LANGUAGE EFFECTS
ON LABOR MARKET OUTCOMES
IN A BILINGUAL ECONOMY:
THE CASE OF UKRAINE

by

Nikolai Kastioukevitch

A thesis submitted in partial fulfilment of
the requirements for the degree of

Master of Arts in Economics

National University of "Kyiv-Mohyla Academy"
Economics Education and Research Consortium
Master's Program in Economics

2003

Approved by ____
Ms. Svitlana Budagovska (Head of the State Examination Committee)

____

____

____

Program Authorized
to Offer Degree Master's Program in Economics, NaUKMA

Date ____

National University of

"Kyiv-Mohyla Academy"

Abstract

LANGUAGE EFFECTS ON
LABOR MARKET OUTCOMES IN
A BILINGUAL ECONOMY:
THE CASE OF UKRAINE

by Nikolai Kastioukevitch

Head of the State Examination Committee: Ms. Svitlana Budagovska,
Economist, World Bank of Ukraine

This research is concerned with identifying language effects on labor market outcomes in the bilingual economy of Ukraine. A unique data set is examined by calculating descriptive statistics and estimating conventional earnings equations, to see whether at present returns to different combinations of Ukrainian and Russian languages skills are symmetric. Both descriptive statistics and initial models calculated by OLS suggest negative rewards for better knowledge of Ukrainian. More parsimonious specifications provide a better fit to the data, and the inclusion of the mother tongue and ethnic identity variables remove negative income bias of the Ukrainian language knowledge per se. The returns to the Ukrainian mother tongue are negative and strongly significant. The models calculated for local labor markets provide similar results. The rewards for bilingual skills are highly asymmetric and strongly dependent on the initial linguistic endowment. The Heckman procedure is used to show the absence of selection bias in the OLS models. The mother tongue identity is interpreted to capture the long-run dependence of labor market outcomes on the initial linguistic endowment, constituting "historically-driven" labor market segmentation. We thus conclude that returns to the language skills used to be significantly asymmetric. However, at present the marginal agent is indifferent between better knowledge in either of the languages. This gives grounds to a claim of successful "ukrainization" policy (reinforced by the increased national awareness of the Ukrainians), which has reached its intermediate goal of equating the remunerations for the Russian and Ukrainian languages knowledge.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis supervisor, Professor Tom Coupe, who has helped me on all stages of work, assisting me with finding valuable papers and providing extremely useful and inspiring comments and advice. I am indebted to Eugene Bolshov, undoubtedly a would-be prominent sociologist, for his aid in getting the data for my research. I would also want to thank Professor Robert Swidinsky for his ingenious suggestions and encouraging remarks.

My warmest gratitude goes to my family and my girl-friend, Alexandra, for the support and love that has been given to me during all these long and hard two years that I have been far away from them.

C h a p t e r 1

INTRODUCTION

Ukraine is a bilingual country. But the Ukrainian language was being suppressed for many years in the country, especially during the Soviet era, so if we consider the initial language situation of the independent Ukraine, the Ukrainian language was not really needed. In short, Russian dominated much of the political and economic life of the country. However, following the Declaration of the Independence in 1991, the Ukrainian language was made the only official language. As the government is trying to implement the policy of "ukrainization", requiring all the official transactions in the economy to be conducted in Ukrainian, and the national awareness of the Ukrainian society is increasing, the role of this language in the economy is becoming more and more important.

In our research we use a representative survey of the Ukrainian population, conducted by the Institute of Sociology of the National Academy of Sciences of Ukraine in 2002 to see if at present the returns to different combinations of Ukrainian and Russian languages skills are symmetric. Both descriptive statistics and initial models calculated by OLS suggest negative rewards for better knowledge of Ukrainian. More parsimonious specifications provide a better fit to the data, and the inclusion of the mother tongue and ethnic identity variables remove negative income bias of the Ukrainian language knowledge per se. The returns to the Ukrainian mother tongue are negative and strongly significant. The models calculated for local labor markets provide similar results. The rewards for bilingual skills are highly asymmetric and strongly dependent on the initial linguistic endowment. The Heckman procedure is used to show the absence of selection bias in the OLS models. The mother tongue identity is interpreted to capture the long-run dependence of labor market outcomes on the initial linguistic endowment, constituting "historically-driven" labor market segmentation.

We thus claim that the "ukrainization" policy, reinforced by the increased national awareness of the Ukrainians, has reached its intermediate goal of equating the remunerations for the Russian and Ukrainian languages. At present, the marginal agent is indifferent between better knowledge in either of the languages. However, the situation is far from any advantage of the Ukrainian language over Russian. Our analysis provides a powerful counterargument against any claims of the newly introduced bias against Russian speakers.

The following section presents the state of the art in the field of the economics of language and research of bilingual economies in particular. Section 3 provides a theoretical framework for our analysis. Section 4 is concerned with data description and variables construction. Section 5 displays estimation findings for different models and sub-samples. Section 6 concludes our analysis as well as outlines policy implications.

Chapter 2

LITERATURE REVIEW

The issue of language is definitely not something explored in length and depth by economists. As it is, they have approached the phenomenon of language in two completely different fashions. The first one is concerned with analyzing the structure of languages using economic analytical tools, and game theory in particular, thus exploring "the magic of the links between the formal language of the mathematical models and natural language" (Rubinstein, 2001). The second one is more strongly related to conventional economic research, treating languages as an integral part of the real world economies that needs to be reflected in the devised economic models. To this end, Grin and Vaillancourt (1999) define the economics of language as "the study of the relationships between linguistic and economic variables". In our work we are going to be concerned with exactly this thing, investigating possible links between the knowledge of Russian and Ukrainian languages and the people's income in Ukraine.

For obvious reasons, the research in the area has mostly been limited to the countries where there is a significant portion of the population not speaking the official language (Hispanics and other immigrant groups in the USA, immigrants in Germany and Israel), speaking one of the official languages (Switzerland; Canada, and Quebec in particular). Actually, "both theoretical and empirical work primarily focus on bilingualism, rather than in multilingualism in general"(Grin and Vaillancourt, 1999).

Grin and Vaillancourt (1999) delineate four definitions of language, as used in the economic literature. The first one regards language as "an ethnic attribute or marker similar to race", as inspired by models of racial discrimination. The second approach relates the properties of language to that of trade ("using the analogy of international trade between countries to stress the role of language as a means of exchange between groups or individuals"), and the third to human capital. As Grin and Vaillancourt point out, the first approach "overlooks the role of language as a communication tool and the following two approaches address this issue, but neglect the ethnic marker dimension of language". However, they opine that Vaillancourt (1980) devises a model that accounts for both of the classical dimensions of language, "distinguishing between the mother tongue, seen as an ethnic attribute and a form of human capital, and the other languages known by an individual, solely defined as elements of human capital". In our work we are going to follow this fourth approach, viz. analyze the influence of the initial "linguistic endowment" on the income level and measure the returns to learning Russian or Ukrainian languages ("language and earnings" area of research in economics of language, as defined by Grin and Vaillancourt (1999)).

Grin and Vaillancourt (1999) summarize the basic principles of the empirical work on the language factor of earnings:

1. the ordinary least squares multivariate analysis technique;
2. public use samples from Censuses or survey data;
3. more or less narrow samples with sex (men) and age (25-64) often used as constraints;
4. the logarithm of earnings, or only wages and salaries as the dependent variable;
5. education, experience and its square and weeks worked (if the sample is not restricted to full year workers) as control factors (independent variables) in accordance with the usual specification

of Mincerian earnings equations. Other variables such as marital status, region, type of employment, etc., are also commonly used, depending on the availability of data.

As Christofides and Swidinsky (1997) indicate, a number of studies on Canada have attempted estimating the economic value of the different official language skills using various time periods, regions, data sets, and measures of language fluency. As the authors opine: "In general they {the studies} fail to find strong language effects on earning outside Quebec. In Quebec, fluency in only the French language is associated with lower labor market earning, but this negative earning effect has become significantly weaker since the 1970's. The returns to bilingualism are generally positive".

In an early work Geoffrey Carliner (1981), using the data from the 1971 Canadian Census, concluded that in Montreal and Quebec there were substantial economic rewards to learning English for native French speakers; however, even such bilinguals earned less than monolingual English speakers (and there were no rewards to the native English speakers deciding to learn French). Outside Quebec, monolingual English workers also earned more than both monolingual and bilingual French speakers (insignificantly), and less than bilingual English workers (again, insignificantly).

Pendakur and Pendakur (1997) concentrate on the returns to speaking non-official languages employing the 1991 Canadian Census data. Using conditional odds from logit regressions they also analyze the employment prospects for different language groups. Remarkably, they discover that the employment prospects are not always symmetrical to the wage premiums, as defined by language skills. Thus, e.g. "being a male unilingual francophone in Montreal actually helps job prospects, but does not necessarily increase wages". All in all, as already stated above, the authors' results differ significantly across the country, "suggesting that the market for official language knowledge is local, and not national".

Christofides and Swidinsky (1997) estimate the effects of bilingual language skills on annual earnings, using the data from the 1971, 1981 and 1991 Censuses. Their OLS results show a substantial change in the impact of official languages skills on annual earnings between 1971 and 1991. But, although "the earnings of bilingual Anglophones and Francophones in Quebec, relative to earnings of unilingual Anglophones in Quebec, have increased sharply between 1971 and 1991, the relative increase is not solely the result of higher absolute earnings paid to bilingual male workers. A major share of the relative increase in the earnings of bilingual workers must be attributed to the severe decline in the labor market status of unilingual English-speaking males in Quebec". This interesting result may be interpreted as monolingual English speakers bearing the costs of the francophonization policy in Quebec. The authors furthermore note that "an analysis of more detailed regression results failed to uncover any definitive age, educational, or occupational patterns to the returns to bilingualism outside Quebec", except (not surprisingly) for the bilinguals working in the public sector, who have significantly higher earnings (as compared to English monolinguals).

The main caveat of analyzing the returns to the language capital is its binary composition. Thus, while the mother tongue is clearly an exogenous asset, the language acquired as a result of a conscious choice on the part of an individual, should be regarded differently. As Pendakur and Pendakur (1997) state: "This categorization is important because, while it can be argued that people who learn an additional language later in life may have other abilities which can affect performance in the labor force, it is harder to envision this being the case for people who start out with two languages". They further conclude that the returns to the mother tongue are the "true" returns to the language capital, as "unpolluted by correlations with unobservable variables". Thus, they separate the bilinguals into "natural" and "learned", and receive quite significant (numerically, statistical significance not demonstrated) difference between wage differentials accruing for "natural" and "learned" bilinguals. The authors opine that: "The true return to official and non-official language knowledge, unpolluted by unobservable variable bias, is somewhere between the two returns".

As Christofides and Swidinsky (1997) admit it is very probable that language variables in their OLS earnings regressions are endogenous, which yields biased and inconsistent results. They address the problem of endogeneity using Heckman (1979) procedure. They "specify separate log-earnings equations for bilingual and unilingual individuals which contain the appropriate inverse Mills ratios; the latter are constructed from a background probit equation which explains the decision to become bilingual. The selection-adjusted log-earnings equations may control for the individual's mother tongue, an attribute which is clearly exogenous".

Henley and Jones (2001) use Heckman procedure and Oaxaca-Ransom (1994) earnings decomposition to analyze the wage differentials between Anglophone and bilingual (those who also know Welsh) workers in Wales. They find that, contrary to the raw data statistical evidence and selectivity-unadjusted OLS results, the bilinguals are "somewhat underpaid, for given characteristics, as compared to Anglophones".

Chapter 3

THEORETICAL FRAMEWORK

We are going to single out at least three different labor markets in Ukraine: Western, Central and Eastern, following the general geographical distribution of Russian and Ukrainian speakers, as well as some other socio-demographic and economic factors that localize these regions. Such approach is quite common in the research work on the issue, especially for the Canadian labor market. Thus, Pendakur and Pendakur (1997) find for 1991 Canadian Census data a different impact of language on earnings for the same type of labor force in different Canadian regions. As the authors confess: "We initially ran regressions for Canada as a whole, however we restricted analysis to three CMAs {Census Metropolitan Areas} because it quickly became apparent that different CMAs have very different pay-offs to language knowledge and are effectively different language markets".

We can further assume the existence of two (in case of bilingualism) separate labor markets for the speakers of the different languages inside each geographically isolated labor market. Christofides and Swidinsky (1997) analyze a basic supply-demand framework (they do it implicitly for Quebec) under this assumption. They consider, following Bloom and Grenier (1992), a local economy with two language communities, French and English, and two corresponding labor markets. Each market is characterized by a supply-demand framework, with the wage rate and the quantity of labor as the defining price-quantity variables. Workers in both markets are identical, except for language skills. Some workers speak only French, some only English, but some are fully bilingual and can function in both labor markets. Fluency in only one language is determined exogenously by mother tongue (or acquired through the educational system), but bilingualism is an acquired language attribute that is endogenous to labor market conditions. Given that the acquisition of a second language is costly, workers will only learn it if the wage differentials in the two language markets are sufficient to provide a competitive rate of return. The supply of workers in each labor markets will depend on the population of each mother tongue group and on the number of individuals with the other mother tongue who are bilingual. Bilinguals can supply labor in either English-language or French-language markets. Shifts in supply can result from natural increases in population, from an increase in the number of individuals who are bilingual, or from migration into or out of the local labor market by individuals with similar language skills. Following Vaillancourt (1992), the authors identify the determinants of the labor demand schedule as the language of the product markets served by employers; the language of the existing instruments of work (technology, instructions, etc.); and the language of the owners of the firms. In the first case, the demand for the speakers of a given language will positively depend on the number of consumers belonging to the same language community (more so for personal products, when consumption involves communication (McManus, 1985)). The second determinant will be influential if the technology of production requires knowledge of the more technologically advanced language (McManus, 1985), or when firms are operating outside their local product market. Finally, higher proportion of the owners and managers speaking a given language will support higher demand for the same-language employees. Christofides and Swidinsky thus conclude that the configuration of labor supply and demand in the two language labor markets will determine the language-wage differential. Varying the demand or supply of workers in the two markets will alter the relative wages of the two language groups.

This framework can be instrumental in analyzing the Ukrainian labor market. However, we should bear in mind that now there are hardly any monolingual Ukrainian speakers in Ukraine, and we should apply some caution while modeling such skills. Ukrainian being the only official language in Ukraine (since 1991) is another peculiarity of this economy as compared to Canada. However, using the above-presented framework, the latter phenomenon can be viewed as an expansion of the demand for Ukrainian speakers, with a simultaneous contraction of the demand for Russian speakers. Quite possibly, this policy has been also reinforced by the increased national awareness of the Ukrainians, embodied in the demand expansion for the products and services in the Ukrainian language, as well as changed employers' linguistic preferences.

On the other hand, the government policies of creating the Ukrainian-dominating educational system, with the Russian language as an optional course of study, as well as some other legislation, which strives to limit the usage of Russian in all spheres of the economy and public life, and especially in mass-media, can be regarded as a sustained effort to increase the supply of Ukrainian speakers. It is clear that the demand effects must show themselves in a much shorter run. Thus, effectively, any current study will mostly capture the results of the demand shift. We might conjecture that if we had panel data or consecutive cross-sections we would find a positive trend for the wage premium of Ukrainian speakers, as compared to Russian monolinguals, through the years of the Ukrainian independence. However, it is still possible to draw meaningful conclusions about the direction of the wage premium change even using one cross-section, as our analysis will demonstrate.

Chapter 4

DATA AND VARIABLES

The data set at our disposal is a representative Ukrainian survey, conducted by the Institute of Sociology (IS) of the National Academy of Sciences of Ukraine in 2002[1]. The polling places included respondents in each Ukrainian oblast[2] (proportional to the number of inhabitants), Crimea and the city of Kyiv. The selection quota reflected the specific regional distribution of the basic socio-demographic characteristics (sex, age, education). Each oblast was represented by the central city, towns and country dwellers (also according to certain proportions). To ensure a representative sample selection within the quota and to achieve randomness, the search for the respondents was carried out by the interviewers following a previously assigned route and maintaining the quota characteristics (the interviewers gave questionnaires only to those respondents who corresponded to the necessary quota features). The adult population (over 17 years old) of Ukraine constituted the general aggregate. The total number of respondents was 1799. The polls were conducted by the method of handout questionnaires (the respondents filled the questionnaires by themselves).

The set of variables in our data set is very rich, with more than 300 questions asked. Although the survey is primarily of a sociological interest, lacking some of the important variables that are conventional for labor force surveys, the data are of immense value for our purposes. They are unique in the number of questions that probe both directly and indirectly into the "linguistic"

1 In what follows we describe the principles of the data collection, as presented in Panina and Golovakha (2001), which is a summary of their previous surveys.

2 This is the term for the largest regional units of administrative division in Ukraine (in total of 24, plus the city of Kyiv and the Autonomous Republic of Crimea).

characteristics of the Ukrainian population. The levels of command of the Ukrainian and Russian languages are solicited separately, with three degrees of self-evaluation for each language ("no knowledge", "partial knowledge" and "full knowledge"). Importantly, the survey also includes the "native tongue" and ethnic identity. In addition, it contains an array of questions that are helpful for investigating the endogeneity between the "linguistic" capital and earnings, viz. the attitude to the official language problem in Ukraine ("Do you consider it necessary to make Russian the second official language?"), place of birth, religion, etc. The survey contains quantitative data on the income ("income received last month"), which greatly facilitates our research, enabling us to use the OLS estimation technique for our main earnings equations. Finally, the data set includes such necessary socio-demographic information as educational level, occupation, age, gender, marital status, place of living (categories of the settlement size and specific region), and the like.

The main problems with the data are a significant number of missing observations (even for such unequivocal questions as gender identification) and clearly understated income. Taking into the account a relatively small size of the data set (as compared to conventional labor surveys), any clearing measures had to be taken with care. Nevertheless, for the means of our research we had to contract our data set quite considerably.

We first selected a pool of the variables of primary importance (income, knowledge and use of Russian and Ukrainian, and the socio-demographic variables mentioned above) and got rid of missing observations for all of them. We then contracted the sample to only include those who held some paid job (including part-time), and were hired employees (we excluded individual businessmen, farmers, full-time students, non-working pensioners, unemployed or out of labor force, and those with no certain place of employment). The idea was to exclude people whose income could not be possibly related to their language skills. We then excluded from the "native tongue" variable and the variables tracking the spheres of language usage ("What language do you speak in your family, with your friends", and the like) observations on individuals who indicated languages other than Russian, Ukrainian or their combination[3]. This procedure enabled us to assess pure Russian-Ukrainian differentials. We finally got rid of zero observations on income, as they are rather unreasonable for the individuals left in our sample, and could not be employed in running regressions on the log of income anyway.

We thus ended up with 656 observations[4]. They include questionably small values of income, but further contraction would be unreasonable[5], as we do not have the information whether the individuals worked full time for the reported period. It may well be that low reported earnings are not grossly understated by the respondents, but simply imply not many hours worked or possibly wage arrears. Fortunately, this does not seem to be a big problem, as nearly 90 percent of the individuals received earnings at least in excess of the minimum wage (118 UH for the time the survey was conducted).

Table 1 shows the distribution of the language skills in the full and contracted samples. One can see that the shares of people possessing different skills of the Ukrainian and Russian languages remain quite stable after the contraction of the initial sample. People having free command of a language make up an overwhelming majority for both Ukrainian and Russian.

---

[3] The number of such individuals was rather insignificant.

[4] We actually lost in the process one "education" category, people with no primary education (mostly non-working pensioners), and one specific settlement type category, "urban-type settlement" (only one observation).

[5] Apart from further contraction of the sample it would truncate the sample distribution, causing estimation problems.

Table 1: Distribution of the Language Skills (Russian and Ukrainian Separately) and Average Income[6] for the Different Levels of Language Knowledge

Language Knowledge	Knowledge of Ukrainian			Knowledge of Russian		
	Full Knowledge	Partial Knowledge	No knowledge	Full Knowledge	Partial knowledge	No knowledge
Full Sample[7]	1323 (73.6 %)	401 (22.3%)	73 (4.1%)	1476 (82.4%)	277 (15.5 %)	39 (2.2%)
Contracted Sample (656 observations)	484 (73.8 %)	154 (23.5 %)	18 (2.7%)	579 (88.3%)	71 (10.8%)	6 (0.9 %)
Average Income (contracted sample)	260.94 (200.64)	286.34 (194.52)	332.78 (254.69)	278.09 (208.66)	204.23 (111.38)	145 (53.2)

The percentages of the respondents who do not have at least basic skills in either of the languages are unsurprisingly very low. These percentages are especially negligible for the contracted sample, which can be explained by the principles of our constructing this data set and also supposedly increased validity of the responses due to inadvertent elimination of measurement errors (e.g. a 19 year old native inhabitant of Ukraine without primary education).

In addition, Table 1 demonstrates average income for the different levels of language knowledge (for the contracted sample). One can readily observe a clear pattern of increasing rewards for better skills in Russian, and negative premiums for higher levels of the Ukrainian language command.

---

[6] Mean and s.e. (in parenthesis).

[7] The bases for computing the full sample percentages were cleared from the missing observations, separately for Ukrainian and Russian, with 1797 and 1792 observations remaining correspondingly. The totals may not add up to 100 due to rounding.

Table 2: Distribution of the Language Skills (Combinations of Russian and Ukrainian)

Combinations of the Language Skills	U3_R3 Full knowledge of Ukrainian/ Full knowledge of Russian	U3_R2 Full knowledge of Ukrainian/ Partial knowledge of Russian	U3_R1 Full knowledge of Ukrainian/ No knowledge of Russian	U2_R3 Partial knowledge of Ukrainian/ Full knowledge of Russian	U2_R2 Partial knowledge of Ukrainian/ Partial knowledge of Russian	U2_R1 Partial knowledge of Ukrainian/ No knowledge of Russian	U1_R3 No knowledge of Ukrainian/ Full knowledge of Russian	U1_R2 No knowledge of Ukrainian/ Partial knowledge of Russian	U1_R1 No knowledge of Ukrainian/ No knowledge of Russian
Full Sample[8] (1792 observations)	1059 (59.1%)	226 (12.6 %)	33 (1.8%)	351 (19.6%)	46 (2.6%)	4 (0.2%)	66 (3.7%)	5 (0.3%)	2 (0.1%)
Contracted Sample (656 observations)	422 (64.3%)	56 (8.5%)	6 (0.9%)	140 (21.3%)	14 (2.1%)	-	17 (2.6%)	1 (0.2%)	-

In order to investigate income differentials as resultant from different language skills in more detail, we compose all possible matches of the levels of knowledge in both languages. Table 2 shows the distributions of the language skills for the full and contracted samples. We can again notice that the shares remain quite stable after the contraction. The hierarchy of the language skills is exactly the same for both samples, with two minor and doubtful combinations disappearing in the contracted sample (and only one observation remaining for a third). The "full" bilinguals (those with free command of both languages) are by far the most numerous category. However, such people do not constitute an overwhelming majority in either of the samples. The other two most numerous categories are "partial" bilinguals (those having partial knowledge in one of the language and a perfect command of the other), with "Russian-speaking" (full knowledge of Russian) partial bilinguals clearly prevailing. The remaining combinations of language knowledge are quite small in the number of the respondents.

[8] The base for computing the full sample percentages was cleared from the missing observations, for Ukrainian and Russian simultaneously, with 1792 observations remaining. The totals may not add up to 100 due to rounding.

Tables 3: Distribution of Income[9] (Monthly Earnings in Ukrainian Hryvnas) by the Language Skills (Combinations of Russian and Ukrainian)

Combinations of the Language Skills	U3_R3	U3_R2	U3_R1	U2_R3	U2_R2	U2_R1	U1_R3	U1_R2	U1_R1	Total
	Full knowledge of Ukrainian/ Full knowledge of Russian	Full knowledge of Ukrainian/ Partial knowledge of Russian	Full knowledge of Ukrainian/ No knowledge of Russian	Partial knowledge of Ukrainian/ Full knowledge of Russian	Partial knowledge of Ukrainian/ Partial knowledge of Russian	Partial knowledge of Ukrainian/ No knowledge of Russian	No knowledge of Ukrainian/ Full knowledge of Russian	No knowledge of Ukrainian/ Partial knowledge of Russian	No knowledge of Ukrainian/ No knowledge of Russian	
Number of Observations in the Full Sample Categories[10]	989 (58.2%)	216 (12.7%)	33 (1.9%)	342 (20.1%)	45 (2.5%)	3 (0.2%)	64 (3.6%)	4 (0.2%)	2 (0.1%)	1698 (100%)
Full Sample (1698 observations)	173.24 (153.59)	175.44 (158)	369.27 (358.58)	227.65 (427.37)	163.4 (120.4)	135.67 (5.77)	202.53 (186.49)	59.75 (53.73)	67.5 (24.75)	188.67 (242.66)
Contracted Sample (656 observations)	271.04 (209.88)	197.23 (98.48)	145 (53.2)	291.01 (198.05)	239.64 (152.87)	-	346.47 (255.6)	100 (-)	-	268.88 (201.09)

We finally present the income distributions by the language skills for the full and contracted samples in Table 3 (the number of observations in the contracted sample categories are not presented as they can be read from Table 2). We can see that the average income is notably higher in the contracted sample, with an exception of a minor category of Ukrainian-speaking monolinguals. This fact probably speaks favorably for our sample constructions techniques, as the severely understated income is somewhat rectified. However, the most interesting are the income distributions across the samples. From the first glance, no clear-cut picture can be derived from the average values of the income across the different language combinations in the full sample. The above-mentioned monolingual Ukrainian-speakers are significantly ahead, but this category is a clear

[9] Mean and s.e. (in parenthesis).

[10] The base for computing the full sample percentages was cleared from the missing observations, for income, Ukrainian and Russian simultaneously, with 1698 observations remaining.

outlier by its size (and plausibility). This is especially evident if we compare their average income to that of their counterparts in the contracted sample. This category is followed by partial and "pure" (no knowledge of Ukrainian) Russian-speaking bilinguals. Then come Ukrainian-speaking partial bilinguals, with full bilinguals (most numerous) being only the fifth. This is especially striking considering that the remaining four categories are quite negligible. The full bilinguals in fact appear to have the lowest income.

The story is quite different for the contracted sample. Russian speakers are at a clear advantage, with Russian-speaking monolinguals occupying the first position, followed by Russian-speaking partial bilinguals and full bilinguals. Then come people with partial knowledge of Russian (people with a worse command of Ukrainian are again in the lead). Ukrainian-speaking monolinguals are closing the line (there is only one person with no command of Ukrainian and partial knowledge of Russian – this "category" takes the literal last place). We can thus see a clear picture of an advantage of Russian speakers vs. a disadvantage of Ukrainian speakers. Roughly speaking, the knowledge of Russian seems to place a person in a higher area of income distribution, whereas better Ukrainian-speaking abilities signal of lower earnings.

However, our preliminary conclusions do not take into account an array of traceable factors that may in fact be responsible for some of the described effects.

Chapter 5

SPECIFICATIONS AND FINDINGS

General OLS Specification

Before considering the local labor markets, we first study the language factor of earnings within the model for the entire Ukrainian labor market. Our main equation, estimated from our contracted sample by OLS looks as:

$$\ln(E) = \alpha + \beta(L) + \gamma(K)+u,$$

where ln(E) is a N*1 vector of logarithms of earnings, L is an N*5 matrix of observations on the language variables, and K is N*X matrix of control variables (N - number of observations and X – number of control variables).

The language variables are simply the dummies, constructed to represent six possible combinations of language knowledge. These are: U1_R3 (no knowledge of Ukrainian and full knowledge of Russian), U2_R3 (partial knowledge of Ukrainian and full knowledge of Russian), U3_R3 (full knowledge of Ukrainian and full knowledge of Russian), U2_R2 (partial knowledge of Ukrainian and partial knowledge of Russian), U3_R2 (full knowledge of Ukrainian and part knowledge of Russian), and U3_R1 (full knowledge of Ukrainian and no knowledge of Russian). We considered it possible to include the sole observation on U1_R2 (no knowledge of Ukrainian and partial knowledge of Russian) into U1_R3 (Russian-speaking monolinguals), taking into account other characteristics of the individual (Russian nationality, born in

Crimea[11], 19 years old, secondary education). The full bilinguals (U3_R3) were chosen as a base dummy.

The controls include 26 regional dummies (the city of Kyiv omitted as a base), 8 dummies for the settlement type (villages omitted as a base), 11 dummies for occupation (low-skilled workers used as a base), 7 educational categories (primary education used as a base), 6 dummies for the marital status (those never married taken as a base), gender (females omitted) and age (the only continuous independent variable). The choice of the base dummies is motivated by a possibility of a more convenient results interpretation. The full description of the control dummies can be found in Table 1 of Appendix 2.

The estimation output is presented in Table 1 of Appendix 3. As White test indicated the presence of heteroscedasticity even at 1 percent significance level, we used White heteroscedasticity consistent variance-covariance matrix[12].

The signs and magnitude of the language dummies coefficients have not diverted much from our expectations based on the descriptive statistics presented in the previous section. There is only one change from our preliminary results: Russian-speaking partial bilinguals (U2_R3) swapped their second place with the fourth place of "pure" partial bilinguals (U2_R2). However, all but one coefficients are statistically insignificant (the Russian-speaking monolinguals' advantage over full bilinguals is significant at 10% significance level), and the exact hierarchy of the language skills income differentials is in fact ambiguous. So far we can only claim that Russian-speaking monolinguals are at a significant advantage, as compared to any other language groups (with an exception of the U2_R2 category), including full bilinguals. This can be seen more clearly from

[11] This is a predominantly Russian-speaking region, which became a part of Ukraine not long ago.

[12] All estimation results should be understood as void of heteroskedasticity problem, unless specifically indicated.

Table 2 of Appendix 3, which presents the estimation output of the same regression, with U1_R3 used as a base. As compared to the Russian monolinguals, full bilinguals on average earn by 22.4% less, and the Ukrainian monolinguals have a 34.8 percent income disadvantage[13].

[13] Throughout our work, we use the Halvorsen and Palmquist (1980) adjustment to calculate the percentage impact of dummy variables. The percentages are only reported for the most important (and producing sufficiently robust effects) variables.

Aggregated Model

It is clear that our estimation suffers from the micronumerosity problem, i.e. a rather low ratio of observations to independent variables. This problem is similar in its consequences to multicollinearity. The t-statistics are unbiased and efficient but they may be too low due to high variances of the coefficients. We cannot increase the number of observations, but we can try to save degrees of freedom by aggregating some of our variables, especially those with grossly insignificant t-ratios and dubious coefficients signs.

We united the categories of marital status into two: married or co-habiting, and single. Only two levels of education were distinguished: those with a higher education and others. The latter re-specification was strongly encouraged by the indication of insignificant and negative returns to education, as compared to its primary level, for all individuals except those with a Specialist's or a Master's degree. We left only four types of settlement, namely major cities (over 1 mln dwellers), cities (100 thousand – 1 mln inhabitants), towns (20 thousand – 100,000 residents), and small towns and villages (less than 20,000 people). Furthermore, we abandoned the oblast division for a more general (but quite conventional and already introduced in our raw data set) separation of Ukraine into West, Center, East and South[14]. Finally, we substituted Age by a constructed variable Experience, using the formula: Age-years of schooling-7 (age of matriculation). The years of schooling were not available directly from the data set, and were defined using the education variables. The exact description of the variables transformations can be found in Table 2 of Appendix 2.

[14] This last aggregation actually allowed to account for the heteroskedasticity, as will be indicated below.

The output of the "aggregated" model, estimated by OLS, can be found in Table 3 of Appendix 3. This specification seems to be much more fortunate as it is void of the heteroscedasticity problem (White test rejects the null hypothesis of heteroscedasticity at any conventional significance level), and does not consume too many degrees of freedom. Most of the coefficients are significant and have the anticipated signs. As could be expected, singles and women tend to have lower earnings, and people are rewarded for living in bigger cities and obtaining higher education. The distribution of earnings across occupations also seems reasonable[15]. The returns to experience are negative and significant, which is not surprising for a transitional economy, in which older age may signal of obsolete skills and low mobility. Regional division does not reveal any pattern of geographically determined income bias.

The returns to the language skills have largely retained their hierarchy[16], which seems to point to a reasonable degree of robustness of our model.

---

[15] Remarkably, CEOs are nearly the worst paid, and politicians reported the lowest earnings in the sample.

[16] Russian-speaking bilinguals overtaking full bilinguals is the only difference between the specifications. However, the income gap between the two groups is statistically insignificant in both cases.

Pure Language Effects Identification

Ideally, we would want to know the pure language effects on earnings, cleared from the influence of such related characteristics as mother tongue and ethnicity. We tried to do this by including the relevant dummies into the aggregated model (UMT is equal to 1 if one's mother tongue is Ukrainian, and to 0 if Russian; Unational is equal to 1 for Ukrainian nationals and 0 for Russians[17]).

The output is presented in Table 4 of Appendix 3 (U1_R3 group was omitted as a base to ease the analysis of income differentials). We can see that the advantage of Russian-speaking pure monolinguals over the rest of the language groups became insignificant, with an exception of Russian-speaking partial bilinguals (and the exclusion of the Unational variable makes this last remaining advantage insignificant). The marginal effects for the two most numerous (after full bilinguals) linguistic groups[18], Russian-speaking partial bilinguals and Ukrainian-speaking partial bilinguals, are nearly identical.

Ukrainian ethnic identity is positively related to income[19], although this reward is statistically insignificant. However, people with Ukrainian mother tongue have a conspicuous disadvantage, which is statistically significant at any conventional level. The income disadvantage of those with the Ukrainian mother tongue as compared to their counterparts is 24.8%.

We can now see that the income gap between Russian-speaking monolinguals and all other language groups (as well as the income differential

[17] 10 observations for the people of other nationalities were excluded to preserve the pure dichotomy of Ukrainian vs. Russian, be it language or ethnic identity.

[18] These groups are also characteristic for identifying the prevalence of either of the languages in a region.

[19] The marginal effect of the Ukrainian ethnic identity is (insignificantly) negative when the mother tongue variable is not included (this result is also suggested by descriptive statistics – see Table 2 of Appendix 1).

between Russian-speaking partial bilinguals and Ukrainian-speaking partial bilinguals) is contingent on the mother tongue identity. Indeed, it seems that it is not per se better knowledge of Ukrainian coupled with worse knowledge of Russian, that drives down one's earnings, but the underlying negative income bias of those whose first language was Ukrainian. This bias is rather strong and is remarkably linguistic, rather than ethnic.

Bilingualism and Earnings

Although our data set allows us to study the language factor of earnings in detail, i.e. analyze different combinations of language skills, there is one such combination that may provide a very insightful dichotomy, once compared versus all the rest. This is the full bilinguals category, U3_R3 in our specification. The studies of bilingualism effects on income are the most widespread in the field of "language and earnings"[20], if not only due to the lack of more specific information on the language skills. However, in the framework of our research, distinguishing the bilingualism factor of earnings may provide some additional colors to the picture of the income distribution due to language knowledge in Ukraine.

We used our aggregated model, controlling for the mother tongue identity, to account for the bilingualism effect. The output is presented in Table 5 of Appendix 3. The marginal effect of the perfect knowledge in both languages is positive and significant[21] at the 10 percent level. The possession of the bilingual skills increases one's earnings by 8.8 percent.

Evidently, the nature of bilingualism is two-fold: one can learn either the Russian or Ukrainian language in addition to the Ukrainian or Russian mother tongue, respectively[22]. We therefore ran two additional regressions for individuals with the Ukrainian and Russian mother tongue separately (see Tables 6 and 7 of

---

[20] See the Literature Review section

[21] The inclusion of the ethnic ethnicity control increases the p-value to nearly 17 percent.

[22] We have to abstract from the possibility of having "natural" bilinguals, i.e. people with two mother tongues, as our data set does not allow for such distinction. However, as we will argue below, it is reasonable to view all bilinguals in Ukraine as "natural", as such language knowledge was imposed exogenously, and not acquired through a conscious effort of an individual (see also the Literature Review section).

Appendix 3)[23]. It turns out that the Ukrainian (mother tongue) bilinguals have almost no advantage over their less educated counterparts, while individuals with the Russian mother tongue are rewarded significantly[24] for learning Ukrainian (their earnings increase by 15.4%).

The returns to the bilingual skills are thus highly asymmetric and strongly dependent on the initial linguistic endowment.

[23] The regression for the Russian mother tongue individuals does not include the occup1 variable, as the appropriate occupation is absent from this sub-sample.

[24] While the inclusion of the Ukrainian ethnic identity has almost no influence on the estimation results for the Ukrainian mother tongue sub-sample, it decreases rather conspicuously the marginal effect significance of the U3_R3 variable (its p-value increases to 13.5%) for the Russian mother tongue sub-sample.

Regional Labor Markets Isolation

It is useful to carry out separate estimations of our model for regional sub-samples, thereby isolating distinct local labor markets, which may have non-identical remuneration distributions across the language skills. We begin by dividing the sample into four parts, namely West, Center, East and South, following a conventional approach to the economic-geographical division of Ukraine. The exact composition of the regional sub-samples can be found in Table 2 of Appendix 2.

We provided the distributions of the language skills and the associated allocation of income for the local labor markets in Table 1 of Appendix 1. As one can readily notice, such geographical division presents a consistent pattern of the income distributions. We may easily compare the sub-samples by contrasting the two representative groups, i.e. Russian-speaking partial bilinguals vs. Ukrainian-speaking partial bilinguals. It is clear that the prevalence of one group over the other will indicate that the primary language of the former is in fact the major language of the district. Thus, using this approach, we can see that Western Ukraine is mainly Ukrainian-speaking, the East and South are largely Russian-speaking, and neither language is dominant in the center of the country.

Unfortunately, it is impossible to contrast income distributions across all of the observant combinations of the language skills for any of the local markets due to insufficient number of observations. However we can still try to compare earnings allocations using only two linguistic variables: know_U and know_R; the former is equal to 1 if a person has a full command of the Ukrainian language, and 0 otherwise, and the latter is equal to 1 for those with a perfect knowledge of Russian, and 0 otherwise.

Table 2 of Appendix 1 presents the average income for the categories of language knowledge (know_U and know_R variables), mother tongue and ethnic identity for local labor markets[25] and the entire Ukrainian labor market. For convenience of comparisons we calculated these statistics both for the variables to be actually included into the regressions, and their base dummies. It can be seen that the usage of know_R and especially know_U variables for the whole sample regressions would be rather problematic, as they do not "pin down" any conspicuous income differentials. For the very same reason they are not very instrumental for the regional analysis either, but as already stated above, any finer specification of language skills would be unreasonable in this case. We may, however, also account for the bilingualism effects, using the U3_R3 variable (see Table 1 of Appendix 1 for the appropriate descriptive statistics).

We first ran our aggregated regression[26] for the Western labor market. As evident form Table 2 of Appendix 1, this sub-sample is very uniform. It seems impossible to account for the effects of the Ukrainian language knowledge, and we cannot include the ethnic and mother tongue identity variables together, as they only differ by one observation. We therefore confined ourselves to the study of the Russian language factor of earnings, controlling for the mother tongue. The output[27] is presented in Table 8 of Appendix 3. The influence of the full knowledge of Russian is negative (which contradicts the descriptive statistics), but negligible from the statistical point of view. The coefficient for the Ukrainian mother tongue surprisingly retains its negative sign for the Western labor market,

---

25 We do not present statistics for Eastern and Southern labor markets separately, as they are very uniform to the extent of our analysis objectives (see Table 1 of Appendix 1) and can be united for all practical purposes.

26 We also do not include any regional dummies in order to save the degrees of freedom. Also, set_aggregate_1 is absent form the regression as there are no such big cities in Western Ukraine.

27 The regression for the Western labor market turned out the only one among the regional ones to suffer from the heteroskedasticity problem, and we only applied White heteroscedasticity consistent variance-covariance matrix for this sub-sample estimation.

and is strongly significant. However, this result should be regarded with considerable caution, as there are only four individuals reporting the Russian language as their mother tongue. Although the marginal effect of the Russian language knowledge is statistically insignificant, it is quite stable within our model, as it is almost uninfluenced by the mother tongue identity.

The returns to the bilingual skills are negative but highly insignificant (see Table 9 of Appendix 3). This result replicates the one we received for the Ukrainian mother tongue sub-sample of the entire Ukrainian labor market, which is natural if we take into the account the nearly uniform composition of the Western labor market in this respect.

The descriptive statistics for the Central labor market closely resemble those of the entire market. We can also see that the Ukrainian and Russian languages are remarkably on par to the extent of the number of speakers who report excellent language skills. It is now possible to estimate the influence of both the Russian and Ukrainian languages knowledge, as well as control for both the ethnicity and mother tongue. However, the specifications estimated for the Central labor market also did not yield significant t-statistics for the income returns to the linguistic skills. The most general specification is presented in Table 10 of Appendix 3. It includes both know_U and know_R variables, as well as the Ukrainian ethnicity and mother tongue dummies. The coefficients resemble the pattern of the full sample regression, viz. positive and insignificant returns to the Russian language knowledge, negative and insignificant returns to the Ukrainian language knowledge, coupled with significant negative returns to the Ukrainian mother tongue (23.3% less of earnings as compared to those with the Russian mother tongue) and insignificant positive returns to the Ukrainian ethnic identity. The less inclusive specifications, calculated for this sub-sample, provided similar results.

The Central labor market is unsurprisingly the leader among the local market in the number of full bilinguals (see Table 1 of Appendix 1). The regression output for the bilingualism effect is presented in Table 11 of Appendix 3. The marginal effect is positive but not really significant.

We finally analyzed the Russian-speaking area of Ukraine, by uniting very similar in this respect Southern and Eastern labor markets. We again present the most general with respect to language characteristics model (see Table 12 of Appendix 3). The results demonstrate no visible difference between the returns to the different languages knowledge. Although the marginal gain of knowing Russian seems to be higher than the one for knowing Ukrainian, both of them are statistically insignificant. Individuals with the Ukrainian mother tongue are persistently at a disadvantage (25.5% less of earnings as compared to those with the Russian mother tongue), while the Ukrainian ethnic identity suggests insignificant positive income reward.

The bilingualism effect for the Southern and Eastern labor markets (see Table 13) closely resembles the one for the Central market. It is positive but insignificant.

The results of the regional regressions strongly reinforce our initial conjectures about the nature of the language factor of earnings, as observed for the whole sample. The returns to the language skills as such (including the bilingual skills)[28] are insignificant for both languages for all local markets. However, the Ukrainian mother tongue is consistently linked to lower income across all of the three regions. On the other hand, the returns to the Ukrainian

[28] The local markets regressions fail to discover any significant effects of bilingualism even when the Ukrainian ethnic identity variable is not included.

ethnic identity are (insignificantly) positive[29]. This pattern presents a challenging puzzle for interpretation. An intuitive explanation will be provided in the concluding section.

[29] This is conditional on the inclusion of the mother tongue variable, same as for the full sample. The descriptive statistics in Table 2 of Appendix 1 also suggest a disadvantage of Ukrainian nationals for all local markets.

Selectivity Analysis

As already stated above (in the Literature Review section) one single important problem of analyzing the returns to the language capital is its binary composition. While the mother tongue is clearly an exogenous asset, we should be cautious with using any other acquired language skills as regressors. As our data set does not allow for a possibility of "natural" bilinguals, one might conclude that we have to view all people with a command of both Russian and Ukrainian from our sample as "learned" bilinguals[30]. If this is true, then our OLS results, obtained in the previous sections, may be biased and inconsistent.

However, we would argue that this problem is negligible for our data set. Intuitively, our constructed sample mostly includes people were educated under the conditions of the Soviet system, which did not really encourage the learning of languages other than Russian, and the latter was imposed exogenously rather than acquired through an individually rational effort. We use the Heckman two-stage procedure to check our hypothesis. We analyze three equations. The first one is the underlying probit selection equation, modeling the bilingual[31] status, and the other two are our conventional semilogarithmic earnings equations, estimated by OLS for the sub-samples of bilinguals and monolinguals separately, and now including the appropriate inverted Mill's ratios from the selection equation[32]. This procedure is designed to account for a possible selection bias that can make the ordinary OLS results biased and inconsistent.

---

[30] See the Literature Review section for more details on this distinction.

[31] The distinction between bilinguals and monolinguals is the same as used in the Bilingualism and Earnings sub-section.

[32] The whole procedure is explained very intuitively in Kennedy (2003). An interested reader can also use Henley and Jones (2001) for a nice exposition of technical details.

As stated in Kennedy (2003), on the minimum standard error criterion the Heckman procedure loses to ordinary OLS when "the errors are not distributed normally, the sample size is small, the amount of censoring is small, the correlation between the errors of the regression and the selection equation is small, and the degree of collinearity between the explanatory variables in the regression and the selection equation is high". It appears that some of these conditions are binding in our case. Indeed, although we may have succeeded in finding the appropriate regressors for the selection equation (as described below), not only our whole constructed sample is rather small, but also the sub-sample of monolinguals is much smaller than that of bilinguals. However, as we have already said, the purpose of this analysis is to see whether the selection bias constitutes a problem for our previous OLS results, and this actually means that for us the whole procedure boils down to a t-test on the coefficients of the inverted Mill's ratios. If these coefficients appear insignificant, then it will mean that there is no correlation between the errors of the (OLS) regressions and the selection equation[33], and thus there is no selection bias (and the Heckman procedure makes no sense[34]).

We have taken great pains to find the regressors for the selection equation such that they would both provide a good fit and be as little as possible correlated with the independent variables from the OLS regressions. We have constructed such variables as no_purist (equal to 1 if a person uses both Russian and Ukrainian either in the family, or with friends, or with neighbors, or at work, or in public; and zero if this person uses only one and the same language in all of the mentioned circumstances), birth_nu[35] (equal to 1 if a person was born outside Ukraine and zero otherwise), official_r (equal to 1 if a person wants Russian to

[33] See Kennedy (2003).

[34] Alternatively speaking, one of the above-cited underlying conditions for the Heckman procedure becomes not simply binding, but prohibiting.

[35] This variable has 2 missing observations.

become an official language in Ukraine, and 0 otherwise), the dummies for the two most numerous religious groups in Ukraine, Greek Catholics (g_catholic) and Russian Orthodox (orthodox), as well as a dummy for atheists (atheist). The rest of the regressors include the Ukrainian ethnic identity and the variables for the extensive (non-aggregated) OLS model.

The regression output is presented in Table 14 of Appendix 3. The coefficients signs are largely as expected. Most notably, people using only one language and those born outside Ukraine are less likely to be bilingual. The same is true for the Russian nationals and those who want the Russian language to gain the official status (the latter effect is insignificant). However, most importantly, the equation provides quite a good fit to the actual values of the dependent variable. As can be read from Table 15 of Appendix 3, the overall prediction success of the regression is more then 78 percent. The equation is particularly good at predicting the bilingual status of an individual (nearly 87% of the correct predictions), but does considerably worse at predicting the monolingual skills (62.5 % of the correct predictions).

We then ran two OLS regressions with the appropriate inverted Mill's ratios (imills_bilingual for the bilingual sub-sample and imills_monolingual for the monolingual sub-sample), making use of our aggregated model and controlling for the exogenous mother tongue asset. As can be seen from the regressions outputs in Tables 16 and 17[36] of Appendix 3, the coefficients for both of the inverted Mill's ratios are insignificant at any conventional significance level. We can therefore accept the hypothesis of no selection bias in our model.

To conclude, we can emphasize that this result justifies our treatment of bilingual skills as an exogenous asset, and in essence no different from the mother

[36] We had to apply White heteroscedasticity consistent variance-covariance matrix for the monolingual sub-sample estimation.

tongue endowment to this end. It also supports our earlier conjectures about the rigidity of the labor supply[37], and the current dominance of the demand-side effects on the equilibrium wage rates within the framework of linguistically isolated labor markets model. It seems that to this extent the labor supply in Ukraine has been very inelastic, exacerbated still further by low geographical labor mobility.

[37] See the Theoretical Framework section.

Chapter 6

CONCLUSION

In our research we are trying to assess the effect of the “ukrainization policy”, which has been conducted by the Ukrainian government since the Declaration of Independence in 1991, by declaring Ukrainian the only official language and requiring all the official economic transactions to be conducted in this language. Assuming existence of a bias against Ukrainian speakers in the former Soviet Ukraine, which should have been reflected in comparatively worse labor outcomes for such people, we want to know whether at present Russian speakers still hold an advantage over their Ukrainian-speaking counterparts.

We use a representative survey of the Ukrainian population, conducted by the Institute of Sociology of the National Academy of Sciences of Ukraine in 2002 to see if the returns to different combinations of Ukrainian and Russian languages skills are symmetric. Our preliminary analysis of the income distribution over the different combinations of the command of Ukrainian and Russian appears to reject this hypothesis. We then estimate by OLS conventional earnings equations and discover that Russian-speaking monolinguals are at a significant advantage, as compared to any other language groups, including full bilinguals (those with a perfect command of both languages).

However, better specifications and, most notably, the inclusion of the variable indicating one’s mother tongue (and the one for ethnic identity), eliminate any systematic advantage of the individuals with better skills in Russian and worse in Ukrainian over those with worse knowledge of Russian and better knowledge of Ukrainian. It appears that the returns to the language skills are quite symmetric, once cleared from the effect of the “initial endowment” (thus, the marginal effects for the two representative groups, Russian-speaking partial bilinguals and Ukrainian-speaking partial bilinguals, are nearly identical[38]). On the other hand, the returns to the Ukrainian ethnic identity are significantly negative. We then estimate the regional regressions for the three representative local markets, Western, Central, and Eastern and Southern combined. The estimation results are essentially similar. The returns to language knowledge do no differ much between the two languages, and are rather insignificant at any rate. However, the Ukrainian mother tongue is persistently linked to lower income across all of the three regions. On the other hand, the returns to the Ukrainian ethnic identity are positive, albeit significant only for the Western market. The question poses as to why is the Ukrainian mother tongue such a consistent predictor of lower earnings[39]?

An intuitive explanation may be that the mother tongue identity is really a "historic" variable that captures a long-run dependence of future income opportunities on one's initial language endowment. It appears that the income we observe now, is not dependent on present language skills, but rather was predetermined quite stringently by one's mother tongue. In a way, the labor market is segmented by the dichotomy of the initial linguistic endowment. But this segmentation is in fact "historically" driven, reflecting the past language factor influence on labor market outcomes. Considering our principles of the sample construction, it is most probably that the Ukrainian mother tongue variable captures the negative income bias against Ukrainian speakers that existed in the Soviet Ukraine (and, evidently, previously as well). As we analyze the sample of individuals with a more or less stable work place, and the average age of more than thirty-nine years old, we can justly suppose that their income status

[38] See Table 4 of Appendix 3.

[39] All OLS results have been proved to be void of the selection bias problem, i.e. unbiased and consistent (see the Selectivity Analysis sub-section).

was determined under the conditions of the old, Russian nationalistic educational and professional system. It seems that Ukrainian-speaking people were doomed to fall behind their more fortunate counterparts, who were endowed with the Russian language initially. The Ukrainian speakers had a bad start, and although they may eventually have mastered the Russian language, as indicated by our cross-section, it is too late. They are already in the "wrong" segment of the labor market. This conclusion supports our initial conjecture of a bias against Ukrainian speakers that existed in the former Soviet Ukraine.

Interesting enough, there have been no incentives for those endowed with the Ukrainian mother tongue to learn Russian[40](see the sub-section Bilingualism and Earnings). This result once again tells us that in the former Soviet Ukraine native Ukrainian speakers had no chance to "catch up" with native Russian speakers. This is not to say that the former were discriminated against on purpose. Rather, the Russian language heavily dominated the whole educational and professional system in the former Soviet Union, and native Ukrainian speakers were constrained in their educational and career choices. The labor force has been very (geographically) immobile in the former USSR, and such additional constraints exacerbated the immobility still further. Moreover, as proved by our selectivity analysis, the language skills used to be imposed exogenously on individuals, and the latter did not really strive consciously to learn the second language in order to increase their earnings.

However, it is evident from our analysis that the linguistic bias is not really conspicuous for the present Ukrainian labor markets. The language skills are remunerated similarly, and for example in the Western labor market, the

[40] There have been, however, positive returns to native Russian speakers learning Ukrainian. This result strongly suggests of the absence of the Ukrainian language discrimination per se in the former Soviet Ukraine. However, it have been native Russian speakers who could benefit form learning Ukrainian, and not vice versa.

Russian language skills are not positively rewarded, although one might expect them to be required rather highly in this predominantly Ukrainian-speaking area.

It goes without saying that it would be interesting to do similar analysis for future data (and preferably with a much higher number of observations), to see the direction of the change of income differentials, as determined by language knowledge. We may already see that the recent generation of Ukrainians faces a situation when the languages are on par, and this should lead to insignificant rewards for the mother tongue indicator. It remains to see whether the Ukrainian language will win even more ground from the Russian language, and eventually surpass it by offering better initial linguistic endowment.

Thus so far we may claim that the "ukrainization" policy, conducted by the government, has reached its intermediate goal of equating the remunerations for the Russian and Ukrainian languages. Quite possibly, this policy has been also reinforced by the increased national awareness of the Ukrainians, embodied in the demand expansion for the products and services in the Ukrainian language, as well as changed employers' linguistic preferences[41].

At present, the marginal agent is indifferent between better knowledge in either of the languages. However, the situation is far from any advantage of the Ukrainian language over Russian. Our analysis provides a powerful counterargument against any claims of the newly introduced bias against Russian speakers.

Last but not least we should emphasize that the importance of better data collection cannot be stated too strongly. The language situation in Ukraine is a burning issue of political and cultural debate. All too often, however, the speculations of the opponents lack any sensible and reliable confirmation by

[41] See the Theoretical Framework for the details on the demand schedule formulation.

empirical evidence. The situation is exacerbated by an unmatched disregard of the official data collecting agencies as to the language variables. It seems apparent that in a bilingual economy labor surveys conducted by sociological and economic research NGOs, and governmental institutions alike, should include questions recording language endowment and skills. In practice, they do not record even mother tongue identity. To our knowledge, the data set used for the present analysis is the first (and by that sociological, not designed for advanced investigation of the labor market) to open opportunities for a comprehensive analysis of the language factors of income determination in Ukraine. Thus our research seems to present a valuable and unique assessment of the language situation in Ukraine, as well as suggest useful policy implications.

BIBLIOGRAPHY

Bloom, D.E. and G. Grenier. *Earnings of the French Minority in Canada and the Spanish Minority in the United States*. In B.R. Chiswick (ed.). Immigration, Language and Ethnicity: Canada and the United States. Washington: American Enterprise Institute. 373-409, 1992.

Christofides L. and R. Swidinsky. *Bilingualism and Earnings: A Study Based on 1971, 1981 and 1991 Census Data*. Ottawa: Official Languages Branch, Department of Canadian Heritage, 1997.

Carliner, G. *Wage Differences by Language Group and the Market for Language Skills in Canada*. The Journal of Human Resources, Volume 16, Issue 3 (Summer, 1981), 384-399.

Grin F. and F. Vaillancourt. *The Economics of Multilingualism: Overview of the Literature and Analytical Framework*. World Bank Group, 1999.

Halvorsen, R. and R. Palmquist *The Interpretation of Dummy Variables in Semilogarithmic Equations*. The American Economic Review, Volume 70, Issue 3. 474-475, June, 1980.

Heckman, J. *Sample Selection as a Specification Error*. Econometrica 47, 1979.

Henley A. and Jones R.E. *Earnings and Linguistic Ability in a Bilingual Economy*. University of Wales: Research Paper No 2001-18, 2001.

Kennedy, P. *A Guide to Econometrics*. Blackwell Publishers, 5th edition, 2003 (forthcoming).

McManus, W.S. *Labor Market Costs of Language Disparity: an Interpretation of Hispanic Earnings Differences*. American Economic Review, Vol. 75, No. 4, 1985.

Oaxaca, R.L. and M.R. Ransom. *On Discrimination and the Decomposition of Wage Differentials*. Journal of Econometrics, 61:5-21, 1994.

Panina, N. and E. Golovakha *Tendencies in the Development of Ukrainian Society (1994-2001). Sociological Indicators*. Kyiv: Institute of Sociology, 2001.

Pendakur K. and R. Pendakur. *Speak and Ye Shall Receive: Language as Human Capital*. Ottawa: Official Languages Branch, Department of Canadian Heritage, 1997.

Rubinstein, A. *Economics and Language*. Cambridge: Cambridge University Press, 2001

Vaillancourt, F. *Difference in Earnings by Language Groups in Quebec. 1970. An Economic Analysis* (Ph.D. Thesis, 1978). Quebec: Centre International de Recherché sur le Bilinguisme. Publication B-90.

Vaillancourt, F. *An Economic Perspective on Language and Public Policy in Canada and the United States*. In B.R. Chiswick (ed.). Immigration, Language and Ethnicity: Canada and the United States. Washington: American Enterprise Institute. 179-228, 1992.

A P P E N D I X 1

Tables 1: Distribution of Income[42] (Monthly Earnings in Ukrainian Hryvnas) by the Language Skills[43] (Combinations of Russian and Ukrainian) for Local Labor Markets

Combinations of the Language Skills	**U3_R3** Full knowledge of Ukrainian/ Full knowledge of Russian	**U3_R2** Full knowledge of Ukrainian/ Partial knowledge of Russian	**U3_R1** Full knowledge of Ukrainian/ No knowledge of Russian	**U2_R3** Partial knowledge of Ukrainian/ Full knowledge of Russian	**U2_R2** Partial knowledge of Ukrainian/ Partial knowledge of Russian	**U1_R3** No knowledge of Ukrainian/ Full knowledge of Russian	**U1_R2** No knowledge of Ukrainian/ Partial knowledge of Russian	**Total**
Number of Observations for the Western Labor Market	**89 (68.5%)**	**35 (26.9 %)**	**4 (3.1%)**	**1 (0.8%)**	**1 (0.8%)**	**-**	**-**	**130 (100 %)**
Income Distribution for the Western Labor Market	200 (106.75)	195.34 (73.97)	155 (23.8)	240 (-)	118 (-)	-	-	212.68 (97.78)
Number of Observations for the Central Labor Market	**136 (80.5 %)**	**12 (7.1 %)**	**1 (0.6 %)**	**14 (8.3 %)**	**6 (3.6 %)**	**-**	**-**	**169 (100 %)**
Income Distribution for the Central Labor Market	283.8 (232.98)	179.83 (81.92)	50 (-)	301 (203.31)	261.67 (194.78)	-	-	275.67 (222.39)
Number of Observations for the Eastern Labor Market	**152 (56.5 %)**	**7 (2.6 %)**	**1 (0.4 %)**	**94 (34.9 %)**	**6 (2.2 %)**	**9 (3.3%)**	**-**	**269 (100 %)**
Income Distribution for the Eastern Labor Market	276.91 (212.77)	198.86 (127.76)	200 (-)	291.54 (191.08)	194.5 (57.28)	415.56 (240.79)	-	282.35 (203.43)
Number of Observations for the Southern Labor Market	**45 (51.1 %)**	**2 (2.3%)**	-	**31 (35.2 %)**	**1 (1.1 %)**	**9 (10.2 %)**	**1 (1.1 %)**	**88 (100 %)**
Income Distribution for the Southern Labor Market	308 (264.7)	350 (353.55)	**-**	286.55 (224.67)	500 (-)	250 (253.86)	200 (-)	297.65 (247.13)
Number of Observations for the Eastern and Southern Labor Markets Combined	**196 (54.9 %)**	**9 (2.5 %)**	**1 (0.3 %)**	**125 (35 %)**	**7 (2 %)**	**18 (5%)**	**1 (0.3 %)**	**357 (100 %)**
Income Distribution for the Eastern and Southern Labor Markets Combined	284.01 (225.34)	227.78 (180.75)	200 (-)	290.3 (199)	238.14 (126.76)	332.78 (254.69)	200 (-)	286.12 (214.76)

[42] Mean and s.e. (in parenthesis) calculated for a particular sub-sample.

[43] The categories U2_R1 and U1_R1 were absent for all sub-samples.

Tables 2: Average Income (Monthly Earnings in Ukrainian Hryvnas) for the Categories of Language Knowledge, Mother Tongue and Ethnic Identity[44] for Local Labor Markets and the Entire Ukrainian Labor Market

Combinations of the Language Skills	KNOW_U		KNOW_R		UMT		UNATIONAL	
	Full Knowledge of Ukrainian	Not full Knowledge of Ukrainian	Full Knowledge of Russian	Not Full Knowledge of Russian	Ukrainian Mother Tongue	Russian Mother Tongue	Ukrainian Ethnic Identity	Russian Ethnic Identity
Number of Observations for the Western Labor Market (129 total)	**127 (98.4%)**	**2 (1.6%)**	**89 (69%)**	**40 (31%)**	**125 (96.9%)**	**4 (3.1%)**	**125 (96.9%)**	**4 (3.1%)**
Average Income for the Western Labor Market	213.7 (98.38)	179 (86.27)	223.85 (106.48)	189.38 (71.4)	208.66 (93.37)	353.75 (149.07)	208.58 (93.42)	356.25 (144.53)
Number of Observations for the Central Labor Market (166 total)	**147 (88.6%)**	**19 (11.4%)**	**147 (88.6%)**	**19 (11.4%)**	**136 (81.9%)**	**30 (18.1 %)**	**152 (91.6%)**	**14 (8.4 %)**
Average Income for the Central Labor Market	267.38 (218.75)	291.26 (201.65)	279.33 (223.69)	198.84 (132.01)	247.49 (164.83)	372.67 (356)	265.04 (215.18)	325.21 (230.42)
Number of Observations for the Eastern and Southern Labor Markets Combined (351 total)	**206 (58.7 %)**	**145 (41.3%)**	**333 (94.9%)**	**18 (5.1%)**	**132 (37.6)**	**219 (62.4%)**	**247 (70.4%)**	**104 (29.6%)**
Average Income for the Eastern and Southern Labor Markets Combined	279.9 (222.84)	291.07 (203.96)	287.83 (217.7)	223.17 (148.57)	215.42 (124.57)	326.16 (245.5)	278.84 (207.27)	298 (232.88)
Number of Observations for the Entire Ukrainian Labor Market (646 total)	**480 (74.3%)**	**166 (25.7%)**	**569 (88.1%)**	**77 (11.9%)**	**393 (60.8%)**	**253 (39.2%)**	**524 (81.1%)**	**122 (18.9%)**
Average Income for the Entire Ukrainian Labor Market	285.55 (197.83)	289.74 (202.33)	275.63 (206.95)	199.61 (108.94)	224.37 (132.63)	332.11 (259.27)	258.08 (190.9)	303.03 (229.26)

[44] 10 observations for the people of other nationalities were excluded from the contracted sample to preserve the pure dichotomy of Ukrainian vs. Russian, be it language or ethnic identity. All calculated statistics apply to the resultant number of observations (indicated specifically for each sub-sample).

A P P E N D I X 2

Table 1: Control Dummies

VARIABLE	DEFINITION	VARIABLE	DEFINITION
Oblast (Region)		**Occupation**	
REG1	Vinnitsa	**OCCUP1**	Politician
REG2	Volyn	**OCCUP2**	CEO
REG3	Dnipropetrovsk	**OCCUP3**	Public administration official
REG4	Donetsk	**OCCUP4**	White-collar worker (industry)
REG5	Zhytomir	**OCCUP5**	White-collar worker (services)
REG6	Transcarpthia	**OCCUP6**	Military or police serviceman
REG7	Zaporizhzhya	**OCCUP7**	Manager in a private business
REG8	Ivano-Frankivsk	**OCCUP8**	Blue-collar worker (services)
REG9	Kyiv (excl. Kyiv City)	**OCCUP9**	Skilled blue-collar worker (industry)
REG10	Kyiv City	**OCCUP10**	Unskilled blue-collar worker (industry)
REG11	Kirovograd	**OCCUP11**	Collective or state farm worker
REG12	Crimea	**Education**	
REG13	Lugansk	**EDU1**	Specialist's or Master's degree
REG14	Lviv	**EDU2**	Bachelor's degree
REG15	Nikolaev	**EDU3**	Technical college
REG16	Odessa	**EDU4**	Incomplete higher education
REG17	Poltava	**EDU5**	High school (general secondary)
REG18	Rivne	**EDU6**	Middle school (base secondary)
REG19	Sumy	**EDU7**	Primary education
REG20	Ternopil	**Settlement Type**	
REG21	Kharkiv		
REG22	Kherson	**SET1**	>1 mln inhabitants
REG23	Hmelnitsk	**SET2**	501,000-1mln inhabitants
REG24	Cherkasy	**SET3**	251,000-500,000 inhabitants
REG25	Chernihiv	**SET4**	101,000-250,000 inhabitants
REG26	Chernivtsi	**SET5**	51,000-100,000 inhabitants
Marital Status		**SET6**	20,000-50,000 inhabitants
MS1	Never married	**SET7**	<20,000 inhabitants
MS2	Married	**SET8**	Village
MS3	Lives with a partner		
MS4	Widow/widower		
MS5	Officially divorced		
MS6	Separated		

Table 2: Variables Aggregation

<table>
<tr><th>AGGREGATED VARIABLE</th><th>DEFINITION</th><th colspan="2">AGGREGATED VARIABLE</th><th>DEFINITION</th></tr>
<tr><td colspan="2">Region</td><td colspan="3">Marital Status</td></tr>
<tr><td rowspan="8">WEST</td><td>Volyn</td><td colspan="2" rowspan="4">SINGLE</td><td>Never married</td></tr>
<tr><td>Transcarpthia</td><td>Widow/widower</td></tr>
<tr><td>Ivano-Frankivsk</td><td>Officially divorced</td></tr>
<tr><td>Lviv</td><td>Separated</td></tr>
<tr><td>Rivne</td><td colspan="2" rowspan="2">MARRIED/CO-HABITING</td><td>Married</td></tr>
<tr><td>Ternopil</td><td>Lives with a partner</td></tr>
<tr><td>Hmelnitsk</td><td colspan="3">Settlement Type</td></tr>
<tr><td>Chernivtsi</td><td colspan="2">SET_AGGREGATE_1</td><td>>1 mln inhabitants</td></tr>
<tr><td rowspan="8">CENTER</td><td>Vinnitsa</td><td colspan="2">SET_AGGREGATE_2</td><td>101,000-1mln inhabitants</td></tr>
<tr><td>Zhytomir</td><td colspan="2">SET_AGGREGATE_3</td><td>20,000-100,000 inhabitants</td></tr>
<tr><td>Kyiv (excl. Kyiv City)</td><td colspan="2">SET_AGGREGATE_4</td><td><20,000 inhabitants</td></tr>
<tr><td>Kyiv City</td><td colspan="3">Education /Years of Schooling</td></tr>
<tr><td>Kirovograd</td><td rowspan="4">EDU_HIGHER</td><td rowspan="2">15 years</td><td rowspan="2">Specialist’s or Master’s degree</td></tr>
<tr><td>Poltava</td></tr>
<tr><td>Cherkasy</td><td rowspan="2">14 years</td><td rowspan="2">Bachelor’s degree</td></tr>
<tr><td>Chernihiv</td></tr>
<tr><td rowspan="4">SOUTH</td><td>Crimea</td><td rowspan="10">OTHER</td><td rowspan="2">12 years</td><td rowspan="2">Technical college</td></tr>
<tr><td>Nikolaev</td></tr>
<tr><td>Odessa</td><td rowspan="2">13 years</td><td rowspan="2">Incomplete higher education</td></tr>
<tr><td>Kherson</td></tr>
<tr><td rowspan="6">EAST</td><td>Dnipropetrovsk</td><td rowspan="2">10 years</td><td rowspan="2">High school (general secondary)</td></tr>
<tr><td>Donetsk</td></tr>
<tr><td>Zaporizhzhya</td><td rowspan="2">8 years</td><td rowspan="2">Middle school (base secondary)</td></tr>
<tr><td>Lugansk</td></tr>
<tr><td>Sumy</td><td rowspan="2">3 years</td><td rowspan="2">Primary education</td></tr>
<tr><td>Kharkiv</td></tr>
</table>

APPENDIX 3

Table 1: Regression Output (Full Bilinguals (U3_R3) Are a Base Dummy)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 1 664
Included observations: 656
Excluded observations: 8 after adjusting endpoints
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.944335	0.289119	17.10136	0.0000
U1_R3	**0.253344**	**0.150481**	**1.683558**	**0.0928**
U2_R3	**-0.023336**	**0.060888**	**-0.383263**	**0.7017**
U2_R2	**0.176146**	**0.119188**	**1.477880**	**0.1400**
U3_R2	**-0.087707**	**0.090857**	**-0.965328**	**0.3348**
U3_R1	**-0.173961**	**0.179786**	**-0.967600**	**0.3336**
AGE	-0.004728	0.002667	-1.772820	0.0768
MALE	0.417064	0.054862	7.601993	0.0000
MS2	0.235015	0.091908	2.557072	0.0108
MS3	0.091609	0.155812	0.587949	0.5568
MS4	0.165190	0.145248	1.137296	0.2559
MS5	0.069217	0.129575	0.534184	0.5934
MS6	0.032707	0.147169	0.222241	0.8242
OCCUP1	-0.417439	0.239297	-1.744434	0.0816
OCCUP2	0.204741	0.351452	0.582558	0.5604
OCCUP3	0.501803	0.222784	2.252416	0.0247
OCCUP4	0.552904	0.182919	3.022664	0.0026
OCCUP5	0.273174	0.189874	1.438710	0.1508
OCCUP6	0.487366	0.215755	2.258885	0.0243
OCCUP7	1.247635	0.280861	4.442182	0.0000
OCCUP8	0.263617	0.185991	1.417362	0.1569
OCCUP9	0.329366	0.176636	1.864663	0.0627
OCCUP10	0.149191	0.185708	0.803363	0.4221
SET1	0.336026	0.112631	2.983428	0.0030
SET2	0.189080	0.097836	1.932616	0.0538
SET3	0.253271	0.082332	3.076210	0.0022
SET4	-0.044692	0.128420	-0.348014	0.7280
SET5	0.153399	0.094475	1.623691	0.1050
SET6	0.170387	0.094228	1.808239	0.0711
SET7	0.166524	0.098366	1.692897	0.0910
EDU1	0.101558	0.201320	0.504462	0.6141
EDU2	-0.211965	0.238844	-0.887464	0.3752

EDU3	-0.005149	0.188736	-0.027282	0.9782
EDU4	-0.041441	0.245485	-0.168813	0.8660
EDU5	-0.113687	0.182422	-0.623208	0.5334
EDU6	-0.090704	0.191933	-0.472583	0.6367
REG1	-0.432857	0.196770	-2.199811	0.0282
REG2	-0.102998	0.174938	-0.588765	0.5562
REG3	-0.257134	0.125589	-2.047427	0.0411
REG4	-0.084788	0.131413	-0.645203	0.5190
REG5	-0.362309	0.195933	-1.849148	0.0649
REG6	-0.243005	0.167165	-1.453677	0.1466
REG7	-0.316378	0.166572	-1.899347	0.0580
REG8	-0.097949	0.190863	-0.513190	0.6080
REG9	0.055623	0.177817	0.312812	0.7545
REG11	-0.534755	0.230110	-2.323912	0.0205
REG12	-0.404106	0.177892	-2.271638	0.0235
REG13	0.055958	0.157474	0.355349	0.7225
REG14	-0.278725	0.146166	-1.906904	0.0570
REG15	-0.238969	0.209056	-1.143084	0.2535
REG16	0.153445	0.148607	1.032557	0.3022
REG17	-0.151677	0.162696	-0.932274	0.3516
REG18	-0.431241	0.192996	-2.234455	0.0258
REG19	-0.569676	0.230772	-2.468568	0.0138
REG20	-0.424981	0.210418	-2.019694	0.0439
REG21	-0.402063	0.139367	-2.884928	0.0041
REG22	-0.500359	0.225169	-2.222147	0.0266
REG23	-0.100444	0.273540	-0.367199	0.7136
REG24	-0.456116	0.181903	-2.507473	0.0124
REG25	-0.405946	0.168930	-2.403046	0.0166
REG26	0.201703	0.207369	0.972678	0.3311

R-squared	0.398973	Mean dependent var	5.382754
Adjusted R-squared	0.338365	S.D. dependent var	0.662713
S.E. of regression	0.539057	Akaike info criterion	1.690385
Sum squared resid	172.8966	Schwarz criterion	2.107544
Log likelihood	-493.4464	F-statistic	6.582858
Durbin-Watson stat	2.062259	Prob(F-statistic)	0.000000

Table 2: Regression Output (Russian Monolinguals (U1_R3) Are a Base Dummy)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 1 664
Included observations: 656
Excluded observations: 8 after adjusting endpoints
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.197679	0.327404	15.87541	0.0000
U3_R3	**-0.253344**	**0.150481**	**-1.683558**	**0.0928**
U2_R3	**-0.276680**	**0.145937**	**-1.895885**	**0.0585**
U2_R2	**-0.077198**	**0.188015**	**-0.410594**	**0.6815**
U3_R2	**-0.341050**	**0.170764**	**-1.997199**	**0.0463**
U3_R1	**-0.427304**	**0.227647**	**-1.877047**	**0.0610**
AGE	-0.004728	0.002667	-1.772820	0.0768
MALE	0.417064	0.054862	7.601993	0.0000
MS2	0.235015	0.091908	2.557072	0.0108
MS3	0.091609	0.155812	0.587949	0.5568
MS4	0.165190	0.145248	1.137296	0.2559
MS5	0.069217	0.129575	0.534184	0.5934
MS6	0.032707	0.147169	0.222241	0.8242
OCCUP1	-0.417439	0.239297	-1.744434	0.0816
OCCUP2	0.204741	0.351452	0.582558	0.5604
OCCUP3	0.501803	0.222784	2.252416	0.0247
OCCUP4	0.552904	0.182919	3.022664	0.0026
OCCUP5	0.273174	0.189874	1.438710	0.1508
OCCUP6	0.487366	0.215755	2.258885	0.0243
OCCUP7	1.247635	0.280861	4.442182	0.0000
OCCUP8	0.263617	0.185991	1.417362	0.1569
OCCUP9	0.329366	0.176636	1.864663	0.0627
OCCUP10	0.149191	0.185708	0.803363	0.4221
SET1	0.336026	0.112631	2.983428	0.0030
SET2	0.189080	0.097836	1.932616	0.0538
SET3	0.253271	0.082332	3.076210	0.0022
SET4	-0.044692	0.128420	-0.348014	0.7280
SET5	0.153399	0.094475	1.623691	0.1050
SET6	0.170387	0.094228	1.808239	0.0711
SET7	0.166524	0.098366	1.692897	0.0910
EDU1	0.101558	0.201320	0.504462	0.6141
EDU2	-0.211965	0.238844	-0.887464	0.3752
EDU3	-0.005149	0.188736	-0.027282	0.9782
EDU4	-0.041441	0.245485	-0.168813	0.8660
EDU5	-0.113687	0.182422	-0.623208	0.5334
EDU6	-0.090704	0.191933	-0.472583	0.6367
REG1	-0.432857	0.196770	-2.199811	0.0282

REG2	-0.102998	0.174938	-0.588765	0.5562
REG3	-0.257134	0.125589	-2.047427	0.0411
REG4	-0.084788	0.131413	-0.645203	0.5190
REG5	-0.362309	0.195933	-1.849148	0.0649
REG6	-0.243005	0.167165	-1.453677	0.1466
REG7	-0.316378	0.166572	-1.899347	0.0580
REG8	-0.097949	0.190863	-0.513190	0.6080
REG9	0.055623	0.177817	0.312812	0.7545
REG11	-0.534755	0.230110	-2.323912	0.0205
REG12	-0.404106	0.177892	-2.271638	0.0235
REG13	0.055958	0.157474	0.355349	0.7225
REG14	-0.278725	0.146166	-1.906904	0.0570
REG15	-0.238969	0.209056	-1.143084	0.2535
REG16	0.153445	0.148607	1.032557	0.3022
REG17	-0.151677	0.162696	-0.932274	0.3516
REG18	-0.431241	0.192996	-2.234455	0.0258
REG19	-0.569676	0.230772	-2.468568	0.0138
REG20	-0.424981	0.210418	-2.019694	0.0439
REG21	-0.402063	0.139367	-2.884928	0.0041
REG22	-0.500359	0.225169	-2.222147	0.0266
REG23	-0.100444	0.273540	-0.367199	0.7136
REG24	-0.456116	0.181903	-2.507473	0.0124
REG25	-0.405946	0.168930	-2.403046	0.0166
REG26	0.201703	0.207369	0.972678	0.3311

R-squared	0.398973	Mean dependent var	5.382754
Adjusted R-squared	0.338365	S.D. dependent var	0.662713
S.E. of regression	0.539057	Akaike info criterion	1.690385
Sum squared resid	172.8966	Schwarz criterion	2.107544
Log likelihood	-493.4464	F-statistic	6.582858
Durbin-Watson stat	2.062259	Prob(F-statistic)	0.000000

Table 3: Regression Output of the Aggregated Model (Full Bilinguals (U3_R3) Are a Base Dummy)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 1 664
Included observations: 656
Excluded observations: 8 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.657013	0.143656	32.41777	0.0000
U1_R3	**0.245298**	**0.144362**	**1.699190**	**0.0898**
U2_R3	**0.005754**	**0.061883**	**0.092976**	**0.9260**
U2_R2	**0.093879**	**0.158532**	**0.592176**	**0.5539**
U3_R2	**-0.112007**	**0.086364**	**-1.296924**	**0.1951**
U3_R1	**-0.234228**	**0.237859**	**-0.984736**	**0.3251**
EXPERIENCE	-0.003955	0.002029	-1.948774	0.0518
MALE	0.413944	0.049763	8.318313	0.0000
SINGLE	-0.186433	0.053513	-3.483890	0.0005
OCCUP1	-0.303395	0.350838	-0.864770	0.3875
OCCUP2	0.323797	0.215554	1.502159	0.1336
OCCUP3	0.690254	0.189360	3.645196	0.0003
OCCUP4	0.730269	0.134526	5.428457	0.0000
OCCUP5	0.423070	0.139685	3.028731	0.0026
OCCUP6	0.591101	0.199602	2.961399	0.0032
OCCUP7	1.341609	0.282906	4.742244	0.0000
OCCUP8	0.422873	0.133592	3.165409	0.0016
OCCUP9	0.432889	0.122890	3.522555	0.0005
OCCUP10	0.250526	0.134348	1.864760	0.0627
EDU_HIGHER	0.121767	0.071142	1.711600	0.0875
SET_AGGREGATE_1	0.400696	0.079238	5.056839	0.0000
SET_AGGREGATE_2	0.183263	0.064292	2.850463	0.0045
SET_AGGREGATE_3	0.097836	0.070782	1.382215	0.1674
EAST	0.044653	0.072756	0.613733	0.5396
CENTER	0.028043	0.080150	0.349884	0.7265
WEST	0.030860	0.087914	0.351026	0.7257

R-squared	0.298378	Mean dependent var	5.382754
Adjusted R-squared	0.270535	S.D. dependent var	0.662713
S.E. of regression	0.566014	Akaike info criterion	1.738433
Sum squared resid	201.8345	Schwarz criterion	1.916238
Log likelihood	-544.2060	F-statistic	10.71676
Durbin-Watson stat	1.810802	Prob(F-statistic)	0.000000

Table 4: Regression Output for the Aggregated Model with the Ukrainian Ethnic Identity and Mother Tongue Included (Russian Monolinguals (U1_R3) Are a Base Dummy)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 1 664
Included observations: 646
Excluded observations: 18 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.924756	0.194953	25.26126	0.0000
U3_R3	**-0.137826**	**0.147881**	**-0.932004**	**0.3517**
U2_R3	**-0.233920**	**0.143923**	**-1.625318**	**0.1046**
U2_R2	**-0.022366**	**0.211109**	**-0.105944**	**0.9157**
U3_R2	**-0.230495**	**0.169278**	**-1.361638**	**0.1738**
U3_R1	**-0.376740**	**0.272744**	**-1.381297**	**0.1677**
UMT	-0.285528	0.068858	-4.146632	0.0000
UNATIONAL	0.092959	0.069384	1.339775	0.1808
EXPERIENCE	-0.003527	0.002011	-1.754121	0.0799
MALE	0.416167	0.049341	8.434438	0.0000
SINGLE	-0.180146	0.052792	-3.412372	0.0007
OCCUP1	-0.298057	0.344638	-0.864841	0.3875
OCCUP2	0.324243	0.211886	1.530272	0.1265
OCCUP3	0.648605	0.187123	3.466202	0.0006
OCCUP4	0.692652	0.132994	5.208158	0.0000
OCCUP5	0.401579	0.138065	2.908629	0.0038
OCCUP6	0.541662	0.200930	2.695772	0.0072
OCCUP7	1.271054	0.278571	4.562759	0.0000
OCCUP8	0.382361	0.131781	2.901485	0.0038
OCCUP9	0.397942	0.120954	3.290032	0.0011
OCCUP10	0.217447	0.132220	1.644579	0.1006
EDU_HIGHER	0.077691	0.071476	1.086951	0.2775
SET_AGGREGATE_1	0.354429	0.079313	4.468762	0.0000
SET_AGGREGATE_2	0.147800	0.064359	2.296506	0.0220
SET_AGGREGATE_3	0.099485	0.070055	1.420102	0.1561
EAST	0.035570	0.072057	0.493641	0.6217
CENTER	0.094165	0.080513	1.169561	0.2426
WEST	0.118847	0.088843	1.337721	0.1815

R-squared	0.321539	Mean dependent var	5.375514
Adjusted R-squared	0.291898	S.D. dependent var	0.660704
S.E. of regression	0.555974	Akaike info criterion	1.706187
Sum squared resid	191.0285	Schwarz criterion	1.899968
Log likelihood	-523.0985	F-statistic	10.84760
Durbin-Watson stat	1.866623	Prob(F-statistic)	0.000000

Table 5: Regression Output for the Bilingualism Effects (Aggregated) Model

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 1 664
Included observations: 656
Excluded observations: 8 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.789427	0.146609	32.66801	0.0000
U3_R3	**0.084336**	**0.050804**	**1.660026**	**0.0974**
UMT	-0.265450	0.058943	-4.503535	0.0000
EXPERIENCE	-0.003353	0.001977	-1.695920	0.0904
MALE	0.410436	0.048878	8.397242	0.0000
SINGLE	-0.183079	0.052591	-3.481199	0.0005
OCCUP1	-0.222727	0.341288	-0.652609	0.5142
OCCUP2	0.326098	0.212502	1.534562	0.1254
OCCUP3	0.636386	0.186937	3.404279	0.0007
OCCUP4	0.694321	0.132586	5.236752	0.0000
OCCUP5	0.406899	0.137175	2.966272	0.0031
OCCUP6	0.563873	0.196309	2.872365	0.0042
OCCUP7	1.267559	0.279206	4.539868	0.0000
OCCUP8	0.411131	0.130625	3.147401	0.0017
OCCUP9	0.415043	0.120788	3.436113	0.0006
OCCUP10	0.235247	0.131865	1.784000	0.0749
EDU_HIGHER	0.095774	0.070374	1.360920	0.1740
SET_AGGREGATE_1	0.331793	0.078553	4.223841	0.0000
SET_AGGREGATE_2	0.133820	0.063897	2.094308	0.0366
SET_AGGREGATE_3	0.074314	0.069478	1.069601	0.2852
EAST	0.019384	0.070836	0.273654	0.7844
CENTER	0.082823	0.079092	1.047170	0.2954
WEST	0.088970	0.084919	1.047698	0.2952

R-squared	0.313811	Mean dependent var	5.382754
Adjusted R-squared	0.289963	S.D. dependent var	0.662713
S.E. of regression	0.558426	Akaike info criterion	1.707044
Sum squared resid	197.3948	Schwarz criterion	1.864333
Log likelihood	-536.9104	F-statistic	13.15848
Durbin-Watson stat	1.836930	Prob(F-statistic)	0.000000

Table 6: Regression Output for the Bilingualism Effects (Aggregated) Model (for the Sub-sample of Individuals with the Ukrainian Mother Tongue)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 1 636 IF UMT=1
Included observations: 394 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.785957	0.178058	26.87858	0.0000
U3_R3	**-0.008201**	**0.071061**	**-0.115403**	**0.9082**
EXPERIENCE	-0.003241	0.002486	-1.303420	0.1932
MALE	0.340172	0.061668	5.516166	0.0000
SINGLE	-0.190286	0.068724	-2.768851	0.0059
OCCUP1	-0.196702	0.339569	-0.579271	0.5628
OCCUP2	0.134572	0.243705	0.552195	0.5811
OCCUP3	0.795460	0.235138	3.382948	0.0008
OCCUP4	0.638116	0.150222	4.247808	0.0000
OCCUP5	0.471077	0.148872	3.164301	0.0017
OCCUP6	0.576699	0.233670	2.468010	0.0140
OCCUP7	1.146466	0.412901	2.776610	0.0058
OCCUP8	0.474295	0.143393	3.307655	0.0010
OCCUP9	0.393250	0.126975	3.097073	0.0021
OCCUP10	0.206538	0.144182	1.432479	0.1528
EDU_HIGHER	0.024167	0.096565	0.250262	0.8025
SET_AGGREGATE_1	0.426658	0.103062	4.139824	0.0000
SET_AGGREGATE_2	0.111578	0.074547	1.496748	0.1353
SET_AGGREGATE_3	0.094977	0.080411	1.181147	0.2383
EAST	-0.204849	0.112923	-1.814065	0.0705
CENTER	-0.059179	0.107628	-0.549848	0.5828
WEST	-0.070483	0.110721	-0.636588	0.5248

R-squared	0.245116	Mean dependent var	5.246558
Adjusted R-squared	0.202502	S.D. dependent var	0.618892
S.E. of regression	0.552688	Akaike info criterion	1.706171
Sum squared resid	113.6326	Schwarz criterion	1.928201
Log likelihood	-314.1158	F-statistic	5.751955
Durbin-Watson stat	1.919369	Prob(F-statistic)	0.000000

Table 7: Regression Output for the Bilingualism Effects (Aggregated) Model (for the Sub-sample of Individuals with the Russian Mother Tongue)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 14 664 IF UMT=0
Included observations: 262 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.827777	0.426285	11.32522	0.0000
U3_R3	**0.142995**	**0.076341**	**1.873118**	**0.0623**
EXPERIENCE	-0.004182	0.003270	-1.278809	0.2022
MALE	0.516615	0.081783	6.316871	0.0000
SINGLE	-0.159615	0.082406	-1.936947	0.0539
OCCUP2	0.421709	0.542612	0.777184	0.4378
OCCUP3	0.328444	0.474381	0.692363	0.4894
OCCUP4	0.588538	0.431734	1.363194	0.1741
OCCUP5	0.096127	0.445682	0.215686	0.8294
OCCUP6	0.446427	0.492646	0.906182	0.3657
OCCUP7	1.246022	0.539241	2.310699	0.0217
OCCUP8	0.156458	0.431795	0.362342	0.7174
OCCUP9	0.293719	0.425609	0.690115	0.4908
OCCUP10	0.076211	0.435392	0.175040	0.8612
EDU_HIGHER	0.124143	0.104020	1.193458	0.2339
SET_AGGREGATE_1	0.183227	0.161202	1.136631	0.2568
SET_AGGREGATE_2	0.098146	0.151364	0.648413	0.5173
SET_AGGREGATE_3	0.046764	0.161598	0.289388	0.7725
EAST	0.162471	0.093170	1.743810	0.0825
CENTER	0.159636	0.133995	1.191357	0.2347
WEST	0.297448	0.307811	0.966333	0.3348

R-squared	0.374738	Mean dependent var	5.587567
Adjusted R-squared	0.322849	S.D. dependent var	0.674777
S.E. of regression	0.555269	Akaike info criterion	1.738028
Sum squared resid	74.30587	Schwarz criterion	2.024040
Log likelihood	-206.6816	F-statistic	7.221911
Durbin-Watson stat	1.632174	Prob(F-statistic)	0.000000

Table 8: Regression Output for the Western Labor Market (Russian Language Knowledge)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample: 442 573
Included observations: 130
Excluded observations: 2
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.986073	0.513420	9.711491	0.0000
KNOW_R	**-0.018271**	**0.109527**	**-0.166820**	**0.8678**
UMT	-0.564014	0.203496	-2.771624	0.0065
EXPERIENCE	-0.004180	0.003577	-1.168609	0.2451
MALE	0.322908	0.119992	2.691071	0.0082
SINGLE	-0.201161	0.134407	-1.496650	0.1373
OCCUP1	-0.015239	0.514349	-0.029627	0.9764
OCCUP2	0.208146	0.597890	0.348135	0.7284
OCCUP3	1.247552	0.551115	2.263686	0.0255
OCCUP4	0.844138	0.529227	1.595040	0.1135
OCCUP5	0.981070	0.528490	1.856364	0.0661
OCCUP6	0.732948	0.655065	1.118893	0.2656
OCCUP7	1.810466	0.546100	3.315267	0.0012
OCCUP8	0.870026	0.552884	1.573614	0.1184
OCCUP9	0.775479	0.521600	1.486732	0.1399
OCCUP10	0.632974	0.532717	1.188200	0.2373
EDU_HIGHER	0.244804	0.133669	1.831424	0.0697
SET_AGGREGATE_2	-0.048271	0.118724	-0.406582	0.6851
SET_AGGREGATE_3	-0.000610	0.108792	-0.005611	0.9955

R-squared	0.340928	Mean dependent var	5.246879
Adjusted R-squared	0.234052	S.D. dependent var	0.526403
S.E. of regression	0.460700	Akaike info criterion	1.422163
Sum squared resid	23.55911	Schwarz criterion	1.841264
Log likelihood	-73.44058	F-statistic	3.189929
Durbin-Watson stat	1.770379	Prob(F-statistic)	0.000091

Table 9: Regression Output for the Western Labor Market (Bilingualism Effect)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample: 442 573
Included observations: 130
Excluded observations: 2
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.002223	0.509148	9.824691	0.0000
U3_R3	**-0.037241**	**0.112626**	**-0.330664**	**0.7415**
UMT	-0.564950	0.204045	-2.768758	0.0066
EXPERIENCE	-0.004365	0.003619	-1.206321	0.2303
MALE	0.321731	0.119908	2.683157	0.0084
SINGLE	-0.203041	0.135867	-1.494408	0.1379
OCCUP1	-0.008331	0.510248	-0.016327	0.9870
OCCUP2	0.218215	0.596224	0.365995	0.7151
OCCUP3	1.254028	0.547584	2.290111	0.0239
OCCUP4	0.845333	0.526227	1.606403	0.1110
OCCUP5	0.982760	0.524885	1.872334	0.0638
OCCUP6	0.739250	0.652527	1.132904	0.2597
OCCUP7	1.821535	0.544606	3.344684	0.0011
OCCUP8	0.875139	0.550125	1.590800	0.1145
OCCUP9	0.778001	0.518032	1.501841	0.1360
OCCUP10	0.629993	0.524765	1.200525	0.2325
EDU_HIGHER	0.251341	0.134160	1.873440	0.0636
SET_AGGREGATE_2	-0.050329	0.119968	-0.419516	0.6756
SET_AGGREGATE_3	0.001524	0.108601	0.014031	0.9888

R-squared	0.341609	Mean dependent var	5.246879
Adjusted R-squared	0.234843	S.D. dependent var	0.526403
S.E. of regression	0.460462	Akaike info criterion	1.421129
Sum squared resid	23.53477	Schwarz criterion	1.840230
Log likelihood	-73.37340	F-statistic	3.199604
Durbin-Watson stat	1.763363	Prob(F-statistic)	0.000088

Table 10: Regression Output for the Central Labor Market (Russian and Ukrainian Knowledge Separately)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample: 1 172
Included observations: 166
Excluded observations: 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.610337	0.334616	13.77800	0.0000
KNOW_R	**0.144371**	**0.149950**	**0.962798**	**0.3373**
KNOW_U	**-0.023537**	**0.162211**	**-0.145104**	**0.8848**
UMT	-0.265687	0.153451	-1.731409	0.0855
UNATIONAL	0.227508	0.208357	1.091918	0.2767
EXPERIENCE	-0.009594	0.004369	-2.195794	0.0297
MALE	0.393906	0.101021	3.899258	0.0001
SINGLE	-0.197591	0.115795	-1.706384	0.0901
OCCUP1	0.185998	0.636821	0.292073	0.7706
OCCUP2	0.987590	0.462942	2.133290	0.0346
OCCUP3	1.054495	0.339770	3.103552	0.0023
OCCUP4	1.079327	0.261266	4.131143	0.0001
OCCUP5	0.755259	0.264366	2.856876	0.0049
OCCUP6	1.049271	0.353050	2.972023	0.0035
OCCUP7	1.138564	0.622180	1.829958	0.0693
OCCUP8	1.013752	0.250116	4.053122	0.0001
OCCUP9	0.786565	0.238660	3.295751	0.0012
OCCUP10	0.393168	0.250581	1.569028	0.1188
EDU_HIGHER	-0.060617	0.139376	-0.434913	0.6643
SET_AGGREGATE_2	-0.161396	0.121453	-1.328876	0.1860
SET_AGGREGATE_3	-0.047171	0.115487	-0.408450	0.6835

R-squared	0.356248	Mean dependent var	5.382329
Adjusted R-squared	0.267455	S.D. dependent var	0.657097
S.E. of regression	0.562402	Akaike info criterion	1.804557
Sum squared resid	45.86284	Schwarz criterion	2.198242
Log likelihood	-128.7782	F-statistic	4.012106
Durbin-Watson stat	1.737768	Prob(F-statistic)	0.000000

Table 11: Regression Output for the Central Labor Market (Bilingualism Effect)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample: 1 172
Included observations: 166
Excluded observations: 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.677679	0.300513	15.56567	0.0000
U3_R3	**0.117231**	**0.119654**	**0.979755**	**0.3288**
UMT	-0.289386	0.149772	-1.932173	0.0553
UNATIONAL	0.190601	0.205863	0.925860	0.3560
EXPERIENCE	-0.009918	0.004364	-2.272432	0.0245
MALE	0.396357	0.100196	3.955833	0.0001
SINGLE	-0.199606	0.115354	-1.730373	0.0857
OCCUP1	0.189107	0.618570	0.305717	0.7603
OCCUP2	0.988357	0.461259	2.142740	0.0338
OCCUP3	1.077098	0.336802	3.198012	0.0017
OCCUP4	1.073834	0.260466	4.122744	0.0001
OCCUP5	0.757079	0.263301	2.875337	0.0046
OCCUP6	1.039417	0.352114	2.951937	0.0037
OCCUP7	1.147449	0.619461	1.852334	0.0660
OCCUP8	1.019895	0.248890	4.097778	0.0001
OCCUP9	0.785959	0.237812	3.304954	0.0012
OCCUP10	0.394885	0.249571	1.582253	0.1158
EDU_HIGHER	-0.070369	0.138079	-0.509629	0.6111
SET_AGGREGATE_2	-0.145457	0.121255	-1.199602	0.2322
SET_AGGREGATE_3	-0.039545	0.114735	-0.344666	0.7308

R-squared	0.356332	Mean dependent var	5.382329
Adjusted R-squared	0.272567	S.D. dependent var	0.657097
S.E. of regression	0.560436	Akaike info criterion	1.792378
Sum squared resid	45.85688	Schwarz criterion	2.167317
Log likelihood	-128.7674	F-statistic	4.253947
Durbin-Watson stat	1.723129	Prob(F-statistic)	0.000000

Table 12: Regression Output for the Eastern and Southern Labor Markets Combined (Russian and Ukrainian Knowledge Separately)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample: 173 441 574 664
Included observations: 351
Excluded observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.841441	0.242128	19.99539	0.0000
KNOW_R	**0.065912**	**0.153637**	**0.429009**	**0.6682**
KNOW_U	**0.018556**	**0.079579**	**0.233171**	**0.8158**
UMT	-0.293834	0.089319	-3.289714	0.0011
UNATIONAL	0.093402	0.081589	1.144795	0.2531
EXPERIENCE	-0.001332	0.002923	-0.455751	0.6489
MALE	0.467230	0.072879	6.411008	0.0000
SINGLE	-0.184051	0.077020	-2.389658	0.0174
OCCUP1	-0.443906	0.625469	-0.709717	0.4784
OCCUP2	0.121433	0.287087	0.422983	0.6726
OCCUP3	0.457185	0.294529	1.552255	0.1216
OCCUP4	0.581218	0.195421	2.974177	0.0032
OCCUP5	0.179361	0.201715	0.889177	0.3746
OCCUP6	0.327049	0.321627	1.016857	0.3100
OCCUP7	1.194945	0.392311	3.045916	0.0025
OCCUP8	0.162053	0.191200	0.847559	0.3973
OCCUP9	0.226095	0.183724	1.230624	0.2193
OCCUP10	0.146664	0.202015	0.726006	0.4684
EDU_HIGHER	0.151579	0.103700	1.461705	0.1448
SET_AGGREGATE_1	0.213361	0.123508	1.727501	0.0850
SET_AGGREGATE_2	0.119475	0.108871	1.097398	0.2733
SET_AGGREGATE_3	0.033878	0.112402	0.301397	0.7633

R-squared	0.321638	Mean dependent var	5.418894
Adjusted R-squared	0.278338	S.D. dependent var	0.700788
S.E. of regression	0.595324	Akaike info criterion	1.861206
Sum squared resid	116.6012	Schwarz criterion	2.103193
Log likelihood	-304.6417	F-statistic	7.428167
Durbin-Watson stat	1.801604	Prob(F-statistic)	0.000000

Table 13: Regression Output for the Eastern and Southern Labor Markets Combined (Bilingualism Effect)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample: 173 441 574 664
Included observations: 351
Excluded observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.890344	0.196183	24.92743	0.0000
U3_R3	**0.056078**	**0.075231**	**0.745415**	**0.4566**
UMT	-0.310890	0.086450	-3.596170	0.0004
UNATIONAL	0.087422	0.081363	1.074472	0.2834
EXPERIENCE	-0.001322	0.002915	-0.453372	0.6506
MALE	0.463920	0.072227	6.423065	0.0000
SINGLE	-0.182873	0.076720	-2.383647	0.0177
OCCUP1	-0.442638	0.624004	-0.709351	0.4786
OCCUP2	0.120187	0.285422	0.421085	0.6740
OCCUP3	0.442800	0.294394	1.504106	0.1335
OCCUP4	0.582983	0.194672	2.994699	0.0030
OCCUP5	0.176969	0.201363	0.878856	0.3801
OCCUP6	0.336606	0.320060	1.051695	0.2937
OCCUP7	1.188667	0.391706	3.034589	0.0026
OCCUP8	0.164870	0.190149	0.867055	0.3865
OCCUP9	0.232344	0.182544	1.272810	0.2040
OCCUP10	0.144915	0.201203	0.720243	0.4719
EDU_HIGHER	0.145804	0.103656	1.406616	0.1605
SET_AGGREGATE_1	0.218921	0.123293	1.775616	0.0767
SET_AGGREGATE_2	0.123730	0.108442	1.140973	0.2547
SET_AGGREGATE_3	0.037989	0.111792	0.339822	0.7342

R-squared	0.322224	Mean dependent var	5.418894
Adjusted R-squared	0.281147	S.D. dependent var	0.700788
S.E. of regression	0.594165	Akaike info criterion	1.854644
Sum squared resid	116.5004	Schwarz criterion	2.085631
Log likelihood	-304.4900	F-statistic	7.844323
Durbin-Watson stat	1.805746	Prob(F-statistic)	0.000000

Table 14: Regression Output of the Probit Selection Equation (U3_R3 Is the Dependent Variable)

Dependent Variable: U3_R3
Method: ML - Binary Probit (Quadratic hill climbing)
Sample (adjusted): 1 664
Included observations: 652
Excluded observations: 12 after adjusting endpoints
Convergence achieved after 5 iterations
QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.383690	0.850511	-0.451129	0.6519
UNATIONAL	0.295521	0.178664	1.654063	0.0981
NO_PURIST	0.916851	0.132292	6.930488	0.0000
BIRTH_NU	-0.756999	0.211329	-3.582090	0.0003
OFFICIAL_R	-0.134989	0.142668	-0.946174	0.3441
G_CATHOLIC	0.670085	0.455982	1.469542	0.1417
ORTHODOX	0.417474	0.336332	1.241257	0.2145
ATHEIST	0.440004	0.368786	1.193114	0.2328
AGE	0.009011	0.005397	1.669691	0.0950
EDU1	1.333529	0.638314	2.089142	0.0367
EDU2	1.374504	0.895673	1.534604	0.1249
EDU3	0.586648	0.613924	0.955570	0.3393
EDU4	0.738061	0.699843	1.054611	0.2916
EDU5	0.561916	0.611753	0.918534	0.3583
EDU6	0.024722	0.618115	0.039996	0.9681
SET1	-0.478870	0.266062	-1.799846	0.0719
SET2	-0.171482	0.294805	-0.581680	0.5608
SET3	-0.417773	0.196184	-2.129498	0.0332
SET4	-0.599234	0.304444	-1.968290	0.0490
SET5	0.154233	0.230799	0.668257	0.5040
SET6	-0.305017	0.271748	-1.122426	0.2617
SET7	0.295426	0.330670	0.893416	0.3716
REG1	-1.360028	0.472724	-2.876999	0.0040
REG2	-0.539849	0.540266	-0.999227	0.3177
REG3	-1.206148	0.381396	-3.162454	0.0016
REG4	-0.861824	0.403281	-2.137034	0.0326
REG5	-0.102410	0.559499	-0.183039	0.8548
REG6	-0.615654	0.619422	-0.993916	0.3203
REG7	-0.829385	0.486598	-1.704458	0.0883
REG8	-1.894094	0.576273	-3.286799	0.0010
REG9	-0.601638	0.483560	-1.244184	0.2134
REG11	-1.472343	0.601253	-2.448789	0.0143
REG12	-1.679946	0.482207	-3.483866	0.0005
REG13	-1.120091	0.432089	-2.592269	0.0095
REG14	-1.115488	0.498885	-2.235963	0.0254
REG15	0.047325	0.577349	0.081970	0.9347

REG16	-1.274089	0.405433	-3.142542	0.0017
REG17	0.094288	0.688531	0.136941	0.8911
REG18	-1.194557	0.513766	-2.325100	0.0201
REG19	-1.439545	0.481903	-2.987211	0.0028
REG20	-0.694271	0.638635	-1.087118	0.2770
REG21	-0.338053	0.419976	-0.804933	0.4209
REG22	-0.644452	0.492550	-1.308399	0.1907
REG23	0.078574	0.613263	0.128124	0.8981
REG24	-0.732792	0.537981	-1.362115	0.1732
REG25	-0.989413	0.566430	-1.746753	0.0807
REG26	-1.028900	0.635197	-1.619812	0.1053

Mean dependent var	0.644172	S.D. dependent var	0.479131
S.E. of regression	0.407857	Akaike info criterion	1.085109
Sum squared resid	100.6401	Schwarz criterion	1.408057
Log likelihood	-306.7456	Hannan-Quinn criter.	1.210355
Restr. log likelihood	-424.4390	Avg. log likelihood	-0.470469
LR statistic (46 df)	235.3869	McFadden R-squared	0.277292
Probability(LR stat)	0.000000		
Obs with Dep=0	232	Total obs	652
Obs with Dep=1	420		

Table 15: Expectation-prediction Table for the Probit Selection Equation

Dependent Variable: U3_R3
Method: ML - Binary Probit (Quadratic hill climbing)
Sample (adjusted): 1 664
Included observations: 652
Excluded observations: 12 after adjusting endpoints
Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
PDep=1)<=C	145	55	200	0	0	0
P(Dep=1)>C	87	365	452	232	420	652
Total	232	420	652	232	420	652
Correct	145	365	510	0	420	420
% Correct	**62.50**	**86.90**	**78.22**	0.00	100.00	64.42
% Incorrect	37.50	13.10	21.78	100.00	0.00	35.58

Table 16: Regression Output of the Selectivity-adjusted OLS Equation (the Bilingual Sub-sample)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Sample (adjusted): 1 657 IF U3_R3=1
Included observations: 417 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.008497	0.200925	24.92722	0.0000
UMT	-0.325030	0.076741	-4.235445	0.0000
EXPERIENCE	-0.005904	0.002604	-2.267149	0.0239
MALE	0.398360	0.062927	6.330521	0.0000
SINGLE	-0.196211	0.070263	-2.792513	0.0055
OCCUP1	-0.173153	0.426970	-0.405538	0.6853
OCCUP2	0.716487	0.261153	2.743555	0.0064
OCCUP3	0.762552	0.212114	3.595005	0.0004
OCCUP4	0.751232	0.160008	4.694967	0.0000
OCCUP5	0.465949	0.165897	2.808668	0.0052
OCCUP6	0.570172	0.237525	2.400476	0.0168
OCCUP7	1.486204	0.321928	4.616568	0.0000
OCCUP8	0.534420	0.159625	3.347983	0.0009
OCCUP9	0.487491	0.143830	3.389359	0.0008
OCCUP10	0.294153	0.162084	1.814813	0.0703
EDU_HIGHER	0.005494	0.087365	0.062889	0.9499
SET_AGGREGATE_1	0.342663	0.099543	3.442367	0.0006
SET_AGGREGATE_2	0.163830	0.083363	1.965252	0.0501
SET_AGGREGATE_3	0.040159	0.085708	0.468557	0.6396
EAST	-0.020265	0.103861	-0.195114	0.8454
CENTER	0.052297	0.105257	0.496855	0.6196
WEST	0.041890	0.111069	0.377157	0.7063
IMILLS_BILINGUAL	**-0.153234**	**0.101245**	**-1.513493**	**0.1310**

R-squared	0.320486	Mean dependent var	5.379019
Adjusted R-squared	0.282544	S.D. dependent var	0.674580
S.E. of regression	0.571388	Akaike info criterion	1.772080
Sum squared resid	128.6349	Schwarz criterion	1.994529
Log likelihood	-346.4788	F-statistic	8.446644
Durbin-Watson stat	1.710092	Prob(F-statistic)	0.000000

Table 17: Regression Output of the Selectivity-adjusted OLS Equation (the Monolingual Sub-sample)

Dependent Variable: LOG(INCOME)
Method: Least Squares
Date: 05/25/03 Time: 17:28
Sample (adjusted): 14 664 IF U3_R3=0
Included observations: 233 after adjusting endpoints
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.815646	0.231758	20.77875	0.0000
UMT	-0.025674	0.109179	-0.235157	0.8143
EXPERIENCE	-0.000698	0.003336	-0.209199	0.8345
MALE	0.534883	0.089141	6.000410	0.0000
SINGLE	-0.138678	0.078203	-1.773301	0.0776
OCCUP1	-0.430063	0.215519	-1.995475	0.0473
OCCUP2	-0.850297	0.650760	-1.306622	0.1928
OCCUP3	0.259573	0.495924	0.523413	0.6012
OCCUP4	0.539603	0.232095	2.324922	0.0210
OCCUP5	0.278244	0.225684	1.232895	0.2190
OCCUP6	0.297044	0.282800	1.050367	0.2948
OCCUP7	0.566353	0.218569	2.591189	0.0102
OCCUP8	0.079810	0.205646	0.388092	0.6983
OCCUP9	0.140628	0.206257	0.681812	0.4961
OCCUP10	0.003567	0.217230	0.016420	0.9869
EDU_HIGHER	0.207994	0.148617	1.399531	0.1631
SET_AGGREGATE_1	0.285980	0.140934	2.029171	0.0437
SET_AGGREGATE_2	0.157783	0.107599	1.466399	0.1440
SET_AGGREGATE_3	0.193166	0.144838	1.333673	0.1838
EAST	0.092451	0.116557	0.793187	0.4286
CENTER	-0.055453	0.154599	-0.358691	0.7202
WEST	-0.065825	0.153659	-0.428383	0.6688
IMILLS_MONOLINGUAL	**0.025499**	**0.089373**	**0.285310**	**0.7757**

R-squared	0.391625	Mean dependent var	5.373029
Adjusted R-squared	0.327891	S.D. dependent var	0.635355
S.E. of regression	0.520878	Akaike info criterion	1.626892
Sum squared resid	56.97591	Schwarz criterion	1.967553
Log likelihood	-166.5329	F-statistic	6.144641
Durbin-Watson stat	1.951591	Prob(F-statistic)	0.000000