Using Oaxaca-Blinder and Machine Learning to Decompose the Gaps of Happiness and Financial Satisfaction by Gender

International Conference on Emerging Challenges:

Smart Business and Digital Economy

Nguyen Thanh Nam[1]; Nguyen Phuong Anh[2*], Ngo Ngoc Lan Linh[1], Pham Quang Huy[1]

[1] School of Applied Mathematics and Informatics, Hanoi University of Science and Technology, Hanoi, Vietnam

[2] Department of Financial Management, School of Economics and Management, Hanoi University of Science and Technology, Hanoi, Vietnam

[*]Corresponding author: anh.nguyenphuong1@hust.edu.vn

Abstract

In this study, we utilize data from the World Values Survey to provide empirical evidence for variations in levels of happiness and financial satisfaction according on gender. While it is generally observed that women tend to report higher levels of happiness than men on a global scale, the distribution of happiness between genders varies at the country level. In this study, we want to analyse the disparity in satisfaction levels based on visible factors, as well as the differential responses of males and females to these qualities.

Research purpose:

The primary objective of this study is to examine the disparity in financial satisfaction and happiness across genders and determine the extent to which this gap may be attributed to varying objective circumstances experienced by men and women, as well as the differential responses exhibited by both genders towards the similar objective conditions.

Research motivation:

In nearly all countries and eras, differences between men and women in relation to the majority of social outcomes have been more prevalent than exceptional. Nevertheless, it is arguable that men and women have never before had such equal access to educational attainment, employment, and civil rights. As these objective aspects of life tend to converge for both sexes, the topic of differences in how they affect personal well-being remains unanswered.

Research design, approach, and method:

To achieve this objective, this topic proposes combining the Supervised Learning model and the Oaxaca-Blinder model. The contribution of our paper is that application of a decomposition technique in the happiness and subjective well-being literature. The Supervised Learning model, a potent supervised learning technique, will be used to develop a classification model of the individual happiness index. This will enable us to better comprehend the relationship between satisfaction factors and indicators, as well as evaluate their effects on men and women. This algorithm will learn from survey data and identify patterns to classify individual contentment. The Oaxaca-Blinder model (Blinder, 1973; Oaxaca, 1973) will assist us in analysing happiness discrepancies based on financial satisfaction, education, social status, and other characteristics. Through a comparative analysis of these characteristics pertaining to males and females, it is possible to ascertain the extent to which each element contributes to the disparity in happiness observed between the two genders.

Main findings:

We conclude that women are more "optimistic" than males and tend to value objective aspects of their lives more positively. When considering all of the aspects of life that are typically measured, it is observed that women tend to experience higher levels of happiness than what may be expected. In terms of financial satisfaction, the results of this section indicate that women are happier than men, despite men have superior financial situations and are more contented with them.

Practical/managerial implications:

This study has the potential to yield significant insights that can drive policymaking and interventions aimed at mitigating the disparity in happiness levels between males and females, so fostering a more egalitarian and successful society for individuals of all genders.

Keywords: Happiness, Financial satisfaction, Gender, Oaxaca-Blinder decomposition, Machine Learning

N. D. Nguyen and P. T. T. Hong (eds.), *Proceedings of the 11th International Conference on Emerging Challenges: Smart Business and Digital Economy 2023 (ICECH 2023)*, Advances in Economics, Business and Management Research 274,
https://doi.org/10.2991/978-94-6463-348-1_6

1. INTRODUCTION

In the modern era, happiness becomes a major consideration in all countries and times. Individuals are not only interested in making a lot of money, having a successful job or building a happy family, but also interested in satisfaction and personal experiences in everyday life. This laid the foundation for the development and application of the personal happiness index. Happiness indicators are considered as a measure of an individual's happiness and satisfaction in all aspects of life, including job satisfaction, personal relationships, health and well-being, a sense of autonomy, and others. Instead of focusing solely on traditional economic indicators such as GDP, the happiness index allows a more comprehensive and profound assessment of the quality of life of a particular country or community. Simultaneously, it is imperative to examine the disparity in levels of happiness experienced by genders within the context of this research. The acknowledgement of socioeconomic inequalities and disparities based on gender is imperative and cannot be disregarded.

Hence, the question arises as: "Do men and women experience happiness differently?". Numerous scholarly publications have examined the disparities in happiness levels among males and females (Blanchflower and Oswald, 2001; Stevenson and Wolfers, 2009; Guven, 2012; Vieira Lima, 2011; Graham and Chattopadhyay, 2012; Zweig, 2014; Arrosa & Gandelman, 2016). Most empirical research indicates that women tend to experience higher levels of happiness compared to males. However, it is important to note that variations in happiness levels between genders exist at the national level. Blanchflower and Oswald (2001) conducted a study that examined the levels of happiness among men and women in the United States and the United Kingdom. The researchers utilized data from the General Social Survey spanning the years 1972 to 1998, as well as the Eurobarometer British Survey conducted between 1975 and 1986. Their findings indicated that men tend to report lower levels of happiness compared to women in both countries. Arrosa and Gandelman (2016) investigated gender variations in happiness and discovered that these disparities might be explained by women's more optimistic attitude on life.

Financial satisfaction and happiness are closely related to an individual's well-being. Research on gender differences in these areas can shed light on the effects of economic factors on different genders. For instance, if women consistently report lower levels of financial satisfaction than men, this may suggest that women encounter unique obstacles to achieving economic security, such as wage gaps or limited access to economic opportunities. The primary objective of this study is to examine the disparity in financial satisfaction and happiness across genders and determine the extent to which this gap may be attributed to varying objective circumstances experienced by men and women, as well as the differential responses exhibited by both genders towards the similar objective conditions.

To achieve this objective, this topic proposes combining the Supervised Learning model and the Oaxaca-Blinder model. The contribution of our paper is that application of a decomposition technique in the happiness and subjective well- being literature. The Supervised Learning model, a potent supervised learning technique, will be used to develop a classification model of the individual happiness index. This will enable us to better comprehend the relationship between satisfaction factors and indicators, as well as evaluate their effects on men and women. This algorithm will learn from survey data and identify patterns to classify individual contentment. The next purpose of this research is to examine and comprehend the difference in happiness between men and women. The Oaxaca-Blinder model (Blinder, 1973; Oaxaca, 1973) will assist us in analysing happiness discrepancies based on financial satisfaction, education, social status, and other characteristics. Through a comparative analysis of these characteristics pertaining to males and females, it is possible to ascertain the extent to which each element contributes to the disparity in happiness observed between the two genders.

We conclude that women are more "optimistic" than males and tend to value objective aspects of their lives more positively. Our use of the term "optimism" is extremely broad. It is more of an assertion of ignorance than of knowledge. When considering all of the aspects of life that are typically measured, it is observed that women tend to experience higher levels of happiness than what may be expected. In terms of financial satisfaction, the results of this section indicate that women are happier than men, despite men have superior financial situations and are more contented with them. This study has the potential to yield significant insights that can drive policy-making and interventions aimed at mitigating the disparity in happiness levels between males and females, so fostering a more egalitarian and successful society for individuals of all genders.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Multiple studies have shown evidence of a consistent disparity in subjective well-being between genders (Blanchflower and Oswald, 2001; Stevenson and Wolfers, 2009; Guven, 2012; Vieira Lima, 2011; Graham and Chattopadhyay, 2012; Zweig, 2014). Numerous studies indicate that women generally experience higher levels of happiness compared to males, however it is important to note that variations exist across different countries. Blanchflower and Oswald (2001) provide empirical support for the notion that males tend to report lower levels of happiness compared to women in both the United States and the United Kingdom. This conclusion is drawn on an analysis of data obtained from the General Social Survey spanning the years 1972 to 1998, as well as the Eurobarometer British Survey conducted between 1975 and 1986. In a study conducted by Guven in 2012, an analysis was performed on the disparities in happiness levels observed among married couples. The findings of this study suggest that the presence of persistent gaps in happiness within these couples is associated with an increased probability of divorce. The researchers interpret this discovery as being indicative of a reluctance towards the unequal distribution of well-being within couples. According to the findings of Stevenson and Wolfers (2009), there is empirical support for the notion that women in the United States exhibit higher levels of happiness compared to men. However, it is important to note that females have encountered a decrease in both absolute and relative measures of happiness over time. They attribute this narrowing disparity to the increased expectations brought about by gender equality as well as the double burden of working and domestic responsibilities placed on women.

According to Vieira Lima (2011), females are happier than males on a global scale. Contrary to the findings of Graham and Chattopadhyay, the authors demonstrate that in the majority of African and developing nations, women are happier than males, while the opposite is true in approximately 15 European and other industrialized nations. The author discovers a negative correlation between women's happiness and the extent of their rights and accomplishments, concluding that women's happiness has a paradoxical component in which improved objective conditions do not bestow them happiness.

Zweig (2014) employs the Gallup World Poll to examine the gender disparity in happiness in 73 countries at various phases of development. The researcher used country-specific ordinary least square regressions, which provide empirical support for the notion that women tend to experience higher levels of happiness compared to males across the majority of nations included in the sample. There is no discernible correlation between the extent of economic growth, women's rights, geographical location, or religious beliefs and the severity of the disparity in happiness levels between females and males. Guven (2012) analyzes happiness gaps within married couples and concludes that persistent happiness gaps increase the likelihood of divorce, particularly when the happiness gap is detrimental to the wife. They interpret this result as an aversion to unequal well-being distribution within couples.

Several variables are related to individual's happiness like marital status, employment status, education, age and income. Numerous international studies have demonstrated the close relationship between marriage and individual health (Dayton 1936; Durkheim 1987; Robins and Regier 1991; Mastekaasa 1993). It is well-documented that married people live longer and generally experience better physical and mental health than unmarried people (e.g., Waite 1995; Miller et al. 2013). Helliwell (2003) asserts that marriage is one of the unambiguously positive and statistically significant correlates of life satisfaction. Helliwell also reports that in most countries, married individuals are satisfied with their lives than those who cohabit with a partner. In addition, empirical studies have found a significant correlation between any type of occupation, whether paid or unpaid, and well-being, primarily via richer social networks and self-esteem (Clark and Oswald 1994; Winkelmann and Winkelmann 1998; Clark 2003, 2007). The literature on happiness generally indicates a U-shaped correlation between age and happiness (Blanchflower and Oswald, 2001, 2007; Clark, 2007; Dear et al., 2002; Di Tella et al., 2003). This styled truth has three possible explanations: (1) increased expectations during the early years of maturity; (2) as people mature, they acquire a greater level of self-acceptance, which helps to lower expectations; (3) happier people live longer (Lopez Ulloa et al., 2013). Consistent with theoretical predictions in economics, empirical studies have demonstrated that there is a positive relationship between income and individual happiness, particularly among individuals with lower incomes (Frey and Stutzer, 2002; Blanchflower and Oswald, 2001; Clark et al., 2008). Finally, numerous scholars have conducted research on the correlation between education and individuals' levels of life satisfaction. On one side, researchers such as Clark and Oswald (1994), Frey and Stutzer (2002), and Veenhoven (2010) did not discover a consistent correlation between the two variables. This first group of authors concludes that despite education enables individuals to adapt more effectively to dynamic environments, this higher predisposition to well-being is offset by a proportional increase in aspiration levels. However, authors such as Helliwell (2003), Lopez and Guijarro (2012) discovered a positive and statistically significant relationship between both variables. Through indirect channels such as income, consumption, employment, occupational status, and health, education appears to promote well-being. This study employs the WVS Trend 1981-2022 data set from the World Values Survey to examine the relationship between education, marital status, age, and income and individual satisfaction levels. We anticipate that by analyzing data from numerous countries and time periods, we will be able to provide a clearer picture and a more in-depth analysis of the factors that influence global human happiness.

3. METHODOLOGY

3.1 Machine Learning for categorizing the happiness index of individuals

Descriptive data

Given the diversity and complexity of the WVS Trend 1981-2022 dataset, we can conduct descriptive statistics on happiness levels, distributions of determinants, and their interrelationships. By analyzing data and presenting statistics, we are able to provide a comprehensive view of global individual happiness and its contributing factors. The describe() function inside the Pandas package is employed to compute descriptive statistics for the given dataset. The tool offers insights into fundamental statistical properties of data columns, including sample size, mean, standard deviation, minimum value, percentiles, and maximum value (Figure 1).

```
df.describe()
```

	studyno	stdyno_w	s001	s002	s002vs	s003	cow_num	s004	s006	s007	...	y022b	y022c	y023	y023a	y02
count	439531.0	439531.000000	439531.0	439531.000000	439531.000000	439531.000000	439531.000000	439531.000000	4.395310e+05	4.395310e+05	...	378585.000000	385749.000000	413893.000000	393757.000000	411631.000
mean	4001.0	425.004791	2.0	4.882950	4.882950	456.779353	456.391035	-3.124938	9.526464e+07	3.711059e+09	...	0.493785	0.655605	0.312377	0.257488	0.270
std	0.0	610.553679	0.0	1.663917	1.663917	257.229794	260.029482	1.923214	2.174096e+08	2.852437e+09	...	0.325571	0.305316	0.281814	0.346283	0.3207
min	4001.0	1.000000	2.0	1.000000	1.000000	8.000000	2.000000	-4.000000	1.000000e+00	2.072000e+07	...	0.000000	0.000000	0.000000	0.000000	0.000
25%	4001.0	2.000000	2.0	3.000000	3.000000	231.000000	230.000000	-4.000000	4.590000e+02	7.603209e+08	...	0.330000	0.660000	0.069000	0.000000	0.000
50%	4001.0	12.000000	2.0	5.000000	5.000000	440.000000	439.000000	-4.000000	9.270000e+02	3.560624e+09	...	0.660000	0.660000	0.259259	0.000000	0.1111
75%	4001.0	345.000000	2.0	6.000000	6.000000	703.000000	703.000000	-4.000000	2.146000e+03	6.430222e+09	...	0.660000	1.000000	0.481482	0.444444	0.4444
max	4001.0	1562.000000	2.0	7.000000	7.000000	909.000000	920.000000	2.000000	9.090704e+08	8.940522e+09	...	1.000000	1.000000	1.000000	1.000000	1.000

8 rows × 717 columns

Figure 1. Descriptive statistics of the dataset

When using the valuecounts() function in the Pandas library is used to count the number of occurrences of values in a column of the DataFrame. It returns a new Series with the values being the unique values in the column and the indices being the number of occurrences of each of those values. Below this function has counted the number of occurrences of different values in our dependent variable (Happiness) column (coded as column a008). The data set includes the question of assessing happiness and satisfaction with life through the question "Feeling of happiness". This question is rated through 4 levels: 1. Very happy, 2. Quite happy, 3. Not very happy, 4. Not happy at all. This allows to consider the happiness level of each individual in the data set (Figure 2). Negative values appearing here are interpreted as individuals participating in the survey who refused or did not give an answer. So rows with negative values in column a008 will be deleted to serve the next steps of the happiness index classification problem (Figure 3). The value of rows has decreased from more than 439,000 to 430,877 rows. Next, in order to understand about the frequency of happiness indicators, we used Python's Numpy and Matplotlib libraries to visualize data based on pie charts (Figure 4).

```
new_data["a008"].value_counts()

Out[11]:  3.0     228195
          4.0     123959
          2.0      66379
          1.0      12344
         -4.0       3999
         -1.0       3426
         -2.0       1090
         -5.0        139
         Name: a008, dtype: int64
```

Figure 2. Statistics on the level of happiness of individuals participating in the survey

```
In [7]:  new_data = df.drop(df[df['a008'] < 0].index)
         new_data.shape

Out[7]:  (430877, 728)
```

```
import seaborn as sns
import matplotlib.pyplot as plt
positive_values = new_data[new_data["a008"] > 0]["a008"]
value_counts = positive_values.value_counts()
plt.figure(figsize=(6, 6))
sns.set_palette("pastel")
# Vẽ biểu đồ tròn
plt.pie(value_counts, labels=value_counts.index, autopct='%1.1f%%')
plt.axis('equal')
plt.title("Tỉ lệ phần trăm của các mức độ hạnh phúc ")
plt.show()
```

Figure 3. Elimination of individuals who did not participate in the survey

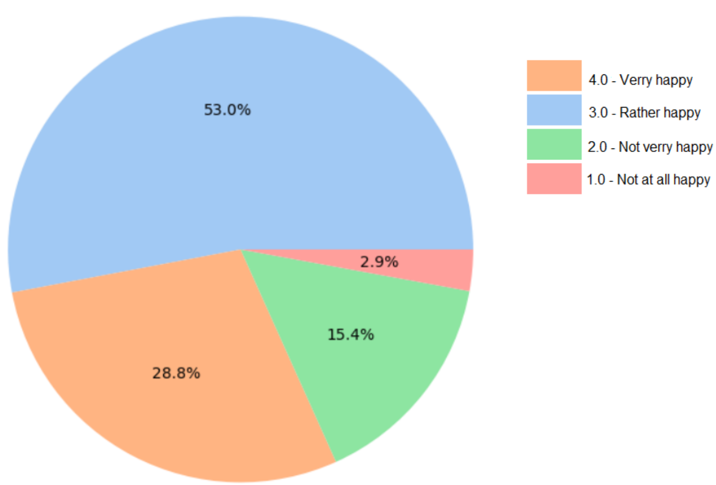

Figure 4. Happiness levels reported by the participants

The pie chart provides a visual representation of the proportions of various values inside the Happiness variable. This facilitates a comprehensive comprehension of the global distribution of the happiness index among the respondents. Additionally, colours and labels are applied to the elements of the pie chart, making it simple to identify and assess the percentage of each value.

Dimensionality Reduction and StandardScaler

Dimensionality Reduction is an essential component of data pre-processing. The primary objective of data dimensionality reduction is to eliminate superfluous features and retain only the most essential ones for training the classification model. This decreases model complexity, accelerates training, and decreases the likelihood of overfitting. Initially, the data set includes a number of columns with substantial negative values, which correspond to study participants who did not respond to the query. By contrasting the number of positive and negative values in each column, we have determined which columns contain negative values. Then, we eliminated these columns using the drop() method of pandas with the axis=1 parameter. The procedure of column removal reduced the data dimension and eliminated superfluous variables, resulting in a new data set containing columns with only positive values. As a result, our dataset contains 248 survey questions with the most responses, reducing the intricacy of the classification model and increasing the accuracy.

In our study, we used the StandardScaler method to normalize the data. The StandardScaler method is a popular and useful data normalization method. It assumes the data are normally distributed (Gaussian) with a mean of 0 and a standard deviation of 1. This normalization contributes to the homogenization of the variable's value range, eradicating deviations. and facilitate the adoption of machine learning algorithms more effectively. We utilized the StandardScaler technique to normalize the data. This method assists in normalizing the data to a normal distribution (Gaussian) with a mean of 0 and standard deviation of 1.

In the process of building a classification model using Machine Learning models, we have evaluated the performance of the model by important measures such as Precision, Recall and F1-score. These metrics provide crucial insight into the classification ability of the model on the test data set. Precision assesses the capability of a model to make accurate predictions for samples belonging to a specific class. It computes the ratio between the number of true predictions and the total number of predictions in the class to be predicted. A greater value indicates that the model can correctly classify samples of the specified class. Recall measures the model's capacity to capture all instances of a given class. It computes the ratio between the number of accurate predictions for the class to be predicted and the total number of actual samples belonging to that class. A greater value indicates that the model can accurately capture more instances of the class. The F1-score is a metric that calculates the harmonic mean of precision and recall. It provides a summary of the model's efficacy in correctly classifying samples belonging to the class to be predicted and capturing all samples belonging to that class. The Precision, Recall, and F1-score results are a method for evaluating the classification model's performance and provide crucial information regarding the model's predictive ability on the test dataset. This result can be used to compare the accuracy and generalizability of the model in classifying new samples with other models or as a basis for further process decisions. program of study. Despite the results of the three indicators of Precision, Recall, and F1-score are virtually identical, the RandomForest model outperforms the other two models. when trained with a large dataset containing numerous features. However, the model evaluation indexes yield low results, indicating that there is no firm relationship between the feature variables and the target class. In addition, as we can see, the happiness index is a complex concept; an individual's happiness index can depend on a variety of factors, some of which cannot be adequately explained or measured in survey question datasets. Factors such as gender inequality, psychological and social factors, and special circumstances cannot be fully reflected in the data, and each individual has unique happiness experiences and assessments. Numerous factors, including education, family, culture, and living conditions, can influence how an individual evaluates and experiences contentment. To address these "unexplained" factors, we will examine the Oaxaca-Blinder model in greater detail in the following section.

3.2 Analyzing the disparity in happiness between males and females using the Oaxaca-Blinder decomposition model

The Oaxaca-Blinder model is an analytical framework used to dissect the differences between two groups by identifying the influence of various factors on those differences. Through this model, the inherent characteristics of the group's differences and the variations caused by the differential impact of distinct factors, including inequalities, are revealed. In this study, we will employ the Oaxaca-Blinder model to analyze differences in happiness between males and females. In the context of a linear regression model, it is posited that there exists a dependent variable Y along with a set of independent variables $X = (X_1, X_2, \dots, X_k)$. Where X is a k-dimensional vector of independent variables, β represents the vector of linear regression coefficients, and ϵ denotes the error term:

$$Y = X\beta + \epsilon$$

In the event of applying a regression model to the male group, indicated as (M), the equation assumes the structure:

$$Y^M = X^M\beta^M + \epsilon^M$$

In a parallel manner, the female group, labeled as (F), is characterized by the equation $Y^F = X^F\beta^F + \epsilon^F$. The vectors X^M and X^F correspondingly denote the independent variables for the male and female groups. The expectation of the dependent variable for each group is defined as follows:

$$\bar{Y}^M = {\beta_0}^M + \sum_{j=1}^{k} {\beta_j}^M {\bar{x}_j}^M$$

$$\bar{Y}^F = {\beta_0}^F + \sum_{j=1}^{k} {\beta_j}^F {\bar{x}_j}^F$$

Where $\bar{x}_j$ stands for the mean of each independent variable x_j. Consequently, the disparity in expectations between the two groups is:

$$\Delta\bar{Y} = \left(\beta_0^M - \beta_0^F\right) + \sum_{j=1}^{k} \left(\beta_j^M \bar{x}_j^M - \beta_j^F \bar{x}_j^F\right) \quad (1)$$

The difference in expectations between the two groups is the sum of the following components:

- The average difference of each x_j
- The disparity of β between the two groups
- The variation of factors not present in the model

So, how significant is the influence of these components? To answer this question, the model assumes all independent variables and coefficients are the same, and then changes each factor one by one to identify the impact of the factor on the difference between the two groups.

$$\Delta\bar{Y} = \underbrace{\left(\beta_0^M - \beta_0^F\right)}_{B} + \underbrace{\sum_{j=1}^{k} \beta_j^F \left(\bar{x}_j^M - \bar{x}_j^F\right)}_{E} + \underbrace{\sum_{j=1}^{k} \bar{x}_j^F \left(\beta_j^M - \beta_j^F\right)}_{C} + \underbrace{\sum_{j=1}^{k} \left(\bar{x}_j^M - \bar{x}_j^F\right)\left(\beta_j^M - \beta_j^F\right)}_{I}$$

The above decomposition process takes group M as the reference. The expectation of the difference can be decomposed into the following four components:

- Component (B) constitutes the inherent disparity. It encompasses unobserved characteristics.
- Component (E) represents the change in group F when utilizing the values of independent variables from group M as a reference. The disparity of this component can be elucidated through the distinction in the independent variables (explanatory variables). Therefore, this difference is termed the "endowment effect."
- Component (C) captures the discrepancy of group F when employing the regression coefficients of group M as a reference:
- Component (I) represents the interaction between the endowment effect and the coefficients effect.

Taking an alternative perspective, we can assume a specific degree of fairness between the two groups. Consequently, under this method, we must establish the defining criteria for this assumption.

Suppose the vector β∗ is indicative of the benchmark vector representing this condition. As a result, the anticipation variance becomes:

$$\Delta\bar{Y} = \sum_{j=1}^{k} \beta_j^* \left(\bar{x}_j^M - \bar{x}_j^F\right) + \left[\sum_{j=1}^{k} \bar{x}_j^M \left(\beta_j^M - \beta_j^*\right) + \sum_{j=1}^{k} \bar{x}_j^F \left(\beta_j^* - \beta_j^F\right)\right]$$

The vector β^* always lies between β^F and β^M. Suppose we possess $\beta^M > \beta^* > \beta^F$, indicating that the model exhibits bias favoring the male group, which is known as "positive discrimination". In contrast, if the configuration is $\beta^M < \beta^* < \beta^F$, it denotes "negative discrimination".

In certain specialized cases, when one group experiences inequality and the vector β^* equates to the vector β of the remaining group, for example $\beta^M = \beta^* > \beta^F$, the equation takes on a new form upon substituting β^* with β^M in the decomposition equation:

$$\Delta\bar{Y} = \sum_{j=1}^{k} \beta_j^M \left(\bar{x}_j^M - \bar{x}_j^F\right) + \sum_{j=1}^{k} \bar{x}_j^F \left(\beta_j^M - \beta_j^F\right)$$

Thus, we arrive at a two-fold decomposition:

- The Explained component (Ec): This amalgamates components (E) and (I) in the three-fold decomposition model. To provide a more comprehensive explanation of the interaction between the magnitude of the independent variable and the regression coefficient, opting for the three-fold model is advisable.
- The Unexplained component (Uc) is analogous to that in the three-fold decomposition..

There are several options for selecting the reference vector, including the following:

- Apply the average regression coefficients between the two groups (Reimers-1983):

$$\beta^* = \frac{\beta^M + \beta^F}{2}$$

- Apply the observations count for each group to (Cotton-1988):

$$\beta^* = \frac{n^M \beta^M + n^F \beta^F}{N}$$

3.3 Data

In this study, we employ the WVS Trend 1981-2022 dataset from the World Values Survey (WVS), a global organization that investigates the attitudes of individuals around the world. This dataset provides information on the world's population's perspectives from 1981 to 2022, as provided by thousands of individuals from numerous nations. WVS collects data through individual surveys, which include a wide range of social, cultural and political questions.

4. Results and discussion

4.1 Statistical Data Analysis

The dataset collects happiness indices across continents. The happiness index is recorded on a scale of 1 to 4, reflecting the happiness level of the individuals interviewed (where a higher score indicates greater happiness).

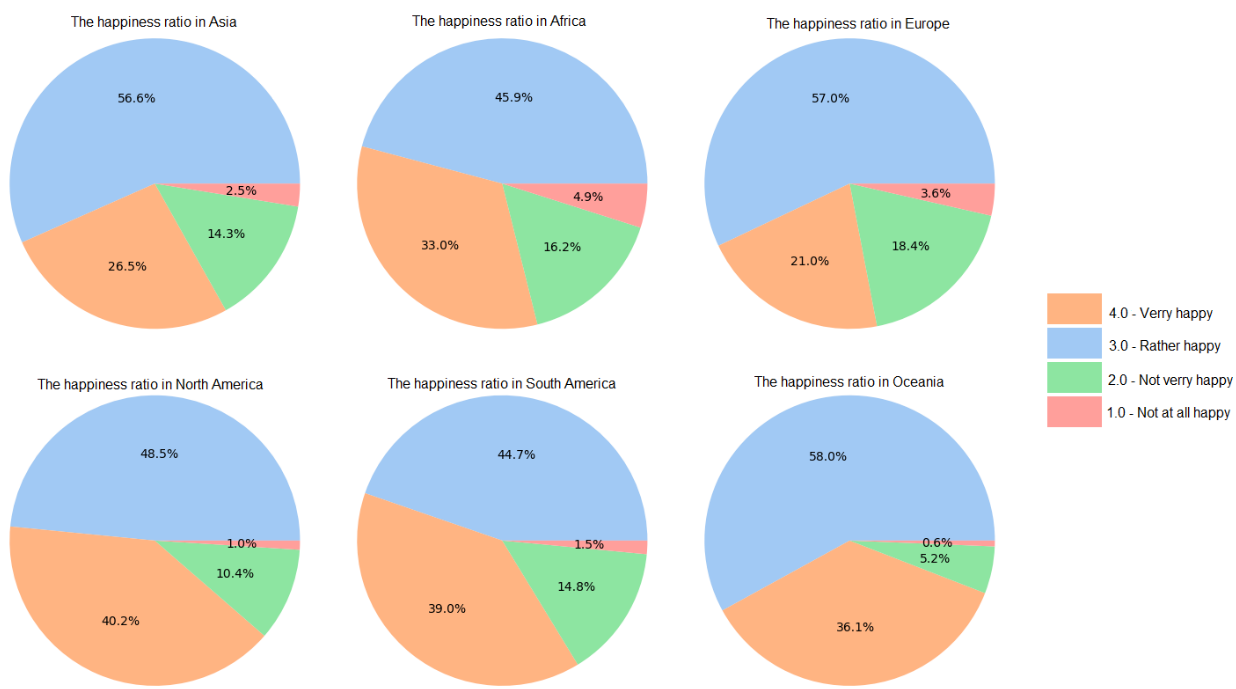

Figure 5. Porpotions of happiness across all continent

The data reveals that the Pacific and North America are the two continents with the highest proportion of happy individuals, with average scores around 3.27 and 3.25 respectively on a scale of 1 to 4. On the other hand, Africa has a significantly high percentage (4.9%) of individuals reporting unhappiness (score 1.0), which is ten times higher than the unhappiness rate in the Pacific, the continent with the lowest proportion of unhappiness.

Other continents also exhibit varying levels of happiness. In Asia and the Pacific, over 50% of surveyed individuals indicated the highest level of happiness (score 4.0). In South America and North America, more than 40% of those surveyed reported a happiness score of 3.0. Europe stands as the continent with the lowest happiness rate, with 22% of the population feeling unhappy.

Below is a summary of the studied factors present in the data. The income index is recorded on a scale of 1 to 10 (higher income corresponds to a higher index), and the happiness index is recorded on a scale from 1 to 4.

	Happiness	Income	Age	Employed	Single	Married	Divorced	Windowed	Financial Satisfaction

Men	3.1270	5.0207	43.4123	70%	66,7%	3,5%	2,6%	27,2%	6.193876
Women	3.1544	5.1491	41.3600	50%	64,7%	5,5%	8,4%	21,2%	6.153327
Total	3.1414	4.9118	43.1310	59,8%	65,7%	4,5%	5,6%	24,2%	6.172611

Table 1. Statistical analysis of data between males and females.

The table above provides information about the average levels of happiness, income, age, percentage of employment, and percentage of marital statuses from the data. The dataset includes both males and females, with varying numbers and proportions within each group.

The male group has, on average, lower happiness and income levels compared to the female group by 0.03 units. The percentage of marriages and divorces among females is significantly higher than that among males.

	Male		Female		*t*	*p* value
	Mean	SE	Mean	SE		
Value	3.1270	0.0035	3.1544	0.0033	-5.6039	2.102e-08

Table 2. t-test for the difference of happiness between males and females

Based on the surveyed data and with a significance level of 5%, it is acceptable to conclude that the female group has a higher happiness index than the male group.

	Male		Women		*t*	*p* value
	Mean	SE	Mean	SE		
Africa	3.0590	0.05282	3.0825	0.0519	3.461	0.0005
North America	3.2744	0.0432	3.2989	0.0427	2.9355	0.0033
Europe	2.9543	0.0454	2.9524	0.0471	-0.4414	0.6589
South America	3.2202	0.0461	3.2050	0.0478	-2.5425	0.0110
Oceania	3.2545	0.0375	3.3302	0.0373	6.9124	5.021e-12
Asia	3.0526	0.0453	3.0887	0.0445	-9.5645	2.2e-16

Table 3. t-test for the difference of happiness between males and females across different regions

In all the surveyed regions, the results consistently indicate that the female group has a higher level of happiness compared to the male group. This is particularly evident in the Pacific, Asian, and African continents. With a significance level of 5%, it can be concluded that the female group is happier than the male group in all surveyed regions except Europe.

In the following section, we will utilize the Oaxaca-Blinder decomposition method to gain a deeper understanding of the factors contributing to the higher happiness index among the male group in comparison to the female group.

4.2 Decomposition of the difference in financial satisfaction between male and female

	Mean difference	Explained	Unexplained	Most important factor
Value	0.0405	0.0761	-0.0355	Income

Table 4. Decomposition of the difference in financial satisfaction

The positive overall result indicates that, on average, men report higher levels of financial satisfaction compared to women. This means that men tend to express a greater sense of contentment or happiness with their financial situation than women.

The positive value of the explained component signifies that a portion of the financial satisfaction difference between men and women can be attributed to observable characteristics, specifically education and income. It suggests that men, on average, have higher levels of education and employment rate, which are positively associated with financial satisfaction. The negative unexplained component show that even after accounting all the effect of education, income and employment rate, it shows that women experience financial satisfaction different from men in a positive way. To simplify it, we can say that men is more satisfied since men have better financial situation than women but if they have the same financial situation, women group would be much more satisfied than men group. All the results implies that there are unique gender-related aspects at play, potentially involving gender norms, societal roles, and economic circumstances, which influence how men and women perceive and experience financial satisfaction differently.

4.3 Decomposition of the difference in happiness between male and female

In the section above, we can clearly observe the disparity in happiness between the male and female groups, as well as the difference in financial circumstances and their perceptions of it between men and women. In this section, we will use the Oaxaca-Blinder decomposition model to analyze the happiness difference between these two groups. The table below presents the decomposition results on the dataset. The mean difference is again decomposed to explained and unexplained part with the coefficient of the reference vector being 0.5.

	Mean difference	Explained	Unexplained	Most important factor
Value	-0.0274	0.00144	-0.0289	Employment rate

Table 5. Decomposition of the difference in happiness

The result is very interesting. In this case, the negative value implies that men has, on average, lower happiness compared to women despite having better financial status. But does financial situation decide happiness?

The positive explained component suggests that part of the negative happiness difference can be attributed to observable factors like marriage, education and also financial status. However, the positive value is quite small, indicating that these observable factors only contribute slightly to the overall difference. The larger negative unexplained component highlights that even after accounting for observable factors, there's still a substantial portion of the financial satisfaction difference that remains unexplained. This component could be indicative of various factors such as societal norms, personal biases, or other hidden influences that are influencing financial satisfaction in ways that go beyond what can be measured by the included variables.

Overall, the results in this section show that, despite men having better status and being more satisfied of it, women is happier. The unexplained part indicates that women are more likely to be more optimistic about life.

5. CONCLUSION

This research presents empirical data supporting the notion that women experience higher levels of happiness compared to males on a global scale. This indicates that while there may be variations at the national level, women generally tend to report greater levels of satisfaction when examining larger geographic or income regions. The findings are consistent with the existing body of research conducted by Blanchflower and Oswald (2001), Stevenson and Wolfers (2009), Guven (2012), Vieira Lima (2011), Graham and Chattopadhyay (2012), and Zweig (2014). This article, however, casts light on a novel aspect of the gender happiness gap: what makes women happier than men? Do they possess superior qualities when evaluated objectively? Do individuals experience improved living conditions? Alternatively, may it be their response to the realities of existence? This research employs a decomposition technique that allows us to isolate objective characteristics and assess the impact of these variables on happiness. In the samples analyzed, women have a lower average level of education, a lower average household income, a lower likelihood of employment, and a greater likelihood of divorce or widowhood than men. This paper provides evidence that females are happier than males due to the manner in which they evaluate their characteristics.

Differences in satisfaction levels between groups may be attributable to observable characteristics (e.g., one group having superior conditions in objective dimensions of life such as income, employment, and education) or to the different ways in which groups respond to these observable characteristics. Our findings indicate that the happiness gap, which is mostly attributed to subjective well-being, cannot be accounted for by observable factors. In fact, the difference in the objective individual determinants of happiness indicates that women should be less happy than males. This indicates that the happiness difference is narrowing due to female optimism. It can be inferred from our analysis that females have a tendency to evaluate factors contributing to happiness in a more positive manner compared to males. This suggests that females possess a more optimistic perspective on life.

6. REFERENCES

Arrosa, M. L., & Gandelman, N. (2016). Happiness decomposition: Female optimism. *Journal of Happiness Studies*, 17, 731-756.

Blanchflower, D., & Oswald, A. (2001). Well-being over time in Britain and the USA. *Journal of Public Economics*, 88(7–8), 1359–1386.

Blanchflower, D., & Oswald, A. (2007). Is well-being U-shaped over the life cycle? *Social Science and Medicine*, 66(8), 1733–1749.

Blinder, A. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(462), 436–455.

Clark, A. (2003). Unemployment as a social norm: Psychological evidence from panel data. *Journal of Labor Economics*, 21, 323–351.

Clark, A. E. (2007). Born to be mild? Cohort effects don't explain why well-being is U-shaped in age. *IZA Discussion Paper No. 3170.*

Clark, A., Frijters, P., & Shields, M. (2008). Relative income, happiness and utility: An explanation of the Easterline paradox and other puzzles. *Journal of Economic Literature*, 46(1), 95–144.

Clark, A., & Oswald, A. (1994). Unhappiness and unemployment. *Economic Journal*, 104, 648–659.

Dayton, N. (1936). Marriage and mental disease. *New England Journal of Medicine*, 215, 153.

Dear, K., Henderson, S., & Korten, A. (2002). Well-being in Australia. *Social Psychiatry and Psychiatric Epidemiology*, 37(11), 503–509.

Di Tella, R., Mac Culloch, R., & Oswald, A. (2003). The macroeconomics of happiness. *The Review of Economics and Statistics,* 85(4), 809-827.

Durkheim, E. ([1897] 1997). Suicide: A study in sociology. *New York: The Free Press*.

Frey, B., & Stutzer, A. (2002). The economics of happiness. *World Economics*, 3(1), 1–17.

Graham, C., & Chattopadhyay, S. (2012). Gender and well being around the world: Some insights from the economics of happiness. *Working Paper Series number 2012-010*.

Guven, C. (2012). You can't be happier than your wife. *Journal of economic behavior and organization*, 82(1), 110–130.

Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective wellbeing. *Economic Modelling*, 20(2), 331.

López, B., & Guijarro, M. (2012). Empirical relationship between education and happiness: Evidence from Share. Retrieved from http://congresoreedes.unican.es/actas/PDFs/196.pdf

López Ulloa, B., Møller, V., & Sousa-Poza, A. (2013). How does subjective well-being evolve with age? A literature reviews. *IZA Discussion Paper No. 7328.*

Mastekaasa, A. (1993). Marital status and subjective well-being: A changing relationship? *Social Indicators Research*, 29, 249–276.

Miller, R., Hollist, C., Olsen, J., & Law, D. (2013). Marital quality and health over 20 years: A growth curve analysis. *Journal of Marriage and Family*, 75(3), 667–680.

Oaxaca, R. L. (1973). Male–female wage differentials in urban labor markets. *International Economic Review, 14*(3), 501–693.

Robins, N., & Regier, D. (1991). Psychiatric disorders in America: The epidemiologic catchment area study. *New York: Free Press.*

Stevenson, B., & Wolfers, J. (2009). The paradox of declining female happiness. *Working Paper Series 2009-11, Federal Reserve Bank of San Francisco.*

Vieira Lima, S. (2011). A cross country investigation of the determination of happiness gender gap. Accessible through: http://www.happinesseconomics.net/ocs/index.php/heirs/ markethappiness/paper/view/345/191.

Veenhoven, R. (2010). Capability and happiness: Conceptual difference and reality links. *Journal of Socio-Economics*, 39(3), 344–350.

Waite, L. (1995). Does marriage matter?. *Demography*, 32(4), 483–407.

Winkelmann, L., & Winkelmann, R. (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica*, 65(257), 1–15.

Zweig, J. S. (2014). Are women happier than men? Evidence from the Gallup World Poll. *Journal of Happiness Studies*, 1389–4978. doi: 10.1007/s10902-014-9521-8