This is the peer reviewed version of the following article: Duncan, A. and Mavisakalyan, A. 2015. Russian language skills and employment in the Former Soviet Union. Economics of Transition. 23 (3): pp. 625-656, which has been published in final form at 10.1111/ecot.12075. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving at http://olabout.wiley.com/WileyCDA/Section/id-828039.html

Russian language skills and employment in the Former Soviet Union*

Alan Duncan† and Astghik Mavisakalyan†‡

† *Bankwest Curtin Economics Centre, Curtin Business School, Curtin University, GPO Box U1987, Perth WA 6845, Australia.*

‡ Corresponding author. E-mail: astghik.mavisakalyan@curtin.edu.au

ABSTRACT: During the Soviet era, proficiency in Russian language was often a ticket to attractive employment opportunities in the member republics. Does it still contribute to securing employment in the former Soviet republics after two decades of transition? Using data from Armenia, Azerbaijan and Georgia in the years 2008-2010, this paper demonstrates that Russian language skills remain economically valuable. The baseline estimates suggest that Russian language skills increase a person's probability of employment by about 6 (males) and 9 (females) percentage points. Our results bear important implications for the ongoing debates on language policies in the post-Soviet countries.

JEL classification: J21, J24, P23

Keywords: Language proficiency; Employment; Former Soviet Union.

*We are grateful to Mike Dockery, Tesfaye Gebremedhin, two anonymous referees and the associate editor for valuable comments. We also thank Juergen Meinecke and Ha Nguyen for helpful discussions.

Russian language skills and employment in the Former Soviet Union

ABSTRACT: During the Soviet era, proficiency in Russian language was often a ticket to attractive employment opportunities in the member republics. Does it still contribute to securing employment in the former Soviet republics after two decades of transition? Using data from Armenia, Azerbaijan and Georgia in the years 2008-2010, this paper demonstrates that Russian language skills remain economically valuable. The baseline estimates suggest that Russian language skills increase a person's probability of employment by about 6 (males) and 9 (females) percentage points. Our results bear important implications for the ongoing debates on language policies in the post-Soviet countries.

JEL classification: J21, J24, P23

Keywords: Language proficiency; Employment; Former Soviet Union.

1. Introduction

Since the collapse of the Soviet Union in 1991, various programs to promote the national languages at the expense of Russian have been initiated in the newly independent states. Among those are, for example, single-language policies that made the titular language the language of the state. In addition, education reforms in some countries phased out the education in Russian language (Dietrich, 2005). Derussification has been seen as an essential means of building independence and national sovereignty. However, the potential costs associated with such programs have been a source of concern. In particular, Russian functions as a second language or *lingua franca* in communication between various ethnic minorities within post-Soviet countries and in interactions between the countries (Pavlenko, 2008). Furthermore, many of the countries remain politically, socially and economically highly dependent on Russia, therefore making Russian language an important medium in which business and public life is conducted. As a result, Russian proficiency, once key to individuals' ability to fit into the system, may still be important in ensuring an easier access to jobs after two decades of transition. Understanding whether and to what extent that is the case is the aim of this paper.

We test the relationship between Russian language skills and employment in three post-Soviet republics of South Caucasus: Armenia, Azerbaijan and Georgia. Based on a recently collected nationally representative individual-level data, the results suggest positive labour market returns associated with Russian language skills in the three countries. After controlling for individuals' education, demographic characteristics and the localities in which they are based, we find there to be around a 6 percentage point increase in the probability of emplyment associated with Russian proficiency for males, and a 9 percentage point increase for females. These results are not only statistically significant, they are also economically meaningful. For males, the return to speaking Russian is over third as much as the effect of completing a university degree (relative to school education). For females, it is quarter as much as the return to a university degree. There is some heterogeneity in the returns to Russian proficiency. In particular, in the sample of males we find that the returns are higher for those with post-school education, suggesting complementarity between language skills and education. Furthermore, Russian proficiency increases the probability of males selecting into jobs in state and foreign organisations, and in government, and trade, services and communications sectors.

We consider the problem of endogeneity in estimating the effect of Russian language skills on employment. Omitted variable bias is of particular concern in the context of this study

since the knowledge of Russian may be indicative of general cognitive skills that are correlated with employment. We adopt an approach proposed by Lang and Siniver (2009) whereby an indicator of an additional foreign language proficiency is added in the estimation. We do not observe significant changes in the estimated marginal effect of Russian proficiency when the two language indicators are simultaneously included. Assuming that the abilities to learn two different foreign languages are similarly correlated with unobservable ability, this evidence suggests that the scope for bias due to unobserved ability is limited in our context. We further demonstrate the validity of our approach through employing propensity score matching estimation which produces similar results. Of course, it is hard to completely rule out all potential concerns around endogeneity under conditions where truly random sources of variation in Russian language proficiency are not available in the data to be exploited for identification. Nevertheless, even if the results may not strongly qualify for interpretation in a causal way, they are valuable for at least two reasons. First, they are informative about the value of Russian language skills relative to other forms of human capital such as the years of education, since estimating the returns to other types of human capital faces similar identification issues as estimating the returns to language skills (Azam et al., 2013). Second, given the lack of research on the economic value of Russian skills in transition countries, even purely indicative results would add to knowledge.[1]

Language skills have all the characteristics of human capital in that they are embodied in a person, are productive in the labour market and are obtained at the cost of time, effort and out-of pocket expenses (Chiswick and Miller, 2014). A large body of literature has studied the returns to language skill acquisition in the context of developed countries. Most of this literature considers labour market outcomes of immigrants in relation to their proficiency in host country languages. Studies by, for example, Dustmann (1994); Chiswick and Miller (1995, 1999, 2002); Dustmann and van Soest (2001, 2002); Shields and Price (2002); Dustmann and Fabbri (2003); Bleakley and Chin (2004), among others, examine this relationship in various countries. Others study the labour market effects of speaking a foreign language by natives (Ginsburgh and Prieto-Rodriguez, 2010) and by immigrants (Isphording, 2013) in developed countries. Finally, there are studies on the labour market returns to language skills of non-immigrants in multilingual labor markets in developed countries. Examples of contributions in this strand of literature include Albouy (2008) in Canada, Rendon (2007) in Spain, Grin and Sfreddo (1998) in Switzerland and Henley and Jones (2005) in Wales.

[1] There is remarkably little economics research done particularly on the three countries we focus on in this paper. From 1985 to 2005, Armenia has been the subject of just 9, Azerbaijan of 2 and Georgia of 26 papers, none of which have appeared in the first or second tier journals of the discipline (Das et al., 2013).

This paper contributes to the nascent literature on labour market returns to language skills in developing countries. This literature is largely concerned with studying the labour market returns to dominant non-indigenous language skills in previous colonies. Contributions include papers by Chiswick et al. (2000) and Godoy et al. (2007) on the returns to Spanish language skills in Bolivia. Angrist and Lavy (1997) study the returns to French language skills in Morocco. More recently, labour market returns to English language skills have been explored in the context of South Africa (Levinsohn, 2007; Casale and Posel, 2011) and India (Azam et al., 2013).[2] This paper is the first attempt to measure the economic value of speaking Russian in the post-Soviet countries.

Studies on labour market returns to human capital in the context of transition countries have focused on the returns to education qualifications. Recent studies demonstrate high education premium in different countries of the Former Soviet Union (see, for example, Pastore and Verashchagina (2006) on Belarus and Arabsheibani and Mussurov (2007) on Kazakhstan). Cross-country studies by Newell and Reilly (1999) and Brainerd (2000) document increasing rates of returns to educational qualifications in transition countries over time. This paper adds to this literature by considering language skills as another form of human capital and allowing to draw comparisons on the labour market values of different forms of human capital.

The next section provides some background on Russian in post-Soviet countries. Section 3 describes the data and reports descriptive statistics. Section 4 presents the empirical approach and section 5 reports the estimation results. Section 6 concludes.

2. Background

The language policy of actively promoting knowledge of Russian among the populations of the Soviet Union dates back to late 1930s. In 1938, Russian was introduced as a compulsory subject in school curricula across the Soviet republics. As a further step towards russification in late 1950s, parents were given the choice of the language of instruction for their children with mother tongue no longer being compulsory in education. The relative status of Russian was further elevated throughout the subsequent decades until the late 1980s as it was promoted not only as a common *lingua franca*, but also as a key component of a common Soviet cultural foundation (Fierman, 2006). According to Kirkwood (1991), there is 'indirect evidence' of a decrease in the number of students who received primary

[2]English has also been analysed in relation to labour market outcomes in the context of developing/transition countries where it is not a former colonial language - see Toomet (2011) for evidence from Latvia and Estonia, Lang and Siniver (2009) for evidence from Israel, Di Paolo and Tansel (2013) for evidence from Turkey.

education in their mother tongue and the number of languages used as media of instruction in the Soviet education system as a whole, since the 1970s.

The growth in the importance of Russian across the republics of Soviet Union apparently was not uniform. With reference to the Soviet education laws of 1958-1959, Bilinsky (1962) argues that resistance to russification was particularly strong in Armenia, Azerbaijan and Georgia compared to some of the other republics, such as Belarus and Ukraine, whose residents had close cultural and racial affinity with Russians. The census data indicate that the three South Caucasian republics had relatively high retention rates of their national languages in the Soviet Union. As of 1979, the share of those classifying the national language as their first language was 90.6% in Armenia, 97.8% in Azerbaijan, 98.2% in Georgia (in contrast to, for example, 74.2% in Belarus and 82.7% in Ukraine) (Kirkwood, 1991). The three republics were also the only ones to declare their national languages as official under the Soviet regime (Fierman, 2006). Nevertheless, knowledge of Russian grew in importance in all parts of the Soviet Union including the South Caucasus, as it was increasingly seen as essential by those who wished to make a career either in administration or in any of the professions which would entail travel beyond one's immediate national territory (Kirkwood, 1991).

The political foundations for Soviet policies of russification disappeared with the disintegration of the Soviet Union in 1991. The three post-Soviet republics of South Caucasus opted for single-language policies that made the titular language the language of the state.[3] Measures of derussification were undertaken in different spheres of public life. In particular, the national languages currently serve as the languages of education in the three countries although instruction in minority languages, including Russian, is permitted. Russian also has a status of a foreign language and is an option for foreign language instruction in schools (Pavlenko, 2008). Some of the popular debate accompanying these changes has considered the implications of derussification on labour market outcomes of individuals living in the countries of South Caucasus.[4] For example, Georgia's former president Mikhail Saakashvili known for his anti-Russian agenda argued: 'To have a career, you don't need to know Russian. You need to study English, Turkish.' (Pavlenko, 2008, p.70). On the other hand, in defence of a recent government initiative to re-open some foreign language schools in Armenia, the education minister Armen Ashotian suggested that the knowledge of Russian is

[3] In the Abkhazian region of Georgia, Abkhazian is also a state language.

[4] Other outcomes discussed in the context of language policies are related to the quality of education systems, the viability of national languages, the nature of cultural identities, the threats to national sovereignties, etc.

simply a matter of Armenians remaining 'competitive' in today's marketplace (Grigoryan, 2010).

A number of benefits from Russian proficiency provide economic incentives to learn the language. First, Russian still commonly maintains its status as a second language and medium of communication between various ethnic minorities within the countries - something of particular relevance to multiethnic Azerbaijan and Georgia. Second, Russian maintains its status as *lingua franca* in interactions with Russia and other post-Soviet countries. Particularly crucial are the three countries' political and economic links to Russia. Due to Russia's strong political influence in the South Caucasus, Russian remains an important 'political language' in the region with wide scope of applications in public sphere. It also plays an important role in the economic life of the region. For example, as of 2012, 29.1% of Armenia's, 26.5% of Georgia's and 10.7% of Azerbaijan's total trade in goods was with a Commonwealth of Independent States member (comprising Armenia, Azerbaijan, Belarus, Kazakhstan, Kyrgyzstan, Moldova, Russia, Tajikistan, Turkmenistan, Ukraine and Uzbekistan) (European Commission, 2012). These circumstances imply that Russian language skills may still be economically valuable in the three countries two decades after transition. In what follows, we provide a quantitative assessment of that possibility.

3. Data and descriptive statistics

We use the annual Data Initiative Survey (DIS) (renamed into Caucasus Barometer in 2010) conducted by the Caucasus Research Resource Centers in Armenia, Azerbaijan, and Georgia. The series began in 2004 with samples from the capital cities only. In 2005, a specimen region in each country was added, while in 2006 the representativeness of the collected data was increased to include all the regions controlled by central governments. In 2007 the partially applied panel data methodology was abolished and unified sampling methodology to ensure the full comparability of data across the three countries was introduced (CRRC, 2007). Since then, the DIS has been carried out based on consistently weighted nationally representative samples in Armenia, Azerbaijan, and Georgia with the survey methodology based on a stratified two stage sampling design (see CRRC, 2008, for further details). DIS covers several important areas of individuals' lives such as demographics and employment, education, skills and health, and political attitudes and economic behavior. Owing to its high quality, the DIS has already been used in other published studies on the region (Habibov and Afandi, 2009, 2011; Charles, 2010; Habibov, 2011, 2012; Roberts and Pollock, 2011).

Overall, the cross-country comparability, representativeness, richness and relevance of the data make DIS the best available source to study the relationship between Russian language

skills and employment in the region. The Life in Transition Survey is another source of similar information on transition countries. However, while it includes information on whether a respondent speaks any of the official languages of the country, it does not have information on the level of language proficiency. Moreover, no information on the knowledge of Russian language is included. An additional advantage of the DIS over the Life in Transition Survey is the larger country samples.

This study is based on repeated cross-section data in the years 2008-2010, comprising 17,673 observations in total, of which 5,979 are from Armenia, 5,810 from Azerbaijan, and 5,884 from Georgia. We restrict the sample to people between the ages of 26 and 60, and drop observations with missing outcome data.[5] The final size of the sample is 10,677 observations, out of which 3,565 are from Armenia, 3,921 from Azerbaijan and 3,191 from Georgia.

Table 1 contains definitions and descriptive statistics for variables used in the analysis. We split the sample into two sub-samples: non-speakers and speakers (of Russian). The DIS asks individuals to describe their Russian language proficiency with response options on a four-point scale of 'advanced', 'intermediate', 'beginner' and 'no basic knowledge'. It is useful to collapse this four-category Russian language proficiency variable into two categories for the statistical analysis. We reduce this information to a dummy variable, RUSSIAN SPEAKER, coded 1 if Russian language proficiency is reported as advanced or intermediate. Thus the reference category is everybody whose response is 'beginner' or 'no basic knowledge'.[6] Across the full sample, 70% are classified as Russian speakers and the remaining 30% as non-speakers.

DIS has no information on the respondents' wages. Therefore, our analysis of the economic value of Russian focuses on employment rather than wages as usually done in the literature.[7] The individuals are specifically asked: 'Do you have a job? This job may be part-time or full-time, or you may be officially employed, informally employed, or self-employed, but it brings you monetary income'. As Table 1 demonstrates, Russian speakers have a higher probability of being employed compared to non-speakers: 51% of speakers

[5] To minimise the concerns over selection effects, estimations based on the populations aged 21-55, 26-65 and 18-65 were also carried out. The results were quantitatively identical.

[6] It is a commonly taken approach in the literature to reduce the language proficiency information recorded in four categories to a dichotomous variable coded 1 for good or very good language skills (Dustmann and van Soest, 2002). In the robustness checks we test the sensitivity of the findings to using alternative definitions of Russian language proficiency.

[7] DIS includes information on 'personal income' in the preceding month reported in 8 categories. We tested a model where continuous income variable defined based on midpoint technique was used as a dependent variable. The coefficient on RUSSIAN SPEAKER was positive and significant.

but only 34% of non-speakers are employed. This gap of 17 percentage points in employment probabilities suggests a positive effect of speaking Russian on employment outcomes.

[Table 1 about here.]

We measure education in levels, distinguishing across school education or below, secondary technical or incomplete university education, and first or higher university degrees. Accordingly, we defined variables EDUC$\leq$ 10 , EDUC11-14 and EDUC$\geq$ 15 based on respondents' reported years of education. Table 1 reveals that speakers are more educated than non-speakers: 33% of speakers have obtained a university education while only 6% of non-speakers have. On the other hand, 60% of non-speakers finish schooling at the lowest education category as compared to 30% among speakers. In terms of demographic characteristics, 47% of speakers and 38% of non-speakers are male. There exists no strong trend in the age distribution for speakers and non-speakers. The vast majority of both speakers and non-speakers come from an ethnic majority background (report Armenian, Azerbaijani or Georgian as their ethnicity in respective countries) and spoke the majority language (Armenian, Azerbaijani or Georgian) as their 'main home language' in childhood.[8] Speakers are 15 percentage points more likely to live in a capital city whereas non-speakers are 21 percentage points more likely to live in rural areas. Finally, speakers are more likely to come from Armenia and Georgia, and less likely to come from Azerbaijan.

In Tables 2 and 3, we analyze the means of the variables RUSSIAN SPEAKER and EMPLOYED separately for males and females, while partially controlling for education, age and country. Speaking Russian is more common for men (75%) than for women (67%). The probability that a male with not more than 10 years of education speaks Russian is 63%. At intermediate levels of education the probability of speaking Russian is 74% and for people with a completed university degree that probability climbs up to 93%. Similar patterns are observed for females, although their proficiency at the lowest education level is lower at 47%. As expected, Russian proficiency appears to decrease with age in the sample of males. However, among females it is largely uniform across age cohorts. Finally, there are marked differences across countries with regards to the propensity to speak Russian. In Armenia, the probability of speaking Russian is 90% among males and 84% among females. In Georgia it is 79% (males) and 74% (females). Azerbaijan has the lowest prevalence of Russian speakers: 60% (males) and 42% (females).

The gender differences in employment probabilities are substantial: 62% of men and 34% of women are employed. The effect of speaking Russian on the probability of employment is

[8]The results are robust to restricting the sample to those coming from a majority language background only.

12 percentage points for males and 18 percentage points for females. The employment probability is increasing in education for both non-speakers and speakers. For males, the gap in the employment probability between non-speakers and speakers starts off at 3 percentage points at the lowest education level before climbing up to 9 percentage points at intermediate and university education levels. For females the gap is 6 percentage points at the lowest education level and 11 percentage points at intermediate education level. It is non-existent among university graduate females. With regard to age, the employment probability of males starts off at 64% in the youngest cohort and stays relatively stable before starting to decline gradually after the age of 45. In contrast, women's probability of employment is 25% in the youngest cohort, perhaps reflecting the patterns of fertility. However, it increases with a higher rate till the age of 45 relative to men. Similarly, the rate of subsequent decline in the probability of employment of women is also lower relative to men. At each age level, speakers have a higher employment probability than non-speakers. Comparing countries, males in Armenia and Azerbaijan exhibit slightly higher employment probabilities (62% and 69%) compared to those in Georgia (52%). In the sample of females, the employment probability is the highest in Armenia at 38%. In Azerbaijan and Georgia it is 31% and 33% respectively. While Georgia has the largest gap in employment probabilities between male non-speakers and speakers (18 percentage points), it has the smallest gap between female non-speakers and speakers (8 percentage points). In Azerbaijan, the gap is 15 percentage points in the samples of both males and females. In Armenia, male Russian speakers have an employment probability that is 13 percentage points above that of non-speakers. This gap is 17 percentage points in the sample of females.

In summary, while heterogeneity in education, age and country explains some of the variation in employment probabilities in the population, differences between non-speakers and speakers persist when keeping these variables constant.

[Table 2 about here.]

[Table 3 about here.]

4. Empirical approach

Basic model. To capture the relationship between language proficiency and employment, we consider a standard model in which employment propensity Y_i^* for an individual i is assumed to depend on Russian language proficiency RUS_i together with series of additional controls X_i for education, age, ethnicity and location characteristics. Unobserved factors ε_i

further contribute to employment propensity, leading to an equation of the form

$$Y_i^* = X_i\beta + \delta RUS_i + \varepsilon_i \text{ for all } i = 1, ..., N. \tag{1}$$

Observed employment status Y_i is assumed to relate to latent propensity through the criterion $Y_i = 1(Y_i^* \geq 0)$, so that the probability of employment under an assumption of normality for ε_i becomes

$$Pr(Y_i = 1|X_i, RUS_i) = \Phi(X_i\beta + \delta RUS_i), \tag{2}$$

with marginal effects of Russian language proficiency derived from the estimated model thus:

$$\frac{\partial Pr(Y_i = 1|X_i, RUS_i)}{\partial RUS_i} = \delta\phi(X_i\beta + \delta RUS_i). \tag{3}$$

Marginal effects such as those described in (3) can be evaluated either at the sample means or (as is the case in our empirical results) for specified values of each explanatory variable.

Testing for ability bias. To estimate the causal effects of Russian proficiency on employment requires that it is exogenously determined and uncorrelated with the error term in (1). However, unobserved ability α_i may simultaneously contribute to general language proficiency and an enhanced propensity for employment. This has the potential to bias upwards the estimated impact of Russian language on employment. To see this, suppose that $\varepsilon_i = \alpha_i + \eta_i$ for some random disturbance η_i and imagine that $RUS_i = \overline{RUS}_i + \omega\alpha_i$, where $\overline{RUS}_i$ represents the exogenous components of the language proficiency variable. Then,

$$Y_i^* = X_i\beta + \delta(\overline{RUS}_i + \omega\alpha_i) + (\alpha_i + \eta_i), \tag{4}$$

which implies that $\text{cov}(RUS_i, \varepsilon_i) = \text{cov}(\overline{RUS}_i + \omega\alpha_i, \alpha_i + \eta_i) = \text{cov}(\omega\alpha_i, \alpha_i) = \omega\text{var}(\alpha_i)$. The bias caused by this ability-induced correlation acts in the direction of the sign of ω, with a reasonably certain prior of being positive to the extent that ω is positive.

Two approaches could deal with this issue - one requires longitudinal data to difference out (time-invariant) ability through a panel estimator of the change in employment, whereas another requires direct measures of ability. Data limitations prevent the implementation of either approach. Instead, we implement a test favoured by Lang and Siniver (2009) in which indicators of different language proficiency are added sequentially in the estimation of (1). The intention in so doing is to reveal the extent to which unobserved ability causes bias in the effects of Russian language proficiency on employment.

Adding a measure of English language proficiency, ENG_i, into (1) and assuming $ENG_i = \overline{ENG}_i + \nu\alpha_i$, $\text{cov}(ENG_i, \varepsilon_i) = \nu\text{var}(\alpha_i)$ and $\text{cov}(RUS_i, ENG_i) = \omega\nu\text{var}(\alpha_i)$, with both also expected to be positive. If ability does indeed introduce upward bias in the estimated impact

of Russian language proficiency on employment, and under the assumption that English language proficiency exhibits similar properties to Russian, the consequence would be that introducing English proficiency as an added control would affect the language parameter δ.[9] This provides us with an informal test of robustness in our empirical results.

Additional approaches to mitigating unobserved heterogeneity bias. One way to mitigate the influence of unobserved heterogeneity is to include additional variables that could be correlated with the hitherto unexplained component of employment. In particular, it could be argued that language proficiency effects are actually capturing broader cultural characteristics and attitudes among Russian speakers that correlate with, or at least partially explain, employment propensity. To test this conjecture, we sequentially introduce into our empirical specification an additional proxy measure, denoted BICULTURALIST, to capture the strength of affinity with the Russian culture. To the extent that such broader attitudes explain differences in employment between speakers and non-speakers, the introduction of this cultural measure will cause some change in the estimated language proficiency effects.

Another way to reduce the bias generated by unobserved heterogeneity is the propensity score matching method proposed by Rosenbaum and Rubin (1983). We use this approach as a robustness test to compare the employment probabilities of Russian speakers with those of the non-speakers who have a similar propensity to speak Russian.

The propensity score for an individual is the probability of speaking Russian conditional on observable characteristics:

$$Pr(RUS_i = 1|X_i) = \Phi(X_i\kappa). \tag{5}$$

We estimate (5) using the probit estimator with the kernel matching method[10].

It should be kept in mind that the bias is eliminated completely only if speaking Russian can be considered purely random among individuals who have the same propensity scores. Since this is unrealistic in most observational studies, what this method achieves is reducing the bias generated by unobservable confounding factors.

In spite of improvements achieved by implementing these approaches, there could still remain unobserved heterogeneity correlated with both speaking Russian and employment. The next step would be to use a bivariate probit model and estimate a joint likelihood of

[9] While there are significant differences between Russian and English, both are quite distant from the native languages spoken in three countries of our study. Armenian, Azerbaijani and Georgian languages belong to three different language families (Armenian (an independent branch of the Indo-European family), Kartvelian and Turkic), all being distinct to those of Russian (Slavic) and English (Germanic). It is likely that the ability to learn Russian and the ability to learn English are similarly correlated with unobserved ability to learn other skills.

[10] The results are very similar when nearest neighbour or radius matching methods are used instead.

employment and Russian skills. However, we were unable to identify a suitable instrument in the DIS data. Without an exclusion restriction, identification of the marginal effects relies only on functional form, which can be problematic (Altonji et al., 2005).[11]

Other potential sources of bias. Survey and census data on language skills almost always rely on self-reported responses or responses provided by an adult household member thereby leaving scope for measurement error (Chiswick and Miller, 2014). To quantify the bias from measurement error, Dustmann and van Soest (2001, 2002) exploit the panel feature of their dataset while Bleakley and Chin (2004) use an objective measure of language proficiency based on a literacy test in addition to a self-reported measure. All three studies find that the returns to language skills are higher after correcting for the measurement error. If the nature of the measurement error in our Russian proficiency variable is similar, we may be estimating the lower bound of the true return to Russian skills. However, we cannot claim that with certainty, since we do not have access to a source of objective information on Russian language skills to check against our self-reported measure.

Reverse causality could be another potential source of bias, if employment opportunities were a determinant of investment in Russian language skills, thereby causing an upward bias in estimates. This is unlikely to be the case in the context of the former Soviet countries, where Russian language acquisition typically takes place well before the entry to the labour market, usually early on during schooling.

5. Estimation results

Main results. Table 4 reports the results of probit estimation of the effect of Russian language proficiency on the probability of employment. The model controls for variables listed in Table 1. To account for the potential correlation among outcomes of individuals at a given locality, we have clustered the standard errors at the country level.

Consistent with the descriptive statistics, the estimated coefficients reported in Table 4 confirm a significant and positive relationship between Russian language skills and employment for both males and females. Higher level of education is associated with an increased probability of employment. According to screening theories of education, this may reflect not only the role of education in enhancing productivity, but also in credentialising it (Arrow, 1973; Spence, 1973). Males in the two oldest cohorts are less likely to be employed relative to the reference individuals of 36-40 years of age. In contrast, women in the two

[11] The results of such estimation (not reported) confirm the finding that Russian language skills have a positive effect on the employment outcome, and also suggest that the univariate probit model tends to produce conservative estimates of the effect of Russian proficiency on employment.

youngest cohorts are less likely to be employed, while those aged 46-50 are more likely to be employed relative to the reference individuals. Being part of the ethnic and linguistic majority does not bear statistically significant implications for the employment probability. With regard to the locality, being based in a capital city is associated with an increased probability of employment for males. Conversely, women in rural and other urban areas are in fact more likely to be employed relative to women in the capital cities. This may potentially reflect the patterns of industrial segregation. Employment probability of Armenian and Georgian men is lower than that of their Azerbaijani fellow-residents. Georgian women, too, have a lower probability of employment relative to women in Azerbaijan.

To aid with the interpretation of the coefficients, Table 4 also reports the marginal effects. We fix the covariates to a non-Russian-speaker, 36-40 years old from ethnic and linguistic majority background, residing in the Azerbaijani capital of Baku as of 2010.[12] The marginal effect of speaking Russian is 6.4 percentage points for males and 8.5 percentage points for females. We also find large education-related marginal effects. Moving a person with tertiary education to intermediate level of education results in a decrease in the employment probability of 12.4 (males) and 22.1 (females) percentage points. Moving that person to at most 10 years of education level instead, results in a decrease in the employment probability of 18.6 (males) and 34 (females) percentage points. This is just 2.9 (males) and 4 (females) times as large as the penalty associated with not speaking Russian.

[Table 4 about here.]

Robustness checks.

Unobserved heterogeneity. Language proficiency may be endogenous to labour market outcomes. In particular, there may be omitted individual characteristics that affect both Russian language proficiency as well as employment. We address this concern by including additional variables that could plausibly be correlated with both language skills and employment, and checking how the marginal effect on RUSSIAN SPEAKER changes due to the inclusion of these additional regressors. Table 5 reports the results. Columns (1) and (6) restate the marginal effects of speaking Russian from the model without additional controls.

Omitted ability is an important source of bias. Individuals with higher ability are more likely to have better Russian language skills as well as jobs - a possibility that will lead to an upward biased estimate of return to Russian. Controlling for the level of education alleviates this problem to a certain extent. Here we additionally control for respondents' English

[12] The results are robust to alternative definitions of a reference individual.

language skills (defined similarly to Russian skills). Around 9.5% of males and 13.1% of females in the sample are categorised as English speakers. As argued earlier, If acquisition of different languages is driven by differences in unobserved ability, we would expect the fluencies to be correlated across languages and observe significant changes in the estimates when language dummies are simultaneously included in the employment equation.

The unconditional correlation between Russian and English language proficiencies is significant at 0.19 for males and 0.26 for females. However, as Table 5 demonstrates, the estimated marginal effect on Russian language skills for males is hardly affected by the inclusion of English proficiency in the model. The marginal effect for females also changes remarkably little after controlling for English proficiency. This suggests that the results are not just capturing the effect of unobserved ability on employment. Additionally, the results suggest a positive significant effect of English language proficiency on the probability of employment. For males, the magnitude of this effect at 5.7 percentage points is comparable to the effect of Russian proficiency on employment. For females, speaking English increases the probability of employment by 12.5 percentage points - a considerably larger effect compared to that of Russian.

Another concern is that people who speak Russian may have associated attitudes and values with broad implications for labour market outcomes. Exploiting the richness of the DIS data, we introduce a dummy variable, BICULTURALIST, equal to 1 if the respondent disagrees to the statement "Our way of life needs to be protected against Russian influences". Not only this measure may be of direct relevance to Russian proficiency and labour market outcomes, it may also be related to other life choices associated with those outcomes. This variable is available in the years 2008 and 2009 only, resulting in a smaller sample (also due to a large number of missing values) in which estimations are carried out. Columns (3) and (8) of Table 5 show the results. The marginal effect of speaking Russian for males is larger as a result of inclusion of this variable, while that for females hardly changes. As a result, the returns to Russian language proficiency are higher for males than for females - a pattern that is consistent with some of the findings by Ginsburgh and Prieto-Rodriguez (2013). BICULTURALIST itself is significantly negatively correlated with the probability of employment suggesting positive labour market returns to nationalism (or its correlates).

Finally, columns (4) and (9) of Table 5 report the marginal effects from a model where the additional variables are included jointly. The marginal effects on RUSSIAN SPEAKER preserve their statistical significance. We estimate marginal effects of Russian proficiency of 10.5 (males) and 8.2 (females) percentage points. The original finding of a positive and

significant effect of Russian language proficiency for both males and females is robust to the inclusion of additional variables.

As an additional check on the sensitivity of our results to unobserved heterogeneity, we apply a propensity score matching estimator. We use the comprehensive set of controls from Table 4 to compute propensity scores using kernel methods and compare the employment probabilities of Russian speakers and non-speakers. The results are reported in Table 6. Columns (1) and (4) restate the marginal effects of speaking Russian from the baseline model, while columns (2) and (5) restate those from the model with the two additional controls included. The estimated treatment effects for both males and females are statistically significant and very similar in magnitude to the marginal effects from the model with additional controls. This serves as a validation to our initial strategy to mitigating unobserved heterogeneity through inclusion of additional controls, and further confirms that Russian proficiency has non-zero and potentially large positive effect on employment probability of individuals.

[Table 5 about here.]

[Table 6 about here.]

Measurement of Russian proficiency. Language proficiency can be defined in a number of ways, bearing implications for the results. In Table 7, we report the marginal effects from regressions using different measures of language skills. Columns (1) and (4) restate the marginal effects of speaking Russian from the baseline model.

The measure of Russian language proficiency used in the baseline regressions is generated based on the ability in Russian language self-reported on a four-point scale of 'advanced', 'intermediate', 'beginner' and 'no basic knowledge'. Following a common approach in the literature, we reduced this information to a dummy variable, coded 1 if Russian language proficiency is reported as advanced or intermediate and 0 otherwise. While this approach has the benefit of minimising the probability of misclassification of language proficiency, it conceals the information on different levels of language proficiency. Columns (2) and (5) of Table 7 report marginal effects from regressions where Russian proficiency is instead measured with dummies for beginner, intermediate and advanced levels (with no basic knowledge being the excluded category). In our sample of males, 17.38% of the respondents report speaking Russian at beginner's, 50.73% at intermediate and 23.80% at advanced level. Similarly, among females, 18.30% report speaking Russian at beginner's, 46.24% at intermediate and 20.72% at advanced level. The remaining respondents have no basic knowledge of the language. According to the results, beginner's level Russian is associated with an increase in

the employment probability by 7.9 (males) and 9.7 (females) percentage points. The returns to intermediate-level Russian are higher at 11.9 percentage points for males and 13.7 for females. Advanced Russian speakers are 13.6 (males) and 19.1 (females) percentage points more likely to be employed relative to those with no basic knowledge of the language.

The use of self-assessed measures of Russian proficiency can be potentially problematic if different respondents consider different factors in assessing their language ability without the benefit of a common base. In addition to self-reported language proficiency, DIS includes information on language usage. In particular, the respondents are asked "Which language did you mainly speak at home when you were a child?". RUSSIAN NATIVE incorporates this information in the measurement of language proficiency. It is set to 1 for individuals who spoke Russian at home as children, or if a language other than Russian was spoken in the home, spoke Russian at an 'advanced' level. It was set to 0 where a language other than Russian was spoken in the home and the respondent had 'no basic knowledge' of Russian or spoke it at a beginner's or intermediate level. This stringent definition of Russian skills has the additional benefit of minimising the problem of measurement error associated with the possibility of people potentially over-reporting their language proficiency (Dustmann and van Soest, 2001, 2002). Based on this measure, 23.80% of males and 20.72% of females are 'native speakers'. According to the results reported in columns (3) and (6) of Table 7, this measure of Russian proficiency is associated with an increase in the probability of employment by 2.9 (males) and 6.9 (females) percentage points. Overall, the results using different measures of Russian proficiency confirm that speaking Russian is positively related to the probability of being employed.

[Table 7 about here.]

Heterogeneity in returns to Russian. Following up on the earlier discussion of descriptive statistics, we explore the heterogeneity in returns to Russian language skills according to several observable characteristics. The results of this exercise are summarised in Tables 8-10.

First, we consider whether individuals with different levels of education gain different returns to their Russian language skills (Table 8). We re-estimate the baseline model separately for individuals with three different levels of education. Considering the males in the lowest education category, Russian language proficiency is associated with increase in the probability of employment by 4.5 percentage points. Russian language returns of males with post-school education are higher (13.0 (intermediate level) and 13.2 (university level) percentage points), suggesting complementarity between language skills and education. For females,

we find a statistically significant relationship between Russian proficiency and employment for those with 11-14 years of education only.

Second, we explore heterogeneity in returns to Russian language skills by age cohort (Table 9). Accordingly, we re-estimate the return to Russian proficiency in the samples of younger (26-35) and older (36-60) individuals. The return to Russian language skills appears to be higher for older males. It is 7.1 percentage points compared to 4.5 percentage points for younger males. This advantage to older males may be due to their greater work experience. According to our estimates, Russian proficiency is statistically significantly related to employment probability of young females only. While we find that a higher level of education increases the employment probability of males and females in both age groups, the return to university-level education (relative to school education or below) is higher for older males and females. In all cases, the returns to education are higher for females than males.

Finally, in Table 10 we consider the returns to Russian language skills by country. The labour market value of speaking Russian is the highest for men in Georgia (14 percentage points), while for women, it appears to be non-existent. Conversely, both in Armenia and Azerbaijan women appear to gain more from Russian proficiency than men do. In Armenia, speaking Russian is associated with 6.8 (males) and 7.4 (females) percentage points increase in the probability of employment. In Azerbaijan, the estimated marginal effects of Russian proficiency are 4.7 (males) and 13.1 (females) percentage points. In all three countries, there are positive returns to university level education, more so for females than males.

Overall, the relevance of Russian language skills for securing employment for males remains clear, regardless of respondents' education, age and location. The results for females, while less robust, nevertheless broadly support the baseline finding on the positive labour market return to Russian language proficiency.

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

Russian proficiency and occupational choice. Russian language skills remain economically valuable since individuals must fit into the existing economic and social systems, where Russian language continues to maintain a strong influence. However, certain activities and types of jobs are more intensive in communication skills and have higher language

requirements than others, and knowledge of Russian language may be an important determinant of individuals' selection into such jobs. Previous studies in the context of multilingual labour markets have identified the contribution of language skills to several patterns of occupational choice. First, given the strict regulation of language requirements in public sector occupations, language-proficient workers are more likely to be selected into the public sector (Di Paolo, 2011). Second, language proficiency also increases the selection into communication-intensive jobs such as those in trade, services, government and education, among others (Quella and Rendon, 2012). The DIS allows us to test whether these observations hold in our context.

First, the respondents are asked on the type of their primary workplace. We use this information to define dummies for employment in: (1) - own business/a small family business, (2) - a medium or large private organisation, (3) - a state organisation and (4) - a foreign organisation, business or an international or local NGO. The probability of being employed in own business or a small family business is 33.9% for males and 25.26% for females. The probability of employment in a medium or large private organisation is lower: 20.57% for males and 15.92% for females. Over half (50.47%) of females are employed in a state organisation compared to 31.35% employment rate among males. Foreign organisations, businesses and NGOs employ 4.3% of males and 2.38% of females, while the rest are employed elsewhere.

The estimated marginal effects from models with these indicators used as dependent variables are reported in Table 11. Among males, Russian speakers are less likely to be found among the self-employed or among those employed in a small family business. Instead, we confirm that the language proficiency contributes to the males' selection into public sector. This is not surprising, given that on average, the requirements for language proficiency, at least formally, are likely to be higher in the public sector. In the Soviet Union, Russian proficiency was key to employment in the state apparatus. Even though the titular languages are the official languages in post-Soviet Armenia, Azerbaijan and Georgia, Russia remains a powerful neighbour influencing the policy-making in the three countries. As a result, Russian remains an important resource used in government affairs particularly in areas such as foreign policy, security, development. As expected, we also find that Russian proficiency is a statistically significant determinant of males' selection into jobs in foreign organisations, businesses or NGOs.

[Table 11 about here.]

For both males and females, below university level education appears to contribute to employment in private sector organisations of different scales. Conversely, university level

education increases the individuals' employment in state organisations, to a particularly large extent for females. For males, it additionally increases the probability of employment in foreign organisations, businesses or NGOs.

Second, the respondents are asked on the sector of their primary workplace. Based on this information we define dummies for employment in: (1) - agriculture, forestry and fishing, (2) - mining, manufacturing and construction, (3) - trade, services and communication, and (4) - government, education and healthcare. The employment rate in agriculture, forestry and fishing is 16.3% for males and 13.2% for females. Mining, manufacturing and construction employ 27.13% of males but only 6.62% of females. The probability of employment in trade, services and communications is 23.95% for males and 18.53% for females. Government, education and healthcare employ close to half (47.61%) of females but only 18.94% of males. The rest are employed in other sectors.

The estimated marginal effects from models with these indicators used as dependent variables are reported in Table 12. Among males, we confirm that language proficiency increases selection into jobs in trade, services and communication. Additionally, consistent with the findings in Table 11, we find that Russian-speaking males are more likely to be employed in government, education and healthcare (largely state organisations). As Quella and Rendon (2012) argue, these are sectors where high level communication skills are particularly needed as individuals need to relate to others not only through reading and speaking but also by producing written documents such as reports and instructions.

[Table 12 about here.]

Human capital characteristics (or demographic characteristics) do not have any explanatory power over males' and females' probabilities of employment in agriculture, forestry and fishing jobs. What does appear to matter for selection into jobs in that sector is one's location: the probability of employment is higher for those outside capital cities, and particularly for those who are based in rural areas (results not reported but available on request). For both males and females, below university level education is associated with a higher probability of employment in jobs in mining, manufacturing, and construction as well as in trade, services, and communication. Those with university level education, on the other hand, have a higher probability of holding jobs in government, education and healthcare.

The analysis summarised in Tables 11 and 12 does not yield statistically significant marginal effects on Russian proficiency for females. In a study on gender bias in the use of

foreign languages, Ginsburgh and Prieto-Rodriguez (2013) find that in some European countries men get higher returns on foreign language use than women. They interpret this finding as suggestive of some form of discrimination against women linked to this specific type of human capital. While we do not observe a similar pattern in our baseline estimates, the absence of a statistically significant relationship between women's language proficiency and their employment, when considered disaggregated by employment type and sector, is puzzling, and warrants further investigation in future research.

6. Conclusion

Two decades after transition, many formerly Soviet republics are still economically and politically dependent on Russia and each other. Russian language, remaining the main *lingua franca* in the post-Soviet territory, continues to be widely used in business and public life. As a result, Russian language skills remain economically valuable in the former Soviet republics. Based on data from Armenia, Azerbaijan and Georgia in the years 2008-2010, this paper shows that Russian proficiency increases a person's probability of employment by about 6 (males) and 9 (females) percentage points. Comparison with returns to other forms of human capital suggests that this effect is economically significant. This result is consistent with findings on positive labour market returns to dominant non-indigenous language skills in the former colonies (Angrist and Lavy, 1997; Chiswick et al., 2000; Godoy et al., 2007; Levinsohn, 2007; Casale and Posel, 2011; Azam et al., 2013). This paper contributes to this nascent literature on returns to language skills in developing countries by providing the first evidence on the economic value of speaking Russian in post-Soviet countries.

The findings in this paper are important in informing the ongoing debates on language policies in the post-Soviet countries. Many of these countries have initiated various programs to promote the national languages at the expense of Russian since the collapse of the Soviet Union. The results of this paper suggest that the resultant drop of Russian language skills may lead to future loss of economic opportunities. However, there is evidence to suggest that promotion of native language policies affects national identity and political preferences, therefore affecting the selection of politicians and the policies they implement (Aspachs-Bracons et al., 2008; Clots-Figueras and Masella, 2013). Assessment of these additional outcomes is required to make conclusive recommendations on language policies.

This paper is a first step in investigating empirically the association between Russian language skills and labour market outcomes in transition countries, however more research is needed. Methodologically, finding random sources of variation in Russian proficiency,

subject to data availability, is required to achieve better identification of the effect of language. Measurement of labour market returns to Russian language skills in terms of wages is an important direction of future research. Furthermore, considering the impact of Russian language skills on social and political outcomes is another area with high potential returns to further analysis. More research based on datasets from other transition countries and contexts would be a valuable addition to knowledge.

REFERENCES

Albouy, D. (2008). The wage gap between Francophones and Anglophones: A Canadian perspective, 1970–2000. *Canadian Journal of Economics/Revue canadienne d'économique 41*(4), 1211–1238.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). An evaluation of instrumental variable strategies for estimating the effects of Catholic schooling. *Journal of Human Resources 40*(4), 791–821.

Angrist, J. D. and V. Lavy (1997). The effect of a change in language of instruction on the returns to schooling in Morocco. *Journal of Labor Economics 15*(1), S48–76.

Arabsheibani, G. R. and A. Mussurov (2007). Returns to schooling in Kazakhstan. *Economics of Transition 15*(2), 341–364.

Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics 2*(3), 193–216.

Aspachs-Bracons, O., I. Clots-Figueras, J. Costa-Font, and P. Masella (2008). Compulsory language education policies and identity formation. *Journal of the European Economic Association 6*(2-3), 434–444.

Azam, M., A. Chin, and N. Prakash (2013). The returns to English-language skills in India. *Economic Development and Cultural Change 61*(2), pp. 335–367.

Bilinsky, Y. (1962). The Soviet education laws of 1958-9 and Soviet nationality policy. *Soviet Studies 14*(2), 138–157.

Bleakley, H. and A. Chin (2004, 2013/12/02). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics 86*(2), 481–496.

Brainerd, E. (2000). Women in transition: Changes in gender wage differentials in Eastern Europe and the Former Soviet Union. *Industrial and Labor Relations Review 54*(1), pp. 138–162.

Casale, D. and D. Posel (2011). English language proficiency and earnings in a developing country: The case of South Africa. *The Journal of Socio-Economics 40*(4), 385 – 393.

Charles, R. (2010). Religiosity and trust in religious institutions: Tales from the South Caucasus (Armenia, Azerbaijan, and Georgia). *Politics and Religion 3*(Special Issue 02), 228–261.

Chiswick, B. R. and P. W. Miller (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics 13*(2), pp. 246–288.

Chiswick, B. R. and P. W. Miller (1999). Language skills and earnings among legalized aliens. *Journal of Population Economics 12*(1), 63–89.

Chiswick, B. R. and P. W. Miller (2002). Immigrant earnings: Language skills, linguistic concentrations and the business cycle. *Journal of Population Economics 15*(1), 31–57.

Chiswick, B. R. and P. W. Miller (2014). International migration and the economics of language. *IZA Discussion Paper 7880*.

Chiswick, B. R., H. A. Patrinos, and M. E. Hurst (2000). Indigenous language skills and the labor market in a developing economy: Bolivia. *Economic Development and Cultural Change 48*(2), pp. 349–367.

Clots-Figueras, I. and P. Masella (2013). Education, language and identity. *The Economic Journal 123*(570), F332–F357.

CRRC (2007). Data initiative 2007 - unified sampling methodology. Technical report, Caucasus Research Resource Centres.

CRRC (2008). Data initiative - 2008. Fieldwork report, Caucasus Research Resource Centres.

Das, J., Q.-T. Do, K. Shaines, and S. Srikant (2013). U.S. and them: The geography of academic research. *Journal of Development Economics 105*(0), 112 – 130.

Di Paolo, A. (2011). Knowledge of Catalan, public/private sector choice and earnings: Evidence from a double sample selection model. *Hacienda Pública Española 197*(2), 9–35.

Di Paolo, A. and A. Tansel (2013). Returns to foreign language skills in a developing country: The case of Turkey. *IZA Discussion Paper 7724*.

Dietrich, A. P. (2005). Language policy and the status of Russian in the Soviet Union and the successor states outside the Russian Federation. *Australian Slavonic and East European Studies 19*(1/2), 1–27.

Dustmann, C. (1994). Speaking fluency, writing fluency and earnings of migrants. *Journal of Population Economics 7*(2), 133–156.

Dustmann, C. and F. Fabbri (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal 113*(489), 695–717.

Dustmann, C. and A. van Soest (2001). Language fluency and earnings: Estimation with misclassified language indicators. *Review of Economics and Statistics 83*(4), 663 – 674.

Dustmann, C. and A. van Soest (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review 55*(3), pp. 473–492.

European Commission (2012). Trade with partner regions. http://ec.europa.eu/trade/policy/countries-and-regions/statistics/index_en.htm.

Fierman, W. (2006). Language and education in post-Soviet Kazakhstan: Kazakh-medium instruction in urban schools. *The Russian Review 65*(1), 98–116.

Ginsburgh, V. and J. Prieto-Rodriguez (2013). Is there a gender bias in the use of foreign languages in Europe? *Kyklos 66*(4), 552–566.

Ginsburgh, V. A. and J. Prieto-Rodriguez (2010). Returns to foreign languages of native workers in the European Union. *Industrial and Labor Relations Review 64*(3), 599–618.

Godoy, R., V. Reyes-García, C. Seyfried, T. Huanca, W. R. Leonard, T. McDade, S. Tanner, and V. Vadez (2007, 6). Language skills and earnings: Evidence from a pre-industrial economy in the Bolivian amazon. *Economics of Education Review 26*(3), 349–360.

Grigoryan, M. (2010). Opposition mounts to proposed foreign-language schools in Armenia. http://www.eurasianet.org/node/61317.

Grin, F. and C. Sfreddo (1998). Language-based earnings differentials on the Swiss labour market: Is Italian a liability? *International Journal of Manpower 19*(7), 520–532.

Habibov, N. (2012). Income inequality and its driving forces in transitional countries: Evidence from Armenia, Azerbaijan and Georgia. *Journal of Comparative Social Welfare 28*(3), 209–221.

Habibov, N. N. (2011). Self-perceived social stratification in low-income transitional countries: Examining the multi-country survey in Armenia, Azerbaijan, and Georgia. *International Journal of Social Economics 38*(1), 5–22.

Habibov, N. N. and E. Afandi (2009). Analysis of subjective wellbeing in low-income transitional countries: Evidence from comparative national surveys in Armenia, Azerbaijan and Georgia. *Journal of Comparative Social Welfare 25*(3), 203–219.

Habibov, N. N. and E. N. Afandi (2011). Self-rated health and social capital in transitional countries: Multilevel analysis of comparative surveys in Armenia, Azerbaijan, and Georgia. *Social Science and Medicine 72*(7), 1193 – 1204.

Henley, A. and R. E. Jones (2005). Earnings and linguistic proficiency in a bilingual economy. *The Manchester School 73*(3), 300–320.

Isphording, I. E. (2013). Returns to foreign language skills of immigrants in Spain. *LABOUR 27*(4), 443–461.

Kirkwood, M. (1991). Glasnost' 'the national question' and Soviet language policy. *Soviet Studies 43*(1), 61–81.

Lang, K. and E. Siniver (2009). The return to English in a non-English speaking country: Russian immigrants and native Israelis in Israel. *The B.E. Journal of Economic Analysis and Policy 9*(1), 1935–1682.

Levinsohn, J. (2007). *Globalization and Poverty*, Chapter Globalization and the Returns to Speaking English in South Africa. University of Chicago Press.

Newell, A. and B. Reilly (1999). Rates of return to educational qualifications in the transitional economies. *Education Economics 7*(1), 67–84.

Pastore, F. and A. Verashchagina (2006). Private returns to human capital over transition: A case study of Belarus. *Economics of Education Review 25*(1), 91 – 107.

Pavlenko, A. (2008). Russian in post-Soviet countries. *Russian Linguistics 32*(1), 59–80.

Quella, N. and S. Rendon (2012). Occupational selection in multilingual labor markets: The case of Catalonia. *International Journal of Manpower 33*(8), 918–937.

Rendon, S. (2007). The Catalan premium: Language and employment in Catalonia. *Journal of Population Economics 20*(3), 669–686.

Roberts, K. and G. Pollock (2011). Class in the South Caucasus. *European Societies 13*(5), 757–782.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Shields, M. A. and S. W. Price (2002). The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. *Journal of Population Economics 15*(1), 137–160.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Toomet, O. (2011). Learn English, not the local language! Ethnic Russians in the Baltic states. *The American Economic Review 101*(3), 526–531.

List of Tables

Table 1: Description of variables

Variable	Definition	Mean (s.d.)		
		Non-speakers	Speakers	All
RUSSIAN SPEAKER	0-1 binary variable; equals 1 if respondent has advanced or intermediate knowledge of Russian language	0	1	0.70 (0.46)
EMPLOYED	0-1 binary variable; equals 1 if respondent reported having a job	0.34 (0.47)	0.51 (0.50)	0.46 (0.50)
MALE	0-1 binary variable; equals 1 if respondent is male	0.38 (0.48)	0.47 (0.50)	0.44 (0.50)
EDUC$\leq$ 10	0-1 binary variable; equals 1 if respondent has no more than 10 years of education	0.60 (0.49)	0.30 (0.46)	0.39 (0.49)
EDUC11-14	0-1 binary variable; equals 1 if respondent has between 11 and 14 years of education	0.34 (0.47)	0.38 (0.48)	0.36 (0.48)
EDUC$\geq$ 15	0-1 binary variable; equals 1 if respondent has 15 or more years of education	0.06 (0.25)	0.33 (0.47)	0.25 (0.43)
AGE 26-30	0-1 binary variable; equals 1 if respondent is between 26 and 30 years old	0.17 (0.37)	0.13 (0.33)	0.14 (0.34)
AGE 31-35	0-1 binary variable; equals 1 if respondent is between 31 and 35 years old	0.14 (0.35)	0.13 (0.34)	0.14 (0.34)
AGE 36-40	0-1 binary variable; equals 1 if respondent is between 36 and 40 years old	0.15 (0.35)	0.15 (0.36)	0.15 (0.36)
AGE 41-45	0-1 binary variable; equals 1 if respondent is between 41 and 45 years old	0.15 (0.36)	0.15 (0.35)	0.15 (0.36)
AGE 46-50	0-1 binary variable; equals 1 if respondent is between 46 and 50 years old	0.16 (0.36)	0.18 (0.38)	0.17 (0.38)
AGE 51-55	0-1 binary variable; equals 1 if respondent is between 51 and 55 years old	0.13 (0.34)	0.16 (0.36)	0.15 (0.36)
AGE 56-60	0-1 binary variable; equals 1 if respondent is between 56 and 60 years old	0.11 (0.31)	0.11 (0.32)	0.11 (031)

Table 1: Description of variables (continued)

Variable	Definition	Mean (s.d.)		
		Non-speakers	Speakers	All
ETHNIC MAJORITY	0-1 binary variable; equals 1 if respondent is from the ethnic majority	0.93 (0.25)	0.94 (0.24)	0.93 (0.25)
LINGUISTIC MAJORITY	0-1 binary variable; equals 1 if respondent is from the linguistic majority	0.94 (0.24)	0.92 (0.27)	0.93 (0.26)
CAPITAL CITY	0-1 binary variable; equals 1 if respondent lives in capital city	0.18 (0.38)	0.33 (0.47)	0.29 (0.45)
OTHER URBAN	0-1 binary variable; equals 1 if respondent lives in urban area	0.31 (0.46)	0.37 (0.48)	0.35 (0.48)
RURAL	0-1 binary variable; equals 1 if respondent lives in rural area	0.51 (0.50)	0.30 (0.46)	0.36 (0.48)
AZERBAIJAN	0-1 binary variable; equals 1 if respondent lives in Azerbaijan	0.61 (0.49)	0.27 (0.44)	0.37 (0.48)
ARMENIA	0-1 binary variable; equals 1 if respondent lives in Armenia	0.15 (0.36)	0.41 (0.49)	0.33 (0.47)
GEORGIA	0-1 binary variable; equals 1 if respondent lives in Georgia	0.24 (0.43)	0.32 (0.47)	0.30 (0.46)

Table 2: Russian speaking and employment probabilities - males

		Mean of RUSSIAN SPEAKER (s.d.)	Mean of EMPLOYED (s.d.)		
			Non-speakers	Speakers	All
Education	EDUC$\leq$ 10	0.63	0.52	0.55	0.54
		(0.48)	(0.50)	(0.50)	(0.50)
	EDUC11-14	0.74	0.54	0.63	0.60
		(0.44)	(0.50)	(0.48)	(0.49)
	EDUC$\geq$ 15	0.93	0.68	0.77	0.76
		(0.25)	(0.47)	(0.42)	(0.43)
Age	AGE 26-30	0.65	0.58	0.67	0.64
		(0.48)	(0.49)	(0.47)	(0.48)
	AGE 31-35	0.68	0.56	0.71	0.66
		(0.47)	(0.50)	(0.45)	(0.47)
	AGE 36-40	0.73	0.54	0.67	0.64
		(0.45)	(0.50)	(0.47)	(0.48)
	AGE 41-45	0.74	0.54	0.68	0.65
		(0.44)	(0.50)	(0.47)	(0.48)
	AGE 46-50	0.79	0.58	0.63	0.62
		(0.40)	(0.50)	(0.48)	(0.49)
	AGE 51-55	0.82	0.43	0.60	0.57
		(0.38)	(0.50)	(0.49)	(0.50)
	AGE 56-60	0.81	0.40	0.56	0.53
		(0.40)	(0.49)	(0.50)	(0.50)
Country	ARMENIA	0.90	0.47	0.63	0.62
		(0.31)	(0.50)	(0.48)	(0.49)
	AZERBAIJAN	0.60	0.60	0.75	0.69
		(0.49)	(0.49)	(0.44)	(0.46)
	GEORGIA	0.79	0.38	0.56	0.52
		(0.41)	(0.49)	(0.50)	(0.50)
Total		0.75	0.53	0.65	0.62
		(0.44)	(0.50)	(0.48)	(0.49)

Table 3: Russian speaking and employment probabilities - females

		Mean of RUSSIAN SPEAKER (s.d.)	Mean of EMPLOYED (s.d.)		
			Non-speakers	Speakers	All
Education	EDUC$\leq$ 10	0.47	0.17	0.23	0.20
		(0.50)	(0.38)	(0.42)	(0.40)
	EDUC11-14	0.72	0.26	0.37	0.34
		(0.45)	(0.44)	(0.48)	(0.47)
	EDUC$\geq$ 15	0.91	0.56	0.56	0.56
		(0.28)	(0.50)	(0.50)	(0.50)
Age	AGE 26-30	0.63	0.17	0.29	0.25
		(0.48)	(0.37)	(0.46)	(0.43)
	AGE 31-35	0.70	0.16	0.37	0.31
		(0.46)	(0.37)	(0.48)	(0.46)
	AGE 36-40	0.69	0.22	0.41	0.35
		(0.46)	(0.41)	(0.49)	(0.48)
	AGE 41-45	0.66	0.24	0.45	0.38
		(0.48)	(0.43)	(0.50)	(0.49)
	AGE 46-50	0.67	0.27	0.41	0.37
		(0.47)	(0.45)	(0.49)	(0.48)
	AGE 51-55	0.68	0.24	0.43	0.37
		(0.47)	(0.43)	(0.50)	(0.48)
	AGE 56-60	0.65	0.24	0.39	0.34
		(0.48)	(0.43)	(0.49)	(0.47)
Country	ARMENIA	0.84	0.23	0.40	0.38
		(0.36)	(0.42)	(0.49)	(0.48)
	AZERBAIJAN	0.42	0.20	0.45	0.31
		(0.49)	(0.40)	(0.50)	(0.46)
	GEORGIA	0.74	0.27	0.35	0.33
		(0.44)	(0.44)	(0.48)	(0.47)
Total		0.67	0.22	0.40	0.34
		(0.47)	(0.42)	(0.49)	(0.47)

Table 4: Baseline probit model - coefficients and marginal effects

	Males		Females	
	Coefficients	Marginal effects	Coefficients	Marginal effects
	(1)	(2)	(3)	(4)
RUSSIAN SPEAKER	0.265***	0.064***	0.214*	0.085*
	(0.026)	(0.014)	(0.115)	(0.045)
EDUC≤10	-0.559***	-0.186***	-1.023***	-0.340***
	(0.053)	(0.005)	(0.128)	(0.045)
EDUC11-14	-0.391***	-0.124***	-0.594***	-0.221***
	(0.079)	(0.018)	(0.074)	(0.031)
AGE 26-30	-0.012	-0.003	-0.388***	-0.150***
	(0.088)	(0.025)	(0.042)	(0.018)
AGE 31-35	0.091	0.024	-0.159**	-0.063**
	(0.079)	(0.017)	(0.067)	(0.027)
AGE 41-45	0.064	0.017	0.113	0.045
	(0.088)	(0.021)	(0.074)	(0.030)
AGE 46-50	-0.062	-0.017	0.092***	0.037***
	(0.086)	(0.026)	(0.024)	(0.010)
AGE 51-55	-0.176***	-0.052***	0.059	0.024
	(0.048)	(0.016)	(0.066)	(0.026)
AGE 56-60	-0.280***	-0.085***	0.015	0.006
	(0.032)	(0.013)	(0.075)	(0.030)
ETHNIC MAJORITY	0.163	0.048	0.001	0.000
	(0.131)	(0.038)	(0.033)	(0.013)
LINGUISTIC MAJORITY	-0.076	-0.020	-0.017	-0.007
	(0.184)	(0.045)	(0.086)	(0.034)
OTHER URBAN	-0.223***	-0.067***	0.167***	0.066***
	(0.083)	(0.020)	(0.030)	(0.012)
RURAL	-0.285**	-0.087**	0.109***	0.044***
	(0.134)	(0.036)	(0.018)	(0.007)
ARMENIA	-0.278***	-0.085***	0.045	0.018
	(0.013)	(0.013)	(0.047)	(0.019)
GEORGIA	-0.519***	-0.170***	-0.206***	-0.081***
	(0.016)	(0.019)	(0.051)	(0.021)
YEARS	Yes	Yes	Yes	Yes
N	4,684	4,684	5,993	5,993

Note.—Marginal effects are calculated for a 36-40 years old university-educated non-Russian speaker from the ethnic and linguistic majority, residing in Baku (Azerbaijan) in 2010; standard errors are clustered at the country level; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level.

Table 5: Probit model including additional controls - marginal effects

	Males				Females			
	(1)	(2)	(3)	(4)	(6)	(7)	(8)	(9)
RUSSIAN SPEAKER	0.064***	0.065***	0.095***	0.105**	0.085*	0.080**	0.086**	0.082**
	(0.014)	(0.018)	(0.035)	(0.047)	(0.045)	(0.040)	(0.039)	(0.033)
ENGLISH SPEAKER		0.057***		0.101***		0.125***		0.113***
		(0.018)		(0.038)		(0.026)		(0.020)
BICULTURALIST			-0.023*	-0.031**			-0.045***	-0.049**
			(0.013)	(0.015)			(0.017)	(0.024)
EDUC≤10	-0.186***	-0.175***	-0.182***	-0.170***	-0.340***	-0.313***	-0.296***	-0.265***
	(0.005)	(0.014)	(0.022)	(0.039)	(0.045)	(0.049)	(0.068)	(0.082)
EDUC11-14	-0.124***	-0.113***	-0.112***	-0.097*	-0.221***	-0.194***	-0.211***	-0.178***
	(0.018)	(0.025)	(0.032)	(0.051)	(0.031)	(0.038)	(0.042)	(0.051)
AGE 26-30	-0.003	-0.007	-0.029	-0.040	-0.150***	-0.153***	-0.151***	-0.154***
	(0.025)	(0.027)	(0.042)	(0.038)	(0.018)	(0.017)	(0.023)	(0.026)
AGE 31-35	0.024	0.015	-0.016	-0.031	-0.063**	-0.068***	-0.062**	-0.067**
	(0.017)	(0.024)	(0.033)	(0.028)	(0.027)	(0.026)	(0.031)	(0.028)
AGE 41-45	0.017	0.022	-0.010	0.001	0.045	0.049	0.034	0.037
	(0.021)	(0.024)	(0.064)	(0.054)	(0.030)	(0.035)	(0.041)	(0.047)
AGE 46-50	-0.017	-0.017	-0.023	-0.030	0.037***	0.040***	0.078***	0.077***
	(0.026)	(0.028)	(0.040)	(0.050)	(0.010)	(0.015)	(0.006)	(0.014)
AGE 51-55	-0.052***	-0.052***	-0.071**	-0.083***	0.024	0.033	0.010	0.010
	(0.016)	(0.016)	(0.032)	(0.026)	(0.026)	(0.039)	(0.044)	(0.049)
AGE 56-60	-0.085***	-0.089***	-0.108	-0.132**	0.006	0.016	0.024	0.048
	(0.013)	(0.015)	(0.072)	(0.058)	(0.030)	(0.029)	(0.046)	(0.053)
Other controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	4,684	4,478	1,505	1,405	5,993	5,790	2,068	1,981

Note.—Marginal effects are calculated for a 36-40 years old university-educated non-Russian speaker from the ethnic and linguistic majority, residing in Baku (Azerbaijan) in 2010; the assumed default values for additional variables are set as: ENGLISH SPEAKER= 0, BICULTURALIST= 0; standard errors are clustered at the country level; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level; other controls set included is identical to Table 4.

Table 6: Propensity score matching - average treatment effects

	Males			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
RUSSIAN SPEAKER	0.064***	0.105**	0.105***	0.085*	0.082**	0.080***
	(0.014)	(0.047)	(0.025)	(0.045)	(0.033)	(0.020)
N	4,684	1,405	4,684	5,993	1,981	5,993

Note.—Columns (1) and (4) restate the marginal effects reported in columns (2) and (4) of Table 4; columns (2) and (5) restate the marginal effects reported in columns (4) and (9) of Table 5; columns (3) and (6) report the average treatment effects from Kernel matching estimator with bandwidth = 0.00009 and with standard errors calculated from bootstrapping with 500 replications; the propensity scores are calculated using the complete set of controls from Table 4; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level.

Table 7: Probit model using different measures of Russian proficiency - marginal effects

	Males			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
RUSSIAN SPEAKER	0.064***			0.085*		
	(0.014)			(0.045)		
RUSSIAN BEGINNER		0.079***			0.097***	
		(0.008)			(0.014)	
RUSSIAN INTERMEDIATE		0.119***			0.137***	
		(0.020)			(0.034)	
RUSSIAN ADVANCED		0.136***			0.191***	
		(0.026)			(0.061)	
RUSSIAN NATIVE			0.029*			0.069*
			(0.016)			(0.042)
EDUC$\leq$10	-0.186***	-0.195***	-0.179***	-0.340***	-0.297***	-0.360***
	(0.005)	(0.014)	(0.004)	(0.045)	(0.036)	(0.060)
EDUC11-14	-0.124***	-0.133***	-0.115***	-0.221***	-0.198***	-0.225***
	(0.018)	(0.026)	(0.015)	(0.031)	(0.025)	(0.030)
AGE 26-30	-0.003	0.001	-0.006	-0.150***	-0.138***	-0.154***
	(0.025)	(0.029)	(0.020)	(0.018)	(0.017)	(0.016)
AGE 31-35	0.024	0.031	0.018	-0.063**	-0.058**	-0.065**
	(0.017)	(0.021)	(0.014)	(0.027)	(0.025)	(0.029)
AGE 41-45	0.017	0.019	0.017	0.045	0.045	0.045
	(0.021)	(0.025)	(0.018)	(0.030)	(0.028)	(0.029)
AGE 46-50	-0.017	-0.019	-0.009	0.037***	0.038***	0.037***
	(0.026)	(0.030)	(0.023)	(0.010)	(0.006)	(0.008)
AGE 51-55	-0.052***	-0.060***	-0.040**	0.024	0.024	0.020
	(0.016)	(0.017)	(0.016)	(0.026)	(0.025)	(0.027)
AGE 56-60	-0.085***	-0.096***	-0.069***	0.006	0.007	-0.000
	(0.013)	(0.013)	(0.011)	(0.030)	(0.030)	(0.032)
Other controls included	Yes	Yes	Yes	Yes	Yes	Yes
N	4,684	4,684	4,684	5,993	5,993	5,993

Note.—Marginal effects are calculated for a 36-40 years old university-educated non-Russian speaker from the ethnic and linguistic majority, residing in Baku (Azerbaijan) in 2010; the assumed default values for new language proficiency measures are set as: RUSSIAN BEGINNER= 0, RUSSIAN INTERMEDIATE= 0, RUSSIAN ADVANCED= 0, RUSSIAN NATIVE= 0; standard errors are clustered at the country level; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level; other controls set included is identical to Table 4.

Table 8: Probit model by education level sub-samples - marginal effects

	Males			Females		
	EDUC≤10	EDUC11-14	EDUC ≥15	EDUC≤10	EDUC11-14	EDUC ≥15
	(1)	(2)	(3)	(4)	(5)	(6)
RUSSIAN SPEAKER	0.045*	0.130***	0.132**	0.017	0.128***	0.037
	(0.024)	(0.028)	(0.053)	(0.015)	(0.026)	(0.048)
AGE 26-30	0.028	-0.009	-0.036	-0.029	-0.150***	-0.179***
	(0.040)	(0.042)	(0.049)	(0.025)	(0.033)	(0.047)
AGE 31-35	0.021	0.001	0.103**	-0.020	-0.110***	-0.021
	(0.040)	(0.043)	(0.048)	(0.024)	(0.034)	(0.047)
AGE 41-45	0.012	0.017	0.057	0.042*	0.002	0.060
	(0.039)	(0.041)	(0.050)	(0.025)	(0.035)	(0.047)
AGE 46-50	0.011	-0.045	-0.044	0.047*	0.021	-0.008
	(0.039)	(0.042)	(0.049)	(0.024)	(0.034)	(0.047)
AGE 51-55	-0.070*	-0.092**	-0.017	0.034	0.017	-0.036
	(0.042)	(0.043)	(0.050)	(0.025)	(0.034)	(0.048)
AGE 56-60	-0.127***	-0.100**	-0.086	0.007	-0.015	0.016
	(0.049)	(0.049)	(0.057)	(0.025)	(0.037)	(0.052)
Other controls included	Yes	Yes	Yes	Yes	Yes	Yes
N	1,756	1,746	1,182	2,366	2,150	1,477

Note.—Marginal effects are calculated for a 36-40 years old non-Russian speaker from the ethnic and linguistic majority, residing in Baku (Azerbaijan) in 2010; standard errors are clustered at the country level; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level; other controls set included is identical to Table 4.

Table 9: Probit model by age group sub-samples - marginal effects

	Males		Females	
	AGE 26-35	AGE 36-60	AGE 26-35	AGE 36-60
	(1)	(2)	(3)	(4)
RUSSIAN SPEAKER	0.045***	0.071***	0.113**	0.075
	(0.011)	(0.017)	(0.050)	(0.046)
EDUC≤10	-0.134***	-0.206***	-0.277***	-0.337***
	(0.005)	(0.006)	(0.018)	(0.051)
EDUC11-14	-0.112***	-0.130***	-0.226***	-0.206***
	(0.018)	(0.027)	(0.025)	(0.034)
AGE 31-35	0.022***		0.086**	
	(0.004)		(0.038)	
AGE 41-45		0.016		0.047
		(0.022)		(0.030)
AGE 46-50		-0.020		0.038***
		(0.028)		(0.009)
AGE 51-55		-0.056***		0.023
		(0.016)		(0.025)
AGE 56-60		-0.092***		0.007
		(0.013)		(0.028)
Other controls included	Yes	Yes	Yes	Yes
N	1,361	3,323	1,547	4,446

Note.—Marginal effects are calculated for a 26-30 years old (columns (1) and (3)) or a 36-40 years old (columns (2) and (4)) university-educated non-Russian speaker from the ethnic and linguistic majority, residing in Baku (Azerbaijan) in 2010; standard errors are clustered at the country level; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level; other controls set included is identical to Table 4.

Table 10: Probit model by country sub-samples - marginal effects

	Males			Females		
	ARMENIA	AZERBAIJAN	GEORGIA	ARMENIA	AZERBAIJAN	GEORGIA
	(1)	(2)	(3)	(4)	(5)	(6)
RUSSIAN SPEAKER	0.068*	0.047***	0.140***	0.074**	0.131***	0.003
	(0.036)	(0.016)	(0.038)	(0.035)	(0.028)	(0.030)
EDUC≤10	-0.218***	-0.197***	-0.189***	-0.361***	-0.441***	-0.245***
	(0.036)	(0.028)	(0.039)	(0.034)	(0.039)	(0.034)
EDUC11-14	-0.180***	-0.140***	-0.105***	-0.258***	-0.273***	-0.170***
	(0.036)	(0.027)	(0.034)	(0.030)	(0.034)	(0.028)
AGE 26-30	-0.037	0.027	-0.049	-0.189***	-0.151***	-0.117***
	(0.042)	(0.023)	(0.054)	(0.046)	(0.045)	(0.044)
AGE 31-35	0.014	0.043*	-0.020	-0.043	-0.134***	-0.035
	(0.040)	(0.023)	(0.054)	(0.044)	(0.048)	(0.044)
AGE 41-45	-0.000	0.042*	-0.030	0.013	0.024	0.107**
	(0.040)	(0.022)	(0.052)	(0.043)	(0.041)	(0.046)
AGE 46-50	-0.060	0.018	-0.062	0.025	0.027	0.052
	(0.042)	(0.023)	(0.049)	(0.040)	(0.042)	(0.043)
AGE 51-55	-0.102**	-0.029	-0.062	0.013	-0.030	0.065
	(0.044)	(0.028)	(0.050)	(0.041)	(0.046)	(0.043)
AGE 56-60	-0.077	-0.092**	-0.129**	-0.026	-0.036	0.048
	(0.050)	(0.037)	(0.052)	(0.044)	(0.051)	(0.045)
Other controls included	Yes	Yes	Yes	Yes	Yes	Yes
N	1,451	1,891	1,342	2,114	2,030	1,849

Note.—Marginal effects are calculated for a 36-40 years old university-educated non-Russian speaker from the ethnic and linguistic majority, residing in a capital city in 2010; standard errors are clustered at the country level; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level; other controls set included is identical to Table 4 with the exception of country dummies that are excluded.

Table 11: Probit model for type of employment - marginal effects

	Males				Females			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RUSSIAN SPEAKER	-0.040***	-0.012	0.137***	0.028**	-0.002	0.001	0.006	-0.001
	(0.008)	(0.016)	(0.022)	(0.013)	(0.006)	(0.008)	(0.022)	(0.014)
EDUC≤10	0.162***	0.063**	-0.218***	-0.032***	0.201***	0.138*	-0.446***	0.000
	(0.024)	(0.027)	(0.024)	(0.005)	(0.035)	(0.074)	(0.015)	(0.012)
EDUC11-14	0.125***	0.067***	-0.153***	-0.026***	0.080***	0.060	-0.201***	-0.000
	(0.024)	(0.022)	(0.013)	(0.009)	(0.013)	(0.044)	(0.036)	(0.011)
AGE 26-30	-0.090***	0.072***	0.011	0.077***	0.012	-0.027	-0.011	-0.020
	(0.010)	(0.010)	(0.031)	(0.005)	(0.017)	(0.020)	(0.009)	(0.018)
AGE 31-35	-0.055***	0.025	0.041***	0.020	-0.000	-0.040	0.023	0.012
	(0.021)	(0.027)	(0.011)	(0.017)	(0.008)	(0.034)	(0.016)	(0.022)
AGE 41-45	-0.036	-0.005	0.062***	0.004	-0.004	-0.029**	0.020	-0.011
	(0.029)	(0.028)	(0.022)	(0.009)	(0.005)	(0.012)	(0.021)	(0.013)
AGE 46-50	-0.044	-0.028	0.074	0.008	0.009	-0.033***	-0.008	-0.013**
	(0.030)	(0.034)	(0.073)	(0.020)	(0.007)	(0.006)	(0.029)	(0.006)
AGE 51-55	-0.019	-0.050**	0.087	-0.024	-0.011	-0.047***	0.092***	-0.034***
	(0.031)	(0.021)	(0.061)	(0.027)	(0.007)	(0.007)	(0.023)	(0.009)
AGE 56-60	-0.060***	-0.031	0.131***	-0.044	-0.009	-0.056*	0.073	-0.039***
	(0.018)	(0.026)	(0.023)	(0.033)	(0.008)	(0.031)	(0.057)	(0.014)
Other controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2,858	2,858	2,858	2,858	2,015	2,015	2,015	2,015

Note.—Dependent variable is employment in: own business/small family business (columns (1) and (5)); medium/large private organisation (columns (2) and (6)), state organisation (columns (3) and (7)) and foreign organisation/business or an international/local NGO (columns (4) and (8)); marginal effects are calculated for a 36-40 years old university-educated non-Russian speaker from the ethnic and linguistic majority, residing in Baku (Azerbaijan) in 2010; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level; other controls set included is identical to Table 4.

Table 12: Probit model for sector of employment - marginal effects

	Males				Females			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RUSSIAN SPEAKER	-0.008	-0.035	0.062***	0.118***	0.000	0.009	-0.014	0.011
	(0.006)	(0.043)	(0.020)	(0.028)	(0.000)	(0.018)	(0.025)	(0.016)
EDUC≤10	0.030	0.111***	0.072*	-0.248***	0.009	0.075***	0.108**	-0.507***
	(0.021)	(0.036)	(0.043)	(0.038)	(0.011)	(0.018)	(0.046)	(0.043)
EDUC11-14	0.020	0.094***	0.051***	-0.182***	0.001	0.033***	0.084***	-0.233***
	(0.014)	(0.024)	(0.005)	(0.033)	(0.002)	(0.011)	(0.019)	(0.048)
AGE 26-30	-0.009	0.057**	0.015***	-0.014	-0.000	-0.006	0.078*	-0.038
	(0.012)	(0.027)	(0.005)	(0.021)	(0.000)	(0.025)	(0.042)	(0.027)
AGE 31-35	0.002	0.021	-0.017*	0.017	-0.000	-0.015	0.017	-0.015
	(0.005)	(0.016)	(0.010)	(0.021)	(0.000)	(0.030)	(0.031)	(0.045)
AGE 41-45	0.002	0.012	-0.017	0.031	-0.000	0.004	0.017***	-0.020*
	(0.005)	(0.058)	(0.024)	(0.042)	(0.000)	(0.023)	(0.006)	(0.011)
AGE 46-50	-0.001	0.051	-0.035***	0.015	0.000	0.001	0.027***	-0.058
	(0.005)	(0.056)	(0.002)	(0.069)	(0.000)	(0.011)	(0.005)	(0.038)
AGE 51-55	-0.001	0.036**	-0.045***	0.045	-0.000	-0.006	0.016	0.025
	(0.008)	(0.018)	(0.011)	(0.050)	(0.000)	(0.011)	(0.018)	(0.019)
AGE 56-60	0.002	0.054	-0.050	-0.014	0.000	-0.007	-0.001	0.016
	(0.007)	(0.085)	(0.042)	(0.037)	(0.000)	(0.025)	(0.005)	(0.048)
Other controls included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	2,835	2,835	2,835	2,835	2,008	2,008	2,008	2,008

Note.—Dependent variable is employment in: agriculture, forestry and fishing (columns (1) and (5)); mining, manufacturing and construction (columns (2) and (6)), trade, services and communication (columns (3) and (7)) and government, education and healthcare (columns (4) and (8)); marginal effects are calculated for a 36-40 years old university-educated non-Russian speaker from the ethnic and linguistic majority, residing in Baku (Azerbaijan) in 2010; * denotes significance at 10 percent level; ** at 5 percent level; *** at 1 percent level; other controls set included is identical to Table 4.