RIETI Discussion Paper Series 25-E-021

Bridging the Gender Well-being Gap: The influence of societal and inherent factors

MI, Jie
ADB Institute

LI, Chao
Kyushu University

KEELEY, Alexander Ryota
Kyushu University

ZHANG, Jiaxu
Kyushu University

SHI, Bo
Kyushu University

MANAGI, Shunsuke
RIETI

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 25-E-021

March 2025

Bridging the Gender Well-Being Gap: The influence of societal and inherent factors*

Jie Mi[1], Chao Li[2], Alexander Ryota Keeley[2,3], Jiaxu Zhang[3], Bo Shi[2], Shunsuke Managi[2,3,4]

[1]ADB Institute, Tokyo, Japan

[2]Urban Institute, Kyushu University, Fukuoka, Japan

[3]Department of Civil Engineering, Kyushu University, Fukuoka, Japan

[4]Research Institute of Economy, Trade and Industry

Abstract

This study explores pervasive gender disparities in subjective well-being (SWB) by analyzing over 2.5 million responses collected from 168 countries between 2004 and 2022. This study uses an exogenous switching treatment effect model (ESTEM) and machine learning techniques to examine both inherent and societal factors that contribute to the gender disparity in SWB. The findings reveal that while men are naturally inclined to report higher well-being, external societal pressures significantly lower their SWB, leading to a paradox: women, despite facing more societal obstacles, often report higher SWB. In addition, the gender gap in societal treatment has widened over time, exacerbating disparities in well-being. This widening gap is primarily fueled by rigid societal norms and unequal treatment of genders across various contexts. This study underscores the urgent need for policy interventions aimed at dismantling these societal norms and promoting inclusive environments where all genders can thrive equally. By addressing both inherent and external factors, such policies can reduce the gap in well-being and foster a more equitable and supportive social framework.

Keywords: Human Well-Being, Gender Disparity, Subjective Well-Being, Machine Learning, Causal Inference, Societal Factor, Inherent Factor

JEL classification: I31, C45, J16

*This study was conducted as a part of the Project "Institutional design to achieve a society of well-being" undertaken at the Research Institute of Economy, Trade and Industry (RIETI).
The draft of this paper was presented at the RIETI DP seminar for the paper. I would like to thank participants of the RIETI DP Seminar for their helpful comments.

Introduction

Improvement in well-being has become increasingly crucial for both individuals and societies, influencing recent public policy decisions (Diener et al., 2018; Jebb et al., 2020; MacKerron, 2012). Subjective well-being (SWB) is an effective and valid indicator of human well-being (Blanchflower and Graham, 2022; Oswald and Wu, 2010). A gender gap in SWB has long been observed; the average SWB of females is generally higher than that of males (Batz and Tay, 2018; Blanchflower, D. and Bryson, A., 2024; Klasen, 2007). Understanding the nuances of the gender gap in SWB can provide valuable insights into the complex interactions among social, biological, and psychological factors that shape individual human well-being (Batz and Tay, 2018; Blanchflower, D. and Bryson, A., 2024; Eckermann, 2012). However, women having a higher SWB status remains unexplained (Blanchflower, D. and Bryson, A., 2024; Blanchflower, D.G. and Bryson, A., 2024; Diener et al., 1999). This study explores the inherent and external factors of the gender gap in SWB by employing a robust methodological framework.

Previous large-scale surveys have consistently observed that females tend to have higher SWB scores. However, considerable evidence contradicts this intuitive observation. Specifically, women frequently report higher levels of life satisfaction and happiness on average despite facing numerous social and economic disadvantages (Batz and Tay, 2018; Blanchflower, D. and Bryson, A., 2024; Blanchflower, D.G. and Bryson, A., 2024). This phenomenon is often called as “female happiness paradox” (Blanchflower, D.G. and Bryson, A., 2024). This paradox can be explained from several perspectives. Wood et al. (1989) suggested that these differences may be rooted in sex-specific social roles that influence emotional expression and processing. Women have more freedom to express their emotions and possess better emotional resilience, leading to a higher evaluation of their lives. This explanation assumes that the differences between men and women are caused by the external environment, resulting in women having relatively higher SWB. However, many findings prove that males should be happier than females. Hormonal variations significantly drive mood and emotional well-being (Kuehner, 2017; Li and Graham, 2017). For instance, progesterone and its effects on depression show how biological predispositions cause mental and physical health challenges in women (Li and Graham, 2017; Oertelt-Prigione, 2012). Additionally, external factors such as work, family, and social environments significantly contribute to the SWB disparity between genders. Gender discrimination and traditional social roles can exacerbate stress and diminish well-being (Napier et al., 2020; Schmitt et al., 2014). External pressures from traditional gender roles, particularly when rigid and restrictive, can lead to significant psychological distress (Seedat et al., 2009). However, there is no consistent explanation of gender disparities in SWB based on quantitative analysis.

Our study follows a dual-framework approach that considers both the inherent and external factors affecting SWB. For example, inherent factors can encompass biological and psychological attributes such as hormonal differences and physical structure, which influence emotional and psychological well-being (Kuehner, 2017; Li and Graham, 2017). External factors could include societal norms, economic status, and familial roles, which collectively shape the environmental context of an individual's life (Batz and Tay, 2018; Seedat et al., 2009). The inherent and external factors were abstract concepts in this study. Limited by the depth of the dataset, we could conduct only a preliminary exploration. Based on both external and inherent factors, we aimed to provide a holistic understanding of gender disparities in SWB.

Our analysis is based on a comprehensive dataset provided by the Gallup World Poll (GWP), spanning 18 years and encapsulating over 2.5 million observations across 168 countries. This rich dataset offers a unique opportunity to examine the long-standing debate about the "female happier paradox" (Blanchflower, D. and Bryson, A., 2024; Blanchflower, D.G. and Bryson, A., 2024), challenging traditional assumptions and uncovering intricate patterns that vary across different cultural and socioeconomic contexts. We employ the exogenous switching treatment effect model (ESTEM), a sophisticated analytical tool, to estimate the causal impacts of gender on human well-being in various societal contexts (e.g., Aryal et al., 2019; Kassie et al., 2014; Liu et al., 2021; Yen et al., 2009). This methodological choice distinguishes between the inherent effects of gender characteristics and the external impacts of societal expectations and roles. Additionally, our study integrates machine learning techniques to enhance the predictive accuracy and robustness of our models, unlike traditional regression approaches that are unable to capture non-linear interactions and complex patterns rooted in SWB data.

Materials and Methodology

Materials

Data Source

Our research used individual-level survey data collected by Gallup, Inc., known as GWP. The present GWP dataset spanned from 2005 to 2022 and comprised 17 surveys. Among the 17, the initial wave covered 2005 and 2006, and each subsequent wave was completed within a single year. The current dataset included 2.594 million individual observations from 168 countries and regions. Each wave of the GWP survey sampled approximately 1,000 individuals per country. In large countries such as China and Russia, Gallup sampled approximately 2000 people in a single wave. The GWP is an extensively utilized global dataset focusing primarily on human well-being (Blanchflower, D. and Bryson, A., 2024; Diener et al., 2010; Jebb et al., 2018; Joshanloo and Jovanović, 2020). Detailed information on the sampling methods and data collection process can be found on Gallup's website (https://www.gallup.com/178667/gallup-world-poll-work.aspx).

During the data cleaning process, some observations were excluded, resulting in 1,911,212 observations used in the analysis. The first three waves, which did not include income questions, were removed entirely. For other waves, if the respondents did not provide their income, the mean income of the respondents' respective countries for that wave was used to fill in missing values. The mean income was calculated by averaging the available income data from the GWP survey for the corresponding country and wave. If a country did not have income data for a particular wave, then the observations for that country in that wave were removed. After this step, we obtained 2,172,297 observations.

As our dependent variable was well-being, only respondents who answered the well-being question were retained, leaving 2,141,833 observations. Additionally, given that disabilities significantly affect human well-being (Diener et al., 2018; Lucas, 2007), observations without answers to the disability question were excluded, reducing the dataset to 2,025,803 observations. Furthermore, age, sex, marital status, and employment status were mandatory variables (Diener et al., 2018; Li and Managi, 2023; Lucas et al., 2004); and observations missing any of these data points were also excluded. Consequently, the final dataset comprised 1,911,212 observations. The details regarding each country and wave are summarized in **Supplementary Material Table S1**.

Variables

This study aimed to explore the underlying causes of gender disparities in human well-being. SWB is a valid measure of human well-being (Oswald and Wu, 2010). A key method to assess SWB is the overall life evaluation, which captures individuals' reflections on the quality of their lives (Diener et al., 2018). The GWP uses an 11-

point Cantril ladder to gauge overall human well-being. Respondents were asked to envision a ladder with 11 steps and select the one they felt represented their current life situation, where the lowest step (0) represented the worst possible life and the highest step (10) represented the best possible life for them. This selected step constitutes the respondents' life evaluation score. The Cantril ladder's simplicity and its extensive use in previous research make it an appropriate choice for the dependent variable (Blanchflower, D. and Bryson, A., 2024; Deaton, 2008; Diener et al., 2018; Joshanloo and Jovanović, 2020).

We considered 63 independents variables in this study based on previous studies with GWP data (Arrosa and Gandelman, 2016; Deaton, 2008; Joshanloo and Jovanović, 2020; Smith and Wesselbaum, 2023) and data availability, including "Wave," "Country," "Household Income," "Health Disability Dummy," "Female Dummy," "Age," "Marital Status," "Employment," "Children Under 15 Dummy," "Feeling of Income," "Income Level," "Having Relatives to Rely on Dummy," "Living Standard Changing Direction," "Not Having Enough Food Dummy," "Not Having Enough Shelter Dummy," "Feeling Well Rested Dummy," "Feeling Respected Dummy," "Smiling Dummy," "Doing Interesting Things Dummy," "Having Enjoyment Dummy," "Suffering Physical Pain Dummy," "Feeling Worry Dummy," "Feeling Sad Dummy," "Feeling Stress Dummy," "Feeling Anger Dummy," "Feeling Satisfied with City Dummy," "Economic Changing Direction," "Thinking Good Time to Find Job Dummy," "Feeling Satisfied with Public Transportation Dummy," "Feeling Satisfied with Road Dummy," "Feeling Satisfied with Education Dummy," "Feeling Satisfied with Air Quality Dummy," "Feeling Satisfied with Water Quality Dummy," "Feeling Satisfied with Healthcare Dummy," "Feeling Satisfied with Affordable House Dummy," "Feeling Satisfied with Opportunity to Make Friends Dummy," "Thinking Good Place for Ethical Minority Dummy," "Thinking Good Place for Gay or Lesbian Dummy," "Thinking Good Place for Immigrants Dummy," "Donated Recently Dummy," "Did Volunteer Recently Dummy," "Helped Stranger

Dummy," "Voiced Opinion to Official Dummy," "Feeling Confident in Local Police Dummy," "Feeling Safe of Alone Night Walking Dummy," "Having Been Stolen Dummy," "Having Been Assaulted Dummy," "Thinking Religion Importance Locally Dummy," "Thinking Children Respected Locally Dummy," "Having Opportunity for Children Learning Locally Dummy," "Feeling Women Respected Dummy," "Feeling Satisfied with Poverty Alleviation Dummy," "Feeling Satisfied with Environmental Efforts Dummy," "Feeling Freedom of Choosing Life Dummy," "Feeling Confidence in Military Dummy," "Feeling Confidence in Judicial System Dummy," "Feeling Confidence in National Government Dummy," "Feeling Confidence in Financial System Dummy," "Feeling Confidence in Election Honesty Dummy," "Feeling Freedom of Media Dummy," "Prevailing Corruption within Business Dummy," "Prevailing Corruption within Government Dummy," and "Approving of Leadership Performance Dummy." "Female Dummy" is typically exogenous. Generally, "Female Dummy" would not change with other variables simply. **Table 1** provides a summary of all variables, excluding "Wave" and "Country." Detailed descriptions of each survey question and the corresponding explanations are provided in **Supplementary Material Table S2**.

Methods

Exogenous Switching Treatment Effect Model (ESTEM)

To explore the treatment effects and base heterogeneity effects on the differences in well-being between male and female populations, we employed ESTEM. The ESTEM estimates causal effects through counterfactual analyses (Kassie et al., 2014; Liu et al., 2021). It divided all observations into two groups based on the exogenous variable gender, creating two sub-datasets corresponding to the male and female populations. Each subdataset was then used to train the respective models. This abstraction is expressed as follows:

$$\begin{cases} model_{ma} = \Theta(\boldsymbol{X}_{ma}, \boldsymbol{y}_{ma}, \pi_{ma}) \\ model_{fe} = \Theta(\boldsymbol{X}_{fe}, \boldsymbol{y}_{fe}, \pi_{fe}) \end{cases} \tag{1}$$

where $\boldsymbol{y}_{ma}$ and $\boldsymbol{y}_{fe}$ represent the dependent variables of each sub-dataset, $\boldsymbol{X}_{ma}$ and $\boldsymbol{X}_{fe}$ represent the independent variables of each sub-dataset, $model_{ma}$ and $model_{fe}$ are corresponding models for two sub-datasets, and π_{ma} and π_{fe} are necessary settings for the model training. The independent variable values of each observation were regarded as individual features and the models were regarded as general external treatments for each group. By treating different groups of people differently, independent variables and human well-being statuses can produce different results. Four prediction scenarios were obtained in this study.

$$\hat{\boldsymbol{y}}_{ma}^{ma} = model_{ma}(\boldsymbol{X}_{ma}) \tag{2}$$

$$\hat{\boldsymbol{y}}_{fe}^{fe} = model_{fe}(\boldsymbol{X}_{fe}) \tag{3}$$

$$\hat{\boldsymbol{y}}_{fe}^{ma} = model_{ma}(\boldsymbol{X}_{fe}) \tag{4}$$

$$\hat{\boldsymbol{y}}_{ma}^{fe} = model_{fe}(\boldsymbol{X}_{ma}) \tag{5}$$

where $\hat{\boldsymbol{y}}_{ma}^{ma}$ represents the prediction of the male population based on the male population model, $\hat{\boldsymbol{y}}_{fe}^{fe}$ represents the prediction of the female population based on the female population model, $\hat{\boldsymbol{y}}_{fe}^{ma}$ represents the prediction of the female population based on the male population model, and $\hat{\boldsymbol{y}}_{ma}^{fe}$ represents the prediction of the male population based on the female population model. Among the four scenarios, **Equations (2)** and **(3)** are factual predictions, whereas **Equations (4)** and **(5)** are counterfactual predictions.

The treatment effects on the two sexes were the same population treated in different ways. These values can be computed as follows:

$$TE_{fe}^{fe-ma} = \overline{\hat{\boldsymbol{y}}_{fe}^{fe}} - \overline{\hat{\boldsymbol{y}}_{fe}^{ma}} \tag{6}$$

$$TE_{ma}^{fe-ma} = \overline{\hat{\boldsymbol{y}}_{ma}^{fe}} - \overline{\hat{\boldsymbol{y}}_{ma}^{ma}} \tag{7}$$

where TE_{fe}^{fe-ma} represents the treatment effect on the female population and TE_{ma}^{fe-ma} represents the treatment effect on the male population. The base heterogeneity effects between the two sexes were different population groups treated in the same way. These values can be estimated as follows:

$$BHE_{fe-ma}^{fe} = \overline{\hat{\boldsymbol{y}}_{fe}^{fe}} - \overline{\hat{\boldsymbol{y}}_{ma}^{fe}} \tag{8}$$

$$BHE_{fe-ma}^{ma} = \overline{\hat{\boldsymbol{y}}_{fe}^{ma}} - \overline{\hat{\boldsymbol{y}}_{ma}^{ma}} \tag{9}$$

where BHE_{fe-ma}^{fe} represents the base heterogeneity effects on the treatments for the female population and BHE_{fe-ma}^{fe} represents that on the treatments for the male population.

Previous studies normally employed linear regression models, mainly ordinary least squares and logistic regression, to achieve the function of the model for each subdataset in the ESTEM (Kassie et al., 2014; Liu et al., 2021). However, these studies rarely verified whether these models have generalization and prediction capabilities. These models were tested using cross-validation. We conducted 10-fold cross-validation based on our dataset. The process first randomly shuffles the entire dataset, divides the dataset into ten parts, and then uses nine parts to train a model and the remaining part to test the performance of the model. The training-test process was conducted ten times, and each time, the training and test datasets were different. The 10-fold cross-validation was used to examine the generalization capability, method stability, and robustness. The test performance for each fold was similar. The mean, minimum, and maximum values of R^2 in the 10-fold cross-validation for the male population subdataset were 33.46%, 30.66%, and 34.04%, respectively. For the female population sub-dataset, the mean, minimum, and maximum values were 33.55%, 30.22%, and 34.15%, respectively. According to the 10-fold cross-validation results, the linear regression models had the ability to predict, but their stability was relatively poor; specifically, the performance variation was almost 4%.

Machine Learning Empowered Improvements

The relatively complex data structure helps use machine learning methods. These methods do not assume linear relationships but minimize residual errors, so they tend to achieve better predictive performance (Breiman, 2001; Chen and Guestrin, 2016). Mainstream machine learning algorithms are divided into connectionism and symbolism. Artificial neural networks are the core technologies of connectionism, whereas models based on decision trees represent symbolism. Artificial neural networks have greater structural freedom and more hyperparameters, which is not conducive to the effective selection of models with better performance. Among the numerous tree-based algorithms, random forest (Breiman, 2001), extreme gradient boosting (XGBoost) (Chen and Guestrin, 2016), light gradient boosting machines (LightGBMs) (Ke et al., 2017), and CatBoost (Prokhorenkova et al., 2018) have gained the most attention because of their performance. Additionally, the potential model must support GPU acceleration. XGBoost, LightGBM, and CatBoost are improved gradient boosting models that solve parallel computing issues. To compare these four algorithms, we fine-tuned their hyperparameters. The best models of random forest, XGBoost, LightGBM, and CatBoost were 36.46%, 37.81%, 37.38%, and 37.53%, respectively, based on a 10-fold cross-validation of the entire dataset. Additionally, the variations in each fold of R^2 for the four algorithms were less than 1%. Considering its predictive ability and stability, we replaced the linear regression with the XGBoost regression in this study.

We adopted a cross-validation method to search for the best hyperparameter sets for the XGBoost models. We used a 10-fold cross-validation as the cross-validation method. However, owing to limited computing resources, we performed cross-validation three times to balance the time consumption and ensure the stability of the search process. The metric is used to evaluate the search process. The XGBoost model training processes for the male and female populations are as follows:

$$\begin{cases} XGB_{ma} = \Theta(\boldsymbol{Xtrain}_{ma}, \boldsymbol{ytrain}_{\boldsymbol{ma}}, \boldsymbol{Hyperparameter}_{ma}) \\ XGB_{fe} = \Theta(\boldsymbol{Xtrain}_{fe}, \boldsymbol{ytrain}_{fe}, \boldsymbol{Hyperparameter}_{fe}) \end{cases} \tag{10}$$

where XGB_{ma} and XGB_{fe} represent the well-trained XGBoost regression models based on the male and female population datasets, respectively; $\boldsymbol{Xtrain}_{fe}$ and $\boldsymbol{Xtrain}_{ma}$ represent the independent variables of the training dataset split from the female and male population sub-datasets, respectively; $\boldsymbol{ytrain}_{fe}$ and $\boldsymbol{ytrain}_{ma}$ represent the real human well-being status of those two sub-datasets; $\boldsymbol{Hyperparameter}_{fe}$ and $\boldsymbol{Hyperparameter}_{ma}$ represent two sets of hyperparameters to train high-accuracy XGBoost models for female and male population sub-dataset; and Θ represents the training process. The R^2 value of the test dataset is computed as follows:

$$\boldsymbol{y\widehat{test}}_{fe} = XGB_{fe}(\boldsymbol{Xtest}_{fe}) \tag{11}$$

$$R^2_{test\ fe} = 1 - \frac{(\boldsymbol{ytest}_{fe} - \boldsymbol{y\widehat{test}}_{fe})^2}{(\boldsymbol{ytest}_{fe} - \overline{\boldsymbol{ytest}_{fe}})^2} \tag{12}$$

$$\boldsymbol{y\widehat{test}}_{ma} = XGB_{ma}(\boldsymbol{Xtest}_{ma}) \tag{13}$$

$$R^2_{test\ ma} = 1 - \frac{(\boldsymbol{ytest}_{ma} - \boldsymbol{y\widehat{test}}_{ma})^2}{(\boldsymbol{ytest}_{ma} - \overline{\boldsymbol{ytest}_{ma}})^2} \tag{14}$$

where $\boldsymbol{y\widehat{test}}_{fe}$ and $\boldsymbol{y\widehat{test}}_{ma}$ represent the predicted values corresponding to the input test datasets, $\boldsymbol{Xtest}_{fe}$ and $\boldsymbol{Xtest}_{ma}$, of the well-trained XGBoost models XGB_{fe} and XGB_{ma}, $\overline{\boldsymbol{ytest}_{fe}}$ and $\overline{\boldsymbol{ytest}_{ma}}$ are the mean real values of human well-being in each sub-dataset, and $R^2_{test\ fe}$ and $R^2_{test\ ma}$ represent the R^2 of the test datasets for the model trained by the training datasets. Combining **Equations (10)–(14)**, R^2 of the test dataset is highly related to the hyperparameters.

The hyperparameter set for the XGBoost model includes the following: the number of trees (“n_estimators”), learning rate (“learning_rate”), maximum depth of each tree (“max_depth”), subsample ratio of training instances (“subsample”), minimum loss reduction required for a new split (“gamma”), minimum sum of instance weight needed in a child leaf (“min_child_weight”), maximum step size that an XGBoost model’s weight can change at each boosting iteration (“max_delta_step”), L1 regularization term on weights (“reg_alpha”), and L2 regularization term on weights (“reg_lambda”). The abbreviations in quotation marks correspond to the XGBoost Python API used by other researchers to facilitate reproduction and imitation. These abbreviations are enclosed in quotation marks to distinguish them from regular text.

We employed Bayesian hyperparameter optimization to identify the optimal hyperparameter set (Wu et al., 2019). This process involved four steps: initializing several sets of hyperparameters, constructing a surrogate function, selecting and evaluating the next set of hyperparameters, and updating the surrogate function. The third and fourth steps were iterated multiple times to obtain a high-performance hyperparameter set. In this study, we performed 20 iterations. In simple terms, the surrogate function used a set of hyperparameters as inputs, and outputs the estimated R^2 of the test dataset. The surrogate functions were continuously optimized through these iterations. The ranges for the hyperparameters were as follows: 100–5000 for “n_estimators,” 0.001–0.1 for “learning_rate,” 3–32 for “max_depth,” 0.5–1.0 for “subsample,” 0.001–10 for “min_child_weight,” 0.001–10 for “max_delta_step,” 0.001–10 for “gamma,” 0.001–10 for “reg_alpha,” and 0.001–10 for “reg_lambda.” We compared the results of Bayesian hyperparameter optimization with 20 iterations of grid search results using over 3,000 possible hyperparameter sets. The Bayesian optimization results were relatively better. While further fine-tuning using grid search could potentially improve performance, the time cost is prohibitively high, and the improvement might be marginal. Therefore, we used Bayesian hyperparameter optimization to fine-tune all the XGBoost models in this study.

Robust Factual and Counterfactual Predictions

Whether it is a factual or counterfactual prediction, well-trained models should never have been trained using the information of an observation that needs to be predicted. In other words, no observations should be included in either the test or the training datasets. Based on the XGBoost models, the predictions of the actual and counterfactual well-being statuses of each gender group were computed as follows:

$$\widehat{\boldsymbol{ytest}_{fe}^{fe}} = XGB_{fe}(\boldsymbol{Xtest}_{fe}) \tag{15}$$

$$\widehat{\boldsymbol{y}_{ma}^{fe}} = XGB_{fe}(\boldsymbol{X}_{ma}) \tag{16}$$

$$\widehat{\boldsymbol{y}_{fe}^{ma}} = XGB_{ma}(\boldsymbol{X}_{fe}) \tag{17}$$

$$\widehat{\boldsymbol{ytest}_{ma}^{ma}} = XGB_{ma}(\boldsymbol{Xtest}_{ma}) \tag{18}$$

where $\widehat{\boldsymbol{ytest}_{fe}^{fe}}$, and $\widehat{\boldsymbol{ytest}_{ma}^{ma}}$ are the predicted well-being status of observations in the test datasets of the female and male population sub-datasets, respectively, which are the actual cases; $\widehat{\boldsymbol{y}_{ma}^{fe}}$ represents the predicted well-being status of the male population treated as female; and $\widehat{\boldsymbol{y}_{fe}^{ma}}$ represents the predicted well-being status of the female population treated as male. $\widehat{\boldsymbol{y}_{ma}^{fe}}$ and $\widehat{\boldsymbol{y}_{fe}^{ma}}$ are counterfactual predicted well-being. Based on this work framework, each training-prediction process could complete 10% of the actual predictions. Hence, we conducted the process 10 times, similar to the 10-fold cross-validation. Because the 10-fold predictions are from 10 and not exactly the same models based on the same hyperparameter setting and different training datasets, the individual-level predictions of each observation may not be stable and robust. To improve the stability, we conducted 10-fold predictions in 10 epochs. In each epoch, the entire dataset was divided into ten parts in different ways. In this computation process, each gender population responds to 100, and not exactly the same models are based on the same hyperparameter setting. The 10-epoch 10-fold training-prediction process can be written as follows:

$$\begin{cases} XGB_{fe@EF} = \Theta(\boldsymbol{Xtrain}_{fe@EF}, \boldsymbol{ytrain}_{fe@EF}, \boldsymbol{Hyperparameter}_{fe}) \\ XGB_{ma@EF} = \Theta(\boldsymbol{Xtrain}_{ma@EF}, \boldsymbol{ytrain}_{ma@EF}, \boldsymbol{Hyperparameter}_{ma}) \end{cases} \tag{19}$$

$$\widehat{\boldsymbol{ytest}_{fe@EF}^{fe@EF}} = XGB_{fe@EF}(\boldsymbol{Xtest}_{fe@EF}) \tag{20}$$

$$\boldsymbol{y}_{ma}^{\widehat{fe@EF}} = XGB_{fe@EF}(\boldsymbol{X}_{ma}) \tag{21}$$

$$\boldsymbol{y}_{fe}^{\widehat{ma@EF}} = XGB_{ma@EF}(\boldsymbol{X}_{fe}) \tag{22}$$

$$\widehat{\boldsymbol{ytest}_{ma@EF}^{ma@EF}} = XGB_{ma@EF}(\boldsymbol{Xtest}_{ma@EF}) \tag{23}$$

where F represents the fold indicator, E represents the epoch indicator, $XGB_{fe@EF}$ represents the XGBoost model trained for the female population in the F fold of the E epoch, $\widehat{\boldsymbol{ytest}_{fe@EF}^{fe@EF}}$ represents the well-being predictions of the test data of the female population sub-dataset based on the corresponding model in the F fold of the E epoch, $\boldsymbol{y}_{ma}^{\widehat{fe@EF}}$ represents the prediction of the male population based on the model trained by the female population sub-dataset in the F fold of the E epoch, and the explanations of other symbols are similar. In **Equations (19)–(23)**, each actual case should be predicted 10 times, and each counterfactual case should be predicted 100 times. We used the individual-level mean predicted values of each observation to obtain the actual and counterfactual predictions used in **Equations (6)–(9)**.

$$\widehat{\boldsymbol{y}}_{fe}^{fe} = ilm(\prod \widehat{\boldsymbol{ytest}_{fe@EF}^{fe@EF}}) \tag{24}$$

$$\widehat{\boldsymbol{y}}_{ma}^{fe} = ilm(\prod \boldsymbol{y}_{ma}^{\widehat{fe@EF}}) \tag{25}$$

$$\widehat{\boldsymbol{y}}_{fe}^{ma} = ilm(\prod \boldsymbol{y}_{fe}^{\widehat{ma@EF}}) \tag{27}$$

$$\widehat{\boldsymbol{y}}_{ma}^{ma} = ilm(\prod \widehat{\boldsymbol{ytest}_{ma@EF}^{ma@EF}}) \tag{28}$$

where $\widehat{\boldsymbol{y}}_{fe}^{fe}$, $\widehat{\boldsymbol{y}}_{ma}^{fe}$, $\widehat{\boldsymbol{y}}_{fe}^{ma}$, and $\widehat{\boldsymbol{y}}_{ma}^{ma}$ represent the individual-level mean predicted well-being status of the female and male populations based on models trained by the female and male populations, and ilm represents the individual-level mean method. As each value was predicted multiple times, the predicted value obtained using the averaging method was stable.

Individual-level Investigation on the Treatment Effects

In this step, each independent variable's contribution to the treatment effects due to gender was investigated. These effects are caused by external environmental factors. The characteristics that increase or decrease the effects of sex require further exploration. We used the absolute value of the treatment effects as the dependent variable of the new XGBoost model. Ideally, the treatment difference between the female and male populations should be smaller, which means a gender-equal society. The independent variables of this model are the same as in previous steps except the exogenous variable, "Gender Female Dummy." Similarly, we employed Bayesian hyperparameter optimization to calibrate the hyperparameter set to determine the relationship. The XGBoost model training process for the individual treatment effects can be expressed as follows:

$$XGB_{TE} = \Theta(\boldsymbol{Xtrain}_{TE}, |\boldsymbol{ytrain}_{TE}|, \boldsymbol{Hyperparameter}_{TE}) \quad (15)$$

where XGB_{TE} represents the well-trained XGBoost regression models for the individual effects, $\boldsymbol{Xtrain}_{TE}$ represents the independent variables of the training dataset, $\boldsymbol{ytrain}_{TE}$ represents the individual-level treatment effect of the training dataset, and $\boldsymbol{Hyperparameter}_{TE}$ represents the hyperparameter set for the best XGBoost to graph the relationship between the independent variables and treatment effects.

Because XGBoost is completely non-parametric, an explanation method is required to compute the contributions of each variable. The Shapley additive explanation (SHAP) method is a novel and effective approach to estimate the individual contribution of each independent variable to the dependent variable in machine learning models. This method uses cooperative game theory and Shapley values to ensure that the contributions of independent variables to the model's predictions are fairly and evenly distributed (Lundberg et al., 2020; Molnar, 2020). The Shapley values were computed using the change in the predictions of a well-trained machine learning model before and after adding a specific independent variable to all possible subsets of other independent variables and then averaging these marginal contributions. The contribution of each independent variable at the individual level can be expressed as

$$\boldsymbol{SHAPtest}_{TE} = SHAP(XGB_{TE}, \boldsymbol{Xtest}_{TE}) \quad (16)$$

where XGB_{TE} represents the XGBoost regression model based on the training dataset from the gender treatment effect investigation dataset; $SHAP$ represents the standard SHAP algorithm; and $\boldsymbol{SHAPtest}_{tot}$ represents the SHAP values of each independent variable and observation in the test dataset. The ratio of the training to test datasets was 9:1. However, the SHAP is computationally expensive. The test dataset contained approximately 200,000 representative observations. Therefore, these calculations were performed only once.

Results

Well-being Differences between Gender Groups

Well-being Differences between Gender Groups

The entire dataset, including 1,911,212 observations, was divided into female and male population groups with 1,017,224 and 893,988 observations, respectively. The mean SWB scores of the female and male populations were 5.569 and 5.466, respectively. We conducted a t-test between the SWB values for each sex group. If the t-test result is significant, the distributions of SWB values for each gender group are significantly different. The p-value of the t-test between the two sex groups was < 0.1%. In other words, the SWB status of the male and female populations was significantly different. Empirically, at first glance, the female population tends to have a better SWB status than the male population.

Models for Two Gender Groups Fine-tuning and Their Performance

We used Bayesian hyperparameter optimization with 20 iterations based on cross-validation to calibrate the best hyperparameter sets for both gender groups. The average R^2 of the best model for the female population was 38.04%, which was derived from three single-fold test R^2 values of 38.30%, 37.77%, and 38.04%. The average training R^2 for this model was 46.84% based on single-fold training R^2 values of 46.82%, 46.88%, and 46.84%. This indicates overfitting in the model for the female population. To address this, 10-epoch 10-fold predictions were necessary. The best hyperparameter set for the female population included "n_estimators" of 1136, "learning_rate" of 0.0252, "max_depth" of 8, "subsample" of 0.797, "min_child_weight" of 0.033, "max_delta_step" of 3.066, "gamma" of 6.636, "reg_alpha" of 0.001, and "reg_lambda" of 0.030. For the male population, the average R^2 of the best model was 37.98% based on three single-fold test R^2 values of 38.30%, 37.84%, and 37.80%. The average training R^2 was 53.29%, derived from single-fold training R^2 values of 53.27%, 53.29%, and 53.31%. The best hyperparameter set for the male population included "n_estimators" of 2553, "learning_rate" of 0.0080, "max_depth" of 10, "subsample" of 0.640, "min_child_weight" of 0.016, "max_delta_step" of 0.992, "gamma" of 0.002, "reg_alpha" of 0.001, and "reg_lambda" of 0.394. **Table 2** presents the R^2 test results from the 10-epoch, 10-fold predictions. Each row in **Table 2** summarizes the 100 test R^2 values computed using a 9:1 training-test ratio. The model for each specific gender group predicts the corresponding observations ten times and the observations in the other gender group 100 times. The performance of each model remains acceptable, as the XGBoost models significantly enhanced prediction accuracy and stability compared to the linear models.

Overall Base Heterogeneity and Treatment Effects

Table 3 illustrates the base heterogeneity and treatment effects of sex on well-being. All populations obtained the lowest well-being status when they were treated as males because the treatment effects of shifting from the treatment for the female population to the treatment for the male population were all positive. Specifically, if a person is treated as female, they are prone to have a better well-being status compared to when they are treated as male. Treatment of the male population significantly reduces human well-being. For the base heterogeneity effects, the relatively male group tended to have higher well-being because, under the same treatment, the difference between the well-being predictions of the female and male populations was significantly negative. The base heterogeneity effects of being male are numerically unable to offset the treatment effects of being treated as male. Thus, empirically, the female appear more likely to achieve higher levels of well-being.

Variations of Treatment and Base Heterogeneity Effects

Temporal Variation of Treatment and Base Heterogeneity Effects

Figure 1 illustrates the temporal variations in treatment and base heterogeneity effects. All values in **Figure 2** are significant, and the p-values for the t-tests are smaller than 0.1%. The treatment effects between the female and male populations gradually increase numerically and temporally. The temporal trend of the base heterogeneity effects between the two sexes was downward, but the trends were not stable.

Country-level Variation of Treatment and Base Heterogeneity Effects

The country-level treatment and base heterogeneity effects are shown in **Figure 2** and summarized in

Supplementary Material Table S3. A total of 184 countries and regions were recorded. The status of most countries is consistent with global patterns. Specifically, in most countries, the treatment effects of shifting from female to male treatment are positive, as shown in **Figures 2.a** and **2.b**, and the base heterogeneity effects between the female and male populations are almost negative, as illustrated in **Figures 2.c** and **2.d**. Some countries and regions have a status that differs from the global pattern. The base heterogeneity effects between female and male populations are significantly positive in Hong Kong, Japan, Korea, the Republic of Indonesia, Singapore, Malaysia, Kuwait, Turkmenistan, Kosovo, Yemen, Gambia, and Algeria. In these countries, the female population tends to have better well-being. Additionally, in Benin, Bhutan, Ghana, Chad, and Togo, treatment of the female population is stricter than that of the male population. Furthermore, in Sierra Leone, the pattern is completely opposite to that of the global status. In this country, the female population is treated more strictly and is more likely to have a better well-being.

Contributions to Treatment Effects of Gender

Model for Variable Contribution Investigation

Similarly, Bayesian hyperparameter optimization with 20 iterations was employed to calibrate the best hyperparameter set. The average test R^2 of the best model was 81.51%, which was derived from three single-fold test R^2 values: 81.37%, 81.64%, and 81.51%. The average training R^2 for this model was 95.72%, based on single-fold training R^2 values of 95.71%, 95.72%, and 95.73%. The best hyperparameter set includes "n_estimators" of 5000, "learning_rate" of 0.0393, "max_depth" of 29, "subsample" of 0.643, "min_child_weight" of 10.0, "max_delta_step" of 29, "gamma" of 0.012, "reg_alpha" of 2.883, and "reg_lambda" of 0.954.

The Contribution Variable Results

The SHAP values represent distributed contributions of each variable for each observation. If the absolute average of the SHAP values for a certain variable is larger, the variable has a larger impact on the dependent variable; in this part, the treatment effects between genders. **Table 4** lists the means of all the 61 potential variables and their corresponding 95% confidence interval. Among the 61 variables, the average contributions of all 57 variables were not equal to 0 at the 95% level. The average contribution of the corresponding variables to gender disparity in external treatment affects human well-being. Larger average absolute values indicate stronger impacts on that disparity and vice versa. **Figure 3** illustrates the relationship between the real values of the top nine variables with the largest impact on the treatment effects between sexes and their SHAP values. According to **Figure 3.a**, the SHAP value of income decreases as income increases. A higher income can reduce the treatment effects between the two sexes. In other words, gender has relatively fewer effects on external treatment in high-income populations. In **Figure 3.b**, if the population does not have food issues, they are prone to stronger treatment effects between the sexes. In other words, without adequate food, the effect of sex on SWB was reduced. Among the six marital statuses, the married and separated populations had no effect on the treatment effects between genders; the single status tended to enlarge the treatment effects; and the divorced, widowed, and domestic partner statuses were likely to reduce the treatment effects between genders, in **Figure 3.c**. Based on **Figure 3.d**, among the six employment statuses, full-time and part-time employed people who do not want full-time employment could reduce the treatment effects between genders, while part-time employed people who want full-time employment, are unemployed, or out of the workforce tend to have larger treatment effects. For the population aged 25–30 years, the external treatments between genders peaked, as shown in

Figure 3.e. As shown in **Figure 3.f**, thinking that living standards are improving would increase the treatment difference between genders compared to thinking of no change or worsening living standards. The sex difference in SWB caused by the external environment is only apparent among those who expect an improvement in their living standards. According to **Figure 3.g**, the treatment effects between sexes gradually increased over time. Based on **Figure 3.h**, thinking that local children are not respected would lead to a higher treatment difference between the genders. In **Figure 3.i**, the income level results are consistent with the findings in **Figure 3.a**; that is, a higher income can reduce external treatment effects.

Discussion

This study examined the relationship between gender and human well-being using a large-scale, internationally representative dataset, GWP, spanning 2005 to 2022, encompassing over 1.9 million observations from 168 countries. Employing a high-accuracy machine-learning-based ESTEM model, the findings revealed that the gender gap in well-being is attributable to both inherent and external factors. This study contributes to existing literature in several ways. First, males are inherently happier than females; however, external environments force males to have lower human well-being. This finding could holistically explain the "female happier paradox." Second, the treatment disparity increases over time; that is, the situation for males gradually worsens. Third, owing to differences in social culture among countries, the impact of sex-related external environments on SWB is not the same. Fourth, among the factors that affect treatment effects between genders, income, age, marital status, and employment status has greatest impact. Fifth, in combination with machine learning technology, we created a paradigm for applying switching models to machine learning models. These findings suggest that policies should focus on dismantling rigid gender norms and creating more inclusive environments to support the emotional and psychological well-being of all genders, thereby reducing disparities in human well-being.

The intuitive data summary shows that females reported higher levels of SWB than males according to the average values from the GWP. The data summary is consistent with mainstream views and the results of previous studies: in terms of three widely used SWB indicators, specifically happiness, life satisfaction, and Cantril ladder, females are prone to obtain relatively higher average scores (Ballas and Thanis, 2022; Blanchflower, D. and Bryson, A., 2024; Diener et al., 2018; MacKerron, 2012; Wood et al., 1989). In the early days, researchers expected that with good controls, they would see no differences between genders (Bartram, 2022; Diener et al., 1999), but subsequent surveys and research repeatedly demonstrated a gender gap in SWB. There are various explanations for this sex gap. Wood et al. (1989) reviewed studies on sex disparities in well-being and found relatively better happiness and life satisfaction evaluations among females. Wood et al. (1989) explained this gender gap from the perspective of the social roles of men and women. Female gender roles typically entail greater emotional responsiveness. In other words, gender role expectations typically allow women to express their emotions more freely than men; this argument has been repeatedly demonstrated in subsequent studies (Batz and Tay, 2018; Diener et al., 1999; Simon and Nath, 2004). Furthermore, women are more receptive to emotional expressions and often consider intense emotional reactions socially acceptable. Emotional resilience among women is also crucial in explaining their higher life satisfaction scores reported by women (Eckermann, 2012). Because women have relatively stronger emotions when asked for happiness and satisfaction evaluations, they are likely to have a higher evaluation. In summary, societal acceptance of women expressing emotions may contribute to reporting higher SWB than men under the same circumstances.

On the other hand, other studies demonstrate that, although having higher SWB, women are also more vulnerable and sensitive to negative emotional and psychological factors, which is regarded as the "female happiness paradox" (Arrosa and Gandelman, 2016; Blanchflower, D. and Bryson, A., 2024; Herbst, 2011). For example, women suffer more from worse mental health (Blanchflower, D. and Bryson, A., 2024), restless sleep (Theorell-Haglöw et al., 2018), self-reported health (Boerma et al., 2016), and depression (Parker and Brotchie, 2010). The potential risk factors for the gender gap in depression and mental disorders include the effect of sex hormones, progesterone's effect (Li and Graham, 2017), women's blunted hypothalamic-pituitary-adrenal axis response to stress, gender-based violence, and a lack of gender equality (Kuehner, 2017) (Seedat et al., 2009). Moreover, due to differences in hormones and physical body, women are more vulnerable to autoimmune conditions (Oertelt-Prigione, 2012) and have relatively poorer physical power performance than men with similar conditions (Ben Mansour et al., 2021). Additionally, men typically have higher testosterone levels, which are associated with greater risk-taking abilities (Apicella et al., 2015), better coping with fatigue, and the maintenance of positive psychological traits (Zitzmann, 2020). In general, in terms of the various gender gaps, especially those based on inherent factors, men should have had better SWB levels.

Essentially, the "female happiness paradox" means women should not have a relatively better SWB status according to their own conditions. Although previous studies have analyzed and explained this issue from multiple perspectives (Blanchflower, D. and Bryson, A., 2024; Diener et al., 1999), these reasons have not been systematically classified. The reasons for this difference in subjective wellbeing can be classified into inherent and external factors. Inherent factors include innate differences between the sexes, such as hormone levels (Li and Graham, 2017) and physical differences (Kuehner, 2017). External factors are acquired from social, family, and work environments such as the acceptance of women's emotional expressions (Batz and Tay, 2018; Diener et al., 1999; Simon and Nath, 2004), gender discrimination (Napier et al., 2020; Schmitt et al., 2014), and social role traditionality (Seedat et al., 2009). Because this study did not attempt to enumerate and categorize all the factors, inherent and external factors should be considered as two abstract concepts. In this study, the base heterogeneous effects were caused by inherent factors, whereas the treatment effects were caused by external factors.

Strict external treatment from society, family, and workplace erases the male advantage in the base heterogeneity difference. Specifically, our treatment effect analyses demonstrated a significant decline in well-being when individuals were treated as males. Despite the inherent advantages, men have shorter life expectancies and are more prone to die from despair-related causes, including drug overdoses (El-Jahel et al., 2023), liver cirrhosis, and suicide (Case and Deaton, 2015). These disparities may be partially explained by the societal consequences of deviating from gender role expectations. Both men and women who engage in behaviors contrary to these expectations often face considerable social disapproval, which can negatively impact their success and happiness (Eagly and Karau, 2002). Gender stereotypes for males, including norms surrounding masculinity, such as the male-breadwinner role, are deeply entrenched in societal and cultural values. Many men are conditioned to embody the traits of toughness, strength, and bravery, often suppressing their vulnerability or feminine traits. Moreover, pressures associated with gender roles, such as husbands and fathers, can lead to significant social stress, which adversely affects mental wellbeing and life satisfaction. For instance, in their analysis of married and cohabiting heterosexual couples across 29 countries from 2004 to 2014, Gonalons-Pons and Gangl (2021) found that adherence to male-breadwinner norms significantly influences the link between male unemployment and the risk of separation or divorce.

According to our results, the “female happiness paradox” could be interpreted by the inherent and external differences between genders. Specifically, women tend to have a poorer human well-being status, as shown by the base heterogeneous effect analysis, due to their innate hormone levels and physical conditions. However, stricter external environments press men to be less happy and satisfied with their lives, as supported by the treatment effect results. Contrary to these intuitive observations, a more detailed analysis of the inherent and external factors influencing gender disparities in SWB suggests that men inherently exhibit higher levels of well-being. Simply put, if the external environment, including society, workplace, and family, treats the female and male populations in the same way, the male population is prone to obtain a relatively higher SWB. This was the base heterogeneity difference between the sexes. Externally, in fact, social norms, like masculine norms and gender roles, force men to be ideals of toughness, emotional restraint, and dominance, which are highly related to mental health disorders (Wong et al., 2017). Gender norms or inequalities can lead to a variety of health and social problems, and victims are never just one side of a simple binary gender (Heise et al., 2019). Promoting gender equality can effectively reduce the treatment effect of gender on SWB. Ongoing technological changes and increased productivity force conservative unfairness to change.

Temporal variations in the treatment effects and base heterogeneous effects were observed. The base heterogeneous effects varied unstably, and the variation did not show a clear trend. However, treatment effects increased monotonically over time. Because of their different calculation mechanisms, base heterogeneous effects are more sensitive to sampling. The base heterogeneous effects were computed from the two groups of populations, females and males, and treated in the same way. Unfortunately, the survey sampling distribution shifts over time, eventually causing temporal variations in the heterogeneous effects. However, the treatment effect change is free from the impact of sampling distribution variations because the treatment effects are estimated as the difference between the two treatments on the same person. Changes in treatment over time were the only explanations. Simply put, the gender gap in SWB caused by external factors is gradually widening, which is consistent with the findings of Blanchflower, D. and Bryson, A. (2024) based on several global databases. In terms of country-level differences in treatment effects, almost all countries were in the same direction except for some countries in sub-Saharan Africa. Long-term political instability and regional conflicts may explain this anomaly (Carmignani and Kler, 2016).

Unlike the base heterogeneous effects, which are caused by inherent factors between genders, treatment effects are more valuable for future policymaking. We used a new model to fit the relationship between the absolute values of individual treatment effects and input variables to predict well-being. The ideal status is zero or close to zero treatment effects because the aim is gender equality in the external environment. Whether society is stricter for men or for women is unequal. An increase in income or a higher income level is related to lower treatment effects between sexes, as shown in **Figures 3.a and 3.i**. This means that the external treatments for richer males and females are similar. Although increased income does not automatically resolve gender disparities, better educational opportunities and socioeconomic environments could improve individuals' feelings of gender equality (Campbell et al., 2021). People with sufficient food intake can be affected by external treatment differences, as shown in **Figure 3.b**. This is consistent with Maslow's hierarchy of needs; that is, when lower-level needs are not met, higher-level needs can be ignored (Maslow, 1943). This is a potential explanation for the abnormal treatment effect status of Sub-Saharan Africa. In terms of marital status, single status contributes more to treatment effects because society and families have significantly different views on single men and women (Apostolou et al., 2020). People who are satisfied with their current job status, specifically full-time and part-time employees, do not want a full-time job and contribute less to the treatment effects, as illustrated in **Figure 3.d**. The impact of age on the treatment effects is the most complex. The contribution of age to the treatment effects peaked at approximately 25. There is no gap in well-being between genders in adolescents is absent (Esteban-Gonzalo et al., 2020). The relationship between age and treatment effects is also an inverted U-shape, but its explanations are different from those of the well-being-age U-shape and midlife crisis. This relationship may be due to the different timing of the two sexes experiencing stress, with one gender experiencing stress relatively earlier. Believing that life contributes more to treatment effects. **Figure 3.g** shows the impact of time on the treatment effects, which is consistent with our previous results. Thinking that children are not respected locally contributes more to the treatment effects between genders. A society that lacks respect for children is likely to experience greater gender inequality (UNICEF, 2023).

This study highlights critical policy implications for addressing gender disparities in SWB based on the adjustment of external factors to reduce treatment effects between genders. Policies should prioritize the creation of social environments with gender equality from an early age. Equitable access to resources and opportunities across social, economic, and educational spheres must be ensured. Governments and policymakers must dismantle entrenched gender norms that restrict emotional expression and impose undue burdens, particularly on men, which perpetuates cycles of inequality and reduced well-being. Gender norms progress with the level of productivity and production relations. However, as the average human lifespan increases, the speed of this concept iteration decreases. Enhancing the ability of people to accept new ideas may be the key to solving this problem. Educational policies should incorporate sex sensitivity to ensure that boys and girls receive equal encouragement and opportunities to pursue their interests and talents without prejudice. Furthermore, comprehensive support systems for singles, considering the diverse challenges they face and accepting their decisions, could help alleviate rendering-related social pressure from their marital status. By adopting a multifaceted approach that addresses both inherent and external factors influencing gender disparities in SWB, policymakers can foster more inclusive societies that enhance the well-being of all genders.

The methodological contributions of this study are noteworthy. First, the ESTEM is an effective causal inference method that can probe an exogenous variable's inherent and external effects. However, previous studies have mainly relied on linear regression technologies with poor predictive abilities (e.g., Kassie et al., 2014; Liu et al., 2021). For switching models, the predictions must be either counterfactual or factual. We employed machine learning technologies as substitutes for linear regression to improve prediction ability. Without illustrating the prediction performance, accepting the results is risky. Second, for complete counterfactual and factual predictions, we propose a robust engineering process, specifically, a multi-epoch multi-fold prediction process. This method guarantees the reliability and stability of predictions. In summary, our study provides a paradigm for studies on applied switching models using machine learning models.

This study has several limitations, and future research directions are proposed. First, although the dataset was multi-year, it was not a panel dataset. Each year, the respondents were different. Therefore, this study had a cross-sectional design. The ESTEM based on counterfactual inference could demonstrate the causal relationship in a way that we could not go further. Panel data would enable a deeper investigation of the reasons for this over time, especially how gender disparity impacts treatment and base heterogeneous effects across life stages. Second, we analyzed the variation across countries, but owing to limited ability and data availability, we could not go further. Comparative studies or multilevel modeling can be used to explore country-level or regional differences more explicitly. Third, most variables in the analysis were binary, which is somewhat uninformative. This limits the power of the model and allows for deeper explanations. Of course, large-scale survey questions, especially in international surveys, should be simple to fit the status of more countries. In future small-scale studies, more informative variables should be included in the analysis.

Conclusions

Our findings underscore the significant impact of gender on human well-being and reveal that gender disparities in SWB are caused by both inherent and external factors. Specifically, inherently males should inherently have higher human well-being; however, owing to external factors, they are less satisfied with their lives. Based on the results of the ESTEM and machine learning techniques, new insights are provided into the nuanced ways in which societal, familial, and workplace norms about gender roles can shape individual well-being, and both males and females could be victims of irrational norms. Improving social norms and promoting gender equality to meet social development levels can mitigate observed disparities in SWB. Therefore, future policies and interventions should concentrate on creating more inclusive environments without rigid traditional gender roles but instead support the emotional and psychological well-being of all genders. This research contributes to evidence calling for a re-evaluation of societal standards and policies to foster a more inclusive environment for all genders to promote human well-being.

Acknowledgments

This study was conducted as part of a project titled "Institutional design to achieve a society of well-being" undertaken at the Research Institute of Economy, Trade, and Industry (RIETI). The authors express their deepest gratitude to the professors and researchers whose invaluable advice and thoughtful feedback greatly enriched the quality of this research.

Figures

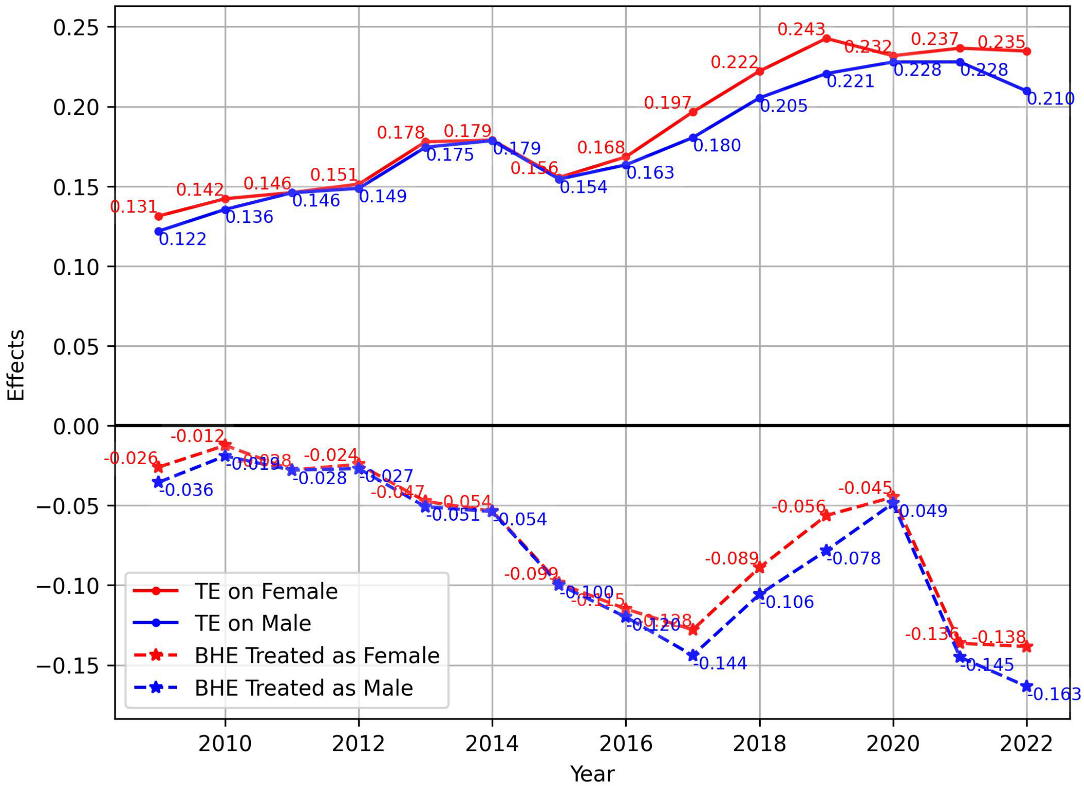

Figure 1: Temporal Variation of Treatment and Base Heterogeneity Effects

(Note: TE on Female means treatment effects on the female population, and TE on Male means treatment effects on the male population. BHE Treated as Female indicates base heterogeneity effects under the treatments for females, and BHE Treated as Male indicates base heterogeneity effects under the treatments for males.**)**

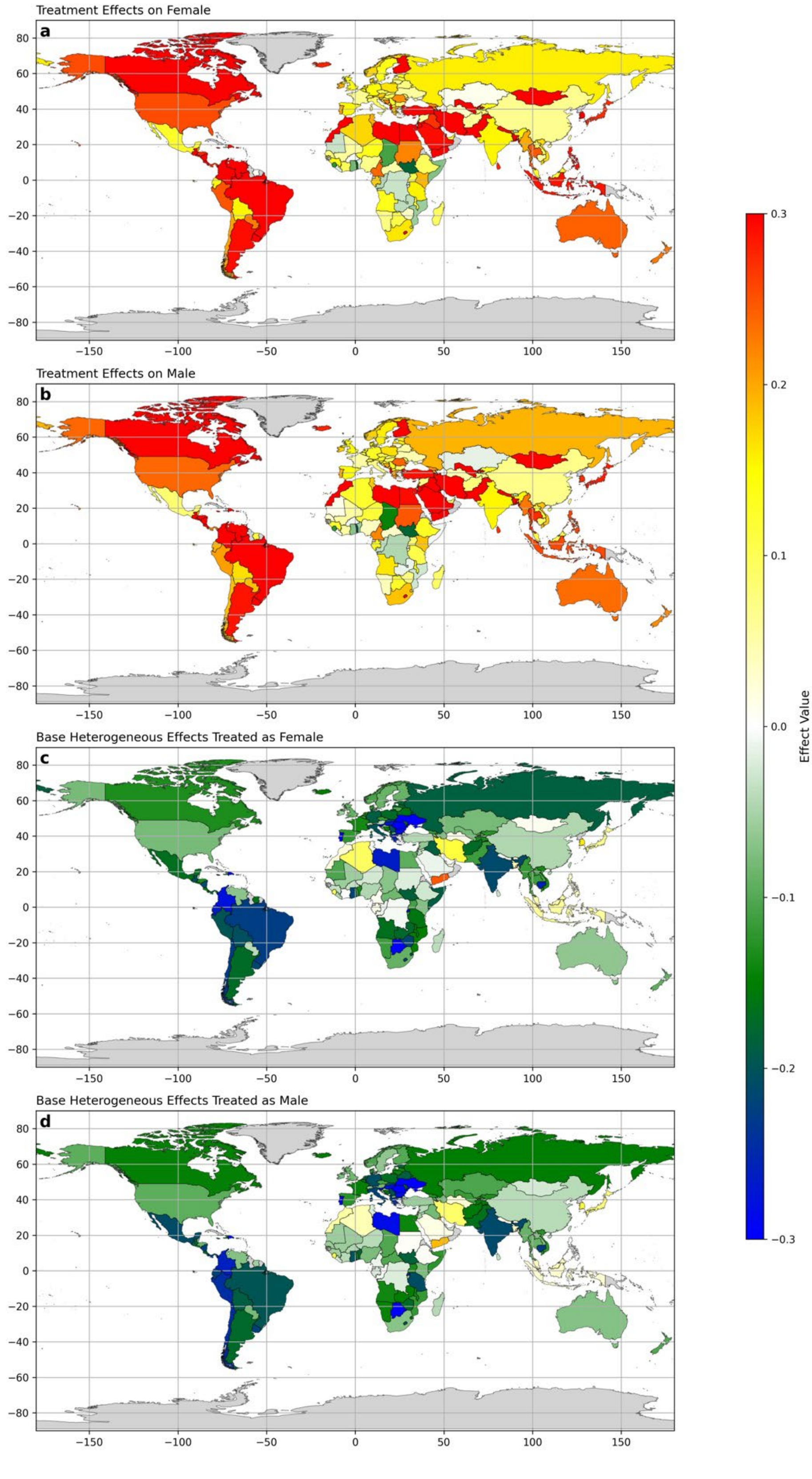

Figure 2: Country-level Base Heterogeneous Effects and Treatment Effects

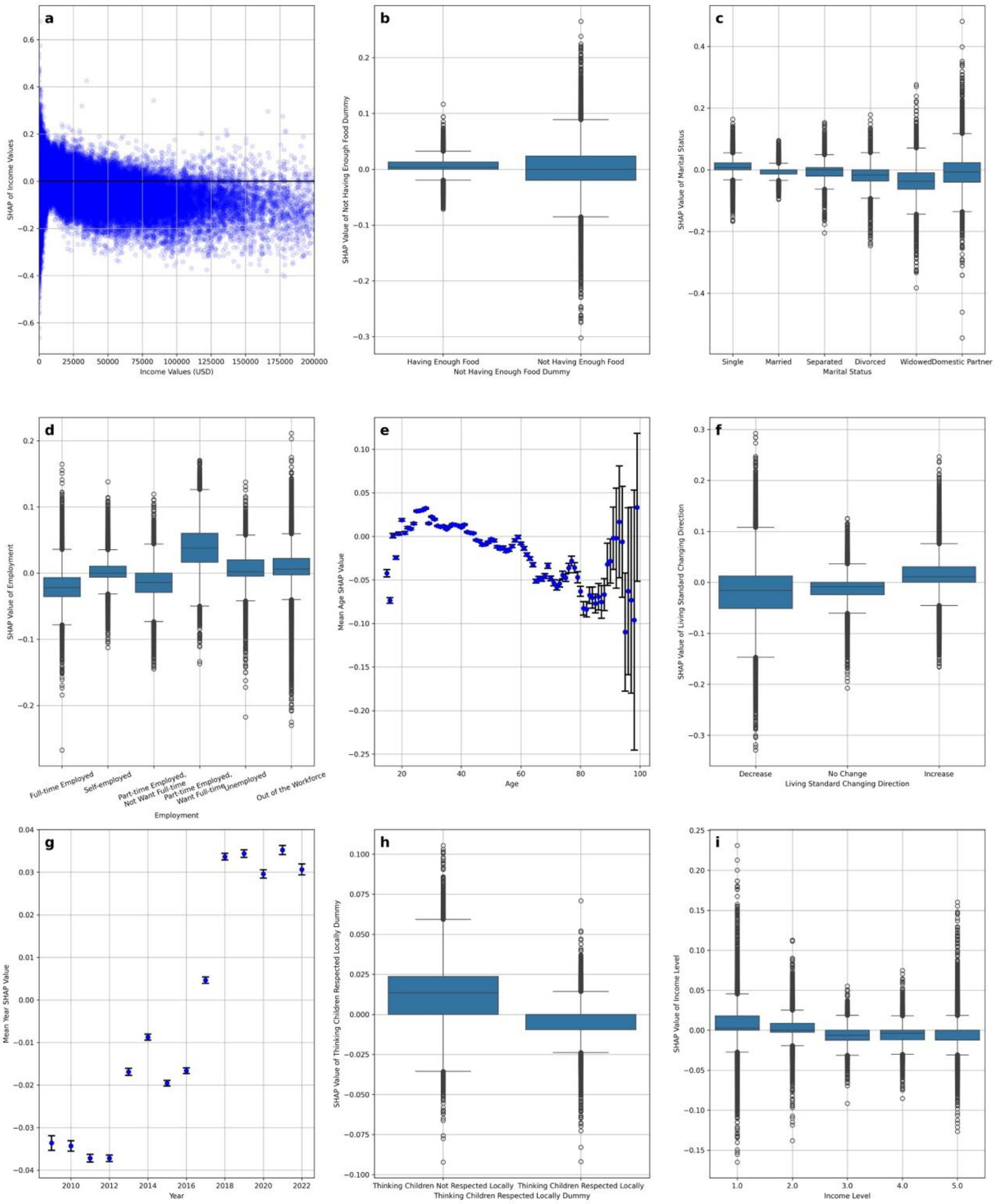

Figure 3: SHAP Value Analysis of Top Nine Variables with Largest Impacts

Tables

Table 1: Data Summary

	mean	std	min	25%	50%	75%	max
SWB	5.521	2.414	0	4	5	7	10
Household Income	25339.30 1	925219.1 04	0.00 0	4190.87 5	10510.19 2	26055.60 0	898033649.9 54
Health Disability Dummy	0.267	0.431	0	1	1	1	1
Female Dummy	0.532	0.499	0	0	1	1	1
Age	41.429	17.552	15	27	39	54	99
Marital Status	2.344	1.710	1	1	2	2	8
Employment	3.646	2.134	1	1	4	6	6
Children Under 15 Dummy	0.522	0.500	0	0	1	1	1
Feeling of Income	2.323	0.965	1	2	2	3	4
Income Level	3.236	1.415	1	2	3	5	5
Having Relatives to Rely on Dummy	0.791	0.407	0	1	1	1	1
Living Standard Changing Direction	0.175	0.820	-1	-1	0	1	1
Not Having Enough Food Dummy	0.322	0.467	0	0	0	1	1
Not Having Enough Shelter Dummy	0.245	0.430	0	0	0	0	1
Feeling Well Rested Dummy	0.674	0.469	0	0	1	1	1
Feeling Respected Dummy	0.865	0.342	0	1	1	1	1
Smiling Dummy	0.703	0.457	0	0	1	1	1
Doing Interesting Things Dummy	0.520	0.500	0	0	1	1	1
Having Enjoyment Dummy	0.689	0.463	0	0	1	1	1
Suffering Physical Pain Dummy	0.307	0.461	0	0	0	1	1
Feeling Worry Dummy	0.380	0.485	0	0	0	1	1
Feeling Sad Dummy	0.234	0.423	0	0	0	0	1
Feeling Stress Dummy	0.329	0.470	0	0	0	1	1
Feeling Anger Dummy	0.197	0.398	0	0	0	0	1

Feeling Satisfied with City Dummy	0.758	0.429	0	1	1	1	1
Economic Changing Direction	0.054	0.851	-1	-1	0	1	1
Thinking Good Time to Find Job Dummy	0.337	0.473	0	0	0	1	1
Feeling Satisfied with Public Transportation Dummy	0.560	0.496	0	0	1	1	1
Feeling Satisfied with Road Dummy	0.523	0.499	0	0	1	1	1
Feeling Satisfied with Education Dummy	0.604	0.489	0	0	1	1	1
Feeling Satisfied with Air Quality Dummy	0.713	0.453	0	0	1	1	1
Feeling Satisfied with Water Quality Dummy	0.668	0.471	0	0	1	1	1
Feeling Satisfied with Healthcare Dummy	0.544	0.498	0	0	1	1	1
Feeling Satisfied with Affordable House Dummy	0.463	0.499	0	0	0	1	1
Feeling Satisfied with Opportunity to Make Friends Dummy	0.684	0.465	0	0	1	1	1
Thinking Good Place for Ethical Minority Dummy	0.575	0.494	0	0	1	1	1
Thinking Good Place for Gay or Lesbian Dummy	0.287	0.452	0	0	0	1	1
Thinking Good Place for Immigrants Dummy	0.569	0.495	0	0	1	1	1
Donated Recently Dummy	0.306	0.461	0	0	0	1	1
Did Volunteer Recently Dummy	0.201	0.401	0	0	0	0	1
Helped Stranger Dummy	0.495	0.500	0	0	0	1	1
Voiced Opinion to Official Dummy	0.159	0.366	0	0	0	0	1
Feeling Confident in Local Police Dummy	0.558	0.497	0	0	1	1	1

Feeling Safe of Alone Night Walking Dummy	0.590	0.492	0	0	1	1	1
Having Been Stolen Dummy	0.139	0.346	0	0	0	0	1
Having Been Assaulted Dummy	0.047	0.211	0	0	0	0	1
Thinking Religion Importance Locally Dummy	0.637	0.481	0	0	1	1	1
Thinking Children Respected Locally Dummy	0.640	0.480	0	0	1	1	1
Having Opportunity for Children Learning Locally Dummy	0.671	0.470	0	0	1	1	1
Feeling Women Respected Dummy	0.566	0.496	0	0	1	1	1
Feeling Satisfied with Poverty Alleviation Dummy	0.357	0.479	0	0	0	1	1
Feeling Satisfied with Environmental Efforts Dummy	0.494	0.500	0	0	0	1	1
Feeling Freedom of Choosing Life Dummy	0.723	0.448	0	0	1	1	1
Feeling Confidence in Military Dummy	0.586	0.493	0	0	1	1	1
Feeling Confidence in Judicial System Dummy	0.431	0.495	0	0	0	1	1
Feeling Confidence in National Government Dummy	0.420	0.494	0	0	0	1	1
Feeling Confidence in Financial System Dummy	0.531	0.499	0	0	1	1	1
Feeling Confidence in Election Honesty Dummy	0.402	0.490	0	0	0	1	1
Feeling Freedom of Media Dummy	0.543	0.498	0	0	1	1	1

Prevailing Corruption within Business Dummy	0.610	0.488	0	0	1	1	1
Prevailing Corruption within Government Dummy	0.581	0.493	0	0	1	1	1
Approving of Leadership Performance Dummy	0.406	0.491	0	0	0	1	1

Table 2: Summary of Test R^2 in 10-epoch 10-fold Prediction

Sub-dataset	Model	Average Test R^2	Maximum Test R^2	Minimum Test R^2	Standard Deviation	Number of Predictions for Each Obs.
Female Population	Female Population	37.23%	38.74%	36.62%	0.49%	10
Male Population	Female Population	36.64%	36.68%	36.60%	0.02%	100
Female Population	Male Population	36.18%	36.23%	36.13%	0.02%	100
Male Population	Male Population	38.03%	38.67%	37.28%	0.23%	10

Table 3: Treatment and Base Heterogeneity Effects

		Treatment		Treatment Effects
		Treated as Female Population	Treated as Male Population	TE_{gender}^{fe-ma}
Gender Group	Female Population	5.569	5.384	0.185***
	Male Population	5.645	5.466	0.178***
Base Heterogeneity Effects	BHE_{fe-ma}^{gender}	-0.075***	-0.082***	

Note: *** p value < 0.1%; ** p value < 5%; * p value < 1%. TE and BHE estimations follow **Equations (6) – (9)**. $gender$ could be the female and male population.

Table 4: Mean Contributions of Variables to Treatment Effects of Gender

Variable	Mean SHAP	Lower Boundary 95% CI	Upper Boundary 95% CI	Variable	Mean SHAP	Lower Boundary 95% CI	Upper Boundary 95% CI
Household Income	-9.974	-10.275	-9.673	Smiling Dummy	0.678	0.634	0.722
Not Having Enough Food Dummy	3.915	3.799	4.030	Thinking Good Place for Gay or Lesbian Dummy	-0.575	-0.642	-0.508
Marital Status	-3.844	-3.978	-3.710	Health Disability Dummy	0.492	0.413	0.570
Having Enjoyment Dummy	3.768	3.693	3.842	Feeling Confident in Local Police Dummy	0.452	0.413	0.490
Age	-3.718	-3.955	-3.481	Helped Stranger Dummy	-0.448	-0.526	-0.369
Living Standard Changing Direction	-2.816	-3.003	-2.629	Feeling Confidence in Election Honesty Dummy	0.438	0.404	0.472
Wave	-2.618	-2.897	-2.338	Feeling Satisfied with City Dummy	0.425	0.355	0.495
Thinking Children Respected Locally Dummy	2.561	2.496	2.627	Feeling Satisfied with Poverty Alleviation Dummy	0.397	0.347	0.447
Income Level	-2.308	-2.378	-2.238	Feeling Satisfied with Road Dummy	0.390	0.350	0.429
Prevailing Corruption within	-2.188	-2.259	-2.117	Feeling Satisfied with Air Quality Dummy	0.337	0.296	0.378

Business Dummy							
Feeling Confidence in Military Dummy	1.867	1.823	1.911	Thinking Good Place for Immigrants Dummy	-0.335	-0.377	-0.293
Voiced Opinion to Official Dummy	-1.610	-1.669	-1.551	Did Volunteer Recently Dummy	0.297	0.242	0.352
Feeling Freedom of Choosing Life Dummy	-1.585	-1.669	-1.502	Feeling Satisfied with Affordable House Dummy	0.293	0.234	0.352
Employment	-1.544	-1.680	-1.409	Feeling Satisfied with Public Transportation Dummy	0.289	0.250	0.328
Feeling Anger Dummy	1.379	1.335	1.423	Doing Interesting Things Dummy	0.276	0.202	0.351
Feeling Satisfied with Opportunity to Make Friends Dummy	-1.315	-1.356	-1.274	Thinking Good Time to Find Job Dummy	0.269	0.212	0.327
Feeling of Income	-1.247	-1.440	-1.054	Having Opportunity for Children Learning Locally Dummy	0.261	0.220	0.302
Feeling Safe of Alone Night Walking Dummy	1.191	1.141	1.240	Suffering Physical Pain Dummy	-0.258	-0.297	-0.219

Having Been Stolen Dummy	1.051	1.006	1.095	Having Relatives to Rely on Dummy	-0.209	-0.327	-0.091
Not Having Enough Shelter Dummy	1.007	0.929	1.085	Donated Recently Dummy	0.195	0.134	0.257
Feeling Sad Dummy	0.938	0.867	1.009	Economic Changing Direction	-0.193	-0.258	-0.128
Thinking Good Place for Ethical Minority Dummy	-0.934	-0.982	-0.886	Feeling Confidence in Financial System Dummy	-0.130	-0.166	-0.094
Feeling Stress Dummy	0.880	0.829	0.931	Feeling Confidence in Judicial System Dummy	0.126	0.093	0.160
Children Under 15 Dummy	-0.873	-0.930	-0.816	Feeling Women Respected Dummy	-0.120	-0.166	-0.073
Feeling Well Rested Dummy	-0.837	-0.900	-0.774	Approving of Leadership Performance Dummy	0.107	0.066	0.149
Feeling Satisfied with Water Quality Dummy	-0.812	-0.847	-0.777	Feeling Worry Dummy	-0.045	-0.113	0.023
Feeling Confidence in National Government Dummy	0.800	0.758	0.842	Thinking Religion Importance Locally Dummy	0.041	0.002	0.079
Feeling Satisfied with Environmental	0.800	0.761	0.839	Feeling Respected Dummy	0.028	-0.013	0.069

Efforts Dummy							
Feeling Satisfied with Healthcare Dummy	0.731	0.683	0.780	Feeling Satisfied with Education Dummy	0.026	-0.006	0.057
Having Been Assaulted Dummy	0.716	0.673	0.760	Feeling Freedom of Media Dummy	0.002	-0.026	0.029
Prevailing Corruption within Government Dummy	-0.711	-0.756	-0.667				

Note: the variables are ranked by the means of each variable's SHAP values; all values in the table are scaled by 10^{-3}; and the 95% CI represents 95% confidence interval.

Reference

Apicella, C.L., Carré, J.M., Dreber, A., 2015. Testosterone and Economic Risk Taking: A Review. Adaptive Human Behavior and Physiology 1(3), 358-385.

Apostolou, M., O, J., Esposito, G., 2020. Singles' Reasons for Being Single: Empirical Evidence From an Evolutionary Perspective. Frontiers in Psychology 11.

Arrosa, M.L., Gandelman, N., 2016. Happiness Decomposition: Female Optimism. J. Happiness Stud. 17(2), 731-756.

Aryal, J.P., Rahut, D.B., Mottaleb, K.A., Ali, A., 2019. Gender and household energy choice using exogenous switching treatment regression: Evidence from Bhutan. Environmental Development 30, 61-75.

Ballas, D., Thanis, I., 2022. Exploring the Geography of Subjective Happiness in Europe During the Years of the Economic Crisis: A Multilevel Modelling Approach. Soc. Indic. Res. 164(1), 105-137.

Bartram, D., 2022. The 'Gender Life-Satisfaction/Depression Paradox' Is an Artefact of Inappropriate Control Variables. Soc. Indic. Res. 164(3), 1061-1072.

Batz, C., Tay, L., 2018. Gender differences in subjective well-being.

Ben Mansour, G., Kacem, A., Ishak, M., Grélot, L., Ftaiti, F., 2021. The effect of body composition on strength and power in male and female students. BMC Sports Science, Medicine and Rehabilitation 13(1).

Blanchflower, D., Bryson, A., 2024. The Gender Well-Being Gap. Soc. Indic. Res. 173(3), 1-45.

Blanchflower, D.G., Bryson, A., 2024. The female happiness paradox. Journal of Population Economics 37(1), 16.

Blanchflower, D.G., Graham, C.L., 2022. The Mid-Life Dip in Well-Being: a Critique. Soc. Indic. Res. 161(1), 287-344.

Boerma, T., Hosseinpoor, A.R., Verdes, E., Chatterji, S., 2016. A global assessment of the gender gap in self-reported health with survey data from 59 countries. BMC Public Health 16(1), 675.

Breiman, L., 2001. Random Forests. Machine Learning 45(1), 5-32.

Campbell, J.A., McIntyre, J., Kucirkova, N., 2021. Gender Equality, Human Development, and PISA Results over Time. Social Sciences 10(12), 480.

Carmignani, F., Kler, P., 2016. The geographical spillover of armed conflict in Sub-Saharan Africa. Economic Systems 40(1), 109-119.

Case, A., Deaton, A., 2015. Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. Proceedings of the National Academy of Sciences 112(49), 15078-15083.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, 22nd SIGKDD Conference on Knowledge Discovery and Data Mining. ACM.

Deaton, A., 2008. Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll. Journal of Economic Perspectives 22(2), 53-72.

Diener, E., Ng, W., Harter, J., Arora, R., 2010. Wealth and happiness across the world: material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. J. Pers. Soc. Psychol. 99(1), 52.

Diener, E., Oishi, S., Tay, L., 2018. Advances in subjective well-being research. Nature Human Behaviour 2(4), 253-260.

Diener, E., Suh, E.M., Lucas, R.E., Smith, H.L., 1999. Subjective well-being: Three decades of progress. Psychol. Bull. 125(2), 276.
Eagly, A.H., Karau, S.J., 2002. Role congruity theory of prejudice toward female leaders. Psychological review 109(3), 573.
Eckermann, E., 2012. The Quality of Life of Adults, in: Land, K.C., Michalos, A.C., Sirgy, M.J. (Eds.), Handbook of Social Indicators and Quality of Life Research. Springer Netherlands, Dordrecht, pp. 373-380.
El-Jahel, L., Macculloch, R., Shafiee, H., 2023. How Does Monetary Policy Affect Welfare? Some New Estimates Using Data on Life Evaluation and Emotional Well-Being. Journal of Money, Credit and Banking 55(8), 2001-2025.
Esteban-Gonzalo, S., Esteban-Gonzalo, L., Cabanas-Sánchez, V., Miret, M., Veiga, O.L., 2020. The Investigation of Gender Differences in Subjective Wellbeing in Children and Adolescents: The UP&DOWN Study. International Journal of Environmental Research and Public Health 17(8), 2732.
Gonalons-Pons, P., Gangl, M., 2021. Marriage and masculinity: Male-breadwinner culture, unemployment, and separation risk in 29 countries. American sociological review 86(3), 465-502.
Heise, L., Greene, M.E., Opper, N., Stavropoulou, M., Harper, C., Nascimento, M., Zewdie, D., Darmstadt, G.L., Greene, M.E., Hawkes, S., Heise, L., Henry, S., Heymann, J., Klugman, J., Levine, R., Raj, A., Rao Gupta, G., 2019. Gender inequality and restrictive gender norms: framing the challenges to health. The Lancet 393(10189), 2440-2454.
Herbst, C.M., 2011. ‘Paradoxical’ decline? Another look at the relative reduction in female happiness. Journal of Economic Psychology 32(5), 773-788.
Jebb, A.T., Morrison, M., Tay, L., Diener, E., 2020. Subjective Well-Being Around the World: Trends and Predictors Across the Life Span. Psychol. Sci. 31(3), 293-305.
Jebb, A.T., Tay, L., Diener, E., Oishi, S., 2018. Happiness, income satiation and turning points around the world. Nature Human Behaviour 2(1), 33-38.
Joshanloo, M., Jovanović, V., 2020. The relationship between gender and life satisfaction: analysis across demographic groups and global regions. Archives of Women's Mental Health 23(3), 331-338.
Kassie, M., Ndiritu, S.W., Stage, J., 2014. What Determines Gender Inequality in Household Food Security in Kenya? Application of Exogenous Switching Treatment Regression. World Development 56, 153-171.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree, Advances in neural information processing systems.
Klasen, S., 2007. Gender-related Indicators of Well-being, Palgrave Macmillan UK, pp. 167-192.
Kuehner, C., 2017. Why is depression more common among women than among men? Lancet Psychiatry 4(2), 146-158.
Li, C., Managi, S., 2023. Income raises human well-being indefinitely, but age consistently slashes it. Scientific Reports 13(1).
Li, S.H., Graham, B.M., 2017. Why are women so vulnerable to anxiety, trauma-related and stress-related disorders? The potential role of sex hormones. The Lancet Psychiatry 4(1), 73-82.
Liu, M., Min, S., Ma, W., Liu, T., 2021. The adoption and impact of E-commerce in rural China: Application of an endogenous switching regression model. Journal of Rural Studies 83, 106-116.
Lucas, R.E., 2007. Long-term disability is associated with lasting changes in subjective well-being: evidence from two nationally representative longitudinal studies. J Pers Soc Psychol 92(4), 717-730.

Lucas, R.E., Clark, A.E., Georgellis, Y., Diener, E., 2004. Unemployment alters the set point for life satisfaction. Psychol Sci 15(1), 8-13.
Lundberg, S.M., Erion, G., Chen, H., Degrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2(1), 56-67.
MacKerron, G., 2012. Happiness Economics from 35 000 Feet. Journal of Economic Surveys 26(4), 705-735.
Maslow, A.H., 1943. A theory of human motivation. Psychological review 50(4), 370.
Molnar, C., 2020. Interpretable machine learning. Lulu. com.
Napier, J.L., Suppes, A., Bettinsoli, M.L., 2020. Denial of gender discrimination is associated with better subjective well-being among women: A system justification account. European Journal of Social Psychology 50(6), 1191-1209.
Oertelt-Prigione, S., 2012. The influence of sex and gender on the immune response. Autoimmunity Reviews 11(6), A479-A485.
Oswald, A.J., Wu, S., 2010. Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the USA. Science 327(5965), 576-579.
Parker, G., Brotchie, H., 2010. Gender differences in depression. International Review of Psychiatry 22(5), 429-436.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features, Advances in neural information processing systems.
Schmitt, M.T., Branscombe, N.R., Postmes, T., Garcia, A., 2014. The consequences of perceived discrimination for psychological well-being: a meta-analytic review. Psychol. Bull. 140(4), 921.
Seedat, S., Scott, K.M., Angermeyer, M.C., Berglund, P., Bromet, E.J., Brugha, T.S., Demyttenaere, K., De Girolamo, G., Haro, J.M., Jin, R., Karam, E.G., Kovess-Masfety, V., Levinson, D., Medina Mora, M.E., Ono, Y., Ormel, J., Pennell, B.-E., Posada-Villa, J., Sampson, N.A., Williams, D., Kessler, R.C., 2009. Cross-National Associations Between Gender and Mental Disorders in the World Health Organization World Mental Health Surveys. Arch. Gen. Psychiatry 66(7), 785.
Simon, R.W., Nath, L.E., 2004. Gender and emotion in the United States: Do men and women differ in self-reports of feelings and expressive behavior? American journal of sociology 109(5), 1137-1176.
Smith, M.D., Wesselbaum, D., 2023. Well-Being and Income Across Space and Time: Evidence from One Million Households. J. Happiness Stud. 24(5), 1813-1840.
Theorell-Haglöw, J., Miller, C.B., Bartlett, D.J., Yee, B.J., Openshaw, H.D., Grunstein, R.R., 2018. Gender differences in obstructive sleep apnoea, insomnia and restless legs syndrome in adults – What do we know? A clinical update. Sleep Medicine Reviews 38, 28-38.
UNICEF, 2023. Global Annual Results Report 2023: Gender Equality. New York.
Wong, Y.J., Ho, M.-H.R., Wang, S.-Y., Miller, I., 2017. Meta-analyses of the relationship between conformity to masculine norms and mental health-related outcomes. Journal of counseling psychology 64(1), 80.
Wood, W., Rhodes, N., Whelan, M., 1989. Sex differences in positive well-being: A consideration of emotional style and marital status. Psychol. Bull. 106(2), 249.
Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., Deng, S.-H., 2019. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. Journal of Electronic Science and Technology 17(1), 26-40.

Yen, S.T., Chen, Z., Eastwood, D.B., 2009. Lifestyles, Demographics, Dietary Behavior, and Obesity: A Switching Regression Analysis. Health Services Research 44(4), 1345-1369.
Zitzmann, M., 2020. Testosterone, mood, behaviour and quality of life. Andrology 8(6), 1598-1605.