Exploring the Relationship Between Income and Life Satisfaction: A Regression Analysis

By MAX GEYER, JIAYIN KRALIK AND HANZE WANG

There has been a long debate about whether money makes us happy amongst economists, psychologists, and philosophers. This paper attempts to answer the question of whether money makes us happy through empirical analysis, using data from the Canadian Community Health Survey, 2017~2018. Through a linear regression model, we measure the marginal effect of income on happiness while controlling for health, marital status, and education, among other variables. We find that income has a positive and statistically significant effect on happiness: a 1% increase in income above the low-income line results in a 0.002 unit increase in satisfaction. Our results solidify the interdisciplinary consensus that money does buy happiness, but more research needs to be done in regression analysis to re-evaluate variables that may explain larger proportions of satisfaction.

I. Introduction

Possibly the most habitual question in the social sciences, does money buy happiness. Money is exchangeable for goods and services that people can generate utility from consuming; but does happiness rather stem from exogenous factors without a price tag? Our paper aims to find out the proper relation between income and the satisfaction of people in Canada, using data gathered from the Canadian Community Health Survey, 2017-2018.

> "Throughout history, philosophers justified happiness to be the highest good and motivation for human action" (Diener, 1984)

As far as we can reason, happiness is the best and most accurate measure of human action. According to neoclassical economic theory, more money will always make consumers

happier. Today, research in economics, psychology, and philosophy has found that as money increases, perceived happiness with life in general will increase (Psychology Today, 2021).

Though a well-researched question, there are issues with the question like "personality bias" that argues people who are "extraverted and resilient are more likely to be happy with life" (Powdthavee, 2010). Our paper does not explore this specific bias, though it is reasonable to keep in mind. Despite this, studies like Howell and Howell 2008 show us that the correlation between income and happiness, though positive, is often measured as relatively small despite neoclassical economic theory. We hypothesize that other potential factors contributing to satisfaction besides income are general health, age, family life - measured through marital status, and education - measured through degree acquisition at a university.

In this regression study, we research how income contributes to satisfaction relative to other explanatory factors. In this paper, we will have a primary interest in the coefficient on our variable of income (I). After using linear multi regression, we found when explanatory factors other than income are held constant; that a 1% increase in income above the low-income (as determined by the Canadian government) line results in a 0.002 unit increase in satisfaction. Thus, from our study, we have confirmed our research question and the studies of peers to find that it is likely that more income will result in higher satisfaction for Canadians.

II. Data Description

We are using the data from Canadian Community Health Survey (CCHS) from 2017-2018. CCHS is a survey that collects information related to health status, health care

utilization, and health determinants for Canadians. CCHS data was collected through telephone interviews (CATI) and personal interviews (CAPI) (Statistics Canada, 2019). The CCHS data is a panel data set, it consists of multiple observations over time, with a total of 113,290 entries. Furthermore, it is an observational dataset. Data was collected by sending questionnaires to participants without adding controlling factors. According to the user guide, the CCHS gathers health-related information from individuals aged 12 and above residing in Canada. However, the survey does not include people living on Aboriginal settlements, full-time members of the Canadian Forces, foster youth aged 12 to 17, institutionalized individuals, and those living in certain health regions in Quebec. Although these exclusions account for less than 3% of the total population that the survey aims to represent, it is important to take into account (Statistics Canada, 2019). Consequently, the research results based on this dataset may not accurately reflect the relationship between happiness and income for children under the age of 12 or other excluded groups.

The CCHS contains theme content and optional content, our variables before manipulation are from the 2-year theme content. Specifically, The data is sampled using a three-step process, with a primary objective of producing reliable estimates for each Health Region (HR). Another goal of this data is ensuring the quality of key characteristics at the provincial level. Distinct sampling frames were employed to select individuals from different age groups, the adult population was chosen using an area frame, while the youth sample was selected through a list frame (Statistics Canada, 2019). The most important variables are general satisfaction (GEN_010) and income (incdghh) because we are going to analyze the effect of income on happiness. These variables are used as dependent and independent variables respectively in our model. Other control variables we chose to use are age

(dhhgage), health (GEN_005), University (a dummy showing if an individual attended university, derived from EHG2DVR3), Married_or_Common_Law (a dummy showing if an individual is married, derived from dhhgms). We also used GEN_015(mental_health) as another indicator of happiness in robustness analysis.

GEN_010 is a quantitative measure of happiness, with discrete values ranging from 1 to 10. A higher value corresponds to greater happiness. We renamed this variable to "satisfaction.

EHG2DVR3 represents individuals' educational attainment, categorized into three levels: 1 for "Less than high school," 2 for "High school," and 3 for "Post-secondary or University." To create a control variable, a dummy variable called "University" was derived. If an individual has attended post-secondary or university education, the value of the dummy variable will be 1; otherwise, it will be 0. By including this control variable in the analysis, the potential for omitted variable bias is reduced, ultimately enhancing the model's accuracy.

Dhhgms is a qualitative variable, values 1, 2, 3, 4 refer to “married”, “common-law”, “divorced”, “single” respectively. Since we are only focusing on whether an individual is married, we created a dummy variable called married_or_common_law. This was also added to our research model as a control variable.

Incdghh is a quantitative variable representing annual income levels, bucketted following guidelines of the CCHS data. Value 1 signifies an annual income range of no income to less than $20,000, value 2 corresponds to $20,000 to $39,999, value 3 represents $40,000 to $59,999, value 4 denotes $60,000 to $79,999, and value 5 indicates an annual

income of $80,000 or higher. To facilitate analysis, we transformed these values to the midpoints of their respective ranges. For example, value 1 becomes $10,000, and notably, value 5 becomes $92,000, which we will discuss further later on.

Dhhgage is a quantitative variable, and its values show the range of individual age. The minimum value 01 means an individual's age is between 12 and 14, the maximum value 16 means that the individual is older than 80. Similar to income, we transformed its value to the midpoint of the age range it represented. For example, value 01 is now transformed to 13, and value 02 is transformed to 16. Particularly, we are using value 82 to represent the age of people who are older than 80. Using a single value instead of a range of values for income and age could make the analysis easier to interpret, though could potentially affect the accuracy of our regression.

GEN_005 is a quantitative variable with values ranging from 1 to 5, where lower values indicate better health. To align the interpretation more closely with common understanding, we transformed the values to establish a positive relationship with health levels. We changed the values as follows: 1 to 5, 2 to 4, 4 to 2, and 5 to 1. 5, now represents an excellent health condition; while 1 denotes a poor health condition. We are utilizing the dataset CCHS_Annual_2017_2018_curated_trimmed as our sample for analysis, provided by Professor Jonathan Graves. The CCHS contains all relevant variables and consists of a random 25% subset of the complete dataset. As a result, it is statistically representative of the full dataset, and our findings based on this sample should lead to analogous conclusions.

III. Summary Statistics

original name	adjusted name	mean	sd	max	min
dhhgage	Age	49.28907	19.90573	82	13
incdghh	Income	64151.48	28083	90000	10000
GEN_010	Satisfaction	8.103168	1.595318	10	1
GEN_005	Health	3.649545	1.019408	5	1
dummy from dhhgms	Married_or_Common_Law	0.5096232	0.49991	1	0
dummy from EHG2DVR3	University	0.5676615	0.495403	1	0

A. Our model

To understand the relationship between money and happiness, we estimate the following linear regression:

$$Y = \beta_0 + \beta_1 I + \beta_2 H + \beta_3 M + \beta_4 A + \beta_5 U + \epsilon$$

common law partner, A is age, and U is a dummy variable indicating whether the individual has attended university. We chose satisfaction with life as our dependent variable because it is the variable in our dataset that is most closely related to happiness. Our explanatory variables: income, health, marital status, age and education were chosen as the main determinants of a person's happiness. We believe these reflect the highest motivation for human action. (Diener, 1984).

B. Our Method

Our method for determining the low-income line is based on the methodology used by Statistics Canada, calculated as 50% of the median household income. According to this calculation, the low-income line is $35,000. As a result, the interpretation for the coefficient β_1 is the change in satisfaction resulting from a 1% increase above the low-income line. Apart from our main specification above, we also consider three variations of the main model:

$$Y = \beta_0 + \beta_1 I + \beta_2 H + \epsilon$$

$$Y = \beta_0 + \beta_1 I + \beta_2 H + \beta_3 M + \epsilon$$

$$Y = \beta_0 + \beta_1 I + \beta_2 H + \beta_3 M + \beta_4 A + \epsilon$$

where the first only includes income and health, the second as well considers marital status, and the third additionally considers the age of our respondents. In our research, we are interested in the coefficient β_1. If the estimated value is positive and significant, then we can conclude that income is positively correlated with satisfaction in life; therefore, money has a positive effect on happiness.

C. Table of Results

Table 1

	Dependent variable:			
	Satisfaction			
	(1)	(2)	(3)	(4)
Health	0.714***	0.717***	0.728***	0.737***
	(0.004)	(0.004)	(0.004)	(0.004)
Income	0.002***	0.001***	0.001***	0.002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Married or Common-Law		0.309***	0.276***	0.303***
		(0.009)	(0.009)	(0.009)
Age			0.003***	0.004***
			(0.0002)	(0.0002)
University				-0.208***
				(0.009)
Constant	5.344***	5.229***	5.041***	5.074***
	(0.016)	(0.016)	(0.021)	(0.021)
Observations	106,981	106,981	106,981	106,981
R2	0.236	0.244	0.245	0.249
Adjusted R2	0.236	0.244	0.245	0.249
Residual Std. Error	1.395 (df = 106978)	1.387 (df = 106977)	1.386 (df = 106976)	1.382 (df = 106975)
F Statistic	16,500.710*** (df = 2; 106978)	11,512.770*** (df = 3; 106977)	8,694.685*** (df = 4; 106976)	7,103.777*** (df = 5; 106975)
Note:				*p<0.1; **p<0.05; ***p<0.01

IV. Discussion

A. Results

Our primary interest, 1 , represents the effect of household income on general satisfaction. Table 1 summarizes the effects of income for four different specifications: holding health constant (column one), holding both health and marital status constant (column two), and holding health, marital status, and age constant (column three), and holding health, marital status, and age, and university constant (column four). The first column shows that while controlling for health, a 1% increase in income above the low-income line results in a 0.002 increase in satisfaction. The second column shows that while

controlling for health and marital status, there is a 0.001 increase in satisfaction due to a 1% increase in income above the low-income line. The third column shows that there is a 0.001 increase in satisfaction from a 1% change in income, when holding health, marital status, and age constant. The fourth column shows the estimates from our main model. After adding the additional control for university education, a 1% increase in income above the low-income line results in a 0.002 increase in satisfaction. In our model, the expected happiness without controls is equal to 5.074 units. We can expect that a 1% increase in income will increase happiness by 0.002 units. Thus, we can expect that for every 5% increase in income, there is a 0.00197% increase in happiness. Therefore, while the effects are statistically significant, they are insignificant economically.

B. Specification Check

I. Testing for Heteroskedasticity— To test for heteroskedasticity, we performed the Breusch-Pagan Test. For all specifications, the results have a p-value of < 2.2e-16, indicating that we can strongly reject the assumption of homoscedasticity. Table 2 shows our results corrected for heteroskedasticity, with adjusted standard errors using the Hubert White robust standard errors. The standard errors for the coefficient on income remained the same, with slight changes in the standard errors for the constant term and health coefficient: from 0.021 to 0.025, and 0.004 to 0.005.

Table 2

	Dependent variable:			
	Satisfaction			
	(1)	(2)	(3)	(4)
Health	0.714***	0.717***	0.728***	0.737***
	(0.005)	(0.005)	(0.005)	(0.005)
Income	0.002***	0.001***	0.001***	0.002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Married or Common-Law		0.309***	0.276***	0.303***
		(0.009)	(0.009)	(0.010)
Age			0.003***	0.004***
			(0.0002)	(0.0002)
University				-0.208***
				(0.009)
Constant	5.344***	5.229***	5.041***	5.074***
	(0.021)	(0.021)	(0.025)	(0.025)
Observations	106,981	106,981	106,981	106,981
R2	0.236	0.244	0.245	0.249
Adjusted R2	0.236	0.244	0.245	0.249
Residual Std. Error	1.395 (df = 106978)	1.387 (df = 106977)	1.386 (df = 106976)	1.382 (df = 106975)
F Statistic	16,500.710*** (df = 2; 106978)	11,512.770*** (df = 3; 106977)	8,694.685*** (df = 4; 106976)	7,103.777*** (df = 5; 106975)
Note:				*p<0.1; **p<0.05; ***p<0.01

C. Robustness Analysis

To assess the robustness of our results, we present two alternative models, one where we replace satisfaction with mental health as the dependent variable and the other where we treat Income as a qualitative variable with five levels, each representing a particular income group.

I. Mental Health as a dependent variable — Since happiness is a rather abstract concept, we want to use an alternative measurement of happiness as a robustness check for our research results. Mental health is a good representation of happiness because a person's happiness is closely related to a person's mental well-being. Thus, our specification is the following regression:

$$Z = \beta_0 + \beta_1 I + \beta_2 H + \beta_3 M + \beta_4 A + \beta_5 U + \epsilon$$

where Z is mental health, and the other variables remain the same as our main model. Our results, adjusted for heteroskedastic errors are depicted in Table 3. In the main model (column 4), we find that a 1% increase in income above the low income line results in a 0.001 increase in mental health. We find a positive and statistically significant relationship between income and mental health, confirming our results from the main model that money has a positive effect on happiness. for the three specifications. Therefore, a Indicating that, on

average, income has a greater effect on mental health when marital status is considered. We find a positive and statistically significant relationship between income and mental health, confirming our results from the main model.

Table 3

	Dependent variable:			
	Mental Health			
	(1)	(2)	(3)	(4)
Health	0.419***	0.420***	0.440***	0.441***
	(0.003)	(0.003)	(0.003)	(0.003)
Income	0.001***	0.0003***	0.001***	0.001***
	(0.00003)	(0.00004)	(0.00004)	(0.00004)
Married or Common-Law		0.127***	0.064***	0.067***
		(0.006)	(0.006)	(0.006)
Age			0.006***	0.006***
			(0.0001)	(0.0001)
University				-0.023***
				(0.005)
Constant	2.329***	2.282***	1.920***	1.924***
	(0.012)	(0.012)	(0.015)	(0.015)
Observations	106,981	106,981	106,981	106,981
R2	0.211	0.215	0.228	0.228
Adjusted R2	0.211	0.215	0.228	0.228
Residual Std. Error	0.852 (df = 106978)	0.850 (df = 106977)	0.843 (df = 106976)	0.843 (df = 106975)
F Statistic	14,286.070*** (df = 2; 106978)	9,745.830*** (df = 3; 106977)	7,887.424*** (df = 4; 106976)	6,314.688*** (df = 5; 106975)
Note:				*p<0.1; **p<0.05; ***p<0.01

II. Measuring the effects of income across income levels — In our main model, we treated the income variable as a quantitative variable by taking the midpoint of the interval for each income group. While this method allows for more precise statistical analysis, and is easier to interpret, it does not accurately capture the variability in the data. Our second robustness analysis treats income as a qualitative variable, with five levels, corresponding to ‘No income or less than $20,000’, ‘$20,000 to $39,999’, ‘$40,000 to $59,999’, ‘$60,000 to $79,999’, and ‘$80,000 or more.’ Table 4 summarizes the results of our regression, with standard errors adjusted for heteroskedasticity.

In our main model, we treated the income variable as a quantitative variable by taking the midpoint of the interval for each income group. While this method allows for more precise statistical analysis, and is easier to interpret, it does not capture the difference in effects across income levels. Our second robustness analysis treats income as a qualitative variable with five levels, corresponding to ‘No income or less than $20,000’, ‘$20,000 to $39,999’, ‘$40,000 to $59,999’, ‘$60,000 to $79,999’, and ‘$80,000 or more.’ Table 4

summarizes the results of our regression, with standard errors adjusted for heteroskedasticity. This method allows us to compare the relative effects of income on happiness across income levels, providing insight on how different income groups value income.

In our main model (column four), we find that the coefficients associated with each income group are positive and statistically significant. Holding health, age, marital status, and education constant, individuals with an income of between $20,000 to $39,999 have an estimated 0.199 higher satisfaction score than those in the income group of 'No income or less than $20,000'. For those belonging to the '40,000 to $59,999' income group, their satisfaction is estimated to be 0.270 higher than those with 'No income or less than $20,000'. Individuals who have incomes between $60,000 to $79,999 have an estimated 0.353 higher satisfaction score. Finally, individuals with an income of '$80,000 or more' have a 0.416 higher satisfaction score. As we move up income groups, the effects of income on satisfaction increases, from 0.199, to 0.270, to 0.353, to 0.416. Hence, we can conclude that the marginal effect of income on satisfaction increases as income increases. These results build on our main results such that increased income leads to more satisfaction.

Table 4

	Dependent variable:			
	Satisfaction			
	(1)	(2)	(3)	(4)
Health		0.717***	0.726***	0.736***
		(0.005)	(0.005)	(0.004)
20,000 to 39,999	0.422***	0.212***	0.200***	0.199***
	(0.025)	(0.022)	(0.022)	(0.018)
40,000 to 59,999	0.636***	0.253***	0.254***	0.270***
	(0.024)	(0.022)	(0.021)	(0.018)
60,000 to 79,999	0.828***	0.320***	0.331***	0.353***
	(0.024)	(0.022)	(0.022)	(0.019)
80,000 or more	1.010***	0.355***	0.383***	0.416***
	(0.022)	(0.020)	(0.020)	(0.017)
Married or Common-Law		0.307***	0.276***	0.303***
		(0.009)	(0.010)	(0.009)
Age			0.003***	0.003***
			(0.0002)	(0.0002)
University				-0.207***
				(0.009)
Constant	7.361***	5.051***	4.876***	4.901***
	(0.021)	(0.026)	(0.029)	(0.024)
Observations	106,981	106,981	106,981	106,981
R2	0.039	0.245	0.246	0.250
Adjusted R2	0.039	0.245	0.246	0.250
Residual Std. Error	1.564 (df = 106976)	1.387 (df = 106974)	1.386 (df = 106973)	1.382 (df = 106972)
F Statistic	1,096.735*** (df = 4; 106976)	5,773.098*** (df = 6; 106974)	4,978.376*** (df = 7; 106973)	4,447.440*** (df = 8; 106972)
Note:				*p<0.1; **p<0.05; ***p<0.01

Conclusion

Our paper attempts to measure the effect of household income on general satisfaction for every dollar above the low-income line. To understand that relationship, we regressed household income on happiness while holding constant: age, marital status, and education. We then ran robustness tests by changing our dependent variable to mental health and measuring adjusted heteroskedasticity.

Using data from the CCHS from 2017-2018, we used Breusch-Pagan Tests, robustness tests, and Hubert White standard error adjustment to adjust for heteroskedasticity. Findings from our data show that the coefficient on income is positive. Furthermore, our regression results show that the effects of income are positive and statistically significant. We see that, for every 1% increase in household income, we can expect an increase in general satisfaction by 0.002 units.

Our study's implication shows that, a higher household income means a higher level of satisfaction. Although income is shown to increase satisfaction, its coefficient of 0.002 is relatively low. Our study contributes to and answers the consensus of the social sciences that money does buy happiness. Something to consider in future research is normalizing exogenous variables (precisely that of health), to make interpreting the model's coefficients more reasonable. Furthermore, future research could add additional explanatory variables like location or average hours spent working. Using dummies in the model already, and added explanatory variables, creating interactions would be a good test of the model's applicability to general research questions.

Attribution

Abstract -- The abstract for this research project was written by Jiayin Kralik in conjunction with Max Geyer. Ideas introduced were discussed by all students.

Introduction -- The introduction was written by Max Geyer

Data Description -- Written by Hanze Wang, and edited by Max Geyer

Summary Statistics -- Written by Hanze Wang, coded in R by Jiayin Kralik

Model -- The regression model for this research project was created by all students. Explanatory and outcome variables were chosen as a group from the CCHS, the paragraph was written by Jiayin Kralik.

Table of results -- Results of the table were coded in R by Jiayin Kralik

Discussion -- written by Jiayin Kralik, edited for grammar by Max Geyer

Results -- written by Jiayin Kralik, edited for grammar by Max Geyer

Specification Check -- written by Jiayin Kralik, edited by Max Geyer

Robustness Analyses -- written by Jiayin Kralik, edited by Max Geyer

Conclusion -- Written by Max Geyer

Formatting -- Done by Max Geyer

References

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575. https://doi.org/10.1037/0033-2909.95.3.542

Howell, R. T., & Howell, C. J. (2008). The relation of economic status to subjective well-being in developing countries: A meta-analysis. *Psychological Bulletin*, *134*(4), 536–560. https://doi.org/10.1037/0033-2909.134.4.536

Powdthavee, N. (2009). How much does money really matter? estimating the causal effects of income on happiness. *Empirical Economics*, *39*(1), 77–92. https://doi.org/10.1007/s00181-009-0295-5

Statistics Canada. (2019). Canadian Community Health Survey (CCHS) Annual Component User Guide: 2018 and 2017-2018 Microdata file., 4-5.

Wu, J. (2021, June 24). *Can money really buy happiness?* Psychology Today. Retrieved April 14, 2023, from https://www.psychologytoday.com/us/blog/the-savvy-psychologist/202106/can-money-really-buy-happiness#:~:text=Key%20points,things%20tend%20to%20predict%20wellbeing.