

Modeling language change in the former Soviet Union

Kyle L. Marquardt*

Paper Presented at the 2019 ASN World Convention
Columbia University
2-4 May 2019

Preliminary draft; not for circulation or citation

Abstract

Language is a key element of post-Soviet politics. However, assessing the role language plays in political outcomes—as well as that which politics play in linguistic outcomes—is difficult given the absence of data that are comparable across regions and time. In this paper, I propose a structural equation model that measures fluency in national languages consistently across most ethnofederal regions of the Soviet Union in the period 1926-2017.

*Varieties of Democracy (V-Dem) Institute, Department of Political Science, University of Gothenburg;
kyle.marquardt@gu.se

Language has been a consistently important element of post-Soviet politics. The Russian government regularly uses the protection of Russian speakers as justification for intervention into the politics of countries in its near abroad (The Moscow Times 2018). On the other hand, activists in Russian republics have rued Russian policy that limits the promotion of their regions' titular languages (Radio Free Europe/Radio Liberty 2018). Indeed, the tension between programs of linguistic revitalization in post-Soviet regions and speakers of other languages (especially Russian speakers) has been at the forefront of vital research on post-Soviet identity and politics (Laitin 1998). Understanding patterns of language change is therefore of clear importance to the study of post-Soviet societies and politics.

However, cross-regionally and cross-temporally consistent data on language change in the former Soviet Union do not exist. While Soviet censuses generally included questions regarding self-identification with relevant languages (e.g. the language a citizen considers her "native" language and a second Soviet language which she spoke fluently), only a few post-Soviet censuses continued with these formulations. Instead, some censuses only asked about native language, while others introduced their own idiosyncratic measures of linguistic capabilities; one country (Uzbekistan) has yet to conduct an official census. As a result, assessing the degree to which linguistic demographics have changed in post-Soviet states is difficult: even in those countries that have had multiple post-Soviet censuses, alterations in the format of language questions (or the absence of such questions) makes comparing change in linguistic demographics over time and space in the former Soviet Union a highly tentative endeavor.

In this paper I propose a model that provides cross-regionally and cross-temporally consistent measures of the proportion of a region's population which is fluent in the region's titular language. The model integrates a structural model of linguistic change with Bayesian factor analysis, allowing for the iterative estimation of the relationship between different measures of linguistic demographics and the underlying concept (titular language fluency).

This model is still a work in progress and, as such, this working paper provides a "warts

and all” analysis of the advantages and disadvantages of the modeling strategy.

1 Soviet and post-Soviet measures of language

Language was a fundamental element of Soviet conceptualizations of national identity, which in turn were a matter of great importance to Soviet politics (Hirsch 1997). As a result, Soviet censuses generally included 1-2 questions regarding citizens’ self identification with relevant languages. Specifically, all censuses included a question regarding the question a citizen considered her “native language” (*rodnoi iazyk*), and the 1970, 1979 and 1989 censuses also included a question regarding self-reported fluency in a second Soviet language (*svobodno vladeet vtorym iazykom*). These data are by no means straightforward measures of linguistic fluency: in particular, there is evidence that many Soviet citizens considered their native language to be the language of their ethnic group, regardless of whether or not they spoke the language.¹

At the same time, these data at least involved consistent questions over time and space. This relative luxury disappeared after the disintegration of the Soviet Union. Only several post-Soviet censuses maintained the Soviet practice of asking about both native and second language; other censuses only asked about native language, while others dropped language altogether and only asked about ethnic identification. Yet other countries yet—perhaps most prominently Kazakhstan—began asking idiosyncratic questions about linguistic capabilities to further political interests (Dave 2004, Dave 2007).²

On the other hand, the increased openness of many countries the former Soviet Union

¹Silver (1974) provides the canonical description of this concern. Equally importantly, all data I present in this paper are self-reported, which means that some respondents may over- or underestimate their abilities or practices. Moreover, spoken proficiency is itself a multidimensional concept, including both comprehension and productive capabilities, as well as accent. In principle, by aggregating over a large number of questions regarding both linguistic abilities and practices it would be possible to develop multidimensional measures of linguistic fluency; in practice, this approach would be very time- and resource-intensive to implement on a cross-national scale. As a result, I use self-reported fluency as an admittedly rough shorthand for the complicated concept I am endeavoring to measure.

²See also Arel (2002) and Herrera (2004) for descriptions of the processes by which language categories were determined in other post-Soviet censuses.

meant that it became possible to conduct more standard survey research in these countries. Importantly, cross-national enterprises such as the World Values Survey began including post-Soviet states in their standard waves, and include questions about language (in the case of the World Values survey, the question regards the language a respondent reports speaking most frequently at home). Similarly, scholars interested in issues regarding language began fielding surveys in specific countries and regions about different aspects of linguistic abilities, providing more fine-grained measures of linguistic proficiency than the blunt Soviet census questions (Colton & Hough 1998, Laitin 2000). As a result, there is a wide variety of data available to measure language and language change in the former Soviet Union. Table 1 illustrates both the data I have incorporated into the model and their coverage.³

³I am in the process of gathering additional data. Specifically, I have access to the 1994 Russian micro-census and the Laitin/Hough surveys of 1995/1996, and am gathering more detailed versions of the Soviet census to fill in some of the missingness. I am also looking into the Eurobarometer and AsiaBarometer surveys for additional post-Soviet data. In principle, I could also extend the timeframe back to the first Russian imperial census (1897), though this will compound issues regarding territorial comparability. I welcome recommendations regarding additional data sources.

Table 1: Data, data sources and coverage

Statistic	Sources ⁴	Coverage
Ethnic identification	Colton/Hough survey, Soviet censuses, post-Soviet censuses, World Values Survey	All ethnofederal units of the Soviet Union (1926, 1939, 1959, 1970, 1979, 1989), Abkhazia (2003, 2011), Armenia (1997, 2005, 2011), Azerbaijan (1997, 1999, 2009), Belarus (1996, 1999, 2009, 2011), Estonia (2000, 2011), Georgia (1996, 2002, 2014), Kazakhstan (1999, 2009, 2011), Kyrgyzstan (1999, 2009, 2011), Latvia (1996, 2000), Lithuania (2001, 2011), Moldova (2003, 2004, 2006, 2014), Nagorno-Karabakh (2005), 16 Russian ASSRs (1993), all ethnofederal units of Russia (2002, 2010), Tajikistan (2000, 2010), Turkmenistan (1995, 2001), Ukraine (2001, 2011), Uzbekistan (1991, 2011, 2017),
Native language	Colton/Hough survey, Soviet censuses, post-Soviet censuses	All ethnofederal units of the Soviet Union (1926, 1970, 1979, 1989), SSSRs (1959), Armenia (2011), Estonia (2000, 2011), Belarus (1999, 2009), Kazakhstan (1999), Kyrgyzstan (1999, 2009), Latvia (2000), Lithuania (2001), Moldova (2004, 2014), 16 Russian ASSRs, all ethnofederal units of Russia (2010), Tajikistan (2000, 2010), Turkmenistan (1995), Ukraine (2001)
Speaks language (Native + second language)	Soviet censuses, post-Soviet censuses	All ethnofederal units of the Soviet Union (1970, 1979, 1989), Armenia (2011), Azerbaijan (1999, 2009), Belarus (2009), Kyrgyzstan (1999, 2009), Latvia (2000), Lithuania (2001), Tajikistan (2000, 2010), Turkmenistan (1995)
Fluent in language (Likert scale)	Colton/Hough survey	16 Russian ASSRs
Fluent in language (no native language)	Russian census	All ethnofederal units of Russia (2002)
Fluent in language (native language question not associated with estimate)	Russian census	All ethnofederal units of Russia (2010)
Most often spoken language (composite measure)	Colton/Hough survey	16 Russian ASSRs
Language most often spoken at home	World Values Survey	Armenia (1997, 2011), Azerbaijan (1997, 2011), Armenia (1997), Azerbaijan (1997), Belarus (1996, 2009, 2011), Estonia (1996, 2011), Georgia (1996, 2009, 2014), Kazakhstan (2011), Kyrgyzstan (2003, 2011), Latvia (1996), Lithuania (1996, 2011), Moldova (1996, 2003, 2006), Ukraine (2006, 2011), Uzbekistan (2011)
Most usually spoken language	Moldovan censuses	Moldova (2004, 2014)

⁴Full citations to be added.

Unfortunately, the available data vary by region and time, hindering comparison.⁵ Figure 1 illustrates this problem, showing trends in different measures of titular linguistic demographics in four post-Soviet regions.⁶ Belarus is one of the few post-Soviet countries that continued the Soviet practice of asking about both citizens' native language and a second language they spoke fluently, providing some continuity in the data. In addition to these data, the World Values Survey has included Belarus in three survey waves, providing additional cross-nationally comparable data on language spoken at home and ethnic demographics. However, these data yield wildly different conclusions about language trends in post-Soviet Belarus. Specifically, the data on ethnic identity⁷ indicate that the proportion of Belarusian speakers is increasing, while the data on native language and the titular speaking population shows a decrease. Moreover, the proportion of Belarusians who speak Belarusian at home is much lower than the other linguistic statistics, indicating that this measure is not directly comparable to the other measures.

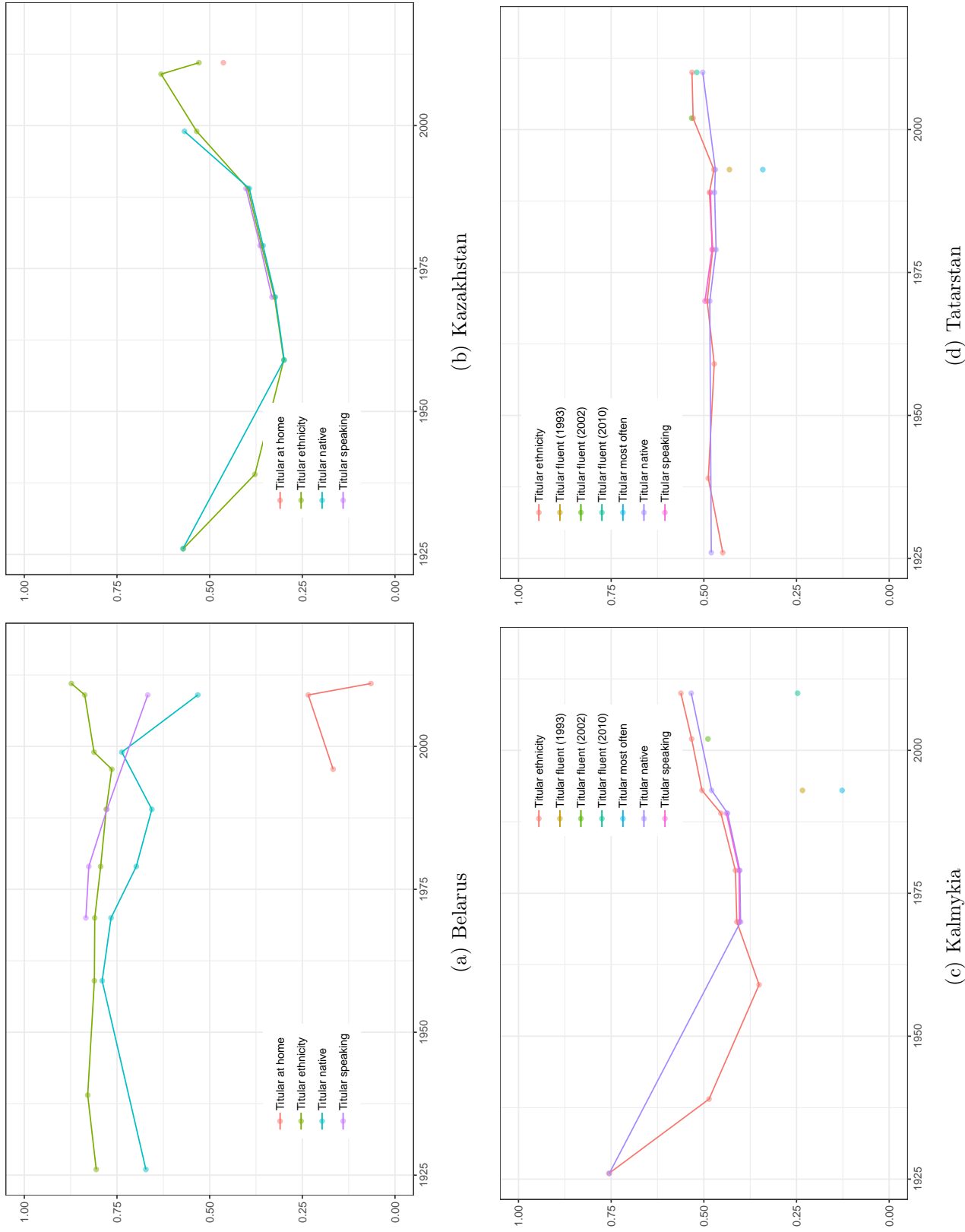
The data from Kazakhstan are perhaps more consistent—showing an increase in the proportion of Kazakh speakers in the post-Soviet period—but illustrate the problem of data sparsity. Specifically, while the Kazakhstani censuses have provided additional data on linguistic capabilities in the post-Soviet period, these measures are unique to Kazakhstan and thus limit cross-national comparability. The remaining measures of linguistic capabilities—native language and ethnicity—are clearly problematic in that they measure concepts closer to ethnic identity than language *per se*.

⁵This concern is not simply an issue of measurement error: the type of linguistic data collected can have substantial implications for analyses of political phenomena (Onuch & Hale 2018).

⁶I create all graphics using *ggplot2* (Wickham 2009).

⁷I include ethnic self-identification as a measure of language given the fact that, in many regions, ethnic identification tracks closely with self-reported linguistic abilities; it is also the measure with the most complete coverage. In the absence of other data, ethnic identification thus provides a rough proxy for linguistic demographics.

Figure 1: Different measures of language in post-Soviet regions



The situation regarding language statistics in the Russian republics of Kalmykia and Tatarstan is different, but equally problematic. Like Kazakhstan, the only consistent measures of linguistic capabilities in the post-Soviet period are ethnic identity and native language. However, the Russian censuses of 2002 and 2010 included questions regarding fluency in regional languages, albeit in a different form than during the Soviet period. Because Russia includes many autonomous regions, even if these questions are idiosyncratic to Russia they are cross-regionally comparable within the Russian Federation. Nevertheless, these different types of data provide contradictory trends, as the case of Kalmykia illustrates. While native language identification and ethnicity indicate an increase in Kalmyk speakers in the region—as does the question regarding fluency in the 2002 census—other data on fluency show a sharp decrease in the proportion of Kalmyk speakers. Comparison of trends over time with these data therefore remains problematic.

2 The model

In this paper, I propose a model to provide cross-regionally and cross-temporally comparable measures of language change in the regions of the former Soviet Union. The model integrates a structural model of language change with a Bayesian factor analytic model. This model iteratively estimates the structural components of the model and the parameters of different factors (i.e. different linguistic statistics) to provide consistent measures of region-level fluency in regional titular languages and accompanying estimates of uncertainty. I discuss these two aspects of the model in turn.

2.1 Structural model of language change

Abrams & Strogatz (2003) propose the following model of language change:

$$\frac{\delta x}{\delta t} = yP_{yx}(x, s) - xP_{xy}(x, s) \tag{1}$$

In this model, there are two languages, which are represented by the proportions x and y . There are no bilinguals in this model, so $y = 1 - x$.

In each time t , some proportion of speakers y change to x , and vice-versa. These proportions (P_{yx} and P_{xy}) are a function of both the proportion of speakers of both languages and the status of language x , s . More specifically, $P_{yx}(x, s) = cx^a s$ and $P_{xy}(x, s) = c(1-x)^a(1-s)$. Here, a represents the peak rate of absorption for a language or, more intuitively, the degree to which the proportion of speakers of a language influences the probability of a language change. c scales this change.

I modify this model both conceptually and substantively. Given the presence of multiple competing languages in most regions of Eurasia—as well as the fact that many citizens are bi- or multilingual—I conceptualize this model as only regarding language x (where y still refers to the proportion of the population that does not speak the language, $1 - x$), essentially turning s into a parameter estimating the relative utility of speaking language x relative to all other possible languages in a territory.

Substantively, I make the following modifications. First, I assume that a is constant across time and space (Abrams & Strogatz (2003) note that empirically this is generally the case, though future iterations of this project will relax this assumption). Second, given that the data on language are not provided at a consistent rate across time and space, and that changes in population size could drastically influence language change, I parameterize c : $c = \phi(\beta + \theta\delta_t + \gamma_1\delta_{Population} + \gamma_2\delta_{Population^2})$, where ϕ is the cumulative distribution function of the standard normal distribution, δ_t is the number of years since the last data point(s) and $\delta_{Population}$ is the ratio of the population in time t and $t - 1$.⁸ Third, I assume that language status s changes over time: specifically, after the disintegration of the Soviet Union; p is the indicator for pre-1993 and post-1993 years.⁹

⁸I assume θ , β and γ are distributed $N(0, 1)$.

⁹My prior specification for s is as follows. I assume that, for each region i , $s_i \sim Beta(\zeta_u + \zeta_{up}, \kappa_u + \kappa_{up})$; u refers to a region's ethnofederal status in 1989 (i.e. Soviet Socialist Republic, Autonomous Soviet Socialist Republic, Autonomous Oblast, and Autonomous Okrug), and up is an additional status-specific parameter for the pre- and post-1993 period. This modeling strategy assumes that ethnofederal status influenced the status of regional languages, and that this influence changed in the pre- and post-Soviet period. I assume

The full model of the proportion of the population in region i ¹⁰ which is fluent in titular language x_i ¹¹ in time t is thus:

$$x_{it} = x_{it-1} + c (y_{it-1} x_{it-1}^a s_{ip} - x_{it-1} y_{it-1}^a [1 - s_{ip}]) \quad (2)$$

Given the mechanical approach to language change, I only need to estimate x_i in $t = 0$, or the first year in which there are data for a given region. I assume $x_{it_0} \sim \text{Beta}(1, 1)$.

2.2 Factor analysis of linguistic fluency

The factor analysis aspect of this model is relatively standard (Jackman 2009), with two alterations.¹² In general,

$$\begin{aligned} y_{itj} &\sim N(\mu_{itj}, \tau_j) \\ \mu_{ijt} &= \lambda_{j1} + \lambda_{j2} x_{it} \end{aligned} \quad (3)$$

In this model, τ represents the precision of linguistic measure j (e.g. self-reported fluency, language spoken at home) for measuring the latent concept of linguistic fluency, while λ represents the estimated intercept and slope parameters for measure j .

all κ and ζ parameters are distributed $\Gamma(1, 1)$.

¹⁰Regions and regional boundaries change over time. In general, I ignore these changes (e.g. in this framework, the Moldavian ASSR is the same unit as the Moldovan SSR and the independent Republic of Moldova) except insofar as these changes influence population size, as measured in c . I do make some exceptions (e.g. I treat Chechnya, Ingushetia and Checheno-Ingushetia as separate regions, since Chechnya and Ingushetia were separate territories in 1926 and again in the post-Soviet period), and I do not account for regions that only became autonomous (or regained autonomous status) following the disintegration of the Soviet Union (e.g. Crimea, Gagauzia and Pridnestrovie). Regions that lose autonomous status (e.g. Ust Orda, Agin-Buryatia) also disappear from the analysis in the periods after their lost autonomy. Future iterations of this project will deal more rigorously with this problem.

¹¹For the sake of simplicity, I only estimate change in the demographically-largest titular language of a territory. This strategy only affects multi-titular regions like Dagestan and Kabardino-Balkaria, where I consider x to be Avar and Kabardin, respectively. Also for the sake of simplicity, I omit Russia proper from my analyses.

¹²For computational ease, I convert both x and y from proportions to unbounded continuous variables using the inverse cumulative distribution function of the normal distribution for the factor analytic portion of the model.

The two alterations of this model are as follows. First, for identification purposes, I establish the measure of self-reported linguistic fluency in the Russian 2010 census as the baseline estimate for the concept of region-level linguistic fluency.¹³ Second, I allow the intercept and slope parameters for the measures of ethnic identification and native language to vary by region, since 1) the relationship between ethnicity, native language and linguistic fluency is idiosyncratic across regions, and 2) these data are the least sparse across time and regions).¹⁴

3 Results

I present results from the model in two ways.¹⁵ First, I compare model estimates to data on language from the four regions which I previously discussed. Second, I compare cross-regional trends using model estimates.

Figure 2 uses the four regions I previously discussed to illustrate how the model estimates compare to actual survey and census data on titular linguistic demographics.¹⁶ The estimates for Belarus largely track those regarding ethnicity and native language during the Soviet period, indicating that the proportion of Belarusian speakers was greater than indicated by those who reported Belarusian as a native language, but less than the Belarusian share of the population. In the post-Soviet period, the data largely track with self-reported proficiency, indicating a decline in the proportion of Belarusian speakers. Estimates for Kazakhstan point to a potential deficiency in the modeling strategy: specifically, the data indicate that the proportion of fluent Kazakh speakers began increasing prior to the disintegration of the Soviet Union, while the model estimates only show an increase in the proportion of fluent Kazakh speakers following the disintegration of the Soviet Union. This discrepancy is likely

¹³More specifically, I set the intercept (λ_1) to 0 and the slope (λ_2) to 1 and precision parameter (τ) to 200 for these data. For all other factors, I assume $\lambda_1 \sim N(0, 1)$ and $\lambda_2 \sim N(1, 1)$; $\tau \sim U^{-2}(0, 1000)$

¹⁴Future iterations of this project will hierarchically cluster the slopes and intercepts by regional ethnofederal status.

¹⁵I use the statistical software *rjags* and *runjags* to estimate the models (Denwood 2016, Plummer 2012).

¹⁶Note that the model estimates include estimates of uncertainty, specifically 90% credible regions, a Bayesian corollary of 90% confidence intervals.

due to language status being modeled as static during the Soviet period; since the proportion of Kazakh speakers drastically decreased between 1926 and 1959, the model cannot account for later Soviet-era revitalization. On the other hand, given sparse data on language in post-Soviet Kazakhstan, the model assigns a relatively high level of uncertainty to estimates in this period.

In Kalmykia, where linguistic data provide divergent trends following the disintegration of the Soviet Union (i.e. native language statistics indicate an increase, while measures of fluency show a decrease), the model estimates that there has been a slight increase in the proportion of fluent Kalmyk speakers. However, the model output also indicates that the data on native language and ethnicity likely drastically overestimates the proportion of fluent Kalmyk speakers, especially in more recent years.

Finally, trends in model estimates for Tatarstan closely track those from census data, though they also indicate that estimates based on ethnicity and native language have historically overestimated the proportion of fluent Tatar speakers.

Figure 2: Comparison of model estimates to data on language in post-Soviet regions

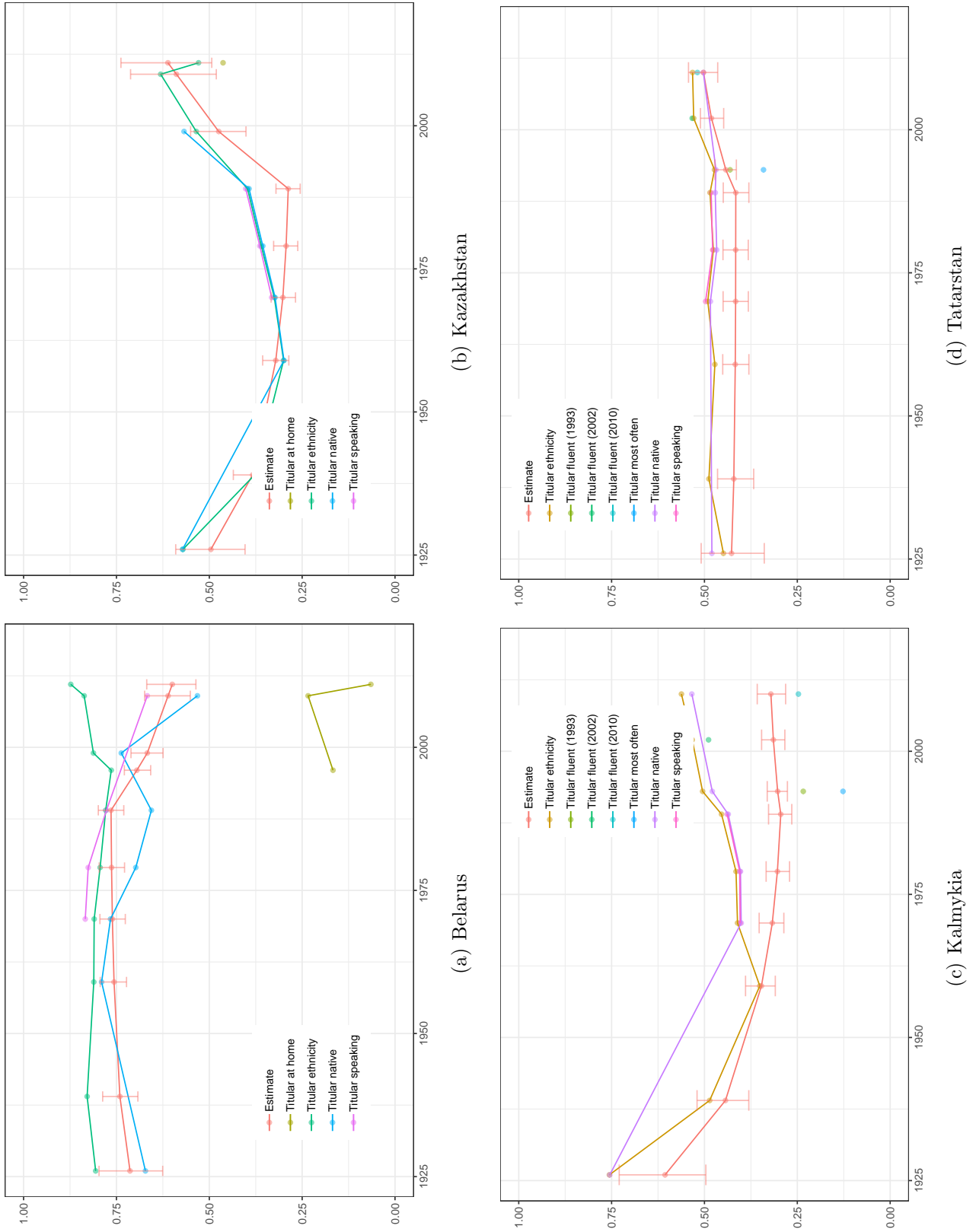


Figure 3 compares linguistic trends in post-Soviet regions using model estimates. The data for Eastern Europe and the Baltics show great discrepancy in trends, both during and after the Soviet period. Specifically, Moldova shows a great increase in the proportion of fluent Moldovan speakers during the Soviet period, especially between 1926 and 1959; followed by a slight increase in the post-Soviet period.¹⁷ In contrast, Belarus, Lithuania and Ukraine show relatively static trends during the Soviet period, while Estonia and Latvia show decreasing proportions of speakers of their titular languages. In the post-Soviet period, Latvia and Lithuania show trends of rapidly increasing proportion of titular speakers, while Estonia and Ukraine show marginal increases. Belarus is the only former Soviet Socialist Republic that shows a decreasing proportion of speakers of the titular language.

Trends in the independent countries of the Caucasus are relatively similar, especially in the post-Soviet period, with proportions of titular-speakers increasing toward a value of 100 percent. The independent states of Central Asia also show rapidly increasing proportions of speakers of their respective titular languages. These trends are in contrast to those during the Soviet period, where the proportion of titular speakers was decreasing, especially in Tajikistan, Kyrgyzstan and Kazakhstan.

Like Eastern Europe and the Baltics, the republics of the Russian Volga region show divergent trends, particularly in the post-Soviet period. The model estimates indicate that the proportion of fluent Tatar and Bashkir speakers in Tatarstan and Bashkortostan, respectively, have increased in the post-Soviet period, albeit to a lesser extent than in the independent states of Central Asia and the Caucasus. On the other hand, the proportion of speakers of Mari and Mordvinian in Mari El and Mordvinia, respectively, have changed minimally in the post-Soviet period; while the proportion of Chuvash and Udmurt speakers in Chuvashia and Udmurtia have continued to decrease.

¹⁷The drastic increase in Moldovan speakers during the Soviet period again points to a problem with the model—specifically, during the interwar period Moldova was a multiethnic autonomous region located in Ukraine; after World War II, this territory was combined with a region that had formerly been part of Romania and was largely ethnically Moldovan. Developing a better method for modeling territorial changes is a priority for future model iterations.

Figure 3: Comparison of trends across post-Soviet regions

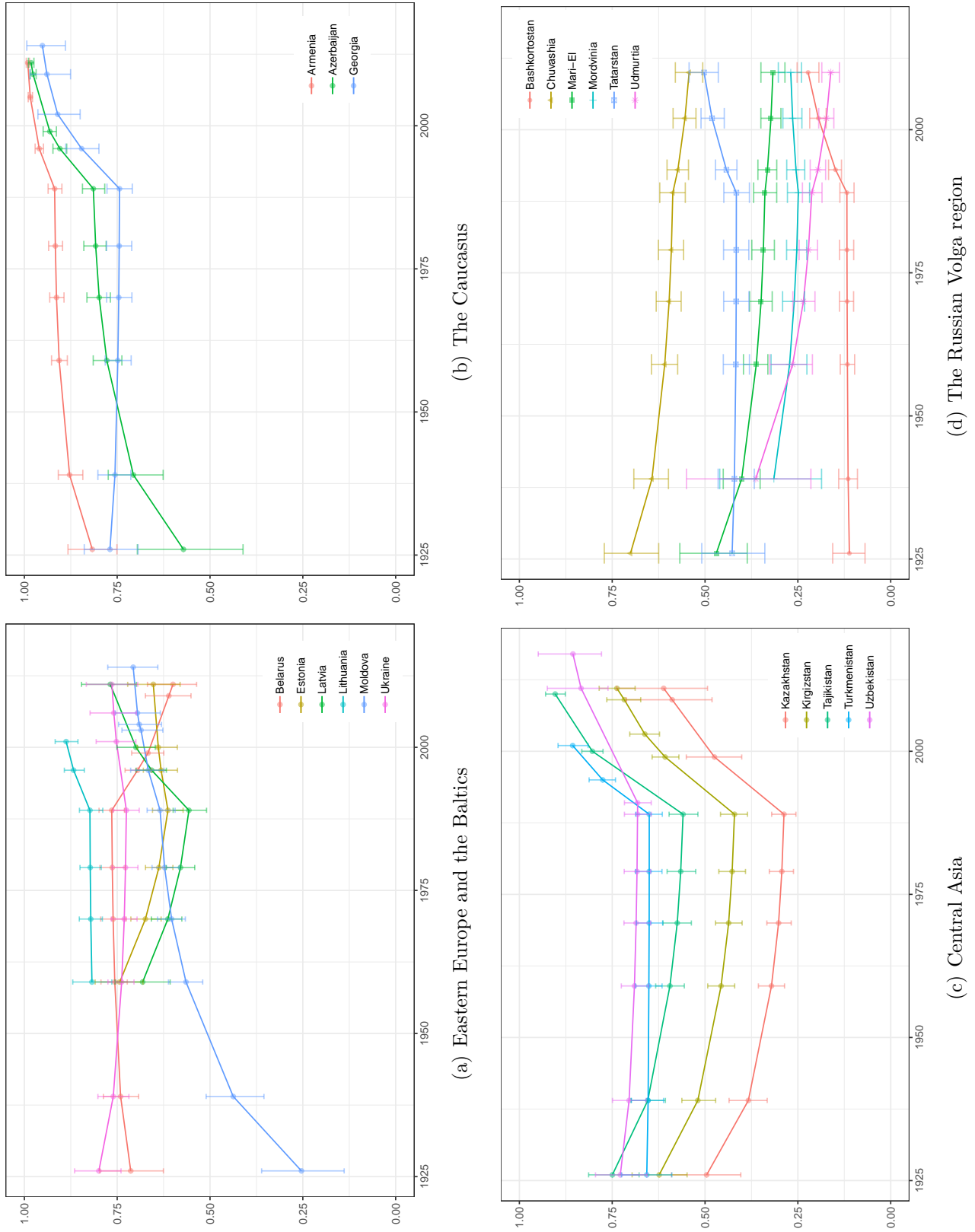
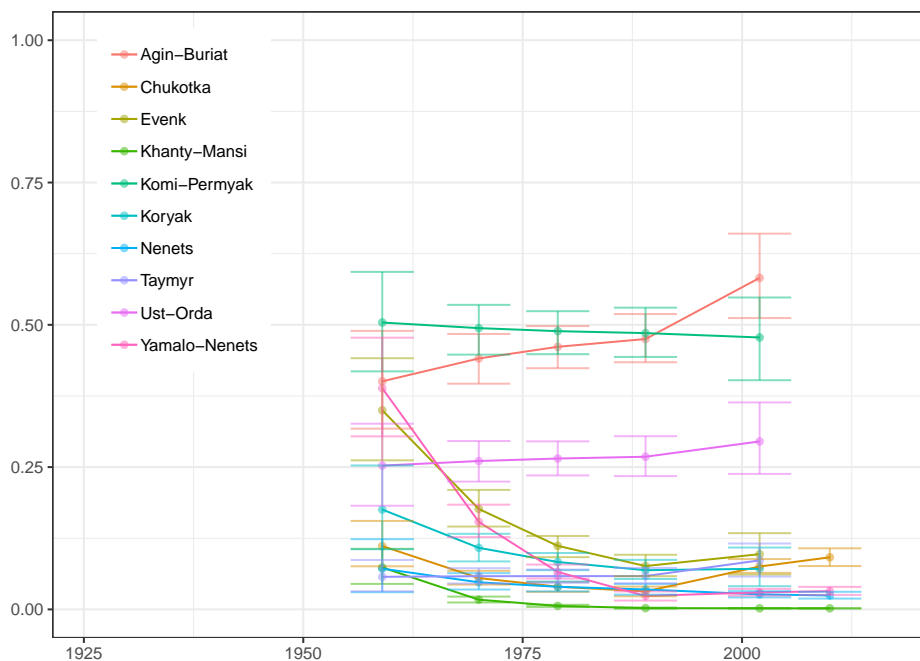


Figure 4: Trends in autonomous okrugs



All of the regions described in the previous sections have estimated proportions of titular speakers greater than 10 percent. However, there are regions with more dire linguistic situations with regard to the survival of their titular languages. Figure 4 presents data from the former autonomous okrugs of the Russian Soviet Federative Socialist Republic (Russia). While trends in these regions are divergent—in particular, the Buryat okrugs of Ust-Orda and Agin-Buriatia showed increasing proportions of Buryat speakers until they were merged with their neighboring oblasts—several of these regions show diminishing proportions of titular speakers. In particular, Khanti-Mansi has a proportion of Khanti speakers approaching zero percent.

4 Conclusion

In this paper I have proposed a model that can provide consistent measures of fluency in titular languages across post-Soviet regions and time; such data will be of great use for scholars interested in issues of language, identity and politics in the former Soviet Union.

Though the present iteration of the model shows great promise, there are numerous issues about which I would greatly appreciate feedback. I have listed the three issues which I consider most apparent, though I obviously welcome additional comments and criticism.

1. Adding additional data sources would greatly improve the measurement—any recommendations regarding cross-national or cross-regional data sources would be much appreciated. At present, my decision rule for including a measure is that it must have two or more associated observations, but ideally any data sources would have 10 or more observations (or would dovetail with existing measures).
2. The territorial boundaries and ethnofederal status of many territories have changed over time. At present, my approach to dealing with these changes is relatively *ad hoc*, and suggestions regarding how I could consistently account for these changes would be very helpful.
3. My measurement of language status, s , is very rough (i.e. I hierarchically cluster these estimates by regional ethnofederal status in 1989, and assume that status changed after 1993 in every region). Ideas about how to better operationalize status would be useful.

References

- Abrams, Daniel M & Steven H Strogatz. 2003. “Modelling the dynamics of language death.” *Nature* 424(6951):900.
- Arel, Dominique. 2002. “Interpreting Nationality and Language in the 2001 Ukrainian Census.” *Post-Soviet Affairs* 18(3):213–249.
- Colton, Timothy J & Jerry F Hough. 1998. *Growing Pains: Russian Democracy and the Election of 1993*. Brookings Institution Press.
- Dave, Bhavna. 2004. “Entitlement through numbers: nationality and language categories in the first post-Soviet census of Kazakhstan.” *Nations and Nationalism* 10(4):439–459.

- Dave, Bhavna. 2007. *Kazakhstan: Ethnicity, Language and Power*. Central Asian Studies
New York: Routledge.
- Denwood, Matthew J. 2016. “runjags: An R Package Providing Interface Utilities, Model
Templates, Parallel Computing Methods and Additional Distributions for MCMC Mod-
els in JAGS.” *Journal of Statistical Software* 71(9):1–25.
- Herrera, Yoshiko M. 2004. “The 2002 Russian Census: Institutional Reform at Goskomstat.”
Post-Soviet Affairs 20(4):350–386.
- Hirsch, Francine. 1997. “The Soviet Union as a work-in-progress: ethnographers and the
category nationality in the 1926, 1937, and 1939 censuses.” *Slavic Review* 56(2):251–
278.
- Jackman, Simon. 2009. *Bayesian analysis for the social sciences*. John Wiley & Sons.
- Laitin, David D. 1998. *Identity in Formation: The Russian-Speaking Populations in the
Near Abroad*. Ithaca: Cornell University Press.
- Laitin, David D. 2000. “What Is a Language Community?” *American Journal of Political
Science* 44(1):142–155.
- Onuch, Olga & Henry E. Hale. 2018. “Capturing ethnicity: the case of Ukraine.” *Post-Soviet
Affairs* 34(2-3):84–106.
- Plummer, Martyn. 2012. “rjags: Bayesian graphical models using MCMC.”.
- Radio Free Europe/Radio Liberty. 2018. “Siberia Activists Urge Government To Keep
Mandatory Yakut Lessons.”
URL: <https://www.rferl.org/a/siberia-activists-urge-government-to-keep-mandatory-yakut-lessons/29278777.html>
- Silver, Brian. 1974. “Social Mobilization and the Russification of Soviet Nationalities.” *The
American Political Science Review* 68(1):45–66.

The Moscow Times. 2018. “Russia Will ‘Vigorously Defend’ Rights of Compatriots Abroad, Putin Says.”.

URL: <https://www.themoscowtimes.com/2018/10/31/russia-will-vigorously-defend-rights-compatriots-abroad-putin-says-a63362>

Wickham, Hadley. 2009. *ggplot2: elegant graphics for data analysis*. Springer New York.