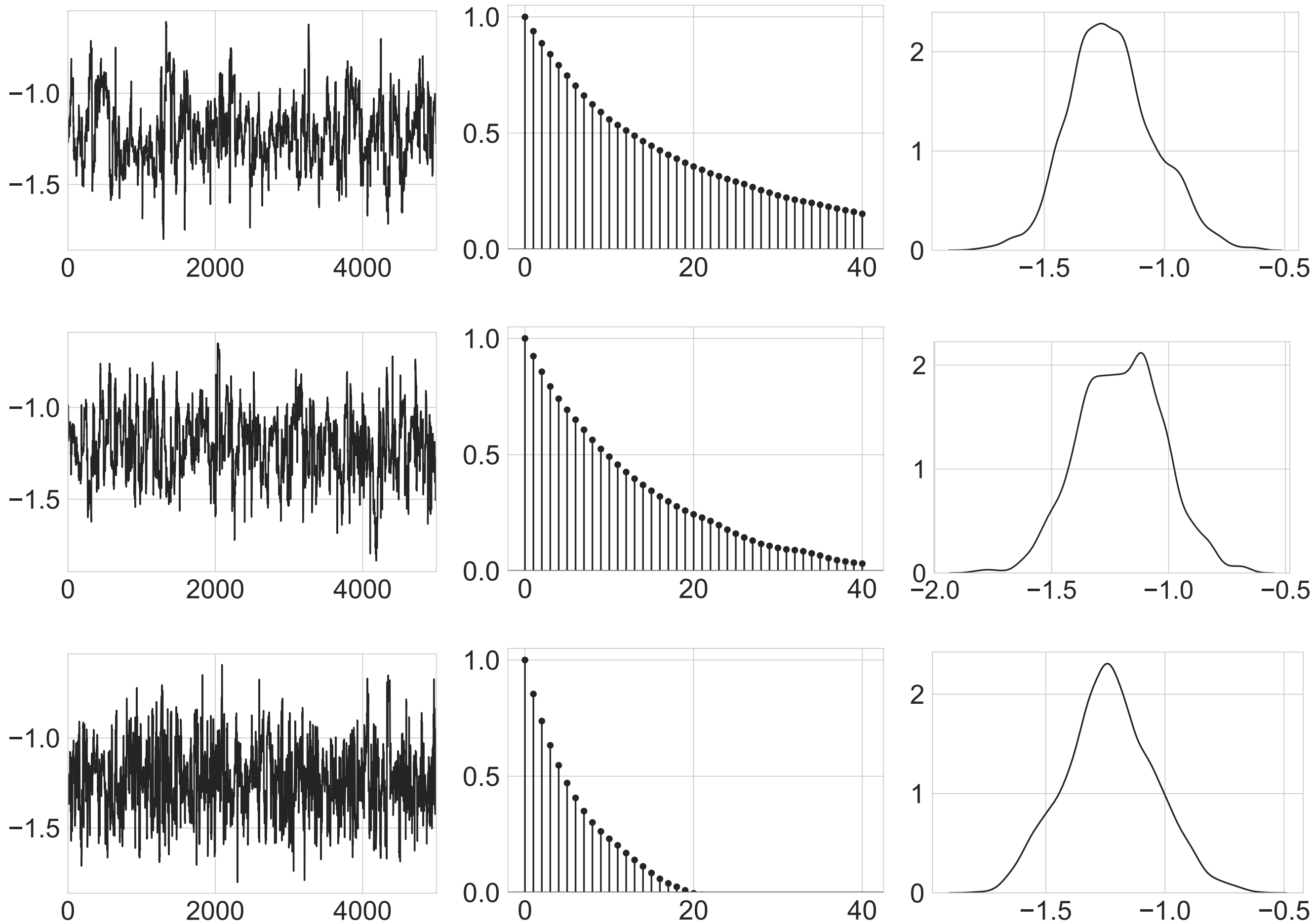


Figure 15: Summary of three Markov Chain Monte Carlo runs for the parameter β_0.

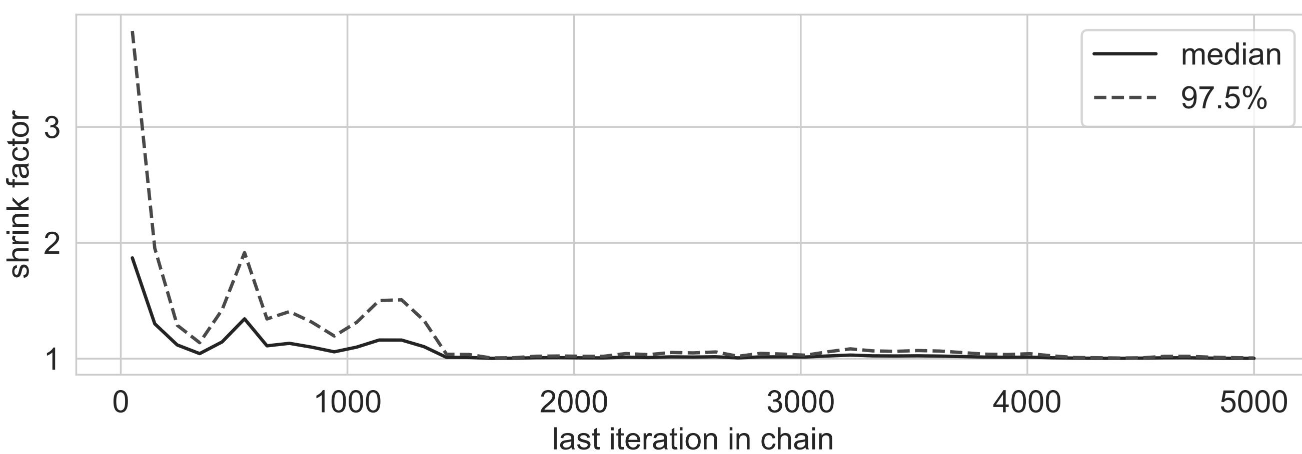


Figure 16: Gelman-Rubin diagnostic for parameter β_0.

MCMC graphical summary for the parameter β_1

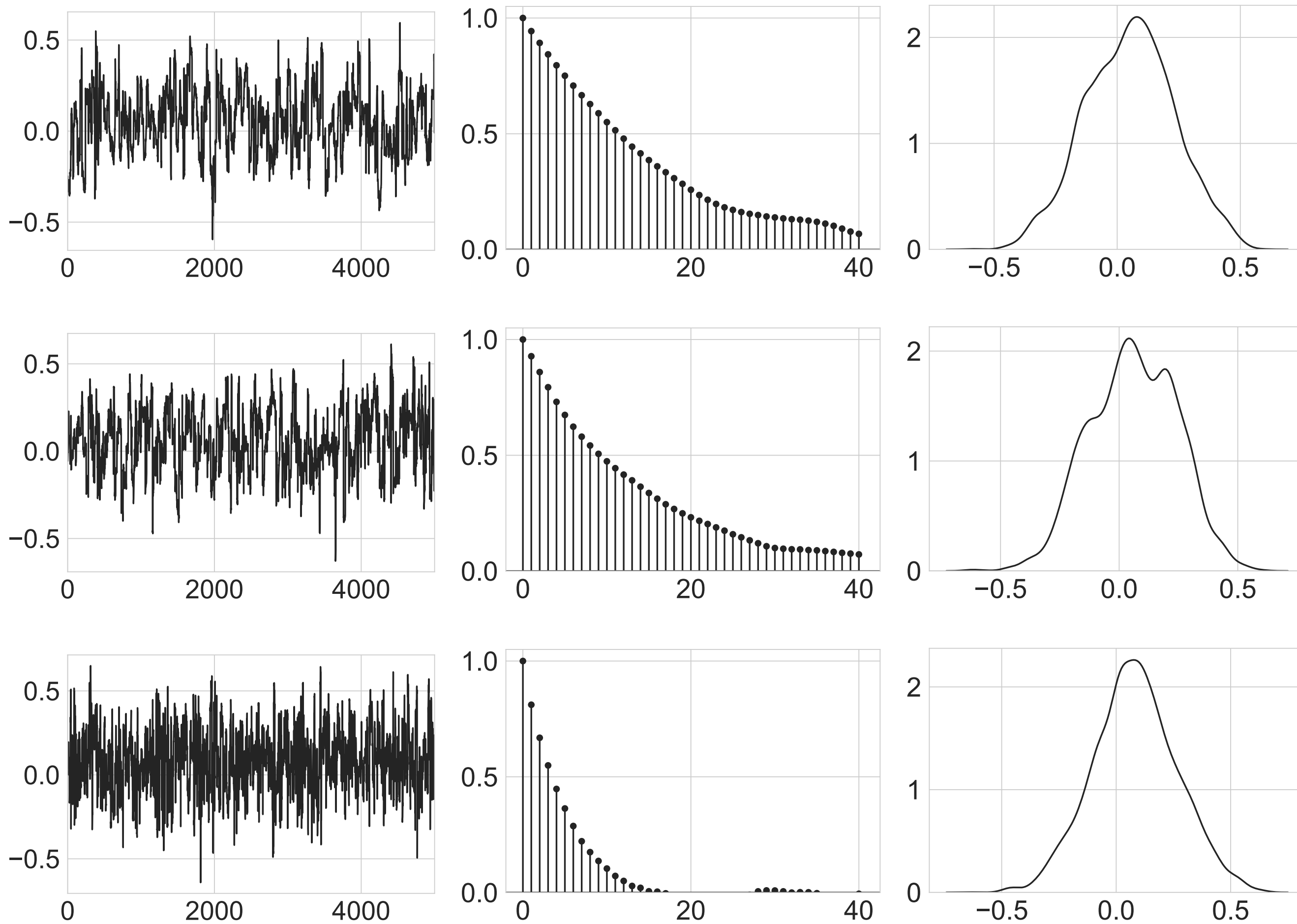


Figure 17: Summary of three Markov Chain Monte Carlo runs for the parameter β_1.

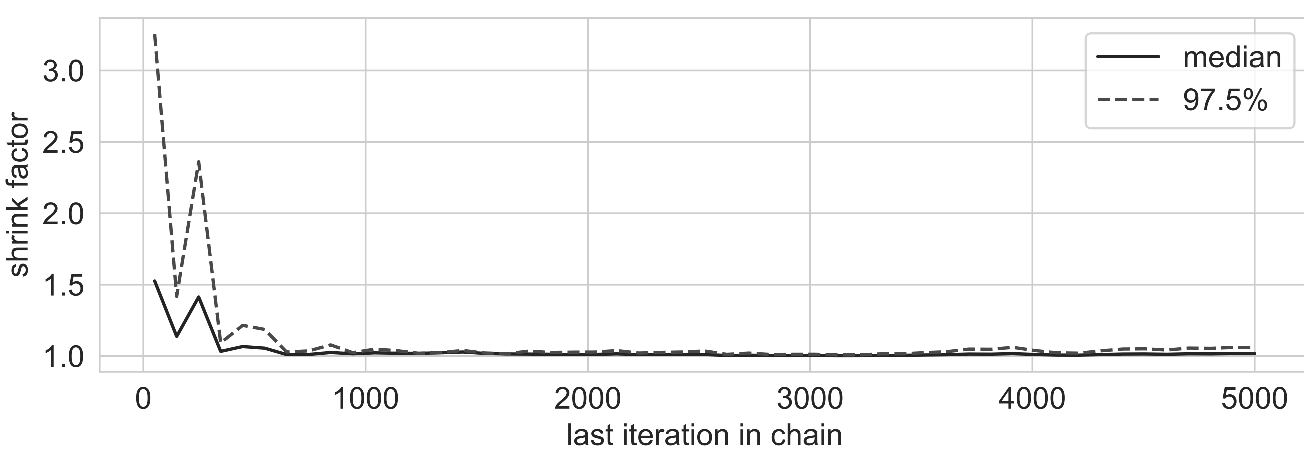


Figure 18: Gelman-Rubin diagnostic for parameter β_1.

MCMC graphical summary for the parameter $\beta_0^{(1)}$

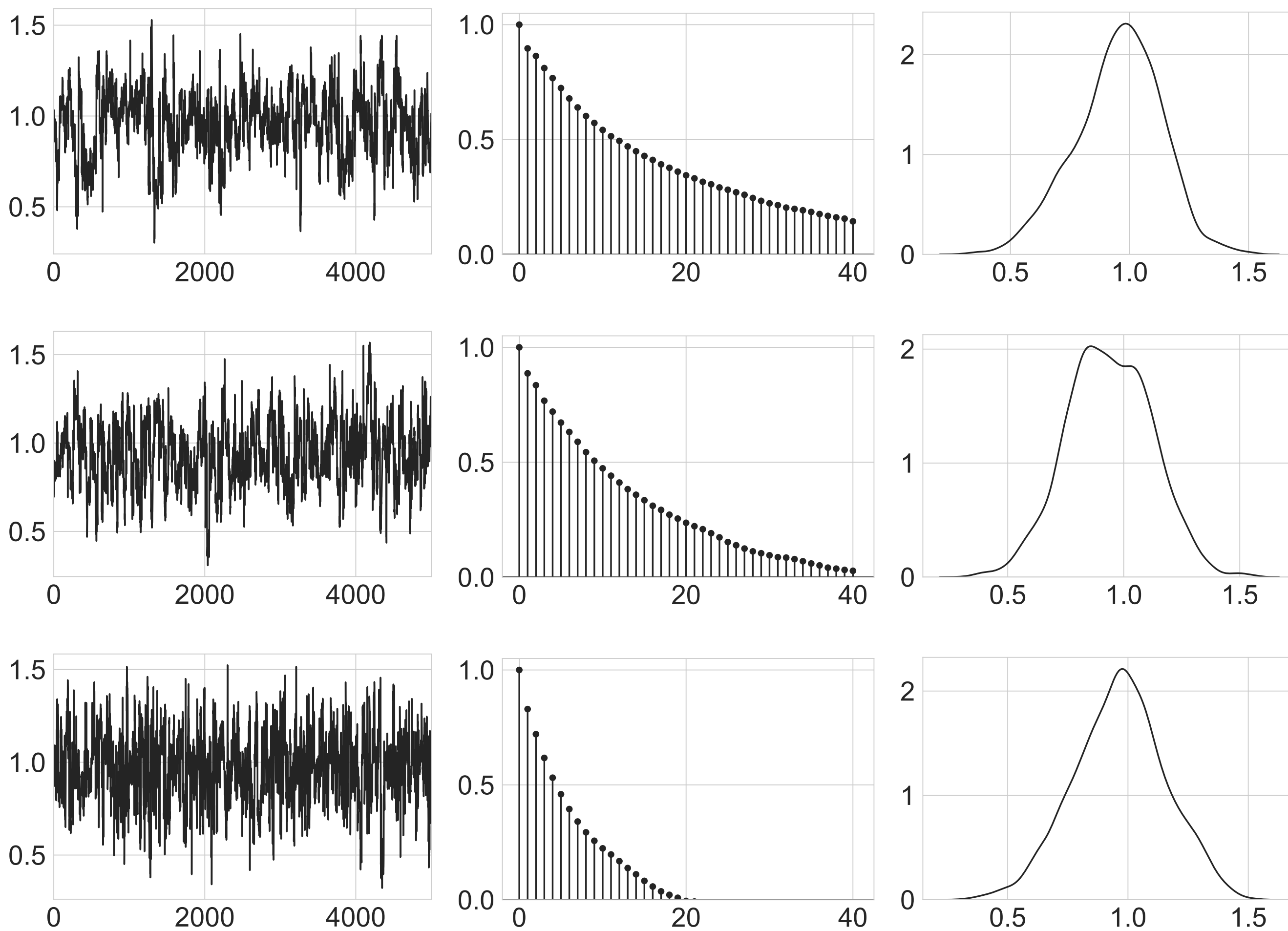


Figure 19: Summary of three Markov Chain Monte Carlo runs for the parameter $\beta_0^{(1)}$.

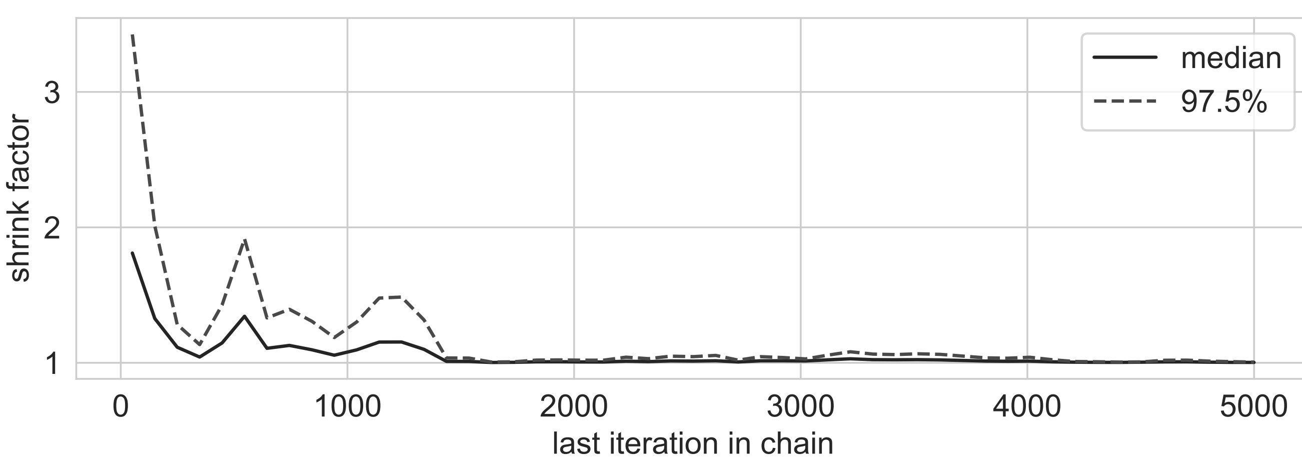


Figure 20: Gelman-Rubin diagnostic for parameter $\beta_0^{(1)}$.

MCMC graphical summary for the parameter $\beta_1^{(1)}$

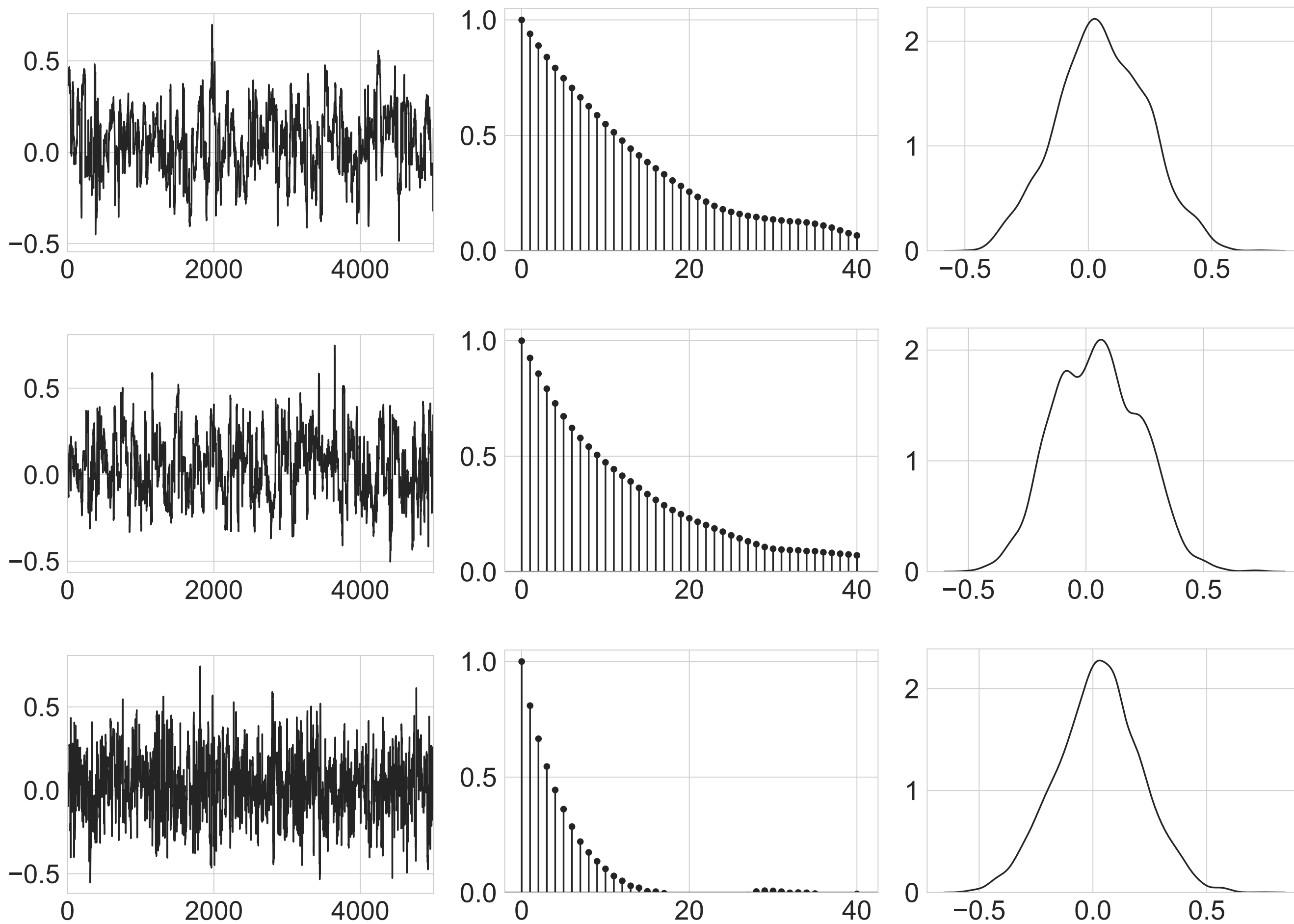


Figure 21: Summary of three Markov Chain Monte Carlo runs for the parameter $\beta_1^{(1)}$.

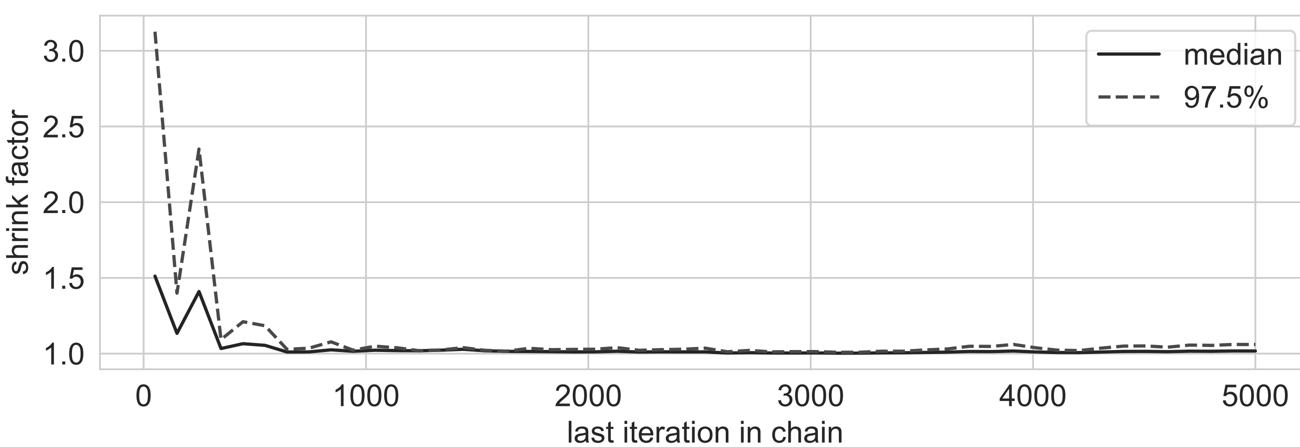


Figure 22: Gelman-Rubin diagnostic for parameter $\beta_1^{(1)}$.

C A hierarchical model

Table 6: Posterior distribution summaries for baseline effects, ϕ, and σ parameters in the hierarchical model (7).

			Quantiles				
Parameter	Mean	Std. Dev.	0.025	0.25	0.5	0.75	0.975
ϕ	368.1	33.9	304.5	345.2	366.5	388.9	437.1
β_0	-1.214	0.195	-1.6	-1.343	-1.209	-1.086	-0.826
β_1	0.063	0.196	-0.335	-0.07	0.07	0.196	0.453
σ	1.031	0.107	0.851	0.957	1.021	1.096	1.266

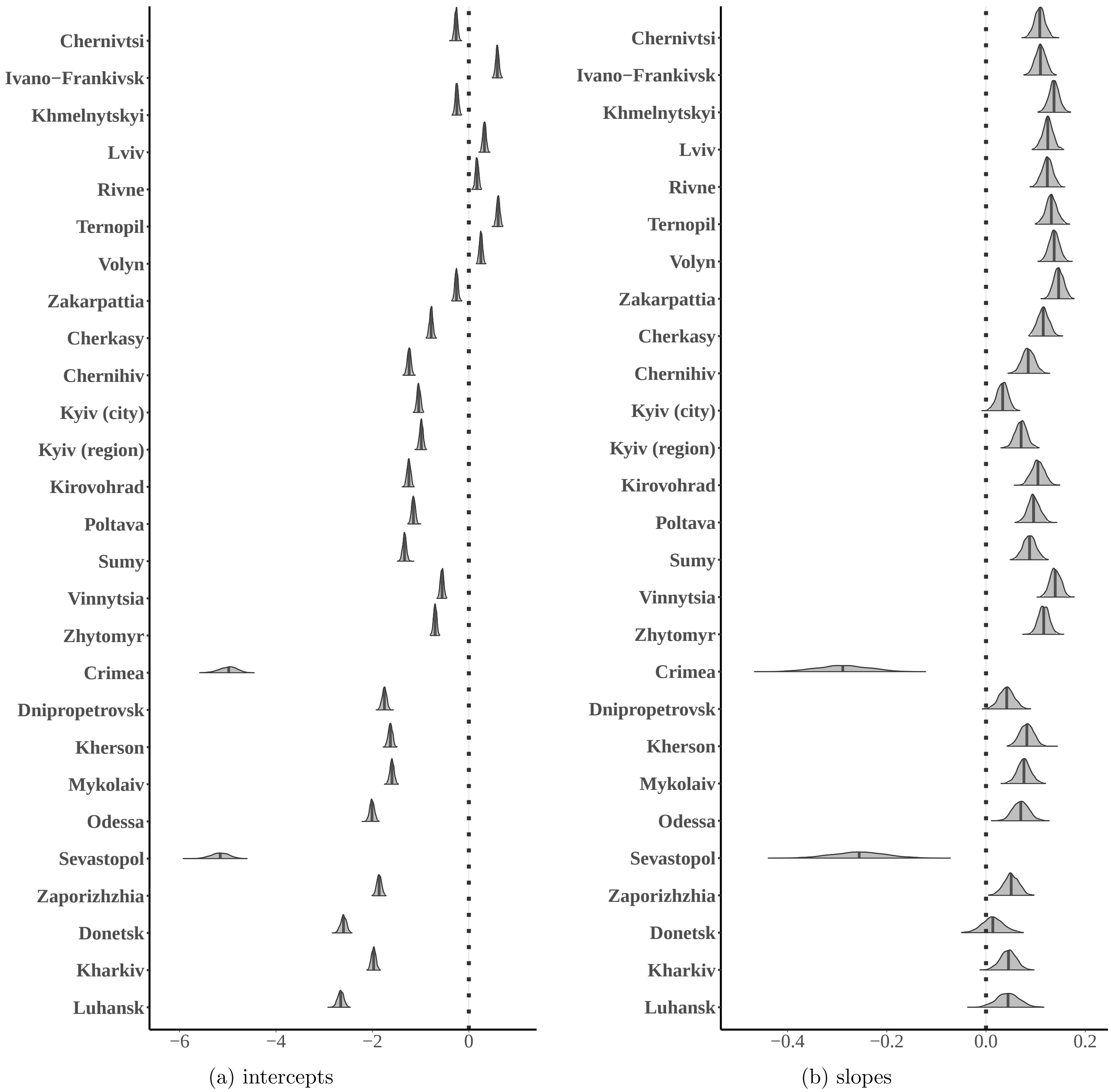


Figure 23: Left panel shows the regional intercepts and right panel shows the regional slopes for the hierarchical model 7.

D Posterior distribution summary tables with respect to $\sigma = 5$ prior parameters

Table 7: Posterior distribution summaries for ϕ and baseline effects β_0 and β_1; the summary is with respect to $\sigma = 5$ prior parameter.

			Quantiles				
Parameter	Mean	Std. Dev.	0.025	0.25	0.5	0.75	0.975
ϕ	375.5	34.0	312.6	352.3	373.9	398.0	444.8
β_0	-1.237	0.972	-3.165	-1.856	-1.213	-0.542	0.623
β_1	0.020	0.906	-1.801	-0.576	-0.006	0.624	1.809

Table 8: Posterior distribution summaries for intercepts; the summary is with respect to $\sigma = 5$ prior parameter.

				Quantiles				
Region	Parameter	Mean	Std. Dev.	0.025	0.25	0.5	0.75	0.975
Chernivtsi	$\beta_0 + \beta_0^{(1)}$	-0.263	0.031	-0.324	-0.283	-0.263	-0.242	-0.202
Ivano-Frankivsk	$\beta_0 + \beta_0^{(2)}$	0.595	0.034	0.528	0.571	0.595	0.618	0.663
Khmelnytskyi	$\beta_0 + \beta_0^{(3)}$	-0.248	0.032	-0.312	-0.270	-0.248	-0.227	-0.186
Lviv	$\beta_0 + \beta_0^{(4)}$	0.325	0.031	0.264	0.305	0.326	0.347	0.386
Rivne	$\beta_0 + \beta_0^{(5)}$	0.169	0.032	0.105	0.147	0.169	0.191	0.231
Ternopil	$\beta_0 + \beta_0^{(6)}$	0.611	0.034	0.549	0.587	0.609	0.634	0.678
Volyn	$\beta_0 + \beta_0^{(7)}$	0.252	0.032	0.190	0.230	0.252	0.272	0.318
Zakarpattia	$\beta_0 + \beta_0^{(8)}$	-0.257	0.032	-0.318	-0.279	-0.258	-0.235	-0.197
Cherkasy	$\beta_0 + \beta_0^{(9)}$	-0.782	0.035	-0.848	-0.806	-0.783	-0.759	-0.711
Chernihiv	$\beta_0 + \beta_0^{(10)}$	-1.241	0.038	-1.318	-1.265	-1.239	-1.216	-1.170
Kyiv (city)	$\beta_0 + \beta_0^{(11)}$	-1.040	0.035	-1.109	-1.063	-1.040	-1.017	-0.971
Kyiv (region)	$\beta_0 + \beta_0^{(12)}$	-0.987	0.036	-1.058	-1.010	-0.987	-0.963	-0.915
Kirovohrad	$\beta_0 + \beta_0^{(13)}$	-1.245	0.039	-1.320	-1.270	-1.244	-1.218	-1.170
Poltava	$\beta_0 + \beta_0^{(14)}$	-1.147	0.038	-1.223	-1.172	-1.147	-1.123	-1.072
Sumy	$\beta_0 + \beta_0^{(15)}$	-1.337	0.038	-1.412	-1.362	-1.337	-1.311	-1.260
Vinnytsia	$\beta_0 + \beta_0^{(16)}$	-0.554	0.034	-0.621	-0.578	-0.554	-0.531	-0.487
Zhytomyr	$\beta_0 + \beta_0^{(17)}$	-0.700	0.034	-0.767	-0.723	-0.700	-0.677	-0.635
Crimea	$\beta_0 + \beta_0^{(18)}$	-5.118	0.202	-5.550	-5.239	-5.111	-4.975	-4.756
Dnipropetrovsk	$\beta_0 + \beta_0^{(19)}$	-1.752	0.044	-1.841	-1.781	-1.751	-1.721	-1.671
Kherson	$\beta_0 + \beta_0^{(20)}$	-1.627	0.042	-1.710	-1.655	-1.626	-1.598	-1.546
Mykolaiv	$\beta_0 + \beta_0^{(21)}$	-1.597	0.043	-1.682	-1.625	-1.595	-1.567	-1.516
Odessa	$\beta_0 + \beta_0^{(22)}$	-2.010	0.049	-2.102	-2.044	-2.009	-1.975	-1.920
Sevastopol	$\beta_0 + \beta_0^{(23)}$	-5.313	0.203	-5.734	-5.449	-5.305	-5.165	-4.949
Zaporizhzhia	$\beta_0 + \beta_0^{(24)}$	-1.862	0.047	-1.952	-1.894	-1.860	-1.831	-1.768
Donetsk	$\beta_0 + \beta_0^{(25)}$	-2.605	0.059	-2.722	-2.644	-2.603	-2.563	-2.493
Kharkiv	$\beta_0 + \beta_0^{(26)}$	-1.977	0.048	-2.071	-2.008	-1.976	-1.945	-1.888
Luhansk	$\beta_0 + \beta_0^{(27)}$	-2.660	0.063	-2.784	-2.700	-2.658	-2.618	-2.537

Table 9: Posterior distribution summaries for slopes; the summary is with respect to $\sigma = 5$ prior parameter.

				Quantiles				
Region	Parameter	Mean	Std. Dev.	0.025	0.25	0.5	0.75	0.975
Chernivtsi	$\beta_1 + \beta_1^{(1)}$	0.108	0.010	0.088	0.101	0.109	0.115	0.129
Ivano-Frankivsk	$\beta_1 + \beta_1^{(2)}$	0.110	0.010	0.090	0.103	0.110	0.117	0.131
Khmelnytskyi	$\beta_1 + \beta_1^{(3)}$	0.137	0.011	0.117	0.130	0.137	0.144	0.158
Lviv	$\beta_1 + \beta_1^{(4)}$	0.124	0.010	0.103	0.118	0.124	0.131	0.145
Rivne	$\beta_1 + \beta_1^{(5)}$	0.124	0.010	0.103	0.117	0.124	0.131	0.144
Ternopil	$\beta_1 + \beta_1^{(6)}$	0.132	0.011	0.110	0.124	0.131	0.139	0.153
Volyn	$\beta_1 + \beta_1^{(7)}$	0.137	0.010	0.118	0.130	0.137	0.144	0.157
Zakarpattia	$\beta_1 + \beta_1^{(8)}$	0.146	0.010	0.125	0.139	0.146	0.153	0.166
Cherkasy	$\beta_1 + \beta_1^{(9)}$	0.115	0.011	0.093	0.107	0.115	0.122	0.136
Chernihiv	$\beta_1 + \beta_1^{(10)}$	0.085	0.013	0.060	0.077	0.085	0.094	0.109
Kyiv (city)	$\beta_1 + \beta_1^{(11)}$	0.033	0.012	0.011	0.026	0.033	0.040	0.056
Kyiv (region)	$\beta_1 + \beta_1^{(12)}$	0.071	0.011	0.049	0.063	0.070	0.078	0.094
Kirovohrad	$\beta_1 + \beta_1^{(13)}$	0.105	0.013	0.081	0.097	0.106	0.114	0.129
Poltava	$\beta_1 + \beta_1^{(14)}$	0.095	0.012	0.073	0.087	0.095	0.104	0.119
Sumy	$\beta_1 + \beta_1^{(15)}$	0.088	0.013	0.064	0.080	0.088	0.096	0.112
Vinnytsia	$\beta_1 + \beta_1^{(16)}$	0.139	0.011	0.119	0.132	0.139	0.146	0.160
Zhytomyr	$\beta_1 + \beta_1^{(17)}$	0.117	0.011	0.095	0.109	0.116	0.124	0.138
Crimea	$\beta_1 + \beta_1^{(18)}$	-0.310	0.058	-0.426	-0.348	-0.308	-0.271	-0.196
Dnipropetrovsk	$\beta_1 + \beta_1^{(19)}$	0.041	0.015	0.012	0.031	0.041	0.050	0.070
Kherson	$\beta_1 + \beta_1^{(20)}$	0.082	0.014	0.053	0.072	0.082	0.091	0.109
Mykolaiv	$\beta_1 + \beta_1^{(21)}$	0.077	0.013	0.052	0.068	0.077	0.086	0.104
Odessa	$\beta_1 + \beta_1^{(22)}$	0.071	0.016	0.039	0.061	0.070	0.081	0.102
Sevastopol	$\beta_1 + \beta_1^{(23)}$	-0.274	0.059	-0.385	-0.315	-0.276	-0.236	-0.159
Zaporizhzhia	$\beta_1 + \beta_1^{(24)}$	0.051	0.015	0.022	0.041	0.052	0.061	0.080
Donetsk	$\beta_1 + \beta_1^{(25)}$	0.013	0.021	-0.030	-0.001	0.014	0.027	0.054
Kharkiv	$\beta_1 + \beta_1^{(26)}$	0.044	0.015	0.015	0.034	0.044	0.053	0.075
Luhansk	$\beta_1 + \beta_1^{(27)}$	0.045	0.022	0.004	0.030	0.044	0.059	0.087