James LaBelle

Economics 312

Reed College

Wage Differentials For Bilingualism in French and English Native Speakers Across French and English Speaking Provinces of Canada

This paper attempts to evaluate the hypothesis that there is a wage premium attached to bilingualism. The hypothesized premium would be for the economic value of bilingualism (an effect on total income) to increase with proximity to Quebec for native English speakers and to decrease with proximity to Quebec for French native speakers. The intuition is that it is more favorable to an individual's income to speak the language of the province of residence the further an individual is from the region in which their native tongue is spoken. I start with a fairly traditional model specification for income controlling for determination:

$$\text{Log(income)} = \beta_1(\text{female}) + \beta_2(\text{minority status}) + \beta_3(\text{age}) + \beta_4(\text{education})$$

The first model applied to the Canadian census data does not include union membership, another element of classic models of income, because the census does not carry information on the individual's membership status. In the later models, terms are added specifying effects for native language, bilingualism, province of residence, and a set of interaction terms. The weak form of the hypothesis is for a positive effect of bilingualism on income in general and especially for bilingualism in provinces where the predominant language is not the individual's first. The strong

form is for a monotonically diminishing effect of the importance of bilingualism on English Canadians' incomes in provinces further and further west of Quebec and a monotonically increasing effect for French Canadians. The Atlantic Provinces and the territories are omitted because of collinearity: no native French speakers are recorded in these regions in the (incomplete) dataset.

The dataset includes just under 550,000 observations across Canada. For purposes of this project, individuals whose first language is neither English nor French are dropped, as well as (for reasons of simplicity) the small minority reporting both English and French as first languages. French or English as a first language is denoted by the dummy variable "french." Minority status is coded according to the Canadian census' "visible minority indicator" field, which reports an individuals' response as either "Chinese," "South Asian," "Black," "other visible minority or multiple visible minority," or "not a visible minority." Interestingly, the last category includes both Ethnic European and Aboriginal Canadians. The model in use here specifies all "visible minorities" and all individuals responding as Aboriginal together in a single dummy variable, "vismin." The variable "age" gives the individual's age, with all observations above 85 reported as 85. Education is covered by a variable called "school" which corresponds roughly to a unit increase in education. Its values go from 1-9, where 1 is less than 5 years of schooling and 9 is 18 years or more of schooling. It is important to note that an increase of 1 in the variable "school" does not necessarily correspond in every case to an additional year of schooling, but rather what the census takers in Canada determined to be a convenient unit of education for census purposes. 5,6,7, and 8 years of schooling are

counted in a single increment, as are 14, 15, 16, and 17 years, though other increments do correspond to a single year's education. Individuals under 15 years of age are coded in the census with a value of 99, and must consequently be omitted.

The first model includes only the simple theoretical controls for the determinants of income. The second includes the correlations of bilingualism, residence in Quebec, and whether one's first language is French with the log of income, while specifying that the effect of one dummy variable is the same whether another is turned on or not. This regression tests whether these dummy variables have effects at all, but do not give the level of nuance necessary to provide satisfactory evidence in favor or against the hypothesis. The third regression adds interaction terms between bilingualism, residence in Quebec, and whether one's first language is French, allowing for eight possible different effects for monolingual and bilingual English and French Canadians inside and outside Quebec. The fourth model allows for a vastly wider spectrum of different effects by adding dummy variables and interaction terms with those provinces, bilingualism, and first language for individual provinces besides Quebec. Note that the baseline category in this regression is British Columbia, and Atlantic Provinces and territories are omitted because of collinearity with some of the dummy variables. Results consistent with the strong hypothesis would be significant positive effects on income such that for native English speakers

effect(Quebec)>effect(Ontario)>effect(Manitoba)>effect(Saskatchewan)
>effect(Alberta)>effect(baseline/British Columbia)

and the reverse for native French speakers.

Results

	ltotinc	ltotinc	ltotinc	ltotinc
female	-0.508	-0.508	-0.505	-0.505
	(142.06)**	(141.97)**	(141.13)**	(141.16)**
vismin	-0.307	-0.317	-0.317	-0.312
	(34.80)**	(35.73)**	(35.75)**	(35.03)**
age	0.025	0.025	0.025	0.025
	(229.15)**	(229.11)**	(227.75)**	(227.33)**
school	0.211	0.210	0.209	0.209
	(232.17)**	(225.06)**	(223.09)**	(222.50)**
qc		-0.095	0.037	0.138
		(12.35)**	(2.31)*	(5.62)**
bilingualenfr		0.007	0.115	0.152
		(1.51)	(3.38)**	(1.43)
french		0.050	0.112	0.812
		(6.30)**	(4.98)**	(1.74)
0b.french#1.bilingualenfr			-0.198	-0.312
			(5.64)**	(2.87)**
0b.bilingualenfr#1.qc			-0.218	-0.295
			(7.97)**	(8.84)**
0b.qc#1.french			-0.160	-0.240
			(4.02)**	(2.32)*
1.bilingualenfr#1.french#1.qc			-0.287	-0.400
			(6.48)**	(3.57)**
on				0.149
				(6.52)**
0b.bilingualenfr#1.on				-0.103
				(4.36)**
0b.on#1.french				-0.027
				(0.25)
1.bilingualenfr#1.french#1.on				-0.089
				(0.76)
mb				0.055
				(1.31)
0b.bilingualenfr#1.mb				-0.047
				(1.08)
0b.mb#1.french				-0.200
				(1.19)
1.bilingualenfr#1.french#1.mb				-0.231
				(1.28)
sk				0.004
				(0.07)
0b.bilingualenfr#1.sk				-0.043
				(0.81)
0b.sk#1.french				-0.150
				(0.81)
1.bilingualenfr#1.french#1.sk				-0.247
				(1.21)
ab				0.033
				(1.07)
0b.bilingualenfr#1.ab				-0.018
				(0.56)
0b.ab#1.french				-0.323
				(2.28)*
1.bilingualenfr#1.french#1.ab				-0.302
				(1.98)*
_cons	7.649	7.667	7.680	7.658
	(869.70)**	(853.09)**	(847.41)**	(742.89)**
R2	0.20	0.20	0.20	0.20
N	404,534	404,534	404,534	404,534

* p<0.05; ** p<0.01

Category **Proportion Relative to Baseline Income**

Regression 2

Resident in Quebec	e^(–0.095) = 0.91
Bilingual	e^(0) = 1 (not statistically significant at 95%)
French Native Speaker	e^(0.050) = 1.05

Regression 3

monolingual English outside QC:	e^0 = 1 (baseline)
monolingual French outside QC:	e^(0.112) = 1.12
monolingual English in QC:	e^(0.037) = 1.04
monolingual French in QC:	e^(0.037 + 0.112 – 0.160) = 0.99
bilingual English outside QC:	e^(0.115) = 1.12
bilingual French outside QC:	e^(0.115 + 0.112 – 0.198) = 1.03
bilingual English in QC:	e^(0.115 + 0.037 – 0.218) = 0.94
bilingual French in QC:	e^(0.115 + 0.112 + 0.037 – 0.198 – 0.218) = 0.86

Regression 4

monolingual English in BC:	e^0 = 1 (baseline)
monolingual English in AB:	e^(0) = 1 (not statistically significant at 95%)
monolingual English in SK:	e^(0) = 1 (not statistically significant at 95%)
monolingual English in MB:	e^(0) = 1 (not statistically significant at 95%)
monolingual English in ON:	e^(0.149) = 1.16
monolingual English in QC:	e^(0.138) = 1.15
monolingual French in BC:	e^(0) = 1 (not statistically significant at 95%)
monolingual French in AB:	e^(0 – 0.323) = 0.72
monolingual French in SK:	e^(0) = 1 (not statistically significant at 95%)
monolingual French in MB:	e^(0) = 1 (not statistically significant at 95%)
monolingual French in ON:	e^(0 + 0.149 + 0) = 1.16
monolingual French in QC:	e^(0 + 0.138 – 0.240) = 0.90
bilingual English in BC:	e^(0) = 1 (not statistically significant at 95%)
bilingual English in AB:	e^(0) = 1 (not statistically significant at 95%)
bilingual English in SK:	e^(0) = 1 (not statistically significant at 95%)
bilingual English in MB:	e^(0) = 1 (not statistically significant at 95%)
bilingual English in ON:	e^(0.149 + 0 – 0.103) = 1.05
bilingual English in QC:	e^(0.138 + 0 – 0.295) = 0.85
bilingual French in BC:	e^(0 + 0 – 0.312) = 0.73
bilingual French in AB:	e^(0 + 0 + 0 + 0 – 0.323 – 0.302 – 0.312) = 0.39
bilingual French in SK:	e^(0 + 0 + 0 + 0 – 0.312+ 0 + 0) = 0.73
bilingual French in MB:	e^(0 + 0 + 0 + 0 – 0.312 + 0 + 0) = 0.73
bilingual French in ON:	e^(0.149 + 0 + 0 – 0.103 – 0.312 + 0 + 0) = 0.77
bilingual French in QC:	e^(0.138 + 0 + 0 – 0.295 – 0.312 – 0.240– 0.400) = 0.33

The results of the most general regression regarding language in Canada, the second regression, show significant effects for all the classical control variables on

log income. There is no statistically significant correlation between income and bilingualism at the 95% level. There is an effect for residence in Quebec, which is associated with 9% less income for individuals at the mean in other variables. Another effect exists for French native speakers, who at the mean make 5% more income than English speakers, controlling for residence in Quebec.

The findings of the third regression are that there are significant differences between the different groups under study, controlling for the other variables. Bilingualism in English Canadians is associated with higher incomes outside of Quebec than inside while monolingualism is associated with higher incomes in Quebec than outside. Monolingual English Canadians in Quebec on average make more than bilingual English Canadians, though bilingual English Canadians make more than monolingual English Canadians outside Quebec. In French Canadians, bilingualism is associated with higher incomes outside of Quebec than inside while there is no statistically significant income differential between monolingual French Canadians in and out of Quebec. Within French Canadians residing in Quebec, on average those who are bilingual had lower incomes than those who are monolingual. For French Canadians outside Quebec, this same effect is seen.

Weak H_1: (referring to mean predicted incomes)

Being bilingual is more important to English Canadians' income in a province speaking a different language:

I. bilingual English in QC > bilingual English outside QC,

Being monolingual is more disadvantageous to English Canadians' income in a province speaking a different language:

II. Monolingual English in QC < monolingual English outside QC

Bilingualism in general is linked with higher income for English Canadians:

III. Bilingual English in QC > monolingual English in QC

IV. Bilingual English outside QC > monolingual English outside QC*

Being bilingual is more important to French Canadians' income in a province speaking a different language:

V. Bilingual French outside QC > bilingual French in QC*

Being monolingual is more disadvantageous to French Canadians' income in a province speaking a different language:

VI. Monolingual French in QC > monolingual French outside QC**

Bilingualism in general is linked with higher income for French Canadians:

VII. Bilingual French in QC > monolingual French in QC

VIII. Bilingual French outside QC > monolingual French outside QC

* Inequality consistent with the results of regression 3

** No statistical difference according to the results of regression 3

Using the results of the third regression, the null hypothesis cannot be rejected. Only conditions IV and V are supported by the data while in all other cases

the results are either that there is no statistically significant difference or that the evidence for the opposite inequality is statistically significant. The data suggest no pattern of language capability-income relation that is consistent with intuitive theory that speaking more languages, particularly the native language of one's place of residence, always has a positive effect on one's economic opportunities and hence income. The results of the fourth regression are equally inconsistent with the strong hypothesis, and likewise the null cannot be rejected. There is not a statistically significant monotonic decrease in the value of bilingualism for English Canadians moving westward through the Canadian provinces from Quebec to British Columbia, nor is there a statistically significant monotonic increase in the value of bilingualism for French Canadians over the same geographic interval. The lack of statistical significance in many of the plains provinces, as well as that the hypothesized relationship is not supported even in a simpler case combine into a failure to reject the null hypothesis when testing the strong case. A further troubling element of the regressions is that the R-squared remains low, at 0.20, rather than increasing with the complexity of specification.

Conclusion

The tests of the strong and weak hypotheses performed here assess whether within the Canadian census data higher incomes are correlated with bilingualism, particularly for individuals residing in provinces where a language other than their first predominates. The results, if not outright falsifying it, suggest that there are probably other important effects at play that outweigh the importance of the theory this paper attempts to test. Either the model is mis-specified or the theory is

incorrect. There are some possible explanations for the inconsistency of the results and the hypothesis. The apparently perverse income premiums on monolingualism for individuals in provinces not speaking their language could be because only very wealthy individuals (perhaps those in very technical careers) can afford not to learn the language of their province of residence. At the same time, lower-income workers, particularly in the all-important service sector, would be more likely to be bilingual (or become so out of necessity) when in provinces speaking a language other than their native tongue. It may simply not be true that a wage premium exists for bilingualism in foreign-speaking provinces because the vast majority of employment opportunities not requiring language skills are high-income work and consequently low-income monolingual individuals either do not migrate or do not remain monolingual.

This would be an endogeneity problem, in which income and provincial location, after controlling for the (presumably positive) effects of education, affected the probability of being bilingual negatively. Endogeneity may also enter the model in the form of income determining decisions to move to other provinces. Further research could attempt to eliminate this endogeneity by finding and implementing appropriate instruments for bilingualism and possibly province of residence. These would need to be characteristics correlated with an individual of a particular language category's status as bilingual or their province, but exogenous and not affecting income independently of those variables. Presumably given strong and valid instruments it would be possible to control for bilingualism and province of residence independent of the reverse causality from income to these variables as

originally measured. This would allow for testing of the wage premium hypothesis independent of the reverse causation from income to province and bilingualism.