Undergraduate Economic Review

Volume 14 | Issue 1 Article 5

2017

Is There Really a Foreign Language Premium in Canada?: Evaluating the Foreign Language Effect on Canadian Wages

Peter C. Ki
University of British Columbia, peterki_95@hotmail.com

Is There Really a Foreign Language Premium in Canada?: Evaluating the Foreign Language Effect on Canadian Wages

Abstract

Given the debates on whether knowing a foreign language has a positive or negative effect on wages, there is little or no research conducted on how it fares in Canada, despite extensive studies on French and English. Using the Public Use Microdata Files of the 2011 National Household Survey by Statistics Canada, I find a 2.84% wage penalty for Canadians who know a foreign language and an even greater penalty for Canadians who use these languages at work. I proceed to decompose the results by language and occupation to find varying results. I also attempt to use a religion proxy to see whether it changes a foreign language's wage effects with mixed results.

Keywords

foreign language, effect on wages, Canadian wages, language penalty, occupation, religion, share of speakers

Cover Page Footnote

I would like to thank Prof. Jonathan Graves for giving me the opportunity to pursue this paper while providing useful commentary on how to proceed.

Accreditation Acknowledgement:

This analysis is based on the Statistics Canada *2011 National Household Survey (NHS) public use microdata file (PUMF) on individuals* which contains anonymized data collected in the 2011 National Household Survey. All computations on these microdata were prepared by Peter Ki at the University of British Columbia and the responsibility for the use and interpretation of these data is entirely that of the author.

Introduction:

Learning a foreign language attracts many people for a variety of reasons. Often, delving deeper into a particular culture, speaking to a new set of people, or travelling are cited as main reasons for learning a new language. To more professionally-oriented individuals, learning a new language may be a way to improve salary figures or signal higher qualifications for employers. These ideas have been a topic of interest for economists in human capital research where different papers have analyzed the income effects of knowing a foreign language. These effects may be specified either as a dominant language within a society, a minority language, or a foreign language not recognized as an official language.

In the Canadian context, knowledge of a foreign language can be especially useful. Given a strong sense of multiculturalism and a large share of immigrants in the population, proficiency in another language opens doors to a wide range of opportunities in certain communities. In the Vancouver context, signs can be found written in Chinese, Punjabi, Vietnamese, Korean, and so forth, indicating a diverse metropolis with many different language communities. From a national and a global standpoint, having skills in these languages may impact your earnings in a significant way.

This paper will investigate the role of knowing a non-official language (outside of French and English) in Canada on salaries and wages. Inspired by other literature done on this topic in the United States, I seek to recreate this exercise in the Canadian context and provide evidence of a strong correlation between wages and knowledge of a foreign language. I also seek to investigate how this effect varies by occupation and language. Finally, I consider how this effect may vary by a supply effect of speakers and by religious affiliation. The main research question to pose is as follows: is there a salary and wage premium to those who speak a foreign language in Canada?

A literature review will first introduce the theoretical background and empirical findings on the topic of a foreign language effect in economics. This review will be followed by a brief explanation of my extension on this topic, involving a variable that identifies religious affiliation. Then, I will present my

methodology and the data used, borrowing many techniques and controls used by two key papers. For the actual empirical work, summary tables and regression results will be presented along with a discussion of the findings. Finally, concluding remarks will discuss any relevant implications of this paper.

Literature Review:

According to Grin, Sfreddo, & Vaillancourt (2010), much of the work concerning language and earnings originates from Canada, and particularly from Québec. The authors mention how, in this province, income inequalities between anglophones and francophones in the 1960s sparked this interest as well as "the search for the reasons behind them and for policies to remedy this state of affairs" (p. 58). Grin et al. (2010) also explain the history of integrating language into statistical studies through three options: treating language as an ethno-racial identity, using language as a form of human capital, and a combination of the previous two ideas. They recognize this last idea as standard practice to this day where a mother tongue indicates ethno-racial attributes and a secondary language indicates human capital attributes (see Vaillancourt, François (1980): *Differences in earnings by language groups in québec, 1970. an economic analysis.* Québec: Centre international de recherche sur le bilinguisme). It is also worth mentioning how the authors acknowledge the potential role of signalling for foreign languages but remain steadfast in their assumption that the actual use of such languages and a higher level of competency remains economically useful for productivity.

Of course, a spectrum exists between those who speak a foreign language as a mother tongue and others who pick it up as a second language. In the immigrant case, Grin et al. (2010) observe how "[t]he economic value of immigrants' skills in their native language… remained surprisingly under-researched" (p. 70), noting how mostly qualitative efforts have been made to address this question. Similarly, on the other spectrum, Saiz & Zoido (2005) note that little research has been done on observing earnings for a non-immigrant who learns a new language. They later proceed to investigate this case for college graduates in the US. Finally, speakers in the middle of the spectrum (i.e. immigrant children and heritage languages) are also considered where Kim (2013) and Shin & Alba (2009) investigate this case in the context of Asian- and Hispanic- Americans as 1.5-generation or US-born.

The theory behind the determination of a second language premium involves an investigation of labour market outcomes and their relevant demand and supply of speakers in the local economy. On the one hand, G. (2014, March) from *The Economist* states that the supply of foreign language speakers (particularly those who learn it) drives the results; due to competition with bilinguals, too many speakers of one language deteriorates a premium. Saiz &

Zoido (2005) seem to confirm this logic by finding a variation of language premiums (German at 3.8%, French at 2.3%, and Spanish at 1.5%), explained by the notion that there is more economic value to those who speak a less-spoken language. They also mention how states with larger Hispanic populations reduce a Spanish language premium.

On the other hand, the demand side is also to be considered. Bloom & Grenier (1992) acknowledge the supply phenomenon but also mention the number of consumers with the same language abilities playing a role as demand. Their theory states that growth in this linguistic community raises the demand side while potentially affecting relative wages by this group in the positive direction. Empirically put together, Bloom & Grenier (1992) find a decrease in an earnings gap between anglophone and francophone men in Canada due to an increase in demand for French speakers in the 1970s, and little changes in its associated supply. They also find that, in the US context, the opposite scenario occurred among Spanish- and English-speaking men where supply grew and demand did not seem to shift, resulting in substantial gaps between the two groups' earnings.

Although some authors have found a foreign language premium on wages, others have found less impressive results. Shin & Alba (2009) find that there is no evidence that bilingualism for Asian and Hispanic workers in the United States have higher wages than monolingual counterparts. Similarly, Coombs & Cebula (2010) observe a frail relationship between bilingual nurses and second-language skills that loses statistical significance as more controls are introduced in their study. On the contrasting end, Gándara's (2015) review on the literature of bilingualism uncovers an earnings penalty where bilingual workers may make less than monolingual workers in similar occupations. Indeed, the literature is mixed on whether foreign language skills have any impact on salaries and wages.

Extension of Two Papers:

This analysis will largely be based on two papers to synthesize a foreign language result in the Canadian context. The first paper, conducted by Saiz & Ziodo (2005), finds a 2% to 3% wage premium for a representative sample of US college graduates that have knowledge of a foreign language. Using interaction terms, they also find that certain languages that are known by less people often reward higher premiums (e.g. German's wage premium is 4.0% as opposed to Spanish's 1.7%). They also interact these language variables by occupation and find that knowledge of a foreign language is particularly useful for personal services, business services, and managerial positions. Finally, the authors assess the theory of supply by interacting the share of Hispanics in a state to knowing Spanish, finding a statistically significant diminishing effect where Spanish pays off less in larger Hispanic states.

The second paper by Christofides & Swidinsky (2010) observes the returns to knowing and using a second official language in Canada, using Census data from 2001. Their analysis restricts the sample to Canadian-born individuals between 15 to 64 years of age whose mother tongue is either French in Quebec, or English in the rest of Canada (ROC). Using a range of controls, which I intend to replicate in my own dataset, the authors find a statistically significant benefit to using English in Quebec while the results are insignificant for French in the ROC. They also make a distinction between knowledge of a language and use of a language at work, observing a higher language premium in the ROC for using French but without significant results. For Quebec, the result is statistically significant between knowledge and use.

This paper seeks to replicate the study done by Saiz & Zoido (2005), but in the Canadian context, and using controls like those used in Christofides & Swidinsky's (2010) paper. Although much research has been conducted on using an official language in Canada, there is little or no research done on the effects of a non-official language on wages in Canada. Furthermore, Saiz & Zoido (2005) only observe the knowledge of a foreign language on wages, whereas Christofides & Swidinsky (2010) distinguish between knowledge and usage. Given available data, this paper will incorporate the latter authors' techniques by observing the knowledge versus use effect of a foreign language in Canada. Indeed, there may be differences to untangle between a signalling effect and a human capital effect.

Furthermore, Saiz & Zoido's (2005) use of Hispanic shares in states only seem to give a proxy for the supply of Spanish speakers. This proxy may be an overestimation, where a portion of the Hispanic population may not know Spanish at all. Thus, I aim to use the same technique of interaction except by using actual shares of speakers with a mother tongue of certain languages.

Finally, an original extension that has not been incorporated by either paper is using a religion proxy as an additional interaction term on foreign languages. Arguably, individuals who attend the same religious institutions as their parents end up creating useful connections in the community that facilitate higher-paying jobs (e.g. Chinese-Canadians involved in the Christian community, Indo-Canadians involved in Sikh institutions, etc.). As a potential confounder, religious participation or the presence of certain religious groups may alter a foreign language effect. In Saiz & Zoido's (2005) case, the authors could have observed the share of Catholics in a state or simply interacted a Catholic variable with Spanish to see whether the results change. It is plausible that attending a religious institution or the share of religious institutions in a region (corresponding to a certain language) may drive the results of a language effect.

Data:

This paper will use the Public Use Microdata Files of the 2011 National Household Survey made available by Statistics Canada and obtained through Abacus. Concerning the characteristics of the dataset itself, the Individuals file has 887 012 observations that represent 2.7% of Canada's population. 102 variables are available in this dataset which includes relevant categories of income, language, religion, and all other control variables for demographics and occupation. The data also provides a sampling weight which is thoroughly used throughout the rest of this paper.

Econometric Methodology and Variables:

All the main regressions employed in this empirical exercise will conform to ordinary least squares (OLS) to investigate the effect of a foreign language on log wages and salaries, represented in equation [1]:

$$Log\ Y_i = \alpha + \beta X_i + \delta x_i Dx_i + \varphi Ix_i X_i + \varepsilon_i$$

[1]

where Y_i represents wages and salaries, α represents a constant, X_i represents a foreign language dummy variable, Dx_i represents demographic and occupational controls, Ix_i represents interaction terms for later regressions, and ε_i represents the standard error term.

Dependent Variable:

The dependent variable will use the wages and salaries variable, where gross income before deductions of tax, insurance, and pensions are calculated for individuals. Taking the log transformation of the variable excludes any observations that report a negative or zero income.

Independent Variables:

For language variables, the dataset offers information on the language first learned and still understood at home during childhood, the language used at home, the language used at work, and general ability in certain languages. The variable of having knowledge of a non-official language will be used along with the variable of using a non-official language regularly at work (as opposed to most often). A foreign language dummy is constructed that indicates whether an individual has knowledge of a non-official language or not. This variable will also be split into different languages and groups, which will be discussed later in more detail.

Controls:

Borrowing and taking inspiration from the controls used by Christofides & Swidinsky* (2010) (a star indicates inspiration or a borrowing), demographic and occupation controls will be constructed as follows (brackets indicate if all dummies equal 0):

Education (high school): *trades, pre-university certificate, bachelor's, graduate/post-bachelor's, and doctorate.*

Experience*: defined as *age - years of education - 6*. *Age* represents the midpoint age in a category (e.g. an age category of 25-29 would yield the age 27). *Years of education* is proxied from education, where *high school = 13, trades = 14, pre-university certificate = 15, bachelor's = 17, graduate/post-bachelor's = 19, and doctorate = 24*. Any negative or zero experience is labelled as missing for that observation.

Transformed experience*: defined as $experience^2/1000$.

Marital status (single): *married and divorced*.

CMA Resident (non-CMA)*: a dummy variable indicating an individual living in a Census Metropolitan Area by Statistics Canada. There are 23 defined areas and a single area devoted to non-CMA residents.

Region (British Columbia)*: *maritimes, quebec, ontario, prairies, and northern*.

Male (female): a dummy variable indicating whether the respondent is male.

White (all other groups): a dummy variable indicating whether the respondent identifies as white in a population group.

Occupation (trades/transport/equipment operator/related)*: *management, business/finance/administration, natural/applied sciences/related, health, education/law/social/community/government services, arts/culture/recreation/sport, sales/services, natural resources/agriculture/related, and manufacturing/utilities*.

Industry (wholesale/retail trade/transport/warehouse*)*: *primary/secondary/construction, financial, government services, education/health, information/cultural/accommodation, and professional/technical/administration/waste/other.*

Weeks worked in a year (1-9): *10-19, 20-29, 30-39, 40-48, and 49-52.*

Extension Variable:

Finally, a single variable of religion proxies for individuals who attend religious institutions. However, there is a limitation on only asking about affiliation and not participation.

Strategy:

Using the potential outcomes framework, this paper will define treatment D as either indicating knowledge of a foreign language or regular use of a foreign language at work (as opposed to most often). The effect received is defined as the average treatment effect which calculates the average for a population of individuals, but the effect itself differs from the more narrowly-defined average treatment effect on the treated. Simply comparing those who have language skills and those who do not is dangerous as there may be differences due to selection bias between treated and untreated groups.

Thus, to reduce this selection bias, several controls will be used to provide better comparisons between treated and untreated groups in terms of covariates. Once controls are put in place, potential outcomes no longer depend on D and differences between treated and untreated groups are minimized to mimic randomization. Furthermore, the sample weight provided in the dataset allows for even more accurate estimations on the Canadian population. However, it should be noted that there is always the possibility of omitted variable bias affecting the results (e.g. cognitive biases, effort, ability, and number of hours worked).

For the sake of simplicity and homogenous comparisons, the sample will always be restricted to individuals who were born in Canada, are between the ages of 25 and 64, and have indicated that they are working full time. More detailed assessments will observe a split sample between Quebec, with French used the most at home, and the rest of Canada, with English used the most at home. This ensures a more equal comparison for anglophones and francophones.

In later analyses, the foreign language variable will also be split up into the following dummies: Spanish, Chinese, Arabic, German/Related, Punjabi, Indo-Iranian/Dravidian, Russian/Slavic, Italian, Other European, Asian (East and Southeast), Tagalog, Other, and Multiple. Note that the Multiple variable contains responses of knowing more than one foreign language, which causes a potential downwards bias on explicitly-stated foreign language outcomes in the dataset.

Interaction terms will also be used between a foreign language and all occupation categories to observe whether a certain sector has a greater effect. Interactions between a share of speakers in a CMA and knowledge of a key language will also be assessed to determine a supply effect. Finally, certain

religions (Christianity, Sikhism, Islam, and Buddhism) will be interacted with a foreign language to confirm a potentially significant effect in two cases: as a stand-alone term and as population shares by CMA.

Summary Tables:

Table i: Summary Statistics Observing the Mean of Variables Between FL and non-FL Speakers:

Foreign Language	0	1		0	1
log_wages	10.64871	10.68452	experience	22.12668	18.96017
49-52 weeks worked	0.732691	0.737615	white	0.950296	0.799417
CMA dummy	0.634138	0.86354	high school	0.26185	0.181139
maritimes	0.092716	0.010315	trades	0.163717	0.091162
quebec	0.272804	0.255074	pre-bachelor's	0.312377	0.2912
ontario	0.33188	0.461863	bachelor's	0.181563	0.277729
prairies	0.189057	0.143266	grad/post-bach	0.074607	0.147248
northern	0.001547	0.002839	doctorate	0.005886	0.011521
bc	0.111996	0.126643	male	0.54153	0.505536
				206622 observations	

Log wages appear to remain quite similar, which is our main dependent variable of interest. A key difference is found in living in a CMA where those speaking a foreign language are more likely to reside there. A substantial amount of foreign language (FL) speakers appears to live in Ontario. There is also a much higher proportion of a white demographic in the non-FL group compared to the FL group. In addition, FL speakers tend to have less experience but better education outcomes than their counterparts. The gender ratio is also much more even for FL speakers than non-FL speakers.

Table ii: Summary Statistics Observing the Average Salary/Wage between FL and non-FL Speakers:

Total Sample						206622 observations		100%
FL		0	1		FL =	16819 observations		8.14%
Wages		62588.71	68858.38					
Excluding Quebec and English Most Spoken at Home						143808 observations		69.60%
FL		0	1		FL =	11418 observations		
Wages:		67267	73630.52			FL % of subsample:		7.94%
Only Quebec and French Most Spoken at Home						52151 observations		25.24%
FL		0	1		FL =	2444 observations		
Wages:		50497	54549.74			FL % of subsample:		4.69%

On average, those who can speak a foreign language earn more than their non-FL counterparts. This is consistent for both francophones in Quebec and anglophones in the rest of Canada.

Main Results and Analyses:

Table 1: OLS Regression Results for Knowledge of a Foreign Language:

Variables	**(1) All**	**(2) ROC**	**(3) Quebec**
	log_wages	**log_wages**	**log_wages**
Foreign Language	**-0.0284****	**-0.0191**	**-0.0284**
	(0.0108)	(0.0131)	(0.0266)
trades	0.0576***	0.0969***	-0.0114
	(0.00989)	(0.0134)	(0.0158)
before_uni_cert	0.167***	0.160***	0.181***
	(0.00771)	(0.00922)	(0.0154)
bachelors	0.446***	0.444***	0.438***
	(0.00919)	(0.0109)	(0.0190)
graduate_post_bach	0.556***	0.576***	0.499***
	(0.0126)	(0.0144)	(0.0287)
doctorate	0.705***	0.769***	0.527***
	(0.0379)	(0.0360)	(0.110)
experience	0.0436***	0.0447***	0.0412***
	(0.00114)	(0.00141)	(0.00215)
experience_transform	-0.796***	-0.821***	-0.753***

	(0.0264)	(0.0326)	(0.0497)
married	0.124***	0.131***	0.0981***
	(0.00695)	(0.00868)	(0.0127)
separated	0.0663***	0.0687***	0.0669**
	(0.0109)	(0.0131)	(0.0217)
CMA_dummy	0.112***	0.104***	0.135***
	(0.00595)	(0.00730)	(0.0109)
maritimes	-0.133***	-0.132***	0
	(0.0115)	(0.0121)	(.)
quebec	-0.161***	0	0
	(0.00920)	(.)	(.)
ontario	-0.0207*	-0.0191*	0
	(0.00904)	(0.00919)	(.)
prairies	0.0762***	0.0765***	0
	(0.0100)	(0.0101)	(.)
northern	0.295***	0.303***	0
	(0.0447)	(0.0470)	(.)
male	0.239***	0.254***	0.206***
	(0.00629)	(0.00763)	(0.0122)
white	0.0808***	0.0811***	0.0502
	(0.0122)	(0.0138)	(0.0346)
occ_dum1	0.203***	0.230***	0.123***
	(0.0130)	(0.0160)	(0.0231)
occ_dum2	0.00710	0.0322*	-0.0634***
	(0.0111)	(0.0140)	(0.0187)
occ_dum3	0.186***	0.211***	0.101***
	(0.0117)	(0.0147)	(0.0201)
occ_dum4	0.164***	0.233***	-0.0369
	(0.0156)	(0.0179)	(0.0340)
occ_dum5	0.0594***	0.0885***	-0.0403
	(0.0136)	(0.0167)	(0.0248)
occ_dum6	-0.233***	-0.212***	-0.287***
	(0.0277)	(0.0320)	(0.0564)
occ_dum7	-0.193***	-0.172***	-0.259***
	(0.0117)	(0.0147)	(0.0196)
occ_dum9	-0.114***	-0.0635*	-0.327***
	(0.0258)	(0.0296)	(0.0594)

occ_dum10	-0.0214	0.0162	-0.111***
	(0.0140)	(0.0171)	(0.0252)
ind_dumm1	0.117***	0.120***	0.124***
	(0.00997)	(0.0124)	(0.0171)
ind_dumm3	0.156***	0.151***	0.169***
	(0.0125)	(0.0154)	(0.0229)
ind_dumm4	0.214***	0.196***	0.247***
	(0.00924)	(0.0114)	(0.0174)
ind_dumm5	-0.0242*	-0.0326*	0.00658
	(0.0108)	(0.0131)	(0.0212)
ind_dumm6	-0.0766***	-0.0875***	-0.0334
	(0.0119)	(0.0139)	(0.0255)
ind_dumm7	-0.0718***	-0.0584***	-0.105***
	(0.00990)	(0.0118)	(0.0200)
10-19 Work	0.463***	0.458***	0.482***
	(0.0419)	(0.0506)	(0.0849)
20-29 Work	0.919***	0.968***	0.830***
	(0.0405)	(0.0491)	(0.0804)
30-39 Work	1.129***	1.177***	1.044***
	(0.0393)	(0.0472)	(0.0792)
40-48 Work	1.454***	1.498***	1.385***
	(0.0381)	(0.0457)	(0.0771)
49-52 Work	1.575***	1.621***	1.482***
	(0.0373)	(0.0448)	(0.0760)
_cons	8.127***	8.038***	8.216***
	(0.0421)	(0.0499)	(0.0901)
N	**206622**	**143808**	**52151**
R-squared	**0.1635**	**0.1643**	**0.1526**
Standard errors are in parentheses *** p<0.05, ** p<0.01, *** p<0.001**			

Legend for occupation dummies (Omitted = Trades/Transport/etc./Related)*: 1 = Management, 2 = Business/Finance/Admin, 3 = Natural/Applied Sciences, 4 = Health, 5 = Education/Law /Social/Communication/Government Services, 6 = Art/Culture/Recreation/Sport, 7 = Sales/Service, 9 = Natural Resources/Agriculture/Related, 10 = Manufacturing/Utilities.*

Legend for industry dummies (Omitted = Wholesale/Retail Trade/Transport/Warehouse):
1 = Primary/Secondary/Construction, 3 = Financial/Related, 4 = Government Services, 5 = Education/Health Services, 6 = Information/Cultural/Accommodation, 7 = Professional/Technical/Administration/ Waste/Other.

Note: *Work refers to the number of weeks worked in a year where 1-9 Work is omitted.*

Conducting the main regressions, the results are quite surprising: there is a statistically significant wage penalty for knowing a foreign language in Canada. For a Canadian-born, full-time worker, there is an associated 2.84% penalty on salaries and wages by having skills in a non-official language. Of course, there may be confounders for those who do not speak an official language at home, so subsamples are regressed for frequent English and French speakers. Analyzing the ROC (with English spoken the most at home) and Quebec (with French spoken the most at home), the results lose statistical significance but maintain negative signs.

It is possible that the effects are driven by ethnic factors, so another regression set is performed on those who identify with a white demographic. To ensure an even cleaner treatment effect, a third regression set is performed that restricts the sample to third-generation citizens (where both parents are Canadian-born) in the white demographic and excludes individuals who have majored in the Humanities. This third subsample removes the possibility of immigrant-parent biases, heritage language speakers, and a potential lower bias on incomes by humanities majors, particularly those who have majored in modern languages.

Table 2: White Only Regression Results:

White	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
FL	**-0.0265***	**-0.0180**	**-0.0201**
St. Errors	(0.0120)	(0.0148)	(0.0281)
*** p<0.05, ** p<0.01, *** p<0.001**			

Table 3: White + 3rd Generation/Beyond + No Humanities Regression Results:

White + 3rd Gen+ + No Hum.	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
FL	**-0.0549****	**-0.0738***	**-0.0554**
St. Errors	(0.0210)	(0.0292)	(0.0354)
*** p<0.05, ** p<0.01, *** p<0.001**			

Even when the possibility of ethnic, immigrant, and humanities confounders are removed, there is still a statistically significant wage penalty on knowing a foreign language. In fact, the effect becomes stronger and more significant as you move from the regression on the white demographic to the additional two restrictions in the third regression.

Of course, as mentioned in the literature review, there is a difference between knowing a language and using a language at work. Thus, OLS regressions are repeated, except the variable of interest is regularly using a foreign language at work (as opposed to using it the most).

Table 4: Regression Results on Using a Foreign Language at Work:

Main	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
FL Work	**-0.0623^**	**-0.0800***	**0.0249**
St. Errors	(0.0340)	(0.0351)	(0.0943)
*** p<0.05, ** p<0.01, *** p<0.001, and note: ^ refers to p<0.10**			

Table 5: Using a Foreign Language at Work Robustness Check:

White + 3rd Gen+ + No Hum.	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
FL Work	**-0.0635**	**-0.0638**	**0.0511**
	(0.0443)	(0.0444)	(0.0882)
*** p<0.05, ** p<0.01, *** p<0.001**			

The wage penalty grows substantially larger for individuals who use a foreign language at work, suggesting a 3.39% additional penalty. Even for English speakers who use these skills in the ROC, a significant penalty of 8% is expressed. Regression statistics with additional restrictions, as well as both Quebec regressions, are more inconclusive due to large standard errors. But it is unclear why there is a major markdown on skills that defies signalling or human capital theories of gaining a wage premium. Perhaps the results indicate that jobs in ethnic communities pay less than in mainstream society, yet it is still an ambiguous hypothesis, given negative coefficients in the white/3rd generation/non-humanities regressions.

Does the Effect Vary by Occupation?

Following the procedure of Saiz & Zoido (2005), I will include interaction terms between occupations and the relevant foreign language variables in separate OLS regressions. It is certainly plausible that a variation exists of the language effect by occupation.

Table 6: Interaction Terms by Occupation for Knowing a Foreign Language:

Main:	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
FL	**-0.171****	**-0.132^**	**-0.122**
	(0.0553)	(0.0676)	(0.0925)
fl_x_occ1	**0.207*****	**0.169***	**0.158**
	(0.0602)	(0.0734)	(0.115)
fl_x_occ2	**0.129***	**0.101**	**0.105**
	(0.0597)	(0.0733)	(0.113)
fl_x_occ3	**0.118^**	**0.0592**	**0.0605**
	(0.0614)	(0.0761)	(0.115)
fl_x_occ4	**0.100**	**0.111**	**-0.0598**
	(0.0668)	(0.0763)	(0.154)
fl_x_occ5	**0.170****	**0.142***	**0.121**
	(0.0590)	(0.0717)	(0.105)
fl_x_occ6	**0.199***	**0.185^**	**0.0629**
	(0.0851)	(0.102)	(0.170)
fl_x_occ7	**0.181****	**0.145^**	**0.113**
	(0.0602)	(0.0746)	(0.110)
fl_x_occ9	**0.0807**	**0.00479**	**0.163**
	(0.114)	(0.134)	(0.168)
fl_x_occ10	**0.153***	**0.134**	**0.0614**
	(0.0709)	(0.0864)	(0.131)
*** p<0.05, ** p<0.01, *** p<0.001, and note: ^ refers to p<0.10.**			

Table 7: Interaction Terms by Occupation for Using a Foreign Language:

Main:	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
FL Work	**-0.401^**	**-0.201**	**-0.633^**
	(0.208)	(0.172)	(0.357)
wk_x_occ1	**0.342**	**0.0589**	**0.896***
	(0.228)	(0.209)	(0.384)
wk_x_occ2	**0.404^**	**0.180**	**0.807***
	(0.214)	(0.182)	(0.371)
wk_x_occ3	**0.217**	**0.236**	**-1.303**
	(0.303)	(0.208)	(1.658)

wk_x_occ4	**0.373**	**0.213**	**0.408**
	(0.231)	(0.194)	(0.528)
wk_x_occ5	**0.379^**	**0.113**	**0.916***
	(0.213)	(0.177)	(0.386)
wk_x_occ6	**-0.172**	**-0.615**	**0.582**
	(0.401)	(0.505)	(0.561)
wk_x_occ7	**0.420**	**0.159**	**0.777***
	(0.216)	(0.184)	(0.374)
wk_x_occ9	**0.891**	**0.784***	**0.889***
	(0.298)	(0.327)	(0.400)
wk_x_occ10	**0.198**	**-0.0273**	**0.386**
	(0.240)	(0.225)	(0.374)
*** p<0.05, ** p<0.01, *** p<0.001, and note: ^ refers to p<0.10.**			

Legend for occupation dummies (Omitted = Trades/Transport/etc./Related)*: 1 = Management, 2 = Business/Finance/Administration, 3 = Natural/Applied Sciences, 4 = Health, 5 = Education/Law/Social/Communication/Government Services, 6 = Art/Culture/Recreation/Sport, 7 = Sales/Service, 9 = Natural Resources/Agriculture/Related, 10 = Manufacturing/Utilities.*

Managerial positions appear to have the greatest benefit when it comes to signalling a foreign language skill, overcoming the wage penalty and providing around a 3% wage boost. The results are more mixed when managers use their language skills (besides Quebec). Those working in the arts, as well as sales and services, also appear to overcome the wage penalty to a significant degree. Some of the more notable results are found in Quebec for the FL-work regressions, where several occupations appear to fare much better and are at a statistically significant level than the rest of Canada. It appears that Quebec rewards its employees significantly more for using languages other than French and English. Finally, it is worth noting that occupations in natural resources and agriculture also reward employees well for using their foreign language skills at work.

Dividing the Effect by Language Groups:

Given that the foreign language variable represents several foreign languages, I will once again follow the procedure of Saiz & Zoido (2005) by observing the individual effects of each language group. I have also included a regression that observes only the white demographic.

Table 8: Knowledge of a Foreign Language Effects by Group:

Main:	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
Spanish	**-0.00255**	**-0.0152**	**-0.0132**
	(0.0223)	(0.0345)	(0.0345)
Chinese	**0.0170**	**0.0567**	**0.0787**
	(0.0386)	(0.0413)	(0.108)
Arabic	**-0.177**	**-0.141**	**-0.300**
	(0.122)	(0.148)	(0.436)
German _other	**-0.0415**	**-0.0406**	**-0.128**
	(0.0254)	(0.0268)	(0.112)
Punjabi	**0.126*** **	**0.177*** **	**0**
	(0.0370)	(0.0426)	(.)
Ind_dr_ir	**-0.0616**	**-0.0905**	**-0.433^**
	(0.0656)	(0.0815)	(0.254)
Russ_slav	**-0.0350**	**-0.0131**	**-0.0258**
	(0.0340)	(0.0356)	(0.144)
Multiple	**-0.0950***	**-0.0829^**	**-0.0565**
	(0.0380)	(0.0448)	(0.0871)
Italian	**-0.00238**	**0.00125**	**0.0430**
	(0.0265)	(0.0347)	(0.0683)
Euro	**-0.0409**	**-0.00597**	**0.0551**
	(0.0271)	(0.0308)	(0.0799)
Asian	**-0.159***	**-0.157**	**-0.312**
	(0.0708)	(0.0823)	(0.201)
Tagalog	**0.00729**	**0.0121**	**0**
	(0.0652)	(0.0727)	(.)
Other	**-0.0952****	**-0.102***	**-0.180**
	(0.0334)	(0.0419)	(0.0928)

Table 9: Knowledge of a Foreign Language Effects by Group in the White Demographic:

Main:	**All: log_wages**	**ROC: log_wages**	**Quebec: log_wages**
Spanish	**0.000168**	**-0.0137**	**-0.0101**
	(0.0238)	(0.0369)	(0.0355)
Chinese	**-0.0307**	**-0.0694**	**0.148**
	(0.119)	(0.151)	(0.181)
Arabic	**-0.176**	**0.0394**	**0.221**
	(0.226)	(0.179)	(0.166)
German _other	**-0.0400**	**-0.0376**	**-0.128**
	(0.0254)	(0.0268)	(0.112)
Punjabi	**0.0980****	**0.0843*****	**0**
	(0.0337)	(0.0191)	(.)
Ind_dr_ir	**0.0979**	**0.185**	**-0.866*****
	(0.250)	(0.331)	(0.0332)
Russ_slav	**-0.0317**	**-0.00865**	**-0.0273**
	(0.0341)	(0.0356)	(0.144)
Multiple	**-0.111***	**-0.104^**	**-0.0421**
	(0.0447)	(0.0533)	(0.0972)
Italian	**-0.000740**	**0.00376**	**0.0417**
	(0.0266)	(0.0349)	(0.0688)
Euro	**-0.0405**	**-0.00505**	**0.0550**
	(0.0273)	(0.0311)	(0.0799)
Asian	**-0.0946**	**-0.107**	**-0.0699**
	(0.0607)	(0.0685)	(0.162)
Tagalog	**-0.323**	**-0.318**	**0**
	(0.329)	(0.330)	(.)
Other	**-0.0608**	**0.00350**	**-0.443***
	(0.0568)	(0.0578)	(0.214)
*** p<0.05, ** p<0.01, *** p<0.001, and note: ^ refers to p<0.10. Standard errors in brackets.**			

Note: *German_other = German and other Germanic Languages, Ind_dr_ir = Indo-Iranian/Dravidian Languages, Russ_slav = Russian/Ukrainian/Polish/Slavic Languages, Multiple = More than one FL (note that (all stated languages are single responses), Euro = Other European Languages, and Asian = East and Southeast Asian Languages.*

Language results are admittedly ambiguous, with most of the effects having large standard errors. Nonetheless, it is conclusive that the trend for several of these languages is to have negative coefficients, which supports the general result of a language wage penalty. Even when an individual speaks multiple languages, the wage penalty is actually intensified.

But there is one remarkable result to note: Punjabi has a statistically significant wage premium of 12.6% that increases to 17.7% when considering English speakers in the ROC. Even when restricting the regression to the white demographic, Punjabi still maintains a very strong result that defies the trend of several other languages. It may be that several Indo-Canadian businesses do remarkably well and compensate employees with higher salaries.

On the other extreme, white individuals in Quebec who speak an Indian or Iranian language have a staggering 86.6% wage penalty. Even with statistical significance, this result should be interpreted with great caution. It is plausible to assume that the individuals selected in the national survey happened to work in occupations that paid significantly less wages than the average salary. Emphasis should be placed on the fact that the results provided do not imply causation.

Investigating Mother Tongue Shares, Religion, and Religion Shares:

Finally, I seek to find whether the effect of a certain language varies by the share of its mother tongue speakers in a CMA, while utilizing my extension of introducing religion as interaction variables. The share of mother tongue speakers can arguably provide a better assessment of the language supply effect than what Saiz & Zoido (2005) has done by using the Hispanic demographic over Spanish speakers. Furthermore, I will test whether certain religions (proxying for religious participation and involvement) will influence a language's effect as well.

Empirically, a variable is constructed that calculates the percentage of a language's mother tongue speakers within a CMA. This variable is then interacted with the language in question to gain an interaction term useful for assessing the supply effect. For religion, two methods will be employed. The first will interact a relevant religion with a language to gain an interaction term. The second will use the same strategy as the mother tongue technique by constructing the shares of a religion's adherents by CMA and interacting the term with a language.

Given that there are only three CMAs strictly in Quebec, these empirical analyses will only analyze the effect of everyone and the ROC. In the dataset, 23 CMAs are recognized and an additional response indicates living outside of a CMA.

Punjabi and Sikhism:

Table 10: Results for the Punjabi Language and Sikhism:

Mother Tongue	All	ROC	Sikh	All	ROC
CMA Share	log_wages	log_wages		log_wages	log_wages
punjabi	0.244**	0.382***	punjabi	0.179*	0.216**
	(0.0839)	(0.104)		(0.0801)	(0.0827)
punj_otherfl	-0.0369***	-0.0328*	punj_otherfl	-0.0319**	-0.0242
	(0.0109)	(0.0133)		(0.0109)	(0.0134)
punj_share	0.0165***	0.0205***	sikh	0.0833	0.0922^
	(0.00217)	(0.00233)		(0.0640)	(0.0694)
cma punj share	-0.0281^	-0.0464*	punjd_sikh	-0.141	-0.136
	(0.0153)	(0.0191)		(0.109)	(0.116)

Sikh	All	ROC
CMA Share	log_wages	log_wages
punjabi	0.245**	0.377***
	(0.0816)	(0.100)
punj_otherfl	-0.0365***	-0.0324*
	(0.0109)	(0.0133)
cma_sikh	0.0156***	0.0196***
	(0.00211)	(0.00226)
cmasikh_punj	-0.0274^	-0.0439*
	(0.0143)	(0.0177)

Note: *The bottom coefficients for each column block refer to the interaction terms with Punjabi. Punj_otherfl refers to other languages. On the left, punj_share = share of individuals reporting a Punjabi mother tongue in a CMA. At the bottom column, cma_sikh = share of Sikhs in a CMA*

For Punjabi, there is a substantial supply effect where an increase in the number of Punjabi speakers in a CMA significantly reduces the language premium of speaking Punjabi in the ROC. As for Sikhism, there appears to be an insignificant effect of the religion on Punjabi, but given a higher negative coefficient, there is at least a suggestion that religious participation may confound Punjabi's wage effect. Interestingly, Sikhs by CMA provide similar results to the

Punjabi-by-CMA effect, but correlating the two interaction terms yields 99.94%. The Indian demographic may be driving the result.

Arabic and Islam:

Table 11: Results for the Arabic Language and Islam:

Mother Tongue	All	ROC	Muslim	All	ROC
CMA Share	log_wages	log_wages		log_wages	log_wages
arabic	-0.372	-0.279	arabic	-0.167	-0.230
	(0.258)	(0.265)		(0.189)	(0.293)
arabic_oth~l	-0.0319**	-0.0189	arabic_oth~l	-0.0280**	-0.0187
	(0.0108)	(0.0132)		(0.0108)	(0.0132)
arab_share	0.0246***	0.0292***	muslim	-0.0749	-0.0758
	(0.00273)	(0.00504)		(0.0656)	(0.0775)
cma_arab_m~h	0.0782	0.0710	musl_arab	0.0383	0.220
	(0.0846)	(0.0871)		(0.249)	(0.329)

Muslim	All	ROC
CMA Share	log_wages	log_wages
arabic	0.331	0.293
	(0.382)	(0.395)
arabic_oth~l	-0.0488***	-0.0420**
	(0.0107)	(0.0131)
musl_share	0.0247***	0.0270***
	(0.00148)	(0.00181)
cmamusl_arab	-0.100	-0.0908
	(0.0850)	(0.0960)

Note: *The bottom coefficients for each column block refer to the interaction terms with Arabic. Arabic_oth~l refers to other languages. On the left, arab_share = share of individuals reporting an Arabic mother tongue in a CMA. At the bottom column, musl_share = share of Muslims in a CMA.*

For Arabic, the results are much less impressive. The CMA interaction term concerning Arabic speakers indicates a positive effect on Arabic itself, but standard errors are too large to make any hard conclusions. Similar trends are found for the Muslim and Muslim-share interaction terms. It seems that the supply and religion effects have no explanatory power over Arabic's effect on wages.

Chinese, Christianity, and Buddhism:

Table 12: Results for the Chinese Language and Christianity:

Mother Tongue	All	ROC	Christian	All	ROC
CMA Share	log_wages	log_wages		log_wages	log_wages
chinese	0.121	0.136	chinese	-0.000731	0.0554
	(0.0743)	(0.0865)		(0.0493)	(0.0523)
chinese_fl	-0.0416***	-0.0403**	chinese_fl	-0.0340**	-0.0276*
	(0.0110)	(0.0136)		(0.0110)	(0.0136)
chin_share	0.0108***	0.0120***	christian	0.0194**	0.0156*
	(0.000914)	(0.000969)		(0.00642)	(0.00713)
cma_chin_m~h	-0.0142^	-0.0112	chris_chin	0.0735	0.0174
	(0.00748)	(0.00789)		(0.0693)	(0.0762)

Christian	All	ROC
CMA Share	log_wages	log_wages
chinese	-0.257	-0.159
	(0.221)	(0.257)
chinese_fl	-0.0378***	-0.0361**
	(0.0110)	(0.0135)
chris_share	-0.00386***	-0.00572***
	(0.000385)	(0.000501)
cmachris_c~n	0.00474	0.00367
	(0.00398)	(0.00489)

Note: *The bottom coefficients for each column block refer to the interaction terms with Chinese. Chinese_fl refers to other languages. On the left, chin_share = share of individuals reporting a Chinese mother tongue in a CMA. At the bottom column, chris_share = share of Christians in a CMA.*

Table 13: Results for the Chinese Language and Buddhism:

Buddhism	All	ROC	Buddhism	All	ROC
	log_wages	log_wages	CMA Share	log_wages	log_wages
chinese	0.0340	0.0633	chinese	0.165	0.162
	(0.0389)	(0.0420)		(0.0978)	(0.109)
chinese_fl	-0.0324**	-0.0257	chinese_fl	-0.0463***	-0.0427**
	(0.0110)	(0.0136)		(0.0110)	(0.0135)
buddhist	-0.00798	-0.0128	budd_share	0.0556***	0.0578***
	(0.0465)	(0.0561)		(0.00392)	(0.00427)
budd_chin	-0.178	-0.0545	cma_budd_c~n	-0.0746^	-0.0550
	(0.170)	(0.182)		(0.0394)	(0.0408)

Note: *The bottom coefficients refer to the interaction terms with Chinese. Chinese_fl refers to other languages. On the right, budd_share = share of Buddhists in a CMA.*

Chinese appears to have a statistically significant relationship with the share of Chinese speakers in a CMA, suggesting a negative supply effect at the 10% significance level. This significance disappears once these same variables are regressed for English speakers in the ROC. Christianity and the share of Christians in a CMA also appear to have little sway over the Chinese language effect. Large standard errors prevent any meaningful conclusions to be drawn.

However, there is more promise with using Buddhism as a religion variable. Although Buddhism by itself does not have any explanatory on a Chinese language effect, the share of Buddhists in a CMA becomes a significant variable at the 10% level, with a coefficient that is even greater than the Sikh-CMA share with Punjabi. This implies that the number of Buddhists negatively impacts a Chinese wage effect. However, like Punjabi, the interaction terms between Chinese speakers and Buddhists correlate by 98.96%. The results should be taken with caution, even though the Buddhist effect is much stronger.

Discussion on the Religion Extension:

Given the empirical exercises with Punjabi, Arabic, and Chinese, religion seems to have little impact on its associated language effect. It is very likely that

religion itself is not a strong enough proxy for religious participation as there may be discrepancies between stating your religious beliefs in a survey and actively participating in community worship. Thus, the hypothesis that religious participation is a potential confounder for a language effect is yet to be truly determined without proper statistics on how frequently an individual participates in religious activity, especially in a community setting.

On the other hand, there is more statistical significance in using the share of religious people in a CMA. As evidenced by Sikhism and Buddhism, a negative supply effect exists for Chinese or Punjabi speakers who live in areas with more Buddhists or Sikhs. But given their incredibly high correlations with mother tongue speakers, there are two suggestions on what drives the results: ethnicity and region. For Sikhs, it is argued that Sikhism and Punjabi speakers are driven by an ethnic Indian effect. For Buddhists, it is harder to make the same argument, given their multiethnic presence in Canada. Thus, the effect may very well be driven by CMA.

Given the limitation of geographic precision in the dataset, the smallest unit to utilize is the CMA. And for both Sikhs and Buddhists, there are two CMAs that arguably drive the results found above: Vancouver and Toronto. In fact, when removing these two CMAs from the regressions, the significance of the religion-by-CMA term disappears in both cases. This instance makes sense: many immigrants and ethnic groups choose to live in metropolitan areas, and so the significance of the results is derived from this case. Finer details on districts within these CMAs may benefit future research on this potentially confounding effect over foreign languages.

General Discussions:

The main results of knowing a foreign language in Canada are certainly surprising. There is an expectation that proficiency in a non-official language would signal to an employer that you are qualified for more pay. There is an equally plausible expectation that foreign language skills are a direct investment to human capital, which in turn contributes to higher wages and salaries. However, there is statistical certainty in several cases that knowledge and use of a foreign language yields a wage penalty.

A hypothesis is proposed on why that is the case: opportunity costs. The time an individual spends learning a foreign language is arguably less productive for gaining a higher salary than other professional activities. It may be that a university graduate faces a trade-off between investing full-time in language immersion or pursuing a professional diploma at a college. Or perhaps the trade-off is between spending evenings at home studying a language or spending those

evenings networking. To generalize, learning a foreign language may not be productive to gaining a higher salary in Canada.

This, in turn, is directly linked to a second hypothesis: lack of business demand. It is likely that many domestic firms in Canada do not have many interactions with foreign clientele or have little exposure to non-English or non-French speakers. Thus, there would be no demand for foreign languages in the workplace and no additional value placed on these skills. This may explain why learning a foreign language may be unproductive for job prospects.

Nonetheless, there are notable exceptions. Occupations such as management, sales, or natural resources offer a higher salary, depending on the region and whether language skills are used. Quebec tends to reward employees more for using foreign languages. Remarkably, Punjabi stands out as a very lucrative language to have, even when eliminating a potential ethnic confounder. The comprehensive nature of this paper shows that the mean results should never be taken as an absolute, given the level of variation that may exist from within.

Conclusion:

This paper attempted to establish an empirical argument that knowledge of a foreign language in Canada is a legitimate treatment effect on income. By using controls to remove potential selection bias, and by showing that covariates are sufficiently balanced, main OLS regressions reveal a statistically significant wage penalty in Canada. These regressions were further decomposed into anglophones in the ROC and francophones in Quebec to ensure more homogenous comparisons. This effort was further pursued by excluding humanities majors and restricting the sample to white, 3rd-generation and above Canadians in a separate regression. Negative coefficients still persisted for log wages and salaries. The same conclusions are drawn with a stronger effect for Canadians who use a foreign language at work.

Occupations were also interacted with foreign language skills to reveal a nuanced picture of a general wage penalty (or premium), which varies by occupation, region, and usage. In addition, the foreign language variable was decomposed into various language groups, revealing less conclusive results but offering a few notable exceptions. Finally, the shares of mother-tongue speakers, religion, and the share of certain religious groups were included with mixed results. These results somewhat confirm the supply effect of speakers in a few languages but also a supply effect of certain religious groups, most likely confounded by ethnic or regional biases. Religion remains to be a weak proxy for religious participation, and does not hold any significance on its own over a foreign language effect.

Works Cited:

Bloom, David E. & Grenier, Gilles. (1992). Earnings of the french minority in canada and the spanish minority in the united states. In Chiswick, B. R. (Ed.), *Immigration, language, and ethnicity: canada and the united states* (pp. 373-409). Washington, D.C: AEI Press.

Christofides, L. N., & Swidinsky, R. (2010). The economic returns to the knowlege and use of a second official language: English in quebec and french in the rest-of-canada. Canadian Public Policy, 36(2), 137.

Coombs, C. K., & Cebula, R. J. (2010). Are there rewards for language skills? evidence from the earnings of registered nurses. The Social Science Journal, 47(3), 659-677.

G., R.L. (2014, March). Johnson: what is a foreign language worth? *The Economist*,. Retrieved from https://www.economist.com/blogs/prospero/2014/03/language-study.

Gándara, P. (2015). Is there really a labor market advantage to being bilingual in the U.S.? ETS Research Report Series, 2015(2), 1-34.

Grin, F., Sfreddo, C., & Vaillancourt, F. (2010). *The economics of the multilingual workplace*. New York: Routledge.

Kim, C. (2013). The economic value of bilingualism among 1.5- and second-generation korean americans. *Development and Society, 42*(1), 89-112.

Saiz, Albert, & Zoido, Elena (2005). "Listening to what the world says: bilingualism and earnings in the united states." *The Review of Economics and Statistics*, vol. 87, no. 3, 2005, pp. 523-538.

Shin, H., & Alba, R. (2009). The economic value of bilingualism for asians and hispanics. *Sociological Forum, 24*(2), 254-275.

Vaillancourt, François. (1980). *Differences in earnings by language groups in québec, 1970. an economic analysis.* Québec: Centre international de recherche sur le bilinguisme.