Effects of Conflict on Ethnic Identity and Trust: Evidence from the 2014 War in Eastern Ukraine

Valerio Sergio Castaldo

RoME Masters in Economics

EIEF and LUISS

Abstract

How do conflicts impact the ethnic identity of individuals living in multi-ethnic societies? We study the impact of the 2014 War in Eastern Ukraine on ethnic identification outside of the conflict areas. We show that Ukrainian citizens living in the regions with an ex-ante higher share of ethnic Russians, identify more as Ukrainian after it. Using survey data from the Kyiv International Institute of Sociology, we find that having a 10% higher share of ethnic Russians in 2001 leads to a 3 percentage point higher probability of self-identifying as only Ukrainian after 2014. On the contrary, in separatist-held areas of Donbas, we find a gradual decrease in Ukrainian identity. Our last set of results focuses on inter-ethnic trust, finding a decrease in trust toward ethnic Russians. These results suggest that a conflict can foster the process of building a national identity, or on the contrary, it can push a multi-ethnic country further into a conflict trap, by increasing polarization between the separatist and government-controlled areas.

Contents

1   Introduction

Ethnicity is a key component of an individual's identity (Lewis and Phoenix, 2004). It is related to one own self-perception, as well as the sense of belonging to the community surrounding him (Baumann et al., 2010). Perhaps, ethnicity has been underlined as one of the key factors leading to nation-building (Smith et al., 1986). While some argue that it is immutable, stemming from objective characteristics such as genetics, others claim that an individual has room to change his own self-perceived ethnic identity, for a variety of reasons (Dahis et al., 2019). In this paper, we contribute to a young branch of the literature studying how conflicts impact the ethnic identity and inter-ethnic trust of individuals living in multi-ethnic societies.

As an event study, we are going to use the 2014 War in Eastern Ukraine, between Russian-backed separatists and Ukrainian government forces. Rich availability of data from the Omnibus Survey, conducted by the Kyiv International Institute of Sociology (KIIS), allows us to study the phenomenon in the territories directly involved in the conflict, as well as in the rest of Ukraine. The survey is representative at the national level and is conducted four to six times a year. Using a difference-in-differences framework, we are going to show that following the outbreak of the war, the regions of Ukraine with a larger ex-ante share of ethnic Russians experience an increase in the probability of self-identifying as being only Ukrainian, as opposed to identifying as partly or only Russian. Moreover, we record a differential impact in the levels of inter-ethnic trust among the different regions of the country.

The dependent variable for our study is the individual's self-reported ethnic identity as asked by the Omnibus Survey. In addition to asking the respondent to choose his ethnicity among the possible nationalities living in Ukraine, the interviewee is also asked to compare how much he identifies as Ukrainian relative to Russian. The options that he can choose are: only Ukrainian, mostly Ukrainian, half Ukrainian and Russian, mostly Russian, or only Russian. Having this rich data allows us to study the dynamics of ethnic identification in a country with long-lasting cultural ties with Russia, and many individuals of mixed ethnic origin.

Our main specification includes as explanatory variables the share of ethnic Russians living

in Ukraine by region, interacted with a post-Crimea indicator, as well as a series of demographic controls including age, education, and characteristics of the place of residence. All the specifications that we use confirm an increase in Ukrainian ethnic identity located in those regions of Ukraine with a larger ex-ante share of ethnic Russians.

Following the outbreak of the 2014 War, the Omnibus Survey was still conducted in separatist-controlled areas of Donetsk oblast for a sufficiently long period of time; thanks to that, we are able to study its differential impact on the basis of which area the respondent was living in at the time of the interview. When doing that, we find that the separatist-held territories experience a gradual decrease in Ukrainian identity relative to government-held ones, while the Russian identity increases. While no pre-war data is available separately for these two areas, from the period-by-period specification it emerges that the difference in ethnic identification between them was small five months after the outbreak of the conflict, and that the gap diverges over the following months. This result leaves space for further research on the role of how media and war propaganda influence the perception of a conflict.

An empirical challenge to the research is potential reverse causality: namely that an increase in Ukrainian ethnic identity has preceded the war, and ultimately was responsible for its eruption. On the one hand, it is true that the 2013-2014 Euromaidan protests had the goal of making Ukraine more independent in its foreign policy from the Russian Federation, but on the other hand, the strongest participation in the protests was reported in the capital Kyiv and in Western Ukraine (Onuch and Sasse, 2016) and the increase in Ukrainian identity that we find is almost entirely concentrated in the southern and eastern regions of the country. What we claim instead, is that the Russian military intervention in Crimea and Eastern Ukraine, which was largely unexpected, has led many Ukrainians who identified to be at least partly Russian to reconsider their ethnic belonging. The exogeneity of the 2014 War to Ukrainian ethnic identification is supported by no pre-existing trends in the latter prior to April 2014.

We perform a series of robustness checks to support our results. First of all, we repeat the estimation by restricting the time frame considered for our analysis and dropping from our sample the regions directly involved in the conflict, for which our estimates may be less robust. Secondly,

we run again our main specification with an alternative measure of ethnicity asking respondents to report their identity from the set of all nationalities living in the country, obtaining similar results. In the last section, we directly address the possible limitations of our study, the main ones being survey-desirability bias and the possibility of a change in the sample composition over time. Specifically, we conduct a series of tests including verifying whether respondents hid their ethnicity more after the outbreak of the conflict and studying the role of war refugees. Lastly, we conduct a pseudo-panel analysis.

Our work is related to the literature on inter-ethnic conflicts. A large part of existing studies investigates how countries which are highly ethnically fragmented may be more likely to experience an armed conflict (Montalvo and Reynal-Querol, 2005). The causal mechanism in our paper is the opposite, claiming that ethnic identity may be endogenous to civil conflicts. Moreover, we contribute to the literature studying how conflicts affect inter-ethnic trust in multicultural societies. Ananyev and Poyker (2018) studies the effect of the 2012 Tuareg Rebellion on the ethnic versus national identity of Mali citizens using data from the Afrobarometer. Similarly, Rohner et al. (2013) study the impact of Uganda's war on terror on ethnic identity and inter-ethnic trust, using data from the same survey. Korovkin and Makarin (2021) use the Omnibus survey to document the sharp rise in anti-Russian attitudes in the less ethnic Russian regions of Ukraine and then uses firm data to study the effects of the Donbas war on trade between Russian and Ukrainian firms. The sudden shift in ethnic identification in Ukraine that we find is a novel result in the literature.

We also contribute to the literature on nation-building; Dell and Querubin (2017) studies how the U.S. bombing campaign in Vietnam lead to an increase in military and political activities, especially among communist militias; Gehring (2020) analyses how the Russian invasion of Crimea has strengthened the European identity of several Eastern European countries. Alesina et al. (2020) studies how countries requiring armies of conscripts, may adopt policies to foster nationalism and homogenize the culture of a country.

The paper is organized as follows. Section 2 provides background information on the situation in Ukraine up until the outbreak of the conflict. Section 3 describes the empirical strategy that we follow as well as the data that we use, with first descriptive evidence. Section 4 displays

the estimates from the difference-in-differences specification, with an interpretation of the results. Sections 5-6 contain robustness checks and a discussion of the possible limitations of our study. Section 7 concludes.

2 Background

2.1 The Ethno-Linguistic division of Ukraine

Historically, Ukraine has always been home to many different ethnic groups, with Ukrainians and Russians being the largest ones. According to the 2001 census, the two groups made up respectively 77.8% and 17.3% of the population of the country. Other minorities include Crimean Tatars, Greeks and Poles. Figure 1 displays the share of ethnic Russians in Ukraine by region. The Russian minority is largely concentrated in the eastern and southern regions of the country, with Crimea being the only region where Russians make up the majority of the population. In the eastern regions up to 40% of the population identifies as Russian, while in the south of the country the same figure stands at around 20%. The central and western regions of the country are those where Russians are less present, with up to 10% in the former and less than 5% in the latter. The Russian language is also widely spoken in Ukraine, with approximately 30% of Ukrainians considering it as their native language and a geographic pattern that follows that of ethnicity. Historically, this ethnolinguistic divide has also reflected a political division of Ukraine, with those living in the south-eastern regions mostly voting for Pro-Russian candidates, and those in the eastern and central regions of the country voting for Pro-European ones.

2.2 The Euromaidan Protests and 2014 Donbas War

In November 2013, the Ukrainian government led by Viktor Yanukovich, refused to sign an Association Agreement with the European Union, in favour of closer ties with Russia. This led to massive protests throughout the country, especially in the capital of Kyiv and the western regions. With instability climbing, in February 2014 Yanukovich left the country, and a pro-European government

Figure 1: Share of ethnic Russians in Ukraine

Notes: The figure shows the share of ethnic Russians in Ukraine by regions (oblasts). The data used to construct the figure come from the 2001 nationwide Ukrainian census

was formed. In response to the events in Kyiv, Russia decided to take control of Crimea in March 2014, with the pretext of protecting the Russian minority living there. Soon, a referendum on the independence of the region was organized, and eventually, Russia annexed it. The referendum was met with condemnation from Ukraine and the international community, which did not recognize the vote.

At the same time in several regions of Southeastern Ukraine, many protested in favour of the old government and of Russia, with rallies often turning violent. This is especially true in the two border regions of Donetsk and Luhansk (the Donbas region) in the East of the country, where pro-Russian separatists went as far as occupying the regional administration buildings. In April 2014, wishing to follow the path of Crimea, a referendum on the independence of the regions was held, but this time the newly-formed Kyiv government announced the launch of the *Anti-Terrorism Operation - ATO*, with the aim of stopping a further spread of the separatist movement. Russia covertly supported the rebels in the Donbas but did not intervene directly (Kofman et al., 2017). After a war that caused thousands of casualties and millions of refugees, a ceasefire was proclaimed in early 2015. The conflict had largely remained frozen with the boundaries shown in Figure 2 until February 2022, when Russia launched a full-scale invasion of Ukraine.

Figure 2: Crimea and the Donbas region

Notes: The figure shows in red Crimea and the territories controlled by pro-Russian separatists as of February 2015. The oblasts of Donetsk and Luhansk (i.e. the Donbas region) are shown in orange, while the regions of Ukraine not involved in the conflict are shown in yellow.

2.3   Attitudes toward Russia

The events in Crimea and the Donbas have led many Ukrainians to reconsider their view of Russia. This is true even for those who self-identified as ethnic Russians. Following Korovkin and Makarin (2021), in Figure 3 we plot the share of Ukrainians with a very negative opinion of Russia, by ethnicity. After the outbreak of the conflict, which corresponds to the vertical line, the share of ethnic Ukrainians with such a view jumped to 20-40%, while the same figure for ethnic Russians to 5-10%. With the cessation of hostilities in early 2015, we observe some reversion to the mean in the attitudes toward Russia, but as of 2017, the level remains very high, especially among ethnic Ukrainians.

Figure 3: Attitudes toward Russia
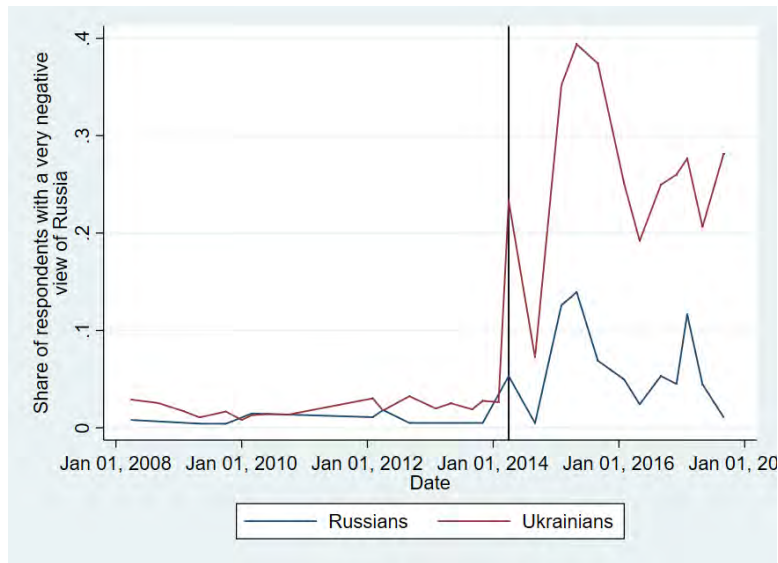
Notes: This figure shows how the attitudes toward Russia have deteriorated as a consequence of the 2014 conflict. The Y-axis plots the share of survey respondents with a very negative view of Russia, and the vertical line corresponds to April 2014, when the Donbas conflict started. Data come from the Omnibus Survey, conducted by the Kyiv International Institute of Sociology.

3 Empirical strategy and data

3.1 Data

Our main data source is the *Omnibus Survey* conducted by the Kyiv International Institute of Sociology - *KIIS*. The survey has been conducted from the late 1990s four to six times a year and asks participants to report their opinions on the current political events shaping the country, as well as the foreign policy that they would wish Ukraine to follow. For each respondent, the survey collects a series of demographic characteristics including age, education, ethnicity, and region of residence. The time span that we use for our analysis is that of 10 years, from February 2007 to the most recent wave of February 2017. The interviews were conducted face-to-face at the respondent's place of residence. Refugees from the Donbas were also included in the Survey, however as we will see in detail in Section 6.2, their share in the sample is negligible and under-represents the

10

official statistics about the share of refugees in Ukraine. Data for the region of Crimea and for the territories of Luhansk oblast controlled by separatist forces are unavailable from March 2014 onwards; therefore we exclude them from our analysis. On the contrary, the survey was conducted also in the separatist-held areas of the Donetsk region for a certain period of time after the outbreak of the conflict. The main variables of interest for our study are ethnicity and inter-ethnic trust. Participants to the survey are asked to report their ethnicity in two separate questions: the first one requires to choose among the various nationalities living in the country, and the second one asks respondents to compare how much they feel Ukrainian relative to Russian. The five possible values for this variable are: only Ukrainian, mostly Ukrainian, half Ukrainian and Russian, mostly Russian, or only Russian. We are going to conduct most of our analysis on ethnicity using this second question, but the results are robust to using this first variable as well. Other data sources that we use include the 2001 National Ukrainian Census which contains the region-by-region share of ethnic Russians living in the country. Lastly, we use data about the number of refugees from the Donbas, and estimates of Ukraine's population after the year 2001[1].

3.2 Baseline specification

The goal of the research is to study the effects of the 2014 conflict on the ethnic identification of Ukrainian citizens, as well as their levels of inter-ethnic trust. In this section we will focus on ethnicity, while in Section 4.4 we will talk about trust. In order to conduct our analysis, we will employ a difference-in-differences methodology, considering as treated units those individuals living in regions with a high share of ethnic Russians (mostly eastern and southern regions, see Figure 1), versus those living where Russians make up a low share of the population (mostly central and western regions). Because of the relatively high homogeneity of the latter regions, we expect the ethnic identification of individuals living there to be largely unaffected by the conflict. The main

---

[1] Data about refugees come from the Ministry of Social Policy of Ukraine, distinguishing whether they come from Crimea or from the Donbas region, while the latter estimates of Ukraine's population come from the State Statistics Service of Ukraine.

specification that we are going to use throughout our analysis is of the type:

$$Y_{irt} = \alpha_r + \gamma_t + \beta_r * Post_t * Rus_r^{2001} + X_{irt} + \epsilon_{irt} \qquad (1)$$

The dependent variable Y_{irt} is a measure of ethnic identification of individual i living in region r in period t. The regression is run five separate times, with the dependent variable being equal to 1 for each of the five possible choices of Ukrainian vs Russian identity, and zero for the others. α_m is a region fixed effect and γ_t is a period fixed effect. Lastly, we add a series of individual controls X_{irt} which include gender, age and education, as well as whether the respondent comes from an urban or rural area. Provided that the parallel trends assumption holds, the coefficient β will measure the impact of having a higher share of ethnic Russians in region r in 2001, on the probability of self-identifying as the outcome variable Y_{irt} after the outbreak of the 2014 conflict. Our main alternative specification directly interacts the variable $Post_t$ with individual Ukrainian regions or macroregions. In this case, the coefficients β_r will capture the differential effect of the conflict on outcome Y_{irt} for individuals living in treated regions with respect to control regions. Lastly, we employ a multiple-periods DiD interacting the variable Rus_r^{2001} with each period considered in our analysis.

3.3   Descriptive Statistics

Table 1 describes the average demographic characteristics of survey respondents, distinguishing the outcome by macroregion. The first five rows focus on the ethnic composition of the sample and show how the vast majority of the respondents in Western and Central Ukraine identify as only Ukrainian. At the same time, the same figure is much smaller in the eastern and southern regions of the country, due to a significant share of respondents identifying as partly or wholly Russian. Next, we notice how respondents from the southern and eastern regions come from relatively more urban areas than their central and especially western counterparts. The last three rows focus on the demographic characteristics of the respondents, showing how the gender composition, age, and education levels are similar across the four regions.

Table 1: Average Characteristics of Survey respondents

	(1) West	(2) Center	(3) East	(4) South
Only Ukrainian	0.91	0.89	0.38	0.58
Mostly Ukrainian	0.02	0.04	0.13	0.10
Half Ukrainian and Russian	0.02	0.03	0.21	0.13
Mostly Russian	0.01	0.01	0.09	0.06
Only Russian	0.01	0.02	0.16	0.07
Urban Rural	0.49	0.65	0.85	0.74
Sex	0.39	0.38	0.37	0.37
Age	47.37	49.68	49.57	49.06
Education	5.51	5.52	5.52	5.52
N	21570	28724	19500	20700

Notes: This table shows the average characteristics of survey respondents distinguishing the outcome by macroregion (west, center, east and south). The first five rows focus on ethnicity, while the last ones report demographic characteristics as well as the type of settlement where the respondent lives.

4 Results

4.1 Before and After Changes in Ethnic Identification

The first piece of evidence that we present is among the biggest motivating factors for our analysis and shows how the fraction of survey respondents identifying as Ukrainian relative to Russian has increased sharply after 2014. In Figure 4 we report the before and after change in the distribution of the above-mentioned five possible outcomes of ethnic identification, at the aggregate level. What we find is that after the outbreak of the conflict, the distribution of the answers is more skewed to the left, meaning that as a whole the respondents have chosen to identify more with a Ukrainian identity, and less with the Russian one.

In addition to looking at the before-and-after change, in Figure 5 we plot the trend over time of the same variable. In particular, we focus on the fraction of individuals identifying as only Ukrainian, but similar results hold for the other four possible outcomes. The graph shows how disaggregating the results by macroregion, the increase in the share of only Ukrainians has occurred almost entirely in the eastern and southern regions of the country. In the former before

13

Figure 4: Change in the ethnic composition of respondents
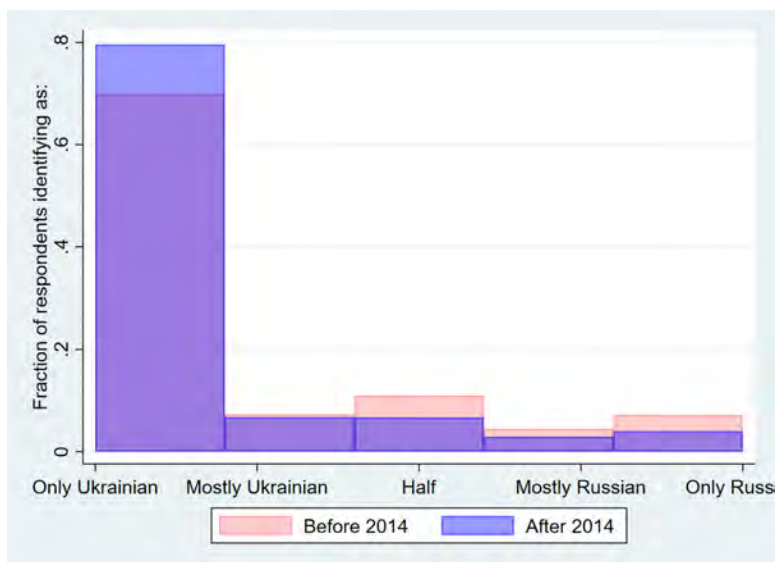
Notes: The figure shows how the distribution of the answers to the question "how do you identify yourself ethnically?" has changed after the outbreak of the 2014 conflict. The distribution in blue is constructed using observations of the period 2014-2017, while the one in red is constructed using the 2007-2014 ones.

the conflict, this share oscillated between 35-40%, while after the conflict the quantity has almost doubled. Similarly for Southern Ukraine, the share jumped from 50% to 70%. On the contrary, Central and Western Ukraine don't observe any jump, since they begin with an already high level of the population identifying as such. Moreover, in this picture we see how the parallel trend assumption is verified across the four macroregions, with no pre-existing trends in ethnic identification prior to the outbreak of the conflict, especially in the eastern, western and central regions. The same holds for the other four possible outcomes of the question (mostly Ukrainian, half Ukrainian and Russian, mostly Russian and only Russian), reported in Figure A.1 in the Appendix. This preliminary evidence confirms that a difference-in-differences specification is appropriate to study this phenomenon. In the next section we will obtain additional evidence on the parallel trends assumption using a regression specification.

14

Figure 5: Trend in the ethnic composition of respondents by macroregion
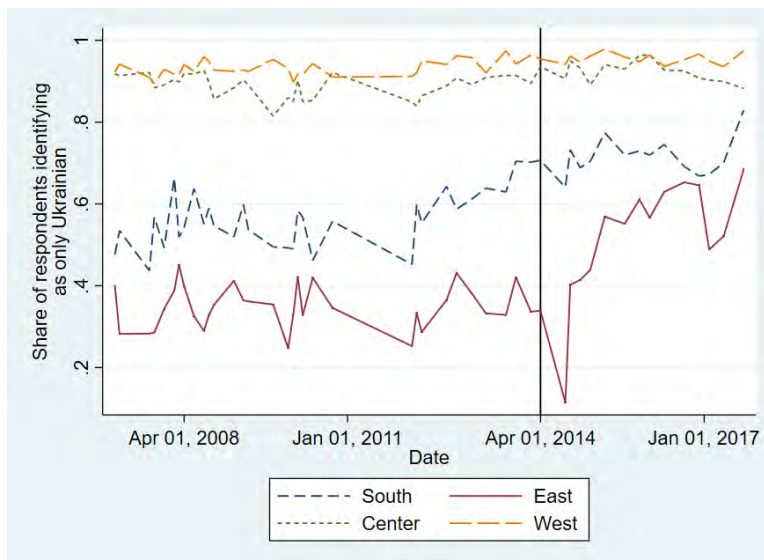
Notes: In the figure we plot the share of respondents identifying as only Ukrainian over time, distinguishing the outcome by macroregion. The vertical line corresponds to April 2014, when Donbas war started.

4.2 Main specification

We begin by estimating Equation (1), reporting the results in Table 2. The main coefficient of interest is the interaction term between $Post$ and Rus_r^{2001} and is displayed in the first row of the Table. As mentioned before, Equation (1) is run five separate times, corresponding to the five possible outcomes of ethnic identification of our main variable of interest. In Column (1) we report the estimated increase in the probability of self-identifying as only Ukrainian: having a 10% higher share of ethnic Russians in the region in 2001, leads to a 3 p.p. higher probability of self-identifying as only Ukrainian after the outbreak of the conflict, controlling for demographic characteristics of the respondent. At the same time, in Columns (3-5) we observe a reduction in the probability of self-identifying as half Russian and Ukrainian, mostly Russian and only Russian. No change is reported on the probability of self-reporting as mostly Ukrainian. Controlling for years of education, it emerges that individuals who spent more years studying are those with a slightly higher probability of self-identifying as Russian, and the same holds for individuals living in urban areas as opposed

15

to rural ones. This is in line with the history of many post-Soviet countries, where big cities had a higher concentration of ethnic Russians, while rural towns and villages were mostly inhabited by local ethnic groups. Region and survey wave fixed effects are included.

The next regression that we run is a multiple-periods DiD, where we interact the variable Rus_r^{2001} with each period considered in our time span. The results of this new estimation with the dependent variable being the probability of self-identifying as only Ukrainian, are shown in Figure 6, and confirm the plausibility of the parallel trends assumption. Indeed, the estimates are largely non-significant before the vertical line, which corresponds to the outbreak of the conflict, while after it, they become positive and statistically significant. This suggests the absence of pre-existing trends in the probability of self-identifying as only Ukrainian, and a significant change in it after the outbreak of the conflict.

Figure 6: Estimates of the multiple-periods DiD
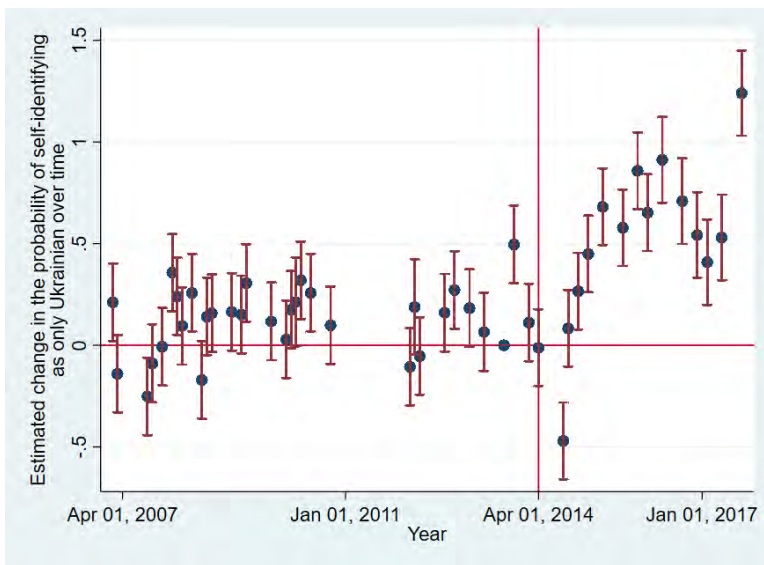
Notes: this figure plots the estimates of a modified version of Equation 1 which includes an interaction of the variable Rus_r^{2001} with individual periods. The omitted period is September 2013, and the vertical line corresponds to April 2014. The estimates show how no pre-existing trends in ethnic identification existed prior to the outbreak of the conflict in the east of the country in 2014.

The last part of this section is dedicated to the estimation of our main alternative specification, in

16

Table 2: Estimated change in the probability of self-identifying as:

	(1) Only Ukrainian	(2) Mostly Ukrainian	(3) Half	(4) Mostly Russian	(5) Only Russian
Post*Rus 2001	0.340***	-0.004	-0.179***	-0.037***	-0.143***
	(0.023)	(0.015)	(0.016)	(0.011)	(0.013)
Sex	-0.003	-0.005**	-0.002	-0.001	0.004*
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
Age	-0.001***	0.000	0.000	0.000***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	-0.010***	0.001	0.003***	0.004***	0.004***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Urban	-0.081***	0.011***	0.030***	0.012***	0.025***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
Region FE	Yes	Yes	Yes	Yes	Yes
Survey Wave FE	Yes	Yes	Yes	Yes	Yes
N	89904	89904	89904	89904	89904
$\bar{y}$	0.713	0.069	0.091	0.038	0.058
$\sigma(y)$	0.452	0.253	0.287	0.191	0.234

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: this Table displays the baseline estimates of Equation (1). The dependent variable of the regression corresponds to the five possible choices of ethnic identification: only Ukrainian, mostly Ukrainian, half Ukrainian and Russian, mostly Russian, only Russian. The regression is therefore run five separate times, with the dependent variable being equal to 1 for each single of the five values, and 0 for the other options. The main independent variable is the share of ethnic Russians in 2001 by region, interacted with an indicator for *Post Crimea*. Other control variables include demographic characteristics of the respondent, as well as the type of settlement where he/she is living in.

which the variable $Post_t$ is interacted with individual Ukrainian regions. The estimates are reported in Figure 7. As before, we decided to report estimates for the probability of self-identifying as only Ukrainian, but the results for the other outcomes are in line with what we have obtained using our main specification. Figure A.2 in the Appendix reports the precise magnitude of coefficients and standard errors for each region. The strongest effect is reported for regions in the eastern and southern parts of the country, which in line with our previous result, are the ones with the highest ex-ante share of ethnic Russians. The regions of Luhansk, Odesa, and Zaporizhzhya have an estimated increase of 0.20-0.40, while the remaining regions in Southeastern Ukraine experience an increase of a magnitude between 0.05-0.20. No significant increase is reported for the western and central regions. Interestingly, the Donetsk oblast reports an estimated coefficient which is non-significant, despite being directly involved in the separatist conflict of 2014, as well as being home to a large Russian minority (38% of the population). At the same time, the Luhansk oblast which was also directly involved in the conflict, experienced one of the strongest estimated increases. A possible explanation for this puzzling result, can originate from the fact that in the latter region the Omnibus Survey was conducted only in the territories controlled by the Ukrainian government, while the KIIS managed for a certain period of time to interview respondents from all the territory of the Donetsk oblast, including the separatist-held areas. Supporting evidence to this claim is obtained by keeping only observations from the government-held part of the Donetsk oblast, and running the same specification again. By doing that, the estimate for that region becomes positive and statistically significant (the result of this new estimation is reported in Figure A.3 in the Appendix). In the next section we are going to investigate how the inclusion of observations from separatist-held areas has affected the overall ethnic identification of respondents coming from Donetsk oblast.

4.3 Separatist vs Loyalist Control

In this section, we are going to try to catch a possible differential impact of the 2014 Conflict on the two halves of the Donetsk region: the one controlled by the Ukrainian government, and the part held by pro-Russian separatist forces. The period that we consider for this section goes from

Figure 7: Region-by-region estimates
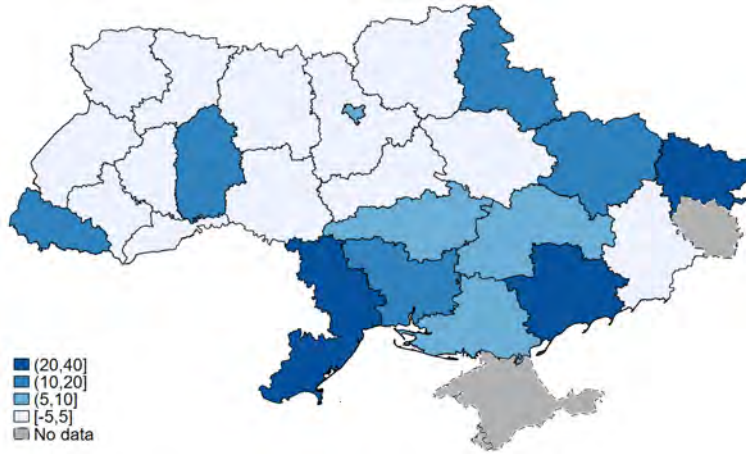
Notes: in this figure we report the estimates of our alternative specification, corresponding to Equation (1) but with the variable $Post_t$ interacted with individual Ukrainian regions. The dependent variable is an indicator variable equal to 1 when the respondent identifies as only Ukrainian, and 0 otherwise. Darker colours correspond to a stronger increase in probability. The region of Crimea and the territory of Luhansk oblast controlled by separatist forces are omitted since the Survey has not been conducted in these areas since April 2014.

September 2014 to February 2016. The former date is given by the fact that observations recorded before it lack information on which half of the Donetsk region they come from,[2] while the latter one is given by the decision of the KIIS to stop conducting interviews in the separatist-held territories of the region. The specification that we are going to employ is the following multiple-periods DiD:

$$Y_{it} = \gamma_t + \beta_t * Period_t * Separatist\ held + X_{it} + \epsilon_{it} \qquad (2)$$

The treated units in this framework are those respondents living in separatist-held areas, whose effect is estimated in deviation from that of individuals living in territories controlled by the Ukrainian government. The coefficients of interest are those multiplying the interaction between a dummy for treated individuals, and each period considered in the analysis. The results of this estimation are reported in Figure 8 and are virtually the opposite of what we have obtained so far, with respondents from the separatist-held areas experience a decrease in the probability of self-

---

[2]Geographical indicators for individual cities are not precise enough to let us recover those areas before 2014.

19

identifying as only Ukrainian. Moreover, the estimated coefficients begins from a relatively small level in September 2014 (-0.137) and decreases in time reaching (-0.460) in February 2016. This decrease suggests a similar level of ethnic identification between the two halves of the region before the outbreak of the conflict, and divergence in time between them. This analysis may be able to explain why on aggregate the estimates for Donetsk oblast reported in Figure 7 are insignificant. The results that we obtain however, have to be interpreted carefully, for a variety of reasons. First of all, the estimates are the result of a comparison that takes place only *after* the outbreak of the conflict, since as explained before, we lack the pre-war values for the variable separatist-held. This implies that we don't have to interpret the decrease in Ukrainian identity in absolute terms, but just show that there has been a divergence with the part of the region controlled by pro-Ukrainian forces. The second possible reason to cast doubts on the estimates is given by risks of change in the sample composition over time. This is because following the outbreak of the conflict, up to one-third of the pre-war Donbas population fled their homes, with a selection rule that is unlikely to be exogenous to ethnicity. Indeed, those individuals fleeing the war zones are the ones who are unwilling to fight for the pro-Russian cause, and hence are likely to identify ex-ante more as Ukrainian. Therefore, we expect a higher concentration of ethnic Russians among the individuals remaining in the separatist-held areas of the region. Nonetheless, despite this last sort of bias, our result is still interesting from a policy point of view, as it would indicate that the Ukrainian government will find it harder to regain control over the region in the future, given that the population which has remained in these areas identifies on average as *less* Ukrainian. Lastly, an additional source of bias comes from the fact that social desirability bias, which leads to the misreporting of one own's real opinion, is likely to be stronger among the individuals living in war zones. In Section 6 we are going to talk in detail about the limitations of our results.

4.4 Inter-Ethnic Trust

Our last set of results focuses on the second part of the research question, which asks whether the 2014 Conflict had an impact on the level of trust among the various ethnic groups living in Ukraine. Respondents to the Omnibus Survey are asked how much they would trust an individual

Figure 8: Decrease in Ukrainian identity over time in separatist-held territories
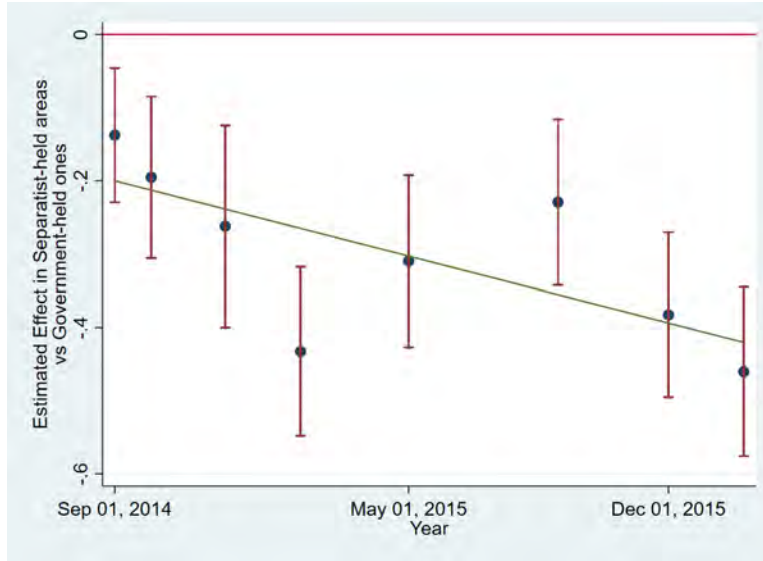
Notes: this Figure shows how respondents from separatist-held territories of Donetsk oblast experienced a decrease in their probability of self-identifying as only Ukrainian, relative to their counterparts living in government-held areas of the same region. The period considered goes from September 2014 to February 2016.

belonging to a certain ethnic group. The possible outcomes in which the respondent can regard such an individual are: as a member of your family, as a close friend, as a neighbor, as a work colleague, as a resident of Ukraine, as a guest of Ukraine, and not allowing such an individual into the country. In our analysis, we are going to focus on the two largest minorities living in Ukraine: ethnic Russians and Russian-speaking Ukrainians. Firstly, in Figure 9 we show how the average opinion that respondents have of ethnic Russians has decreased in all the macroregions of Ukraine. Figure A.4 in the Appendix replicates the result for trust toward Russian-speaking Ukrainians.

In addition to providing a descriptive evidence, we investigated the impact of the conflict on inter-ethnic trust with a modified version of our main regression specification:

$$Y_{irt}^e = \alpha_r + \gamma_t + \beta^e * Post_t * (1 - Rus_r^{2001}) + X_{irt} + \epsilon_{irt} \tag{3}$$

In this framework the outcome variable Y_{irt}^e measures the degree of trust that individual i living

21

Figure 9: Decrease in trust toward ethnic Russians
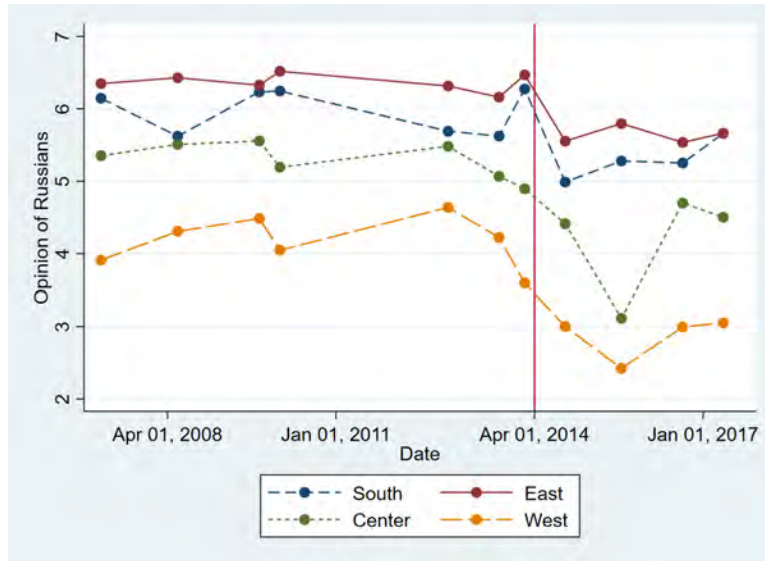
Notes: this Figure shows how in the aftermath of the 2014 Conflict trust levels of Ukrainian citizens toward ethnic Russians has decreased in all Ukrainian macroregions. The vertical line corresponds to the outbreak of the conflict.

in region r at time t has toward ethnicity e. The variable $Post_t$ is interacted with the inverse of the share of ethnic Russians living in region r as of the 2001 Census. This modification is due to the fact that our goal is to show how regions with a lower share of Russians in 2001, have experienced the largest decrease in trust toward ethnic Russians. The results are shown in Table 3, and are in line with this intuition. The first row reports estimates for the interaction coefficient, with a significant drop in trust toward ethnic Russians reported in Column (1). At the same time, in Column (2) we show that the decrease in trust toward Russian-speaking Ukrainians is not statistically significant.

In order to better grasp the territorial evolution of the impact, we report estimates for our main alternative specification, where we interact the variable $Post_t$ with individual Ukrainian macroregions. The results are shown in Table A.1 in the Appendix. The omitted macroregion in this case, is the eastern one, since as seen in Figure 9, it was the part of the country where the levels of trust changed the least of all. The same holds for trust toward Russian-speaking Ukrainians, see Figure A.4 in the Appendix. What we expect from the following analysis is that the western part

22

Table 3: Change in trust towards people identifying as

	(1) Russians	(2) Russian-speaking Ukrainians
Post*(1-Rus 2001)	-2.444***	-0.353
	(0.209)	(0.189)
Sex	-0.054*	-0.074***
	(0.025)	(0.022)
Age	0.004***	0.001
	(0.001)	(0.001)
Education	0.028***	0.019**
	(0.007)	(0.006)
Urban	0.135***	0.129***
	(0.028)	(0.026)
Region FE	Yes	Yes
Survey Wave FE	Yes	Yes
N	21339	21384
$\overline{y}$	5.043	5.690
$\sigma(y)$	2.086	1.751

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: this table reports estimates for Equation 3 which aims at estimating the impact of the 2014 conflict on inter-ethnic trust in Ukraine, for regions with a smaller fraction of ethnic Russians in 2001. The dependent variable is a measure of trust toward ethnicity e, while the independent variables include the usual demographic characteristics included in previous equations

of the country, which was already the one with the lowest levels of trust toward ethnic Russians, has further polarized its opinion in a negative direction, experiencing a divergence with the rest of the country. Table A.1 contains the results of this estimation; Column (1) reports the estimated change in trust toward ethnic Russians and shows that the Central and Western parts of the country experienced the strongest decrease in magnitude. Southern Ukraine on the other hand, experienced a decrease in trust levels of the same magnitude as the eastern one. Column (2) reports the estimated change in trust levels toward Russian-speaking Ukrainians and again, the estimates are non-significant.

4.5 Interpretation of the results

Overall, these results confirm that after the outbreak of the 2014 War, the probability for a respondent of self-identifying as only Ukrainian has increased, while the probability of identifying as wholly or partly Russian has decreased. The estimates are stronger in the regions with an ex-ante higher share of ethnic Russians. On the contrary, in separatist-held areas, we have found a *decrease* in Ukrainian identity, which suggests a polarization between the latter territories, and ones held by Ukrainian government forces. Lastly, we have seen how the levels of trust towards ethnic Russians have decreased in the whole country, while we don't find evidence for a causal impact of the conflict on levels of trust toward Ukrainians of Russian language.

In order to be able to interpret the estimates in a causal sense, in the next Section we are going to conduct a series of robustness checks, while in Section 6, we are going to talk about the possible limitations of our data, and what this implies for the interpretability of the results.

5   Robustness Checks

In this section we are going to replicate the results obtained in Section 4 by changing the set of observations considered in our analysis. The variations include a restriction of the sample in time, and the dropping of regions for which the results may be less robust. Moreover, we will try to replicate the results by using an alternative measure of ethnic identification from the survey.

5.1 Restricting the time frame

The first robustness check that we perform involves dropping all the observations after February 2016. The rationale behind this choice is given by the decision of the KIIS to stop conducting surveys in the territories controlled by separatist forces after that date. Recalling the results found in Section 4.3, claiming that the conflict had a differential impact on ethnic identification in the two halves of the Donetsk oblast, by dropping only the separatist-held part of the region from our sample, we are likely to obtain a bias in our estimates. Therefore, we repeat our analysis keeping only the observations up to February 2016. By doing that, we are unlikely to lose too much valuable information, given that the time frame would still contain observations from the period of highest fighting intensity[3], and we can assume most individuals to have already changed their ethnic identity by that time. Having said that, the results of this new estimation are reported in Table A.2 in the Appendix and are largely consistent with what we have found before. Interestingly, the DiD coefficient gets reduced in magnitude (from 0.34 to 0.22). A possible explanation for this reduction, is given by the fact that the observations which we dropped included only the government-held part of Donetsk oblast, which as we have seen contained a higher share of ethnic Ukrainian respondents. Moreover, by looking at Figure 5, we see that the share of individuals defining as only Ukrainian keeps increasing after February 2016, and therefore by dropping observations after that date we fail to capture this further increase in Ukrainian identity. Therefore, if we neglect these phenomena we obtain a positive bias in the coefficient.

5.2 Dropping the conflict regions

The second test that we perform consists of dropping from our sample the whole regions of Donetsk and Luhansk. The reason for doing that is the possibility of biases in the responses coming from those territories, as a consequence of the ongoing conflict. The main limitations, as we will discuss in detail in the next section, are the social desirability bias and changes in sample composition over time. In fact, as already mentioned before, up to one-third of the pre-war Donbas population has

---

[3]The Donbas war officially started on April 15th 2014, when the Ukrainian Government announced the launch of the *Anti-Terror Operation* in the Donbas. The majority of clashes ended in February 2015, with the signing of the Minsk II memorandum and the implementation of a demilitarized zone, and a ceasefire.

fled the area due to the conflict, likely causing the sample composition of respondents coming from those regions to be changing over time. Moreover, respondents from the area may be unlikely to express their true opinion due to the fear of negative repercussions. For these reasons, we decided to run again our main specification excluding from the sample the regions directly involved in the conflict. The results of this further check are shown in Table A.3 in the Appendix. The estimates of the DiD coefficient more than doubles in size, from 0.34 to 0.76. This last result suggests a stronger increase in Ukrainian ethnic identity *outside* of conflict areas; indeed the Donbas region, despite hosting a much larger fraction of ethnic Russians compared to the rest of mainland Ukraine, has not experienced an increase in Ukrainian ethnic identity that is larger than other regions in the southeastern part of the country. This is especially true for respondents coming from the separatist-held areas of the Donetsk oblast that as we have seen, have even experienced a decrease in Ukrainian identity.

5.3 Alternative measure of ethnicity

The last robustness check that we perform involves using an alternative measure of ethnic identification i.e., the question of the survey asking respondents to choose their own ethnicity from a set of choices which includes all the nationalities living in Ukraine. In addition to Ukrainian and Russian, the respondent can choose to identify as Belarusian, Crimean Tatar, Polish, and so forth... Since we are interested in the pattern of the Ukrainian and Russian ethnicities, we will focus our analysis only on them. The dependent variable in this framework is a binary indicator equal to 1 if the respondent identifies as Ukrainian, and 0 if Russian. The region-by-region estimates are shown in Figure A.5 in the Appendix and are largely consistent with the results obtained until now. The magnitude of the coefficients is significantly reduced relative to the baseline case in Figure A.2, with the largest estimated increase in probability being at most 0.15. Moreover, the variability of the data is increased, with larger standard errors. Interestingly, estimates from our main specification are insignificant, unless we drop the conflict regions of Donetsk and Luhansk.

6   Possible limitations

The last section of the paper is dedicated to commenting on the nature of the data, and the possible limitations that it can imply. Moreover in this part, we will conduct a series of final tests and verifying assumptions with the goal of making our results more robust and credible. In particular, the aggregate increase in the share of respondents defining as only Ukrainian, which we showed in Figure 5 can actually be the result of three distinct mechanisms: an inner change in ethnic identification of the respondents, a social-desirability bias and a change in sample composition. The first mechanism claims that as a result of the Russian military intervention in Ukraine, and their encroachment on Ukrainian affairs, a part of those Ukrainian citizens who identified as ethnic Russian (or at least partly Russian), have reconsidered their inner identity, up to a point where they changed their self-perceived ethnicity. Ideally, this first mechanism is the one that we would like to isolate in our analysis, however, we will be able to do it only partially due to the possible biases coming from the two other mechanisms.

6.1   Social-desirability bias

The first of these possible limitations is the social-desirability bias, which is defined as a sort of response bias where participants of a survey avoid answering truthfully to certain questions and instead decide to do that in a way that would be socially acceptable at the moment of the interview (Krumpal, 2013). In the case of the Omnibus Survey, a possible factor contributing to this sort of bias is the face-to-face nature of the survey. On the one hand, this factor reassures us of the quality of the data, but on the other hand, it can lead respondents to hide or misreport their true beliefs regarding burning topics at the moment of the interview. In the case of Ukraine, after the beginning of the pro-Russian insurgency in the east of the country, it is possible that identifying as ethnic Russian may have become a sort of *taboo* in the country. As a consequence of that, some of the participants in the survey may have decided to hide or misreport their ethnicity due to the fear of negative consequences.

As already mentioned earlier, this misreporting phenomenon is likely to be stronger in the

27

Donetsk and Luhansk oblasts, given that territories involved in armed clashes are more likely to experience cases of politically-motivated violence. Indeed, several cases of human rights violations including arbitrary arrests, have been reported in the Donbas region since the start of the first armed clashes in mid-2014 (OHCHR, 2015). The extreme case of survey desirability bias in this situation is given by a large fraction of ethnic Russians misreporting their ethnicity, or abstaining from participating in the survey after the outbreak of the 2014 conflict. While we can't control for the first, without having available data which would allow the respondent a higher degree of anonymity, we could ideally control for the second one by looking at the evolution of the survey response rate over time. Indeed, if it's true that following the outbreak of the Donbas conflict in 2014 the ethnic Russian population of Eastern and Southern Ukraine has refused to participate in the survey, then the response rate for these regions would drop significantly. In that case, we would have to reconsider the causal interpretation of our results. Despite this being the ideal situation, due to the ongoing 2022 War in Ukraine, the KIIS has limited access to its databases, and hence we are not able to use them at the current moment. Any attempt at testing in this framework has to be limited to the Omnibus Survey data collected before the 2022 Russo-Ukrainian war.

A more feasible version of the social-desirability bias which we can control for is the possibility of respondents trying to hide their nationality. We can verify that, by looking at whether the number of missing values for the Omnibus Survey question about ethnicity has gone up after the outbreak of the conflict. This is what we do in Figure 10 for the regions where we have previously found a stronger increase in the probability of self-identifying as Ukrainian. A visual inspection of the trends, rejects this possibility, finding no evidence of spikes in the number of missing values after the outbreak of the 2014 conflict. This evidence, despite being limited by data availability, is partially reassuring about the possibility of strong social-desirability bias among respondents to the Omnibus Survey.

6.2 Change in sample composition over time

The second main limitation in our analysis derives from the lack of a panel structure in our dataset. Perhaps, the way in which the latter has been constructed was by appending together the results of

Figure 10: Time-series of missing responses to the question: "How do you define yourself ethnically?"
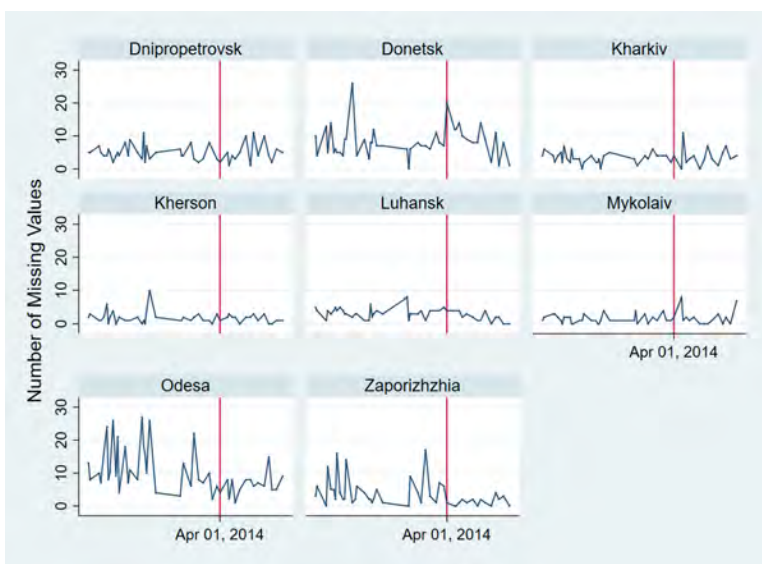

Notes: the figure plots the number of missing values to the question from the survey asking the individual to report his/her own ethnic identity. The 8 panels correspond to the regions with a significant share of ethnic Russian population in 2001. The red line corresponds to the outbreak of the 2014 Conflict.

the different waves of the Omnibus Survey, implying that respondents were likely not the same across survey waves. As discussed above, we cannot directly verify whether ethnic Russian individuals have decided to participate less in the survey after the outbreak of the conflict, but we can still test the homogeneity of the sample composition in other ways.

In Figure 11 we show how the share of respondents from each macroregion has remained constant over time, and the same holds for the type of settlement, as well as the demographic characteristics of age and education. The second vertical line in the top-left panel corresponds to February 2016, when the KIIS decided to stop conducting interviews in the separatist-held territories of the Donbas region, and this is why following that date the share of individuals coming from the various macroregions changes. However, as we have seen in Section 5.1, the results are robust even by dropping all the observations after that date.

The second part of this section is dedicated to studying the role of refugees from the Donbas,

29

Figure 11: Share of respondents by macroregion, age, education and urban/rural settlement
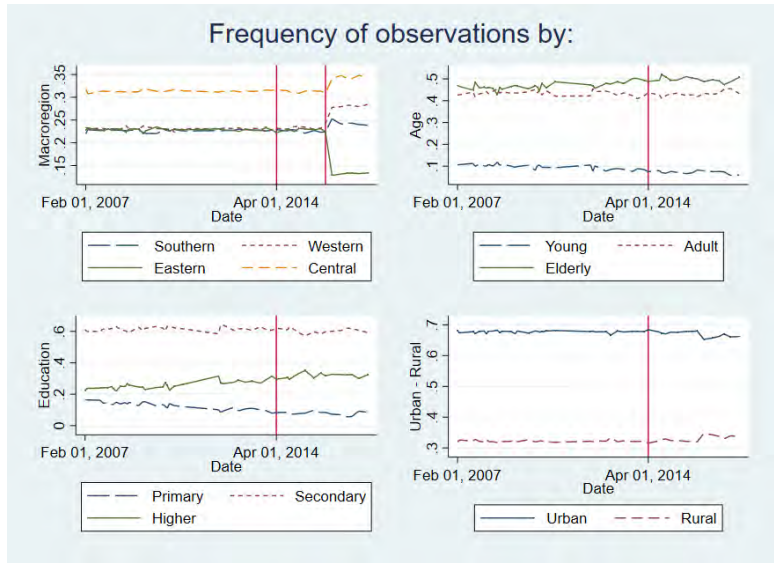

Notes: this Figure shows how the sample has remained balanced over time on the basis of demographic characteristics of respondents (age, education), as well as by the region of residence. The Y-axis plots the share of respondents from a certain category over time, and the vertical line corresponds to the outbreak of the conflict.

and their possible impact on the sample composition of the various regions. This is because refugees have been included in the Omnibus Survey, and were registered as residents of the region where they are currently living. As already mentioned before, after April 2014 up to one-third of the pre-war Donbas population has fled their homes; of them, one-third took shelter in the territory of the Russian Federation, and two-thirds became internal refugees. The above distinction is important as it's unlikely that the selection mechanism of where to relocate as refugees, is random; perhaps, we can presumably assume that those who chose to remain inside the territory of Ukraine, as opposed to fleeing to Russia, are those individuals who disagreed the most with the pro-Russian rebellion, and hence were also more likely to identify ex-ante as Ukrainian. In our analysis, we are going to focus on this latter group of refugees.

We have two simultaneous movements of refugees *inside* Ukraine: an outflow of refugees from the Donbas region, and an inflow of refugees to the rest of the country. The relevant question in

this framework is: since refugees are included in the survey, can this relocation across Ukrainian regions account for the increase in the share of respondents identifying as only Ukrainian? In order to answer this question, we collected data about the share of Donbas refugees living in Ukraine by region in each period[4]. In Figure 12 we plot this quantity as of December 2015, showing how the majority of Donbas refugees took shelter inside the same regions of Donetsk and Luhansk, with up to 10-15% of their population being refugees. The bordering regions of Kharkiv and Zaporizhzhya, as well as the capital Kyiv, host a number of refugees equal to 5-10% of their population, while the rest of the Ukrainian regions have a negligible share.

Figure 12: Share of refugees from the Donbas in Ukraine, by region
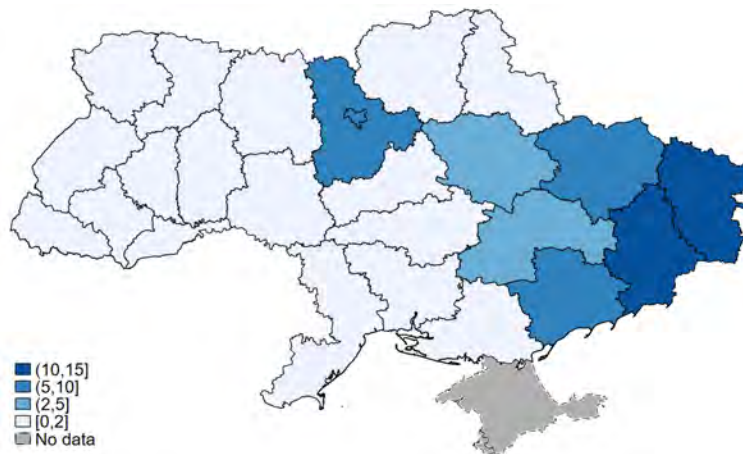
Notes: the Figure plots the share of refugees living in Ukraine by region as a fraction of the estimated region population as of December 2015. Official statistics about refugees come from the Ministry of Social Policy of Ukraine, while the estimates about the region population come from the State Statistics Service of Ukraine.

For what concerns the share of refugees among the interviewees of the Omnibus Survey, the amount of information that we have available is limited. The only survey wave explicitly asking respondents whether they are refugees, is that of May 2015, whose tabulation is shown in Table 4. As shown in the third row of the Table, the share of respondents who declared that they have been living in Donetsk or Luhansk in early 2014 was 1-2% of the total amount, outside of the same

---

[4]The measure was constructed using estimates of the number of refugees from the Ministry of Social Policy of Ukraine, merged with region-by-region population estimates from the State Statistics Service of Ukraine.

Donbas region.

This descriptive evidence alone would be enough to reassure us of the negligible impact that refugees have on the survey composition. Nonetheless, for the sake of completeness we decided to re-run a modified version of Equation (1) which includes the share of refugees over region population that we have just constructed:

$$Y_{irt} = \alpha_r + \gamma_t + \beta_r * Post_t * Rus_r^{2001} + Refugees/Population_r + X_{irt} + \epsilon_{irt} \tag{4}$$

The results of this new estimation are shown in Figure 13, where we compare the coefficients of our baseline model (not controlling for refugees), and the new estimates from Equation (4). For most of the coefficients, the results prove to be robust, as the magnitude remains virtually unchanged. The only estimates which are affected are the ones of those regions with a high share of refugees, with the coefficients getting reduced in size. This happens in the regions of Zaporizhzhia, Luhansk, Kharkiv, Kyiv, and Donetsk, which all host a share of refugees higher than 5%. The magnitude of these results has to be interpreted carefully, as the control variable that we include, comes from official data about Ukrainian refugees, which does not correspond to the true share of refugees interviewed (as it emerges by comparing Figure 12 and Table 4. Therefore, we don't have to interpret the magnitude of the coefficients as causal, but rather observe how the latter get reduced when including refugees as a control variable. In conclusion, we found that refugees can possibly have a role in the increase in the share of ethnic Ukrainians among respondents, but this impact is not enough to explain all the variation.

6.3 Pseudo-panel Study

In the last section of our study, we deal specifically with the lack of a panel structure in our dataset. In order to directly address this issue, we have built a pseudo-panel dataset based on the year of birth of the respondent. We did that by constructing region-specific cohorts of 10 years and considered each of them as a separate observation. In this way, we re-organized our observations as if they came from a fictitious individual of which we keep track over time. Descriptive evidence of

Table 4: Percentage of refugees among the survey Respondents in May 2015

Q: Where did you live in early 2014?	Donetsk	Luhansk	Zhytomyr	Kyiv	Zaporizhzhya	Odesa	Kharkiv
Here, at the same address	79,4	91,3	96,6	95,8	98,8	99,1	99,2
In this city (village), but at a different address	8,3	4,4	0	2,5	0	0	0
In another city (village) of Donetsk or Luhansk region	*6*	*4,4*	*3,5*	*1,7*	*1,2*	*0,9*	*0,8*
In another city (village) of the Autonomous Republic of Crimea	1,4	0	0	0	0	0	0
In another city (village) of another region	5,1	0	0	0	0	0	0

Notes: this Table shows the results of the question asking to the respondents where did they live in early 2014, that is before the outbreak of the conflict. Of particular interest for our analysis is the third row, which shows the share of respondents which stated to be living the regions of Donetsk and Luhansk at the time.

Figure 13: Estimated change in the coefficients after controlling for refugees
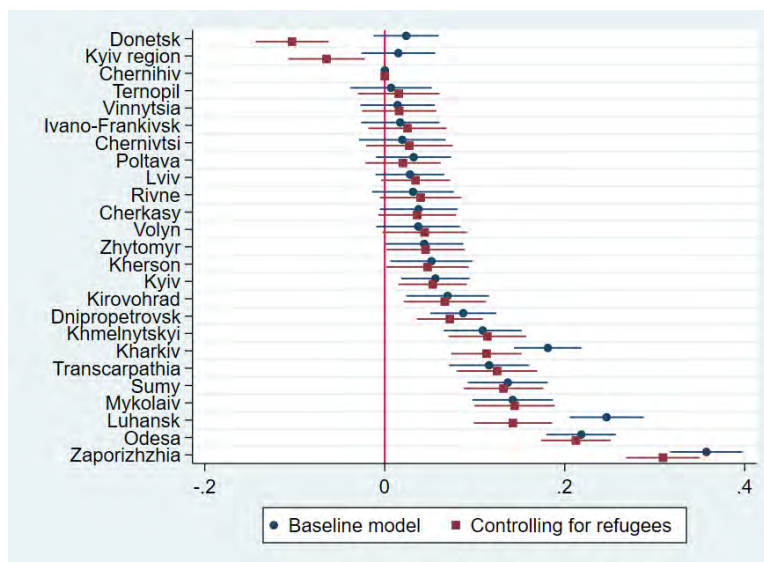

Notes: this Figure compares the estimates of the DiD coefficients when the share of refugees over region population is included, and when it is not.

the increase in Ukrainian identity by cohort, is shown in Figures 14 and 15. The first one shows the change in the fraction of individuals defining as only Ukrainian in levels, while the second shows the change in percentage increase. By looking at them we grasp how the increase in Ukrainian identity has been common to all the cohorts of Eastern and Southern Ukraine, with the change being stronger for the cohorts born between 1950 and 1979. A possible reason for this pattern is that the cohorts born after 1980 were raised when Ukraine was already an independent state, and indeed we find that they begin with an already high share of the population identifying as only Ukrainian before the outbreak of the conflict. On the opposite hand, older people have lived the majority of their life in Soviet Ukraine and hence may feel less attached to the Ukrainian identity, relatively to the Russian one.

To conclude, we performed a regression analysis with the pseudo-panel dataset just constructed, finding results which are largely consistent with what we have found before. They are shown in Table A.4 in the Appendix.

Figure 14: Change in the share of respondents identifying as only Ukrainian by cohort
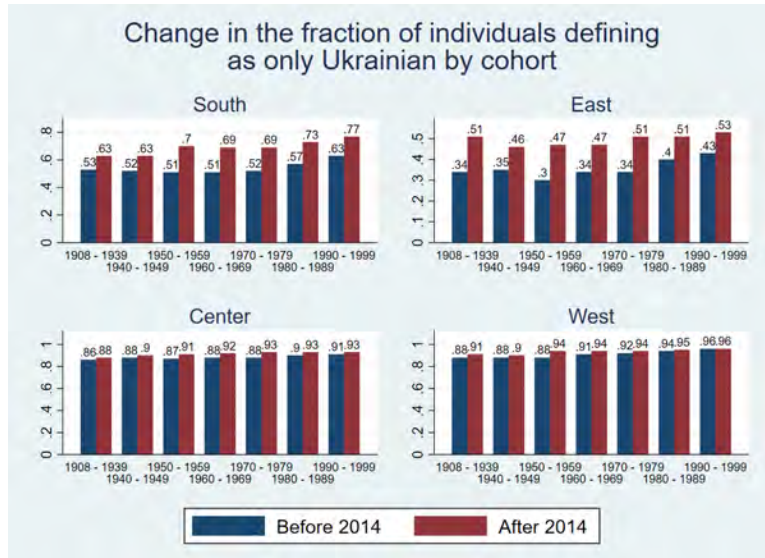
Notes: the Figure shows the increase in the fraction of individuals defining as only Ukrainian by cohort and macroregion, using the pseudo-panel dataset constructed.

Figure 15: Percentage increase in the share of respondents identifying as only Ukrainian by cohort
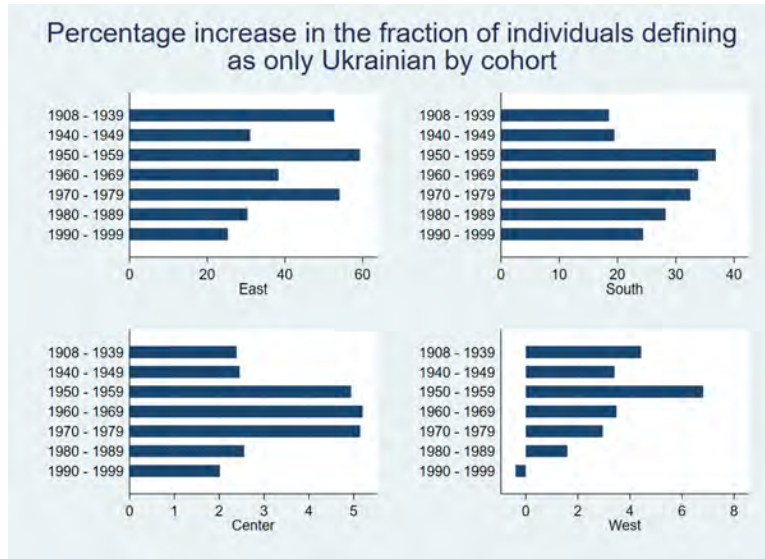
Notes: the Figure shows the percentage increase in the fraction of individuals defining as only Ukrainian by cohort and macroregion, using the pseudo-panel constructed.

7   Conclusion

In this paper we have investigated the effects of conflicts on multiethnic societies, using as a case study the 2014 War in Eastern Ukraine. In particular, we have studied their impact on ethnic identification and on the level of trust among the various groups living in the country. Our results confirm that as a consequence of the conflict, the probability of self-identifying as only Ukrainian has increased for individuals living in regions with an ex-ante higher share of ethnic Russians. At the same time, the probability of identifying as partly or wholly Russian has decreased. This result is robust to a series of robustness checks, including the reduction of the sample size, and of the time frame considered. Focusing our attention on the territories directly involved in the conflict, our evidence suggests an opposite pattern in the areas controlled by pro-Russian separatists. In particular, the probability of self-identifying as Ukrainian has decreased in these areas, relative to government-held ones. We have also found evidence that the conflict has led to a reduction in trust levels toward non-Ukrainian minorities living in the country, in particular toward ethnic Russians. This decrease is common to all regions but stronger in the western and central ones. Our paper contributes to the literature on nation-building, showing how Ukraine started from being a country deeply divided along ethnolinguistic lines before the conflict, and how Ukrainians have become more cohesive as a nation in response to it. On the other hand, if the results that we obtained for the separatist-held areas are credible, this shows how Ukraine and the Donbas region may have become *less* cohesive in the aftermath of it. This last aspect, together with a decrease in trust toward Russians, speaks to the conflict trap literature, suggesting that the Ukrainian government may find it harder to reintegrate those territories in the future.

References

Alberto Alesina, Bryony Reich, and Alessandro Riboni. Nation-building, nationalism, and wars. *Journal of Economic Growth*, 25(4):381–430, 2020.

Maxim Ananyev and Michael Poyker. Nation-building and civil conflict: Theory and evidence from political instability in mali. 2018.

Gerd Baumann, Knott Kim, and McLoughlin Seán. Nation, ethnicity and community. *Diasporas: Concepts, intersections, identities*, pages 45–49, 2010.

Ricardo Dahis, Emily Nix, and Nancy Qian. Choosing racial identity in the united states, 1880-1940. Technical report, National Bureau of Economic Research, 2019.

Melissa Dell and Pablo Querubin. Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies*. *The Quarterly Journal of Economics*, 133(2):701–764, 09 2017. ISSN 0033-5533. doi: 10.1093/qje/qjx037. URL https://doi.org/10.1093/qje/qjx037.

Kai Gehring. External threat, group identity, and support for common policies-the effect of the russian invasion in ukraine on european union identity. 2020.

Michael Kofman, Katya Migacheva, Brian Nichiporuk, Andrew Radin, Jenny Oberholtzer, et al. *Lessons from Russia's operations in Crimea and Eastern Ukraine*. Rand Corporation, 2017.

Vasily Korovkin and Alexey Makarin. Conflict and inter-group trade: Evidence from the 2014 russia-ukraine crisis. *Available at SSRN 3397276*, 2021.

Ivar Krumpal. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & quantity*, 47(4):2025–2047, 2013.

Gail Lewis and Ann Phoenix. 'race','ethnicity'and identity. In *Questioning Identity*, pages 123–158. Routledge, 2004.

José G Montalvo and Marta Reynal-Querol. Ethnic polarization, potential conflict, and civil wars. *American economic review*, 95(3):796–816, 2005.

UN OHCHR. Report on the human rights situation in ukraine. *Office of the United Nations High Commissioner for Human Rights*, 2015.

Olga Onuch and Gwendolyn Sasse. The maidan in movement: Diversity and the cycles of protest. *Europe-Asia Studies*, 68(4):556–587, 2016.

Dominic Rohner, Mathias Thoenig, and Fabrizio Zilibotti. Seeds of distrust: Conflict in uganda. *Journal of Economic Growth*, 18(3):217–252, 2013.

Anthony D Smith et al. The ethnic origins of nations. 1986.

A   Appendix

Figure A.1: Trend in the responses to the question: "how do you define yourself ethically?" by Macroregion
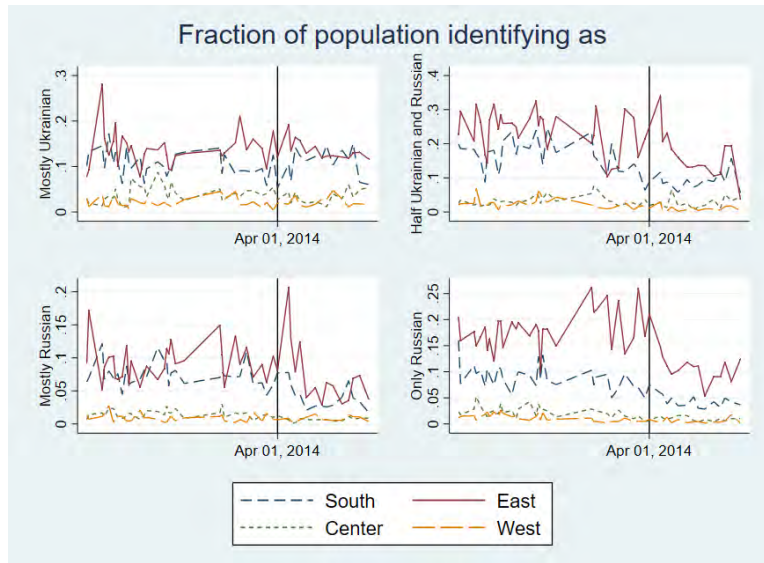
Notes: this Figure shows the trend in the other four possible options that the respondent can choose, when responding to the survey question about his self-perceived ethnicity. Clockwise they are: mostly Ukrainian, half Ukrainian and Russian, only Russian and mostly Russian.

Figure A.2: Region-by-region estimates
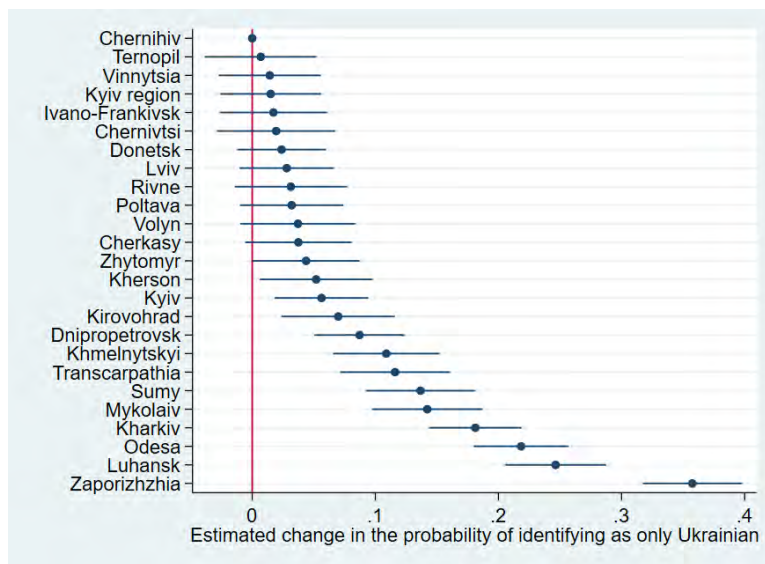
Notes: this figure reports the estimates of our main alternative specification, which corresponds to Equation 1, but interacting the variable *Post* with individual Ukrainian regions. The omitted region is the region of Chernihiv. The region of Crimea and the territories controlled by sepearatist forces in the Luhansk region are omitted.

Figure A.3: Region-by-region estimates excluding separatist-held areas
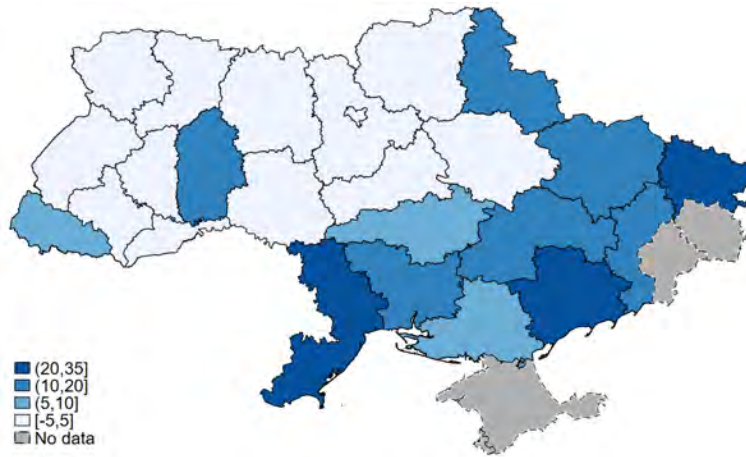

Notes: in this figure we report estimates from our main alternative specification, in which we interact the variable $Post_t$ with individual Ukrainian regions. Darker colours correspond to a stronger increase in probability. The estimation is conducted excluding from the sample the territories of the Donetsk oblast controlled by pro-Russian separatists. The latter territories, together with the ones held by separatist forces in the Luhansk oblast, and the region of Crimea, are shown in grey.

Figure A.4: Decrease in trust levels toward Russian-speaking Ukrainians
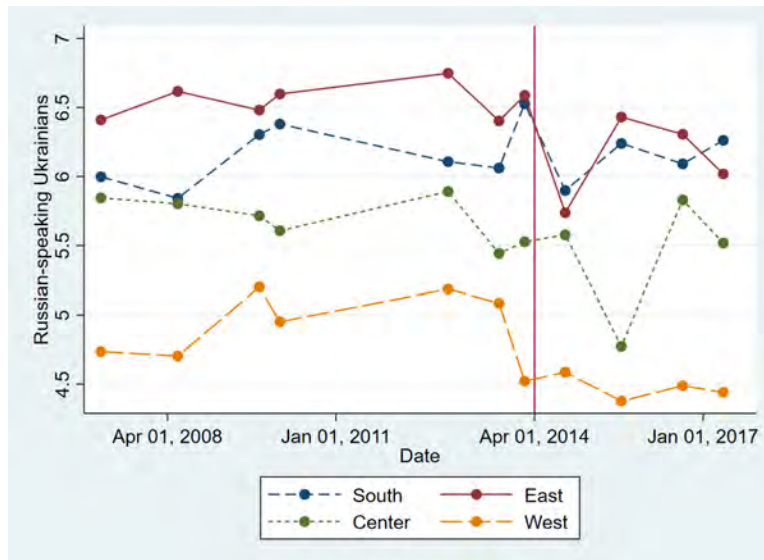
Notes: this Figure shows how trust levels towards Russian-speaking Ukrainians have evolved in the aftermath of the 2014 Conflict, distinguishing the outcome by macroregion.

Table A.1: Estimated change in trust toward people identifying as:

	(1) Russians	(2) Russian-speaking Ukrainians
Post*Southern	-0.035	0.352***
	(0.078)	(0.071)
Post*Central	-0.462***	0.142*
	(0.073)	(0.066)
Post*Western	-0.665***	-0.045
	(0.077)	(0.070)
Sex	-0.055*	-0.068**
	(0.026)	(0.023)
Age	0.004***	0.001
	(0.001)	(0.001)
Education	0.016*	0.010
	(0.007)	(0.006)
Urban	0.141***	0.129***
	(0.028)	(0.026)
Macroregion FE	Yes	Yes
Survey Wave FE	Yes	Yes
N	21339	21384
$\bar{y}$	5.043	5.690
$\sigma(y)$	2.086	1.751

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: this Table reports estimates from a modified version of Equation (3), in which we interact the variable $Post_t$ with individual Ukrainian macroregions. The specification aims at capturing the divergence in inter-ethnic trust levels, across the different parts of Ukraine. The dependent variable is a measure of trust toward individuals belonging to a certain ethnic group, and the independent variables are the same demographic characteristics of the respondent included in previous equations.

Table A.2: Estimated change in the probability of self-identifying as:

	(1) Only Ukrainian	(2) Mostly Ukrainian	(3) Half Ukrainian and Russian	(4) Mostly Russian	(5) Only Russian
Post*Rus 2001	0.224***	0.012	-0.140***	-0.002	-0.129***
	(0.026)	(0.017)	(0.019)	(0.013)	(0.016)
Sex	-0.003	-0.005*	-0.002	-0.001	0.005**
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
Age	-0.001***	0.000	0.000	0.000***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	-0.010***	0.001	0.002***	0.004***	0.005***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Urban	-0.079***	0.009***	0.030***	0.011***	0.026***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Region FE	Yes	Yes	Yes	Yes	Yes
Survey Wave FE	Yes	Yes	Yes	Yes	Yes
N	77771	77771	77771	77771	77771
$\bar{y}$	0.698	0.070	0.096	0.040	0.063
$\sigma(y)$	0.459	0.254	0.295	0.197	0.243

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: this Table reports the estimates of our main specification, with a time span considered that stops at February 2016. The rationale behind such choice, is that the KIIS has stopped conducting interviews in the separatist-held areas of Donetsk region, impacting thus the comparability of the sample over time.

Table A.3: Estimated change in the probability of self-identifying as:

	(1) Only Ukrainian	(2) Mostly Ukrainian	(3) Half Ukrainian and Russian	(4) Mostly Russian	(5) Only Russian
Post*Rus 2001	0.757***	0.011	-0.481***	-0.110***	-0.147***
	(0.028)	(0.017)	(0.019)	(0.013)	(0.015)
Sex	-0.005	-0.003	-0.001	-0.001	0.004**
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
Age	-0.001***	0.000	0.000**	0.000***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education	-0.010***	0.001**	0.002***	0.003***	0.004***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Urban	-0.076***	0.011***	0.030***	0.010***	0.022***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)
Region FE	Yes	Yes	Yes	Yes	Yes
Survey Wave FE	Yes	Yes	Yes	Yes	Yes
N	80830	80830	80830	80830	80830
$\bar{y}$	0.756	0.058	0.081	0.031	0.044
$\sigma(y)$	0.430	0.233	0.273	0.173	0.205

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: this Table reports the estimates from our main specification but excluding from the sample the regions of Donetsk and Luhansk. The motivating fact behind such choice, is that the interviews conducted in areas directly involved in armed clashes may suffer from a series of bias and limitations inlcudind: survey-desirability bias and sample selection.

Figure A.5: Estimated increase in the Ukrainian self-identification by region
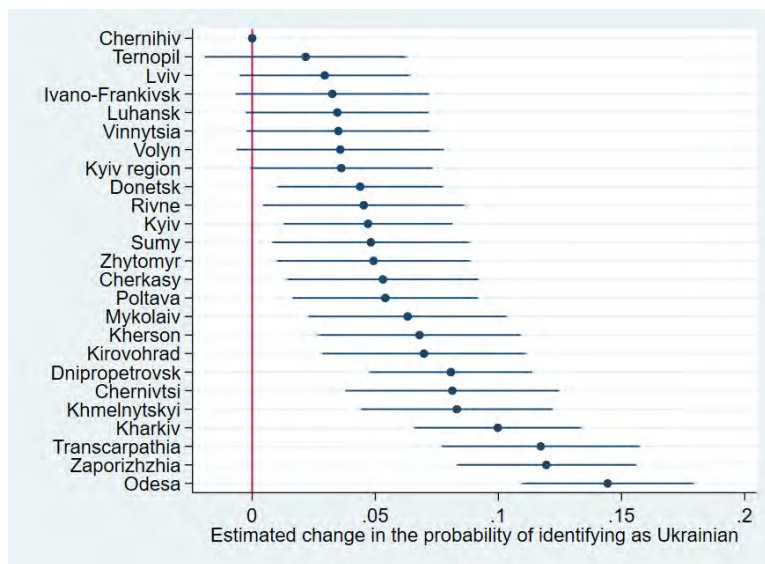
Notes: this Figure reports the increase in the probability of identifying as ethnic Ukrainian, by using as alternative dependent variable the survey question asking the respondent to report his/her own ethnicity from a choice set that includes all the possible nationalities living in the country. The region of Crimea and the separatist-held territories of the Donbas are omitted from the analysis.

Table A.4: Effect on the probability of self-identifying as

	(1) Only Ukrainian	(2) Mostly Ukrainian	(3) Half Ukrainian and Russian	(4) Mostly Russian	(5) Only Russian
Post*Central	-0.002	-0.011	0.012	-0.005	-0.004
	(0.013)	(0.008)	(0.008)	(0.005)	(0.006)
Post*Eastern	0.106***	-0.009	-0.050***	-0.016**	-0.051***
	(0.013)	(0.008)	(0.008)	(0.006)	(0.007)
Post*Southern	0.117***	0.001	-0.050***	-0.025***	-0.031***
	(0.013)	(0.008)	(0.008)	(0.005)	(0.006)
Sex	-0.005	-0.023	0.043*	-0.004	-0.007
	(0.031)	(0.019)	(0.020)	(0.014)	(0.016)
Age	0.004	0.001	-0.003	0.001	-0.003
	(0.004)	(0.002)	(0.003)	(0.002)	(0.002)
Education	-0.006	0.000	-0.004	0.004	-0.001
	(0.007)	(0.004)	(0.005)	(0.003)	(0.004)
Urban	-0.141***	0.018	0.061**	0.039**	0.026
	(0.031)	(0.019)	(0.020)	(0.013)	(0.016)
Macroregion FE	Yes	Yes	Yes	Yes	Yes
Survey Wave FE	Yes	Yes	Yes	Yes	Yes
N	1316	1316	1316	1316	1316
$\bar{y}$	0.692	0.073	0.097	0.040	0.063
$\sigma(y)$	0.238	0.066	0.097	0.047	0.073

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: this Table reports estimates from the regression run using the pseudo-panel constructed using 10-year cohorts based on the year of birth. As before, the variable *Post* is interacted with individual Ukrainian macroregions, with the Western one being omitted

Summary

This thesis studies the impact of conflicts on ethnic identity and inter-ethnic trust of individuals living in multi-ethnic societies. As a case study for this analysis, we used the 2014 War in Eastern Ukraine, between pro-Russian separatists, and Ukrainian government forces. The country of Ukraine, according to the most recent census of 2001, was home to 48 million people, 79% of which identified as ethnic Ukrainian, and 17% as ethnic Russian, with the latter group being largely concentrated in the Eastern and Southern regions of the country (see Figure 1). The results that we obtain suggest that as a consequence of the conflict, a part of Ukrainian citizens who identified as being partly or wholly Russian have changed their identification to only Ukrainian.

The chain of events that preceded the eruption of the conflict began in November 2013, with the refusal of former Ukrainian president Viktor Yanukovich to sign an association agreement with the European Union, in favor of closer ties with Russia. This choice sparked massive protests in the capital of Kyiv and in the rest of the country, which will eventually lead to his downfall, and to the formation of a pro-European government in February 2014. The Russian government condemned the events in Kyiv as a *coup* and took advantage of the situation to annex the Crimean Peninsula, with the pretext of protecting ethnic Russians living there. On the one hand, this move was met with condemnation from the Ukrainian government, but on the other hand, it led to further regions in the East of the country wanting to follow the Crimean experience, to declare independence. This time, the Ukrainian government, declared the launch of an *Anti-Terrorism Operation - ATO* in the Donetsk and Luhansk regions in mid-2014, with the aim of stopping a further spread of the separatist movement. After causing the death of thousands of civilians and millions of refugees, a ceasefire was signed in early 2015. The conflict remained frozen with the boundaries shown in Figure 2 until February 2022, when Russia decided to launch a full-scale invasion of Ukraine.

The data that we use to conduct our analysis comes from the Omnibus Survey, which is representative at the national level and is conducted 4-6 times a year by the Kyiv International Institute of Sociology. The survey asks respondents to state their opinions on a series of political events shaping the country, both at the national and international levels. Demographic characteristics are

also recorded, including gender, age, education, and region of residence of the respondent. The main variables of interest for our study are ethnicity and inter-ethnic trust. Participants to the survey are asked to report their ethnic identity in two separate questions: the first one requires to choose among the various nationalities living in the country, and the second one asks respondents to compare how much they feel Ukrainian relative to Russian. The five possible values for this variable are: only Ukrainian, mostly Ukrainian, half Ukrainian and Russian, mostly Russian, and only Russian. The time span that we use is of 10 years, from February 2007 to February 2017. Interviews are conducted face-to-face. The region of Crimea and the separatist-held territories of the Luhansk region are excluded from our analysis, since the survey has not been conducted in those areas since March 2014. On the contrary, the KIIS managed to conduct the survey in the whole territory of the Donetsk oblast, including the separatist-held ones, until February 2016. Lastly, refugees from the Donbas are included in the survey.

The methodology that we employ is a difference-in-difference regression with the baseline specification reported in Equation (1). In this framework, treated units are those respondents coming from regions with an ex-ante higher share of ethnic Russians, while control units are those individuals living where such a share is small. The logic behind this choice is that the western and central regions of Ukraine, already had a high level of respondents identifying as only Ukrainian before the conflict (more than 90%), and therefore we don't expect a significant impact of the conflict in terms of ethnic identification to happen there (see Figure 5). Our main specification contains an interaction term between the variable $Post_t$ and Rus_r^{2001}. The former is an indicator variable equal to 1 for interviews conducted after April 2014, while the second measures the share of ethnic Russians living in region r in 2001. When estimating the impact on ethnic identification, the dependent variable is a binary indicator equal to 1 for each of the five possible choices of Ukrainian relative to Russian identity, and 0 for the others. Therefore, the interaction coefficient β measures the impact of having a higher share of ethnic Russians in region r in 2001, on the probability of self-identifying as outcome Y, after the outbreak of the 2014 conflict. In addition to Equation (1), we run a series of alternative specifications to reinforce the robustness of our results. The main alternative specification that we run interacts the variable $Post_t$ with individual Ukrainian regions.

In this case, the estimated coefficients β_r measure the impact on the probability of self-identifying as outcome Y for a respondent coming from region r, after the outbreak of the 2014 Conflict. Lastly, we employ a *multiple-periods* difference-in-differences, interacting the variable Rus_r^{2001}, with each period considered in our analysis.

The results of the different specifications all confirm that as a consequence of the 2014 Conflict, the probability of identifying as only Ukrainian has increased in those regions with an ex-ante higher share of ethnic Russians before the outbreak of the conflict. The magnitude of the results suggests that having a 10% higher share of ethnic Russians in 2001 implies a 2.2 - 3.4 percentage points increase in the probability of identifying as only Ukrainian after the eruption of the conflict. Similarly, in the same regions, we find a significant decrease in the probability of reporting as partly or wholly Russian. When observing the region-by-region estimates (see Figure 7) we notice an interesting exception in that geographical pattern, involving the region of Donetsk. The latter region despite hosting a significant share of ethnic Russians before the outcome of the conflict (38%), and being one of the regions directly involved in the 2014 War, has an estimated coefficient which is not significant. An explanation to this puzzle can be found in the decision of the *KIIS* to keep conducting interviews in the territories of Donetsk oblast controlled by pro-Russian forces after the outbreak of armed clashes. Indeed, when comparing the differential impact of the conflict on the separatist-held part of the region, as opposed to the government-held one, we find that the former experiences a *decrease* in Ukrainian identity, and an increase in the Russian one. This result is able to explain why on aggregate the Donetsk region has a coefficient which is non-significant. Moreover, when dropping from the sample observations from separatist-held areas, the estimated increase in the probability of identifying as only Ukrainian becomes positive and statistically significant also for the Donetsk oblast. Our last set of results gives an answer to the second part of the research question, asking whether the 2014 Conflict had an impact on the levels of trust among the various ethnicities living in the country. What we find is a statistically significant decrease in trust toward ethnic Russians, common to all the regions of the country, but stronger in the western and central ones. On the contrary, no impact is recorded on the levels of trust toward Russian-speaking Ukrainians.

The last sections of the thesis are dedicated to conducting a series of robustness checks, as well as commenting on the possible limitations of our results. First of all, we find that the results are robust to a reduction in the time frame considered, and to the exclusion of conflict regions from our sample. The latter robustness check, leads the estimates to double in magnitude (from 0.34 to 0.76), suggesting a stronger impact of the conflict *outside* of conflict regions. Moreover, we run the same specification using an alternative measure of ethnic identity, asking respondents to choose their ethnicity among the various nationalities living in the country. When doing that, we obtain results which are in line with our previous ones. Regarding the possible limitations of our results, we argue that the increase in the share of respondents identifying as only Ukrainian, on which we base our whole analysis, can actually be the results of 3 separate mechanisms; the first one is an inner change in the ethnic identity of respondents and is ideally what we would like to isolate in our analysis. The other two channels are social desirability bias and the possibility of a change in sample composition over time. The first mechanism is a sort of response bias where participants of a survey avoid answering truthfully to certain questions and instead decide to do that in a way that would be socially acceptable at the moment of the interview. In the case of Ukraine, this bias is related to the possibility that ethnic Russians living in Ukraine, in a period of high tensions between the two countries may try to hide their ethnicity. In the thesis, we are able to partially control for it by looking at whether missing values in the question asking respondents to report their own ethnicity have increased after the beginning of the conflict. The evidence shows that this is not the case, therefore reassuring us of the possibility of such bias contaminating our results. Regarding the second main source of bias, involving the possibility of a change in sample composition over time, we first verify that the fraction of respondents by age, education, microregion, and type of settlement has not changed over time. Secondly, we directly investigate the possible impact of refugees from the Donbas on our analysis. A possible source of concern in this framework is given by the fact that refugees are included in the survey and recorded as residents of the region where they are interviewed. By directly controlling for the share of refugees living in the region at the time of the interview in Equation (4), we find evidence that the inflow of refugees had a very small impact on the estimated probability of self-identifying as only Ukrainian, which is not enough to

50

explain all the variability in the data. Lastly, we conduct a pseudo-panel analysis, to account for the lack of panel structure in our dataset. After constructing 10-year-wide cohorts, based on the year of birth, we find that the increase in Ukrainian identity has been common to individuals of all ages, with the strongest increase being reported for those born between 1950 and 1980.

In conclusion, our paper has found empirical evidence of how the 2014 War in Eastern Ukraine has led to a part of its citizens who identified as partly or wholly Russian before the conflict, to change their self-perceived ethnicity, feeling more Ukrainian afterward. This result contributes to the nation-building literature, showing how an ethnically fragmented country may actually become more cohesive when experiencing a conflict. On the other hand, the decrease in Ukrainian identity in the separatist-held areas, as well as a decrease in trust toward Russians in the rest of the country, contribute to the conflict-trap literature, suggesting that Ukraine will find it harder to reintegrate those regions in the future.