

Role of Developing Home Country in Female Immigrants' Probability of Being Self-Employed: Evidence from Canada

Atakan Bakiskan 100392774 Min Ho Kim 1003056266

March 19th 2020

Introduction

In this paper, we will be estimating the effect of being a female immigrant born in a developing country on their probability of being self-employed in Canada.¹ Specifically, the comparison will be made between female immigrants from developing countries and female immigrants from developed countries. We investigate the role of home country in a female immigrant's self-employment decision in Canada by classifying self-employment into two categories: incorporated and unincorporated. We conjecture that a self-employed individual's choice to incorporate their business or not will have different implications on their economic assimilation in Canada. By analyzing the differences in regressing self-employment, incorporated and unincorporated self-employment, we suggest new ways for testing the arguments against the *home country self-employment hypothesis*.

Background

In his recent study, Fields claimed that individuals in developing countries choose to self-employ as a mean to make a modest living out of it (Fields 2019, 4). His study shows that developing nations have a higher rate of self-employment compared to developed nations (Fields 2019, 3). We hypothesize that self-employment is a mechanism that allow immigrants from developing countries to enter the mainstream Canadian economy. Our focus is on female immigrants from developing countries, who may face socio-economic barriers in the labour market and be forced to enter self-employment, according to the blocked mobility thesis (Li 1997, 116).

A crucial part of our study is in differentiating the two types of self-employment. Besides from separating the personal assets from the business, incorporating a business can raise large amounts of capital easier than a sole proprietorship. However, incorporation requires time, large amount of legal paperwork, formation and filing fees. Thus, in the short-term, forming an unincorporated business may be more cost-efficient than incorporating their business for immigrants. Given the opportunity costs and tedious legal work, we assume that immigrants' familiarity with Canada's economic institutions is positively correlated with their decision to incorporate their businesses.

1. We use the pooled Canadian census data for 2011 and 2006, and our categorization follows the IMF's 2011 World Economic Outlook database information in their categorization of advanced, and emerging and developing economies.

Theory

Our model contributes to the labour economics literature by proposing a new way to test the arguments against the home-country self-employment hypothesis. The hypothesis claims that self-employment rate is higher among immigrants from countries with larger self-employed sectors and results from Yuengert (1995, 195) comply with this hypothesis. However, other studies such as Hammarstedt and Shukur's (2009, 745), which was conducted using a sample from Sweden, found no support for the hypothesis. Belton and Uwaifo Oyelere (2012, 538) argued that it is the similarity of economic institutions between nations that affects one's choice to enter self-employment by showing that immigrants from developed countries had higher self-employment rates in the United States than those from developing countries. We test the latter counterargument by studying incorporated and unincorporated self-employment separately. Under the assumption that economic institutions are similar between developed countries and Canada, one would expect lower rate of incorporated self-employment among immigrants from developing countries.

Methodology and Results

Below, we present a few summary statistics on the sample.

selfemp	coh2		Total
	0	1	
0	183,576	67,890	251,466
1	217,783	67,218	285,001
Total	401,359	135,108	536,467

Figure 1: Cross Tabulation of Self-Employment and Immigrants from Developing Countries

selfemp	female		Total
	0	1	
0	128,172	127,164	255,336
1	150,909	134,650	285,559
Total	279,081	261,814	540,895

Figure 2: Cross Tabulation of Self-Employment and Females

se_inc	se_uninc		Total
	0	1	
0	255,336	36,548	291,884
1	249,011	0	249,011
Total	504,347	36,548	540,895

Figure 3: Distribution of the Two Types of Self-Employment.

Inspiring from Li (2001, 1121), we will use a logistic regression model:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_m X_{mi}$$

$$\implies p_i = \frac{e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_m X_{mi}}}{1 + e^{\beta_0 + \beta_1 X_{1i} + \dots + \beta_m X_{mi}}}$$

where p_i is the probability of individual i being self-employed (incorporated, unincorporated, or both). We analyze the differences in estimates and their outcomes on the respective probabilities between the three different self-employment cases. Because our main analysis is on the comparison between female immigrants from developing countries and female immigrants from developed countries, the focus will be on the significance and the coefficient estimates of these variables. Given an abbreviated model:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 female_i + \beta_2 coh2_i + \beta_3 (female_i \times coh2_i)$$

the difference between the two groups would be: ²

$$\beta_0 + \beta_1 + \beta_2 + \beta_3 - (\beta_0 + \beta_1) = \beta_2 + \beta_3 \quad (1)$$

We have included variables such as education level, age, sex, household size, years since migration as controls. In addition, we have an indicator variable that signals whether an individual is from a developed country or a developing country (*coh2*). We also utilize investment income as a proxy for economic uncertainty in our regression to avoid the "push" hypothesis of self-employment where it is claimed that employees are pushed to become self-employed due to worsening economic conditions (Dawson, Henley and Latreille 2009, 5). We also include *efdecile*, variable indicating individual's income decile, to control for the socioeconomic opportunities coming from one's economic status. One can make arguments for both positive and negative relationship between education and self-employment (Dawson, Henley and Latreille 2009, 19). In our case, we have found that across all levels of education, their impact on the probability of self-employment is positive, and the effect increases with the level of education.

First, we estimate the probability of being self-employed without differentiating incorporated and unincorporated businesses (Figure 4). The coefficient on *age2* and *age3* are 0.6560277 and 0.991659 respectively, which is consistent with Li's finding (2001, 1122) that older immigrants have higher odds of self-employment. Furthermore, we can see that the estimates for the variables *coh2*, *female* and their interaction are all significant and negative meaning that being a female and being an immigrant from a developing country are complementary. Ceteris paribus, the odds of being self-employed for a female immigrant from developing country is much lower than the odds of a female immigrants from a developed country.

Now, we run two new regressions, this time differentiating the two different types of self-employment. For incorporated self-employment, the interaction term is no longer significant at $\alpha = 0.05$ (Figure 5), while the

2. In calculating this difference, we use the estimated marginal effects at the average, using **margins, dydx(.) atmeans** command on STATA.

interaction term in the other regression remains significant (Figure 6). The importance lies on the difference in the log odds between female immigrants from developing countries and female immigrants from developed countries, which is calculated using Equation (1). The logistic regression of incorporated self-employment calculated this difference to be -0.005 whereas unincorporated self-employment calculated to be -0.03. On average, female immigrants are less likely to be self-employed, and the odds get worse for unincorporated businesses which contradicts with the Belton and Uwaifo Oyeler's (2012, 538) argument against the home country self-employment hypothesis.

Conclusion

The main finding of this paper tells us that female immigrants from developing countries are 3 percent less likely to be unincorporated self-employed than female immigrants outside developing countries, at the average of the covariates. The likelihood, in the same direction, drops down to 0.5 percent in the case of incorporated self-employment. This contradicts with the Uwaifo Oyelere and Belton's (2012, 538) similarity of economic institutions counterargument against the home country self-employment hypothesis. Female immigrants from developing countries, despite the low similarity in the economic institutions of their home country and Canada, are more likely to enter the Canadian economy through incorporated self-employment than unincorporated self-employment. This finding is crucial for government programs aiming to integrate female immigrants from developing countries into the economy. Our analysis on female immigrants leaves room for future research where the model can focus on male immigrants born in a developing country. The results may be different given there exists empirical regularities such as fertility which results in gender differences in self-employment selection processes (Boden 1996, 671-682).

Logistic regression	Number of obs	=	224,083
	Wald chi2(25)	=	8743.60
	Prob > chi2	=	0.0000
Log pseudolikelihood = -68982.843	Pseudo R2	=	0.0783

selfemp	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
educ1	.0548571	.0273235	2.01	0.045	.001304	.1084102
educ2	.1248553	.0257009	4.86	0.000	.0744825	.1752281
educ3	.1965603	.0294597	6.67	0.000	.1388203	.2543003
educ4	.3553345	.03116	11.40	0.000	.294262	.416407
age2	.6560277	.0228107	28.76	0.000	.6113195	.7007359
age3	.991659	.0235816	42.05	0.000	.94544	1.037878
mar	.6715219	.0268634	25.00	0.000	.6188707	.7241731
ysm	.0099696	.0005768	17.29	0.000	.0088391	.0111
hhsze	.0165147	.0074717	2.21	0.027	.0018704	.031159
coh2	-.3774961	.024185	-15.61	0.000	-.4248979	-.3300943
female	-.5325771	.0182684	-29.15	0.000	-.5683824	-.4967717
int1	-.1337032	.0339543	-3.94	0.000	-.2002523	-.067154
pkid	.0128803	.0208827	0.62	0.537	-.028049	.0538097
hloff	-.0153167	.022251	-0.69	0.491	-.0589279	.0282945
prihm	-.0397358	.0163065	-2.44	0.015	-.0716959	-.0077757
efdecile						
Second decile	.1173365	.0342766	3.42	0.001	.0501556	.1845174
Third decile	-.2853654	.0337374	-8.46	0.000	-.3514895	-.2192413
Fourth decile	-.5279742	.033827	-15.61	0.000	-.5942739	-.4616744
Fifth decile	-.7790549	.0342642	-22.74	0.000	-.8462116	-.7118983
Sixth decile	-.9080488	.0341194	-26.61	0.000	-.9749217	-.8411759
Seventh decile	-1.103423	.0345469	-31.94	0.000	-1.171133	-1.035712
Eighth decile	-1.190987	.034026	-35.00	0.000	-1.257677	-1.124297
Ninth decile	-1.276394	.0339275	-37.62	0.000	-1.342891	-1.209898
Highest decile	-1.065869	.0329684	-32.33	0.000	-1.130486	-1.001252
invst	.000018	1.01e-06	17.93	0.000	.0000161	.00002
_cons	-2.513958	.0503824	-49.90	0.000	-2.612706	-2.41521

Figure 4: Regression Output on Self-Employment

Logistic regression	Number of obs	=	224,083
	Wald chi2(25)	=	4544.96
	Prob > chi2	=	0.0000
Log pseudolikelihood = -34841.181	Pseudo R2	=	0.0897

se_inc	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
educ1	.0835349	.0427405	1.95	0.051	-.000235	.1673048
educ2	.0473284	.040495	1.17	0.243	-.0320404	.1266972
educ3	.1337911	.0458201	2.92	0.004	.0439853	.2235969
educ4	.253948	.0477317	5.32	0.000	.1603954	.3475005
age2	.7448844	.0391444	19.03	0.000	.6681628	.8216059
age3	.9583351	.0402406	23.82	0.000	.879465	1.037205
mar	1.01059	.0520539	19.41	0.000	.9085662	1.112614
ysm	.0100915	.0008427	11.98	0.000	.0084399	.0117432
hhsze	.0179135	.0112752	1.59	0.112	-.0041855	.0400125
coh2	-.2590272	.0351165	-7.38	0.000	-.3278542	-.1902003
female	-.9146733	.030895	-29.61	0.000	-.9752263	-.8541204
int1	.0784293	.0544451	1.44	0.150	-.0282812	.1851397
pkid	.0846714	.0313834	2.70	0.007	.0231612	.1461817
hloff	-.0513336	.0336705	-1.52	0.127	-.1173266	.0146594
prihm	.0807803	.0260463	3.10	0.002	.0297305	.1318301
efdecile						
Second decile	.1761014	.0580183	3.04	0.002	.0623876	.2898152
Third decile	-.1547771	.0575797	-2.69	0.007	-.2676313	-.0419229
Fourth decile	-.3093543	.0574283	-5.39	0.000	-.4219118	-.1967968
Fifth decile	-.4873039	.0578119	-8.43	0.000	-.6006132	-.3739946
Sixth decile	-.5219962	.0568218	-9.19	0.000	-.6333649	-.4106275
Seventh decile	-.6016296	.0564158	-10.66	0.000	-.7122026	-.4910566
Eighth decile	-.5630447	.0545705	-10.32	0.000	-.6700008	-.4560885
Ninth decile	-.5505361	.0537083	-10.25	0.000	-.6558025	-.4452697
Highest decile	-.2645051	.0520675	-5.08	0.000	-.3665555	-.1624547
invst	.000017	9.46e-07	17.92	0.000	.0000151	.0000188
_cons	-4.309726	.0838666	-51.39	0.000	-4.474102	-4.145351

Figure 5: Regression Output on Self-Employment, Incorporated

Logistic regression	Number of obs	=	224,083
	Wald chi2(25)	=	5495.07
	Prob > chi2	=	0.0000
Log pseudolikelihood = -49853.699	Pseudo R2	=	0.0559

se_uninc	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
educ1	.0282354	.0334306	0.84	0.398	-.0372874	.0937581
educ2	.1541302	.0312465	4.93	0.000	.0928882	.2153723
educ3	.2119719	.0362251	5.85	0.000	.140972	.2829717
educ4	.373624	.038343	9.74	0.000	.2984731	.4487749
age2	.5659945	.027352	20.69	0.000	.5123856	.6196035
age3	.9413821	.0281798	33.41	0.000	.8861506	.9966135
mar	.4789765	.0311335	15.38	0.000	.4179559	.539997
ysm	.0080286	.0007182	11.18	0.000	.0066211	.0094362
hhsze	.0136334	.0092978	1.47	0.143	-.0045901	.0318568
coh2	-.4072101	.0305182	-13.34	0.000	-.4670246	-.3473956
female	-.261998	.0220205	-11.90	0.000	-.3051573	-.2188387
int1	-.2191553	.041831	-5.24	0.000	-.3011425	-.1371681
pkid	-.0330481	.0260397	-1.27	0.204	-.084085	.0179887
hloff	.0059665	.0275218	0.22	0.828	-.0479753	.0599082
prihm	-.1254876	.0199317	-6.30	0.000	-.164553	-.0864222
efdecile						
Second decile	.0801489	.0392115	2.04	0.041	.0032958	.157002
Third decile	-.2981724	.038766	-7.69	0.000	-.3741524	-.2221925
Fourth decile	-.5567536	.039157	-14.22	0.000	-.6334999	-.4800073
Fifth decile	-.8190724	.0399176	-20.52	0.000	-.8973094	-.7408354
Sixth decile	-.9885362	.040155	-24.62	0.000	-1.067239	-.9098338
Seventh decile	-1.235361	.0413557	-29.87	0.000	-1.316417	-1.154306
Eighth decile	-1.400105	.041459	-33.77	0.000	-1.481364	-1.318847
Ninth decile	-1.550399	.0419248	-36.98	0.000	-1.63257	-1.468228
Highest decile	-1.393505	.0396164	-35.17	0.000	-1.471151	-1.315858
invst	4.64e-06	4.01e-07	11.58	0.000	3.86e-06	5.43e-06
_cons	-2.709336	.0605027	-44.78	0.000	-2.827919	-2.590753

Figure 6: Regression Output on Self-Employment, Unincorporated

Data Appendix

In our paper, we have used the pooled 2006 and 2011 Canadian census data - *census0611_new.dta*. From the data, we have a total of 12 different variables, with some being manipulated into categorical variables. Below is the list of the original variables extracted from the data set, as well as new variables that were constructed for the purpose of our study.

1. ***agegrp*** - Variable consisting of nine different levels of age group, from 20-24 years old to 60-64 years old. We have separated this variable into three different dummy variables: ***age1***: 20 to 34 years old, ***age2***: 35 to 49 years old, and ***age3***: 50 to 64 years old. These are indicator variables and is equal to 1 if an individual belongs to that age group, or 0 otherwise.
2. ***coh2*** - Variable indicating whether an individual is from a developing country (1), or a developed country (0). This was done through looking at the variable ***pob***, which indicates individual's place of birth. We consequently categorized countries following the IMF's 2011 World Economic Outlook database information in their categorization of advanced, and emerging and developing economies.
3. ***cow*** - Variable containing information on the class of worker. The original variable has six different levels, with four of those belonging under self-employment. We took this information and defined new variables ***se_inc*** and ***se_uninc*** as indicators, taking on value of 1 or 0, for self-employment with incorporated business and self-employment with unincorporated business, respectively. We also have the variable ***selfemp*** that does not differentiate the two types of unemployment.
4. ***efdecile*** - Variable on the income decile group providing a rough ranking of the economic situation of a person based on his or her relative position in the economic families adjusted after-tax income distribution.
5. ***educ*** - A variable containing information on individual's highest education level. Originally, the variable contains five different levels, which we have reconstructed into five distinct dummy variables – 1 if variable corresponds to the individual's highest education level, 0 otherwise:
 - (a) ***educ0***: No degree
 - (b) ***educ1***: High school or equivalent
 - (c) ***educ2***: Some university, college, trade certificate, registered apprenticeship
 - (d) ***educ3***: University Bachelor's degree
 - (e) ***educ4***: University: above bachelors (graduate, diploma, medicine)
6. ***female*** - Dummy variable indicating whether the individual is female (1) or not (0).
7. ***hhsze*** - Variable containing information on an individual's household size. Minimum number is one.
8. ***hloff*** - Variable indicating whether the most spoken language at home is either English or French (1), or otherwise (0). The information was derived by merging pre-existing variables *hlaen* and *hlafr*.

9. *int1* - Variable serving as the interaction term between *female* and *coh2*. Takes on a value of either 1 or 0.
10. *invst* - Continuous variable referring to individual's interest received in previous calendar year from deposits in banks, trust companies, co-operatives, credit unions, caisses populaires, etc.
11. *marst* - Variable containing information on individual's marital status. Originally included five different levels – divorced, never legally married, widowed, separated, legally married, and living common law. We merged legally married and living common law, and created a variable *mar*, a dummy variable indicating whether an individual is either legally married or living common law (1), or currently alone (0).
12. *pkids* - Variable indicating whether an individual has one or more kids (1) or none (0). This was derived using previously existing *pkid** variables.
13. *prihm* - Dummy variable identifying the first (primary) household maintainer.
14. *ysm* - Variable indicating an individual's years in Canada since migration. Those who were born in Canada are assigned value of 0.

References

Boden, Richard. "Gender and self-employment selection: An empirical assessment." *Journal of Behavioral and Experimental Economics* 25, no. 6 (1996): 671–682. doi:[https://doi.org/10.1016/S1053-5357\(96\)90046-3](https://doi.org/10.1016/S1053-5357(96)90046-3).

This paper proposes that there exists empirical regularities which causes gender differences in self-employment. Therefore, we use this paper in order to support our argument that future research can focus on male immigrants born in developing country.

Canada, Government of. "Setting up your business." <https://www.canada.ca/en/revenue-agency/services/tax/businesses/small-businesses-self-employed-income/setting-your-business.html>.

This is where we learn incorporation versus not incorporation, existence of formation fees and etc. Given that incorporation is a matter of time and relying on the legal documents necessary to incorporate businesses provided in this website, we shape our argument against the home-country self-employment hypothesis.

Dawson, Christopher J., Andrew Henley, and Paul L. Latreille. "Why Do Individuals Choose Self-Employment?" *IZA Discussion Paper*, no. 3974 (2009).

We use this paper while discussing the "push" and "pull" hypothesis and it guided as to control for the variables *edeficile* and *invst* as a proxy for worsening economics conditions. Also, we use it as an example of the currently-mixed literature of age on self-employment.

Fields, Gary. "Self-employment and poverty in developing countries." *IZA World of Labor* 60 (2019). doi:<https://doi.org/10.2307/146196>.

We used this paper while providing a background on self-employment in developing countries and showing that developing countries have higher self-employment rate than developed countries. Also, we have used the paper in our discussion on self-employment and entrepreneurship.

Hammarstedt, Matz, and Ghazi Shukur. "Testing the home-country self-employment hypothesis on immigrants in Sweden." *Applied Economic Letters* 16, no. 7 (2009): 745–748. doi:<https://doi.org/10.1080/13504850701221907>.

By testing the home country self-employment hypothesis in Sweden, this paper finds no evidence using traditional estimation methods. We cite this paper to show that home country self-employment hypothesis could vary across countries and may not even be true in general.

Li, Peter S. “Immigrants’ Propensity to Self-Employment: Evidence from Canada.” *The International Migration Review* 35, no. 4 (2001). doi:<https://doi.org/10.1111/j.1747-7379.2001.tb00054.x>.

We use this paper while referring to the blocked mobility thesis. It shapes the audience’s understanding in self-employment, shifting it from entrepreneurship towards an approach where immigrants may be forced to be self-employed due to discrimination in open market.

———. “Self-employment among visible minority immigrants, white immigrants, and native-born persons in secondary and tertiary industries of Canada.” *Canadian Journal of Regional Science* 20, nos. 1-2 (1997): 103–117.

We use this paper while referring to the blocked mobility thesis. It shapes the audience’s understanding in self-employment, shifting it from entrepreneurship towards an approach where immigrants may be forced to be self-employed due to discrimination in open market.

Oyelere, Ruth Uwaifo, and Willie Belton. “Coming to America: Does Having a Developed Home Country Matter for Self-Employment in the United States?” *American Economic Review* 102, no. 3 (2019): 538–542. doi:<https://doi.org/10.2307/146196>.

This is the paper our critique is based on, Oyelere and Belton’s argument against the home country self-employment hypothesis and their approach to self-employment gave the structure to our paper (such as running two separate regressions, incorporated and unincorporated). Also, our classification of countries into developing and developed countries following the IMF’s classification was inspired from this paper.

Yuengert, Andrew M. “Testing Hypotheses of Immigrant Self-Employment.” *The Journal of Human Resources* 30, no. 1 (1995): 194–204. doi:<https://doi.org/10.2307/146196>.

Yuengert’s paper, where we came up with our research question, shows evidence in support of “home-country self employment hypothesis. Yuengert defines the hypothesis, formulates the earnings differential between immigrant’s in terms of human capital theory. We take Yuengert’s paper as a benchmark when referring to “home country self-employment hypothesis”.