THE SOUTHERN BUSINESS AND ECONOMIC JOURNAL

Volume 48, Number 1 2025

Published by
AUBURN UNIVERSITY at MONTGOMERY
COLLEGE OF BUSINESS

AUM
COLLEGE OF BUSINESS

THE SOUTHERN BUSINESS AND ECONOMIC JOURNAL

Editorial Policy: The editors prefer to publish papers describing applied academic studies using data-based research relevant to individuals, businesses, and non-profit institutions located in the southern region of the United States. However, papers with a national or international focus will also be considered, along with theoretical and pedagogical studies that demonstrate relevance to decision making. Papers covering topics from all areas of business and economics are appropriate. Submissions typically get a double-blind review. Past issues of the journal are available by going to http://www.business.aum.edu/SBEJ. There is no submission fee; however, accepted articles require a publication fee of $50 plus an additional $10 per page after formatting for publication. Authors of accepted papers will be sent formatting instructions that include the requirement to have a page layout orientation of portrait and a page layout size of statement (5.5" x 8.5"). Submit papers for publication consideration in a Word file attached to an email to editor-sbej@aum.edu. The editors use publication guidelines found in the Publication Manual of the American Psychological Association, Sixth Edition.

Editors
The Southern Business and Economic Journal
Department of Business Administration
Auburn University at Montgomery
PO Box 244023
Montgomery, AL 36124-4023

THE SOUTHERN BUSINESS AND ECONOMIC JOURNAL

Indexed in PAIS Bulletin and EBSCO Publishers Online ISSN: 0743779X

Volume 48, Number 1 2025

Articles

Published by the Auburn University at Montgomery College of Business

Dynamics of Healthcare Expenditure and GDP in South Asia: Exploring Luxury vs Necessity

Bhavana Kanumalla[1]
Auburn University at Montgomery

Dong-Yop Oh[2]
Auburn University at Montgomery

Abstract

This paper studies the intricate relationship between healthcare expenditure (HE) and Gross Domestic Product (GDP) for 8 South Asian countries for the period 2002 to 2020. In the presence of smooth structural breaks in the low-frequency components of a Fourier expansion, the cointegration test determines the existence of a long-term equilibrium relationship between HE and GDP. We examined income elasticity of healthcare expenditure for the countries where we found cointegration relationship between the two series. Empirical results support the existence of stochastic co-movement between HE and GDP in three countries. Our analysis of income elasticity of HE showcased that Afghanistan and Maldives have an income elasticity rounded off to 1 which implies that HE is a normal good, whereas Pakistan considers HE to be a luxury good since income elasticity resulted in around one-fourth increase from 1.

Keywords: Cointegration; Elasticity; GDP; Healthcare expenditure; Nonlinear structure break; South Asia

1. Introduction

In recent times, there has been a notable surge in healthcare expenditure globally, drawing attention to how different economies perceive healthcare expenditure and the factors drive the growth of HE. Previous research has provided insight that various factors such as population age structure, cost of care, technology (Dieleman et al. 2017), and income (Kleiman 1974; Newhouse 1977) contribute to an increase in HE. Based on research by Kleiman (1974) and Newhouse (1977) income emerged as the leading factor in explaining variances among countries concerning both the level and progression of healthcare expenditure.

Spending on healthcare varies from developing and developed nations; although health economists have been studying factors contributing to healthcare expenditure patterns of these nations keenly, there are few studies dedicated to investigating HE in South Asian countries. The average HE per capita of South Asian countries has more than tripled from $45.26 in the year 2000 to $174.29 in 2020 showcasing an overall economic growth that has fueled greater demand for healthcare services. This study analyzes the relationship between healthcare expenditure (HE) and Gross Domestic Product (GDP) of South Asian economies to determine whether these developing nations consider HE to be a luxury good or a necessity which in turn helps us understand the interplay between economic development and health spending. As living standards rise, there tends to be a corresponding increase in the income allocated towards healthcare expenses (Blomqvist and Carter, 1997). If healthcare is indeed categorized as a luxury good, its demand is expected to rise significantly with increasing wealth, indicating an income elasticity higher than 1. Since healthcare plays a vital role in meeting basic human needs, classifying it as a luxury good may not be ideal. Conventionally, healthcare services are presumed to be given to address the requirements of society with a focus on necessity as opposed to luxury. The idea that everyone needs access to healthcare services to preserve their health and well-being, regardless of their financial situation, serves to support this viewpoint even more. As per Blomqvist and Carter (1997) given that the occurrence of illnesses and the per unit cost of care are the same across different population groups, there is a high probability that lower-income groups would allocate a greater portion of their income to healthcare compared to

their richer counterparts. Farag et al. (2012) derived an income elasticity for healthcare expenditure (HE) that fell below one across 173 developed and developing nations between 1995 and 2006. Additionally, they observed that this elasticity is lower in low-income countries compared to high-income ones.

Drawing on empirical evidence from eight South Asian countries, this paper provides insights into variations in income elasticity of HE and examines the relationship between HE and GDP across diverse contexts where HE can be found to have an impact on GDP. A situation wherein increased levels of HE results in a healthier population could therefore boost GDP by improving worker productivity. Our analysis adopts a robust approach to evaluate the stationarity of HE and GDP data, using Unit Root test to discern underlying trends and disruptions in healthcare expenditure dynamics. Further, we employ cointegration test to explore the long-term equilibrium relationship between non-stationary time series variables, accounting for potential structural breaks that may influence healthcare spending patterns over time.

In the subsequent sections of this paper, we present a review of relevant literature, outline our methodological framework, present data, discuss empirical findings, and discuss concluding results. This paper aims to study the long-run economic relationship between HE and GDP to evaluate income elasticity of HE for 8 South Asian countries and eventually determine whether health care expenditure is a luxury or necessity.

2. Literature Review

Early research employed by McCoskey and Selden (1998) on OECD countries found that HE and GDP are stationary since unit root hypothesis could be rejected. Unlike previous studies which did not account for structural breaks in unit root tests, a study by Jewell et al. (2003) used a panel LM unit root test that allows for heterogeneous level shifts. The findings of the study reject the null hypothesis of unit roots for both HE and GDP, indicating that these variables are stationary. These results add to the concerns of robustness of empirical results of time series properties of HE and GDP. Similarly, Carrion-i-Silvestre (2005) similarly discovered substantiating evidence indicating the stationarity of HE and GDP, even in the presence of multiple breaks affecting both their levels and slopes.

Gerdtham, Löthgren (2000) examined stationarity and cointegration of HE and GDP for 21 OECD countries which involved

the use of a test battery that includes new country-by-country and panel test, it suggests that both HE and GDP are non-stationary. These findings contribute to the highlighting the importance of considering both unit root and cointegration analysis when investigating the relationship between health expenditure and economic variables. On the contrary, Clemente et al. (2004) found evidence of a cointegrating relationship between HE and GDP by applying the test of Gregory and Hansen (1996a) which accounts for structural breaks and Roberts (2000) also found such a relationship whereas French (2012) found evidence of long-run causal relationship between HE and GDP in both directions. Tomoko and Shigeyuki (2015) employed cointegration analysis to test non-stationarity in a previous panel cointegration study on Japan's HE and GDP which resulted in findings highlighting a cointegration relationship among the variables considered, suggesting a long-term association.

A previous study investigating income elasticity of HE of South-East Asian Regional (SEAR) countries by Khan and Mahumud (2015) found that healthcare expenditure is perceived differently depending on whether it is provided through the public or private sector. The results indicated that the citizens of SEAR countries viewed HE as a necessity when provided through public sector whereas HE provided by private sector as a luxury.

Beylik et al. (2021) studied the relationship between healthcare expenditure and economic growth in OECD countries using the Driscoll-Kraay standard error approach (1998) which concluded that increase in HE positively correlates with economic growth, particularly GDP growth. This dynamic relationship between HE and GDP is further explored in 'Health at a Glance': OECD Indicators (2021) using evidence from periods of economic shocks – the 2008 economic crisis and the COVID-19 pandemic, indicating that there has been a continual rise in the ratio of HE and GDP showcasing that more countries have started treating HE as a necessity. Stationarity and cointegration were used to investigate the relationship between HE and GDP in OECD countries using nonlinear structural breaks based on of a Fourier expansion by Lee et al. (2019), which found that around 70 percent of the countries considered for the study showcased the presence of stochastic co-movement between the series, an analysis which is further solidified in this research carried out on South Asian economies.

Traditionally, structural breaks in time series data are assumed to occur abruptly and concurrently. However, recent studies have acknowledged that structural changes can occur gradually, affecting the level or slope of a series over time. By employing tests developed by Enders and Lee (2012) and Banerjee et al. (2017), this study can to capture these smooth structural shifts in the data. In this study, we examine the unit root and cointegration characteristics of HE and GDP while considering the presence of gradual structural changes using the low-frequency components of a Fourier expansion.

The Fourier-based tests utilized in this research do not rely on assumptions or estimations of break-related parameters. Instead, they incorporate a time-varying deterministic term that includes trigonometric functions to capture the essential characteristics of a time series affected by multiple structural breaks without the need for additional estimation procedures to determine the number and location of breaks.

As highlighted by Enders and Lee (2012a), the aggregation of trigonometric functions typically provides a satisfactory fit for various types of nonlinear breaks, even when the specific form of the break in the data is uncertain. This versatility in performance across different types of breaks is particularly valuable in practical applications where the exact nature of the break is often unidentified.

3. Methodology

3.1 Unit Root Test

To determine if a time series is stationary, Dickey-Fuller type of regression is used, which is denoted as follows:

$$x_i = d(i) + \rho x_{i-1} + x_i + \epsilon_i \quad (1)$$

Where ϵ_i represents a stationary disturbance with variance σ_i^2 and x_i represents $ln\text{HE}_i$ or $ln\text{GDP}_i$. Previously, Becker et al. (2004) demonstrated that a time series with structural breaks (sudden changes in the data trend) can often be effectively captured by adapting Gallent's flexible Fourier form. In continuation of this, Enders and Lee (2012a, b) introduce a time-varying deterministic term d(i), which integrates Fourier expansion.

The Fourier expansion is represented as:

$$d(i) = \alpha_0 + \sum_{s=1}^{c} \left(\alpha_s \sin\frac{2\pi si}{N} + \beta_s \sin\frac{2\pi si}{N}\right), c \leq N/2 \qquad (2)$$

Here s is a specific frequency, c indicates the number of frequencies in the approximation, and N represents the number of observations. Based on Lagrange multiplier (LM) unit root tests (Schmidt and Phillips 1992; Amsler and Lee 1995; Enders and Lee 2012b), a Fourier LM unit root test is used to accommodate the time-varying deterministic term which accounts for gradual changes in the data. This test facilitates modeling smooth, evolving breaks over sharp ones.

For simplicity, the following analysis focuses on a single frequency s. The initial step in conducting the Fourier LM unit root test involves estimating a first-differenced regression.

$$\Delta x_i = \mu_0 + \mu_1 \Delta \sin\left(\frac{2\pi si}{N}\right) + \mu_2 \Delta \cos\left(\frac{2\pi si}{N}\right) + u_i \qquad (3)$$

Estimated coefficients $\tilde{\mu}_0$, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are used to build the following detrended series:

$$\tilde{Y}_i = x_i - \tilde{\lambda} - \tilde{\mu}_0 i - \tilde{\mu}_1 \sin\left(\frac{2\pi si}{N}\right) - \tilde{\mu}_2 \cos\left(\frac{2\pi si}{N}\right), i = 2, \ldots, N \quad (4)$$

In the above series, $\tilde{\lambda} = x_1 - \tilde{\mu}_0 - \tilde{\mu}_1 \sin\left(\frac{2\pi s}{N}\right) - \tilde{\mu}_2 \cos\left(\frac{2\pi s}{N}\right)$, and x_1 is the first observation of x_i. We determine the presence of a unit root in x_i by testing another regression using the detrended μ series of $\tilde{Y}_i$:

$$\Delta x_i = \psi \tilde{Y}_{i-1} + d_0 + d_1 \Delta \sin\left(\frac{2\pi si}{N}\right) + d_2 \Delta \cos\left(\frac{2\pi si}{N}\right) + \sum_{t=1}^{p} \xi_t \Delta \tilde{Y}_{i-t} + \varepsilon_i \quad (5)$$

p represents the number of lagged dependent variables included to account for possible serial correlation. Including a time trend does not affect the testing results as LM unit root tests are conducted on detrended data. The null hypothesis of unit root (H_0: $\psi = 0$) tests if the series has a unit root, while the alternative hypothesis (H_a: $\psi < 0$) suggests the series is stationary. κ_{LMi} denotes the test statistic for each country. Since the distribution of κ_{LMi} depends on the frequency s, the critical values of κ_{LMi} become a function of $\hat{s}$ denoted by:

$$\kappa_{LMi} = \kappa_{LMi}(\hat{s})$$

In the above, $\hat{s}$ represents the frequency s that minimizes the sum of squared residuals (SSR) obtained from testing regression (5). $\hat{s}$ is determined by testing different frequency values and selecting the one that gives the smallest SSR. The critical values for $\kappa_{LMi}(\hat{s})$ are given in Table 1 of Enders and Lee (2012a).

If the data does not display any nonlinear shifts, a simpler unit root test without trigonometric terms might be more appropriate. Enders and Lee (2012a) recommend first testing whether the data follows a linear trend (no sudden changes) or a nonlinear trend (with changes) with a specific frequency s using the N statistic below:

$$N(s) = \frac{(SSR_0 - SSR_1(s))/2}{SSR_1(s)/(N-c)} \quad (6)$$

$SSR_1(s)$ represents the sum of squared residuals from the regression that includes nonlinear deterministic function, where c denotes the number of regressors. SSR_0, on the other hand, is obtained from the regression that excludes the nonlinear terms. After determining the most appropriate single frequency from Table 1 of Enders and Lee (2012a), a test is conducted to see if the data has a nonlinear trend. Additionally, Enders and Lee (2012a, b) suggest that combining multiple frequencies might capture any remaining nonlinearity. In this case, you should choose a small number of frequencies (between 1 and 5) to prevent the potential loss of statistical power due to overfitting.

3.2 Cointegration test in the presence of smooth breaks

Banerjee et al. (1998) introduced an unrestricted dynamic representation of a conditional Error Correction Model (ECM) known as the Autoregressive distributed Lag (ADL) model:

$$\Delta x_i = \gamma + \alpha x_{i-1} + \theta z_{i-1} + \psi\ \Delta z_i + \sum_{t=1}^{P_1} \eta 1_t \Delta x_{i-t} + \sum_{t=1}^{p_2} \eta_{2t} \Delta y_{i-t} + e_i \quad (7)$$

In this model, p_1 and p_2 represent the number of lagged differenced regressors and dependent variables included to ensure no serial correlation in the error term e_i. In our estimation approach, x_i and z_i denote the natural logarithm of healthcare expenditure (HE_i) and gross domestic product (GDP_i), respectively. Similar to Gertham

and Löthgren (2000), we use a single-equation approach to model the relationship between variables without including a time trend. In this model, we focus on the first difference of $\ln HE_{\mathrm{i}}$ as the dependent variable. To test whether the variables are cointegrated (i.e., have a stable long-term relationship), we use the t-statistic for α (t_{ADLi}). The ADL (Autoregressive Distributed Lag) model we use is known for being accurate and powerful compared to other similar tests. Additionally, the ADL test avoids complications that arise in other models, like the conditional ECM (Error Correction Model), which includes a term that corrects errors over time. Furthermore, the ADL test does not suffer from the nuisance parameter issue present in the conditional ECM model, which includes the error-correction term ($x_{i-1} - \beta z_{i-1}$). Given these advantages, the ADL test has proven to be effective in addressing various econometric challenges.

Banerjee et al. (2017) introduced a modified version of the ADL cointegration test, that uses a simpler approach, building on Gallant's (1981) flexible Fourier form. This updated test accounts for smooth, gradual changes in the data, rather than sudden breaks. By incorporating low-frequency elements from a Fourier expansion into the model, this method avoids the need for an extra step to determine where and how many breaks occur. The modified testing regression is characterized as:

$$\Delta x_i = d(i) + \mu x_{i-1} + \theta z_{i-1} + \psi \Delta z_i + \sum_{t=1}^{c_1} \eta t \Delta x_{i-t} + \sum_{t=1}^{c} \eta_{2t} \Delta z_{i-t} + \epsilon_i \quad (8)$$

Where,

$$d(i) = \gamma_0 + \sum_{s=1}^{c} \gamma_1 , s \sin\left(\frac{2\pi si}{N}\right) + \sum_{s=1}^{q} \gamma_2 , s \cos\left(\frac{2\pi si}{N}\right), \leq N/2 \quad (9)$$

and $\epsilon_i \sim iid\ (0, \sigma^2_{\epsilon t})$. In Eq. (9), s refers to a specific frequency, c signifies the number of frequencies included in the approximation, and N denotes the total number of observations. To ascertain the appropriate values for c_1, c_2, and s, Banerjee et al. (2017) utilize Akaike's Information Criterion (AIC).

$$\mathrm{AIC} = F \ln\left(\frac{SSR}{N}\right) + 2r \quad (10)$$

In this context, N represents the total number of observations, r denotes the number of estimated parameters, and SSR stands for the

sum of squared residuals obtained from Equation (8). After selecting a single frequency, denoted as ŝ, the test statistic $t^{F}_{ADLi}(\hat{s})$ is derived from the t statistic for the null hypothesis of $\mu = 0$ in Equation (8). To examine the hypothesis of no cointegration (μ=0) versus the alternative ($\mu < 0$), one can compare the value of $t^{F}_{ADLi}(\hat{s})$ with critical values provided in Table 1 of Banerjee et al. (2017). Since using a single frequency (c=1) usually captures the key characteristics of a time series, another method is proposed that includes multiple frequencies. In this method, all frequencies from 1 to c are included in the equation. The number of frequencies (c) is chosen using the Akaike Information Criterion (AIC). It's recommended to keep c small, between 1 and 5, to avoid overfitting, which could reduce the test's accuracy. The resulting test statistic for multiple frequencies, $t^{F}_{ADLi}(\hat{c})$, is then compared to critical values reported in Table 2 by Banerjee et al. (2017).

4. Data and Results

4.1 Data

For the empirical analysis, this study utilized recent data spanning from 2002 to 2020[3] for eight South Asian countries: Afghanistan, Bangladesh, Bhutan, India, Sri Lanka, Maldives, Nepal, and Pakistan. Gross Domestic Product (GDP) and population data were sourced from the World Development Indicators (WDI), while healthcare expenditure per capita was gathered from the South Asia Health Expenditure Database of the World Bank. Both sets of data underwent a logarithmic transformation[4].

4.2 Empirical evidence

Unit Root Test

Table 1 displays the findings from panels (a) and (b), showcasing the application of the methodology proposed by Enders and Lee (2012a). This methodology aims to investigate the presence of a unit root in the two series, accounting for smooth time-varying shifts. Enders and Lee (2012a) show that the distribution of the test statistic depends on the value of the frequency (c) and the number of observations (N). The

[3] The period from 2002 to 2020 was chosen due to the availability of consistent and reliable data.

[4] $\ln (HE)_{it}$ and $\ln (GDP)_{it}$ denote the logarithm of HE and GDP, respectively, for each country i measured in the year t.

critical values at 1, 5, and 10 percent significance levels are reported in Table 1.

Table 1 – Unit Root Test using a single frequency

Country	(a) Healthcare Expenditure				(b) GDP			
	t-stat	*s*	Lags	F(*s*)	t-stat	*s*	Lags	F(*s*)
Afghanistan	-2.317	1	3	33.601	-2.741	1	3	62.839
Bangladesh	-0.780	3	0	22.683	-0.436	3	0	52.119
Bhutan	-4.159*	1	0	44.917	-3.035	1	0	65.321
India	-3.084	1	3	39.682	-2.632	1	0	60.725
Sri Lanka	-3.148	1	3	41.448	-2.731	1	3	56.308
Maldives	-1.857	5	2	56.355	-2.484	1	0	84.133
Nepal	-1.655	4	3	22.575	-2.726	1	0	48.341
Pakistan	-3.078	1	1	39.773	-3.328	1	1	99.724

(a) *The 1, 5, and 10% critical values of Banerjee et al. (2017) tests with single frequency are: −4.86, −4.15, and −3.79 ($s = 1$); −4.56, −3.79, and −3.39 ($s = 2$); −4.27, −3.51, and −3.13 ($s = 3$); −4.12, −3.37, and −3.00 ($s = 4$); −4.04, and −3.30, −2.94 ($s = 5$)
(b) ***, **, and * represent 1, 5, and 10% rejection, respectively

In Table 1, the first columns of panels (a) and (b) represent a test statistic for each country. Based on the single frequency value, *s*, in Table 1, the unit root test makes use of two periodic functions to address potential smooth breaks that may exist within the series. The selected value for the single frequency *s* which minimizes the SSR is reported in the second columns of both the panels.

The results show that the null hypothesis of unit root for HE is rejected for the country Bhutan and cannot be rejected for all the other countries at a 10 percent significance level. Whereas the null hypothesis is not rejected for all countries at the same significance. Since we have used trigonometric components to control nonlinearities the results indicate that HE and GDP can be categorized as $I(1)$ processes.

Cointegration Test

In this section, cointegration tests of Banerjee et al. and Engle and Granger (1987) are discussed in detail which allows us to focus on a case where two or more series are cointegrated with an unknown number of smooth shifts. We examine residual-based Engle and Granger test without structural breaks and the ADL test which allows

for linear structural breaks. The maximum lag is set to be 5. The number of lags for the EG test is determined through a general-to-specific approach: we assess the significance of the final lagged variable by comparing it with a threshold of 1.65. The most suitable lag for the ADL model is chosen based on the Akaike Information Criterion (AIC), with a maximum lag of 5. In the case of the Engle-Granger (EG) test, both a constant and a time trend are included in the deterministic function to capture the notable upward trend observed in healthcare expenditure (HE) and GDP.

When results are examined, we found that the EG test rejects the null of no cointegration for only one country – India at a 10 percent significance level as shown in Table 2. This implies that HE and GDP are cointegrated around a linear time trend only for one country. Whereas, in the ADL test the null hypothesis is not rejected for any country.

Table 2 – cointegration test with/without linear break

Country	**EG** (without structural breaks)		**ADL** (with linear structural break)			
	t-stat	EG_lag	t-stat	Lag: Y	Lag: X	Best AIC
Afghanistan	-1.965	1	1.216	3	3	-11.300
Bangladesh	-2.229	1	2.102	2	2	-8.942
Bhutan	-0.937	3	-0.845	3	1	-9.424
India	-4.012*	3	-0.241	3	3	-10.212
Sri Lanka	-2.705	2	2.270	1	3	-11.343
Maldives	-2.183	0	-0.148	3	3	-13.307
Nepal	-1.321	3	1.391	2	2	-8.038
Pakistan	-2.735	3	2.462	3	2	-10.491

(a) *The 1, 5, and 10% critical values for the EG test from the response surface approach of MacKinnon are −4.76, −4.04, and −3.69, respectively.
(b) The 1%, 5%, and 10% critical values for the ADL test ($N = 50$) are −3.95, −3.29, and −2.94, respectively
(c) ***, **, and * represent 1, 5, and 10% rejection, respectively

Fourier Cointegration

We implement the Fourier ADL test with single and cumulative frequencies considering smooth breaks to investigate the long-term relationship between the series across the 8 South Asian countries; the

results are shown in Table 3. Both the single and cumulative frequency sections present FADL, lag specifications, and Akaike Information Criterion (AIC) values for both dependent and independent variables. The maximum number of frequencies is set to be 5 and the number of frequencies is selected using minimum AIC.

As per the results, in 3 out of 8 countries, the null of no cointegration is rejected, namely - Afghanistan, Maldives, and Pakistan. These countries reject the null hypothesis at a 1 percent rejection level, which is considered potent and further supports the notion of stochastic co-movement between HE and GDP. Countries such as Bangladesh, Bhutan, India, Sri Lanka, and Nepal present varying degrees of ADL results which are less pronounced compared to Afghanistan, Maldives, or Pakistan.

Table 3 – Fourier cointegration with smooth breaks

Country	Single Frequency					Cumulative Frequency				
	FADL	Lags			AIC	FADL	Lags			AIC
	t-stat	*s*	DY	DX		t-stat	*c*	DY	DX	
Afghanistan	-6.861***	2	3	3	-18.940	0.000	3	2	3	-35.142
Bangladesh	1.242	2	3	2	-11.752	1.466	3	3	2	-11.362
Bhutan	1.295	1	2	2	-12.782	-0.896	3	3	3	-13.227
India	0.541	1	2	2	-12.635	-3.752	3	3	2	-12.733
Sri Lanka	-1.378	1	1	1	-14.025	-3.362	3	3	3	-14.510
Maldives	-9.966***	3	3	2	-17.325	0.258	3	3	3	-17.324
Nepal	-2.271	2	3	3	-11.982	-0.466	3	3	2	-11.177
Pakistan	-6.424***	2	3	3	-12.292	-0.217	3	3	3	-12.014

(a) *The 1, 5, and 10% critical values of Banerjee et al. (2017) tests with single frequency are: −4.86, −4.15, and −3.79 (s = 1); −4.56, −3.79, and −3.39 (s = 2); −4.27, −3.51, and −3.13 (s = 3); −4.12, −3.37, and −3.00 (s = 4); −4.04, and −3.30, −2.94 (s = 5)
(b) The 1, 5, and 10% critical values of Banerjee et al. (2017) tests with multiple frequencies are: −5.63, −4.88, and −4.50 (c = 2); −6.34, −5.54, and −5.13 (c = 3); −7.03, −6.17, and −5.74 (c = 4); −7.70, −6.77, and −6.30 (c = 5)
(c) ***, **, and * represent 1, 5, and 10% rejection, respectively

To further validate our results, we employed an additional variable, inflation, in the Fourier cointegration model. The findings from the modified Fourier cointegration model, which includes inflation, are consistent with those from the original model without

inflation, with the exception of Afghanistan[5]. The results[6] of the Unit Root and Cointegration tests are provided in the Appendix.

Income Elasticity

In this section, we examine whether healthcare expenditure is a luxury good or a necessity. 3 out of 8 countries reject the null of no cointegration – here, the income elasticity of HE that is displayed below notably shows Pakistan having an elasticity of 1.236, which indicates that Pakistan considers HE to be a luxury good. Afghanistan and Maldives have an income elasticity that can be rounded off to a 1, which is at a less significant level than Pakistan's income elasticity of HE, through which we can infer that HE is a normal good for these 2 countries.

Overall, it can be said that South Asian countries regard healthcare expenditure as a necessity rather than a luxury good, and these nations tend to spend more on healthcare as their income increases.

Table 4 – Income Elasticity of HE Analysis

Country	μ	θ	Income elasticity of HE
Afghanistan	-0.679	0.688	1.012
Maldives	-0.405	0.408	1.009
Pakistan	-0.442	0.546	1.236

5. Conclusion and Discussion

In this paper, we analyzed the long-run relationship between healthcare expenditure and GDP for 8 South Asian countries. Unit root test (Enders and Lee 2012a) and cointegration test (Banerjee et al. 2017) were used to examine the stationarity of the data, accounting

[5] Inflation data for Afghanistan is only available for the years 2005 to 2019 (Source: World Bank). The highly unstable inflation rate during this period may have impacted its co-movement relationship with healthcare expenditure.

[6] Sri Lanka and Nepal are not considered for checking the robustness of the Fourier cointegration model because inflations of these countries are stationary at 5 percent and 1 percent significant levels based on Table A1.

for gradual structural changes and smooth shifts in the non-stationarity of HE and GDP among South Asian countries.

The cointegration analysis, based on the methods of Banerjee et al. (2017), provided evidence of a long-term equilibrium relationship between HE and GDP in 3 of the 8 South Asia countries studied. This suggests that despite short-term fluctuations, there is a stable relationship between economic growth and healthcare spending in the long run.

One of the significant findings of this study is the variation in income elasticity of HE across the countries. For Afghanistan and the Maldives, the income elasticity of HE was found to be approximately 1, indicating that healthcare is considered a normal good in these countries. This implies that as income increases, healthcare expenditure increases proportionally, reflecting its importance in the consumption basket. The elasticity values suggest that in these economies, healthcare is a necessity rather than a luxury, likely due to the basic level of healthcare infrastructure and services that need to be met as incomes rise.

In contrast, Pakistan exhibited a higher income elasticity of HE, greater than 1, suggesting that healthcare is viewed as a luxury good. This could be attributed to the disparities in income distribution and healthcare access within the country. As incomes rise, individuals may allocate a larger portion of their income to healthcare, possibly due to increased access to private healthcare services, which are perceived as higher quality but come at a higher cost.

This analysis offers a macroeconomic perspective on the dynamics between healthcare spending and GDP in selected South Asian countries. However, several limitations should be acknowledged. First, using healthcare expenditure as a percentage of GDP, while useful for cross-country comparisons, oversimplifies the multifaceted nature of healthcare financing, efficiency, and outcomes. Second, significant differences exist among the countries' healthcare systems in terms of structure, funding mechanisms, and access, all of which can influence the relationship between income and healthcare spending. Third, whether healthcare is classified as a ‘necessity’ or ‘luxury’ good is context-dependent, shaped by factors such as the availability of public services, socioeconomic disparities, cultural expectations, and, critically, government policies. These policies influence the degree to which healthcare is treated as a public right versus a market-driven

service. Future research should explore these micro-level factors to provide a more comprehensive understanding of how healthcare expenditure patterns evolve in relation to economic development across South Asia.

References

Khan JA and Mahumud RA (2015) Is healthcare a 'Necessity' or 'Luxury'? an empirical evidence from public and private sector analyses of South-East Asian countries. *Health Economics Review* 5:1-14

Banerjee A, Dolado J, and Mestre R (1998) Error-correction mechanism tests for cointegration in a single equation framework. *Journal of Time Series Analysis* 19:283–297

Banerjee P, Arčabić V, and Lee H (2017) Fourier ADL cointegration test to approximate smooth breaks with new evidence from crude oil market. *Economic Modelling* 67:114–124

Beylik U, Cirakli U, Cetin M, Ecevit E, and Senol O (2021) The Relationship between Health Expenditure Indicators and Economic Growth in OECD Countries: A Driscoll-Kraay Approach. *Frontiers in Public Health* 10:1-10

Blomqvist AG and Carter R (1997) Is health care really a luxury? *Journal of Health Economics* 16:207–229

Carrion-i-Silvestre JL (2005) Health care expenditure and GDP: are they broken stationary? *Journal of Health Economics* 24(5):839–854

Clemente J, Marcuello C, Montañés A, and Pueyo F (2004) On the international stability of health care expenditure functions: are government and private functions similar? *Journal of Health Economics* 23(3):589–613

Dieleman JL, Squires E, Bui AL, Campbell M, Chapin A, Hamavid H, Horst C, Li Z, Matyasz T, Reynolds A, Sadat N, Schneider MT, and Murray CJL (2017) Factors Associated with Increases in US Health Care Spending. *Journal of the American Medical Association* 318(17):1625-1728

Enders W and Lee J (2012a) The flexible Fourier form and Dickey–Fuller type unit root test. Economics Letters 117:196–199

Enders W and Lee J (2012b) A unit root test using a Fourier series to approximate smooth breaks. Oxford Bulletin of Economics and Statistics 74:574–599

Engle RF and Granger CWJ (1987) Cointegration and error correction: representation, estimation, and testing. Econometrica 55(2):251–276

Farag M, NandaKumar A, Wallack S, Hodgkin D, Gaumer G, and Erbil C (2012) The income

elasticity of health care spending in developing and developed countries. International Journal of Health Care Finance and Economics 12:145–162

French D (2012) Causation between health and income: A need to panic. Empirical Economics 42:583–601

Gerdtham Ulf-G, and Löthgren M (2000) On Stationarity and Cointegration of International

Health Expenditure and GDP. Journal of Health Economics 19(4):461–475

Gregory AW and Hansen BE (1996a) Residual-based tests for cointegration in models with regime shifts. Journal of Econometrics 70(1):99–126

Hansen P, and King A (1996) The determinants of health care expenditure: A cointegration approach. Journal of Health Economics 15:127–137

Jewell T, Lee J, Tieslau M, and Strazicich MC (2003) Stationarity of health expenditures and GDP: evidence from panel unit root tests with heterogeneous structural breaks. *Journal of Health Economics* 22(2):313-323

Kleiman E (1974) The determinants of national outlay on health. The economics of health and medical care. *International Economic Association Series*. 66-68

Lee H, Oh D, and Meng M (2019) Stationarity and cointegration of health care expenditure and GDP: evidence from tests with smooth structural shifts. *Empirical Economics* 57: 631–652

McCoskey S and Selden TM (1998) Health care expenditures and GDP: panel data unit root test results. *Journal of Health Economics* 17:369–376

Newhouse JP (1977) Medical care expenditure: a cross-national survey. *The Journal of Human Resources* 12:115–125

Roberts, J (2000). Spurious regression problems in the determinants of health care expenditure: A comment on Hitiris (1997). *Applied Economics Letters* 7:279–283

Tomakoshi T and Hamori S (2015) Health-care expenditure, GDP and share of the elderly in Japan: a panel cointegration analysis. *Japanese Economic Review* 66(2): 248-259

Appendix

Table A.1 – Unit Root Test using a single frequency for Inflation

Country	Inflation			
	t-stat	*s*	Lags	F(*s*)
Afghanistan	-3.095	1	2	3.186
Bangladesh	-2.943	1	3	13.205
Bhutan	-3.329	1	0	10.063
India	-3.289*	5	3	10.495
Sri Lanka	-4.260**	1	0	7.360
Maldives	-1.802	4	0	3.387
Nepal	-4.972***	1	0	10.168
Pakistan	-3.226	1	0	8.111

(a) The 1, 5, and 10% critical values of Banerjee et al. (2017) tests with single frequency are: −4.86, −4.15, and −3.79 (*s* = 1); −4.56, −3.79, and −3.39 (*s* = 2); −4.27, −3.51, and −3.13 (*s* = 3); −4.12, −3.37, and −3.00 (*s* = 4); −4.04, and −3.30, −2.94 (*s* = 5)
(b) ***, **, and * represent 1, 5, and 10% rejection, respectively

Table A.2 – Fourier cointegration with smooth breaks

Country	(a) Single Frequency					
	FADL	Lags				AIC
	t-stat	*s*	DY	DX1	DX2	
Afghanistan	-0.137	1	1	1	1	-6.186
Bangladesh	-0.061	1	1	2	1	-7.999
Bhutan	-1.950	1	1	1	1	-4.699
India	-0.079	1	2	2	2	-7.636
Maldives	-5.575***	1	2	1	1	-4.856
Nepal	-2.566	1	2	1	2	-6.32
Pakistan	-6.371***	1	1	2	1	-6.634

Country	(b) Cumulative Frequency					
	FADL	Lags				AIC
	t-stat	*c*	DY	DX1	DX2	
Afghanistan	0.000	2	1	1	1	-43.669
Bangladesh	-1.699	2	2	1	1	-8.211
Bhutan	-1.519	2	2	1	2	-5.975
India	-0.237	2	2	2	2	-7.458
Maldives	-3.586	2	2	1	1	-4.972
Nepal	-1.524	2	2	2	1	-6.685
Pakistan	-5.462***	1	1	1	2	-6.842

(a) The 1, 5, and 10% critical values of Banerjee et al. (2017) tests with single frequency are: −4.86, −4.15, and −3.79 (s = 1); −4.56, −3.79, and −3.39 (s = 2); −4.27, −3.51, and −3.13 (s = 3); −4.12, −3.37, and −3.00 (s = 4); −4.04, and −3.30, −2.94 (s = 5)
(b) The 1, 5, and 10% critical values of Banerjee et al. (2017) tests with multiple frequencies are: −5.63, −4.88, and −4.50 (c = 2); −6.34, −5.54, and −5.13 (c = 3); −7.03, −6.17, and −5.74 (c = 4); −7.70, −6.77, and −6.30 (c = 5)
(c) ***, **, and * represent 1, 5, and 10% rejection, respectively
(d) The dependent variable is Y (Healthcare Expenditure), while the independent variables are X1 (GDP) and X2 (Inflation).

Impact of Gender on Raises: Evidence from the Survey of Household Economic Decisionmaking 2017 – 2021

James Farrell
Florida Southern College

Abstract

This paper explores the relationship between gender and certain labor market choices and outcomes based on survey results from the Federal Reserve's Survey of Household Economic Decisionmaking (SHED) from the years 2017 through 2021. Using binary regression analysis, this paper finds that when considering each of the years separately, female respondents were less likely than male respondents to ask for a raise in the 2017 – 2018 years, but not less likely in the later years (2019 – 2021). Diving deeper into the results showed that it was only the non-married female respondents that had statistically different odds of asking for a raise in the later survey period, with the odds improving from 0.826 to 1.059 (Z-test of difference: -2.53). This finding, in combination with the impact of asking for a raise on receiving a raise (63.1 percent vs 42.0 percent for asking and not asking, respectively), suggests that there was an economic improvement for non-married females in the later years. While this paper cannot show a causal relationship between the willingness to negotiate, the social movements of the period, including the #MeToo movement, appear to be coinciding factors to greater parity in negotiation attempts.

1. Introduction

When examining the gender wage gap, many studies have focused on a mix of characteristics including field of employment, schooling, age, and work history, but newer studies have also pointed to psychological norms and preferences as contributing factors to the persistent gap. Among preferences including competition and risk aversion, the willingness to negotiate stands out for its role on wages (Blau and Khan 2017; Bertrand 2011).

Negotiations may have a meaningful impact on labor market outcomes. Small et al. (2007) found that when offered the lowest cash payment for participation in an experiment, 23 percent of men asked for more money, while only 3 percent of women asked for more. Leibbrandt and List (2015) found that the clarity on whether negotiations were expected can influence the gender gap in job applications, with women preferring jobs that explicitly listed the wages as negotiable. This notion fits with work from Babcock et al. (2003) that suggested women may have been socialized to believe that negotiating could be perceived as being pushy. However, if negotiating is explicitly listed as acceptable, women are more likely to initiate negotiations, i.e., cues matter. Croson and Gneezy (2009) highlighted that while several studies focused on the results of negotiations and that there was only a small gender gap, the larger and more important gap existed in whether negotiations were initiated at all. A more recent working paper (Cortes et al., 2024), which looked at college graduates that were surveyed about their negotiating behavior, found that the proportion of female students negotiating behavior increased for the 2018 - 2020 cohorts, while male negotiating behavior remained similar to previous years.

Hendrix et al. (1998) discussed the perceptions of gender on sexual harassment and found that female respondents were more likely to perceive being sexually harassed by male initiators and that the initiator having a position of power increased the perception of sexual harassment as well. This begs the question of whether a wide-spread "calling-out" of sexual harassment from bosses would have an impact on the gender gap in willingness to negotiate. The social changes, such

as the #MeToo movement, which were reaching a peak in late 2018, may have had influence, among other factors, on individual's willingness to negotiate.

Due to the complexity of these issues, looking towards intersectionality literature adds a bit of insight into how gender, race and other characteristics may influence one's willingness to negotiate. Giritli Nygren et al. (2020) discuss how the role of identities shape experiences with risk, and how risk is embedded within power structures. As we see changes in the power structures, we may see changes in the experiences with risk. Mazei et al. (2015) performed a meta-analysis of the literature on negotiations across genders and found the differences to be context based, where certain situations, e.g., being provided salary ranges, result in smaller gender differences. They also pointed to predictions from role congruity theory (Eagly & Karau, 2002), which argues that the cultural norms expected of women and the biases towards them leave them misaligned for labor negotiations, as a reasonable predictor of behaviors.

This paper analyzes the gender gap of responses to the questions relating to asking for a raise from the Federal Reserve Board's Survey of Household Economic Decisionmaking (SHED) over the years 2017 – 2021. The first section will discuss and summarize the data, the second section will briefly discuss motivation, the third section will review the model used to study the results. The fourth section will discuss the results along with any caveats. The fifth section will summarize the findings and layout areas for future research and improvements in the analysis.

2. SHED Data Summary

The Federal Reserve Board has conducted the SHED survey since 2013 with the goal of measuring the economic well-being of U.S. households and gather information on their finances, behaviors, and potential risks. This survey combines key demographic characteristics with a wide variety of preference, financial literacy, and economic health questions.

The SHED data provides a mix of respondents across age, race, marital status, educational attainment, and income levels. The response rate remained consistent over time, with the number of responses ranging from a low of 11,158 in 2018 to a high of 12,294 in 2017. In addition to these variables, geographical regions[1] were also used as controls, but will be suppressed in the results tables. The survey was conducted during the Fall of each year and the question regarding asking for a raise references the prior 12 months, so, for the 2017 data, the reference period would be Fall 2016 through Fall 2017.

This analysis is not considered representative of the U.S. population as population weights are not used. While there is value in generalizing the results to the broader population, the goal of the research is to identify the impact between two multi-year time periods. Additionally, the analysis includes only those respondents that were employed. While the weights may be scalable across imbalanced respondent sizes when combining multiple years of data, they are no longer applicable when the data is subset by a characteristic not included in the weight generation.

The percentage of respondents that asked for a raise each year is summarized in Table 1, where we can see that there was a notable spike in 2019 (10.1 percent) and 2021 (8.9 percent) compared to other years.

Table 1: Summary of Demographic Variables by Year

Year	Asked for a Raise
2017	7.1%
2018	7.5%
2019	10.1%
2020	7.0%
2021	8.9%

[1] This analysis used the 9-category regional control variable.

The remaining summary data is included in Appendix Table A1, where we can see that most of the other key demographic variables (gender, age, race, education, and income) had slight variation from year to year. Marital status had one minor reporting change in 2021, as "Living with a Partner" was no longer an answer choice.

It is also important to point out that the percentages were based on the total number of respondents for each year, e.g., 873 out of 12294 (7.1 percent) of the respondents asked for a raise in 2017, however 7090 of the survey respondents replied that the question was not applicable because they were not employed. In 2018 – 2021, unemployed people were no longer asked the raise question. The analysis that follows will limit the data to the participants for which the raise question was applicable and answered with either affirmative or negative and exclude those that were not employed or that responded that it was not applicable. Additionally, there are a small number (7) of cases where participants selected "NA" for an answer to question included in the analysis.[2] Given the small number of occurrences, those observations were excluded from the analysis.

3. Motivation

Before analyzing this data, it is important to consider the relationship between asking for a raise and receiving one as the economic value of a willingness to ask for a raise fundamentally lies in whether that translates into receiving a raise. Across all years in this analysis, only 42.0 percent of respondents received a raise if they did not ask for one, while 63.1 percent of respondents received a raise if they asked for one, as shown in Table 2. While there are many more complicating factors that would impact whether a respondent received a raise, this simple relationship stands out and supports the argument that improving willingness to ask for a raise should improve outcomes overall.

[2] There were 7 cases where "NA" was selected as an Income category among employed respondents.

Table 2: Relationship between Asking for and Receiving a Raise

	Received a Raise		
Asked for a Raise	No	Yes	Total Count
No	58.0%	42.0%	27806
Yes	36.9%	63.1%	4785

4. Model Specification

The survey question regarding asking for a raise has two possibilities for each respondent (Yes or No), so modeling this result requires use of a binary variable model. In this case, a logistic model is appropriate with the following model (Pampel 2000):

$$ln\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta_1 female_i + \beta_2 demographics_i + \beta_3 education_i + \beta_4 region_i + \beta_4 income_i$$

where the dependent variables are a dummy variable ($female_i$) for gender, a vector of basic demographic categorical variables ($demographics_i$) which include age, race and marital status, an education categorical variable ($education_i$), a region categorical variable ($region_i$), and a household income categorical variable ($income_i$). For ease of interpretation, this paper will report the odds, rather than the log-odds.

There are several limitations to this analysis as the SHED data only contains the responses from one member of the household and may not present a full picture of the household's labor dynamics. It also lacks insight into a general willingness to take risks, which may have an impact on the willingness to ask for a raise. There are also detailed job characteristics which are not available, including industry and occupation[3], that may have both stand-alone and interacting impacts on the willingness to ask for a raise. Robustness tests are

[3] SHED does include these survey questions, but they are too broadly defined to add meaningfulness to the analysis.

conducted to rule out certain omitted variable biases following the analysis.

5. Results

To analyze the data, this paper looks at the issue from a couple of angles:

1. Separate the regressions by year to identify if any patterns emerge among the key variables.
2. Divide the years into two categories, the early (2017 – 2018) and later (2019 – 2021) periods and take a deeper look into the role of marital status by gender.
3. Z-test for comparing the coefficients of across the two periods.

In Table 3, we can see that Female respondents had statistically significant lower odds than Male respondents to ask for a raise in 2017 and 2018 (90 percent and 99 percent level of significance, respectively), while they no longer had statistically significant lower odds in 2019 – 2021. We also see the expected relationship between age and the odds of asking for a raise, with the odds falling as age increases beyond the typical peak earning years. What also stands out in this analysis is the role of the "Never Married" marital status, which follows the pattern found among Female respondents of having statistically lower odds in the 2017 – 2018 surveys, but not statistically different results in the 2019 – 2021 period. It is that result that motivated the deeper look into the interaction between gender and marital status.

Table 3: Impact (Odds Ratios) of Gender on Willingness to Ask for a Raise, By Year[4]

	(1) 2017	(2) 2018	(3) 2019	(4) 2020	(5) 2021
Female	0.863*	0.795***	0.921	0.928	0.923
	(0.0680)	(0.0618)	(0.0606)	(0.0725)	(0.0642)
Age					
25 - 34	1.665***	0.925	1.387**	1.035	1.436**
	(0.319)	(0.165)	(0.204)	(0.155)	(0.234)
35 – 44	1.058	0.854	1.068	0.904	0.934
	(0.217)	(0.160)	(0.169)	(0.151)	(0.165)
45 – 54	0.975	0.584***	0.751*	0.595***	0.685**
	(0.203)	(0.113)	(0.125)	(0.108)	(0.126)
55 – 64	0.808	0.464***	0.525***	0.507***	0.537***
	(0.170)	(0.0907)	(0.0892)	(0.0917)	(0.0999)
65 – 74	0.544**	0.196***	0.231***	0.319***	0.414***
	(0.154)	(0.0523)	(0.0550)	(0.0790)	(0.0922)
75+	1.111	0.187***	0.283***	0.314**	0.0757***
	(0.595)	(0.0929)	(0.114)	(0.143)	(0.0556)
Race					
Black	1.154	1.524***	1.473***	1.093	0.998
	(0.159)	(0.183)	(0.154)	(0.143)	(0.119)
Hispanic	1.417***	1.149	1.077	0.919	0.916
	(0.155)	(0.133)	(0.109)	(0.113)	(0.101)
Asian	1.380	1.314	1.392**	1.096	1.179
	(0.287)	(0.236)	(0.205)	(0.193)	(0.185)
Other	1.045	1.557***	1.285	0.961	1.173
	(0.223)	(0.267)	(0.196)	(0.190)	(0.199)
Marital Status					
Widowed	1.107	0.864	1.532	1.200	1.174
	(0.350)	(0.286)	(0.401)	(0.387)	(0.361)
Divorced	1.099	1.260*	1.209	1.081	1.348**
	(0.143)	(0.160)	(0.145)	(0.166)	(0.171)
Separated	0.992	0.884	1.052	1.438	2.123***
	(0.290)	(0.266)	(0.280)	(0.430)	(0.536)
Never Married	0.767**	0.752**	0.858	1.011	1.035
	(0.0880)	(0.0878)	(0.0818)	(0.117)	(0.0956)
Live w/ Partner	1.149	1.250*	1.166	1.485***	N/A[5]
	(0.168)	(0.169)	(0.128)	(0.192)	
Constant	0.257***	0.382**	0.443**	0.179***	0.268***
	(0.103)	(0.156)	(0.141)	(0.0706)	(0.0994)
Observations	5,189	6,323	7,318	6,759	7,010

Standard Errors of Odds in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[4] The Regions, Income and Education results were suppressed, they offered limited additional insight into the behavior.
[5] "Live w/ Partner" was no longer an option in the 2021 survey.

To take a deeper look into some of the variable interactions, we will adjust the specification by defining Married as respondents that selected Married and Single as all other respondents. Further, we will look at the interactions between marital status and gender. The original model is respecified to be:

$$ln\left(\frac{P_i}{1-P_i}\right) = \alpha + \beta_1 singlefemale_i + \beta_2 marriedfemale_i + \beta_3 singlemale_i + \beta_2 demographics_i + \beta_3 education_i + \beta_4 region_i + \beta_4 income_i$$

where *singlefemale*, *marriedfemale* and *singlemale* are now dummy variables for their respective cases and marital status as a demographic variable is no longer included. To further consider the impact of the potential shift from the 2017-2018 period to the 2019 – 2021 period, the observations will be grouped together by those years, rather than analyzed by individual year.

In Table 4 we can see the relationships a bit more clearly, with both Single Female and Single Male respondents statistically less likely (at the 95 percent and 90 percent levels, respectively) than Married Males to ask for a raise in the 2017 - 2018 period and no longer statistically less likely in the 2019 - 2021 period, while Married Females remained less likely to ask for a raise across both periods.

Table 4: Analysis by Marital Status and Year Grouping

	2017 - 2018	2019 - 2021
Single Female	0.826**	1.059
	(0.0667)	(0.0634)
Married Female	0.768***	0.869**
	(0.0575)	(0.0488)
Single Male	0.874*	1.037
	(0.0711)	(0.0614)
Age		
25 - 34	1.324**	1.362***
	(0.170)	(0.118)
35 – 44	1.037	1.064
	(0.139)	(0.0986)
45 – 54	0.833	0.750***
	(0.113)	(0.0723)
55 – 64	0.669***	0.577***
	(0.0914)	(0.0560)
65 – 74	0.340***	0.351***
	(0.0635)	(0.0452)
75+	0.392***	0.247***
	(0.138)	(0.0666)
Race		
Black	1.286***	1.182**
	(0.115)	(0.0787)
Hispanic	1.292***	0.976
	(0.101)	(0.0617)
Asian	1.282*	1.210**
	(0.173)	(0.110)
Other	1.278*	1.164
	(0.170)	(0.114)
Constant	0.315***	0.290***
	(0.0884)	(0.0581)
Observations	11,512	21,087

Standard Errors of Odds in parentheses
*** p<0.01, ** p<0.05, * p<0.1

It is important, however, to confirm this by checking for the statistical difference between the coefficients using a Z-test as described in Clogg et al. (1995) with:

$$Z = \frac{\beta_1 - \beta_2}{\sqrt{\left(SE_{\beta_1}\right)^2 + \left(SE_{\beta_2}\right)^2}}$$

Computing these values for each of the gender-marital status variables we get Z scores of -2.53, -1.34 and -1.74 for the Single Female, Married Female and Single Male respondents, respectively, as shown in Table 5. The Single Female respondents are the only group that had a statistically different result at the 95 percent level from the 2017 - 2018 period to the 2019 - 2021 period. This is an interesting finding, particularly in comparison with Cortés et al (2024), whose work focused on graduating students, who are primarily young and unmarried, also showing an increase in willingness to negotiate.

Table 5: Z-scores for Difference in Coefficients

	Z-score
Single Female	-2.53
Married Female	-1.34
Single Male	-1.74

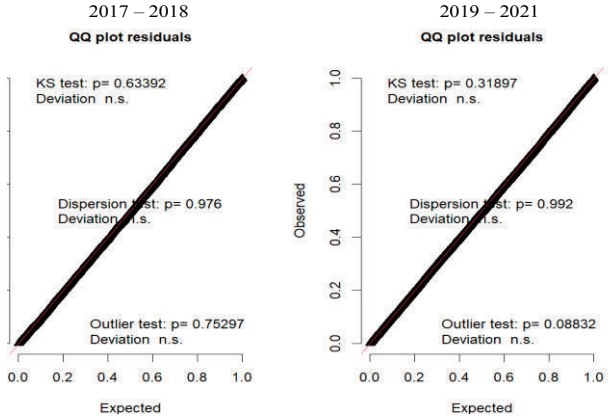
Robustness checks were performed as part of the analysis. To check for multi-collinearity in the models, a Variance Inflation Factor (VIF) test was run on both the 2017-2018 and 2019 – 2021 periods. We can see in Table 6 below, all variables show little multi-collinearity, with VIFs ranging between 1.2 and 2.1 (O'brien, 2007).

Table 6: Testing for Multi-Collinearity (VIF)

Variable	**2017 - 2018**	**2019 - 2020**
Single Female	1.789873	1.723294
Married Female	1.333805	1.314390
Single Male	1.544538	1.560975
Age Categories	1.347315	1.410896
Racial Categories	1.325860	1.296061
Education Categories	1.483048	1.488716
Region Categories	1.233994	1.209673
Income Categories	2.072490	1.963085

To check the dispersion of the results, a DHARMa nonparametric dispersion test (Hartig, 2024) was run on the two models. The results below show that the results are well dispersed (Dispersion test p>.05) and have insignificant deviations (KS test >.05), for both model periods.

Figure 1: DHARMa nonparametric dispersion test via sd of residuals vs. simulated



To check for misspecification among the included variables, we can look at the results of a Link test (Pregibon, 1980) to show that the model does not have a link error. While this test has limited scope, it provides some additional reassurance that the model is not improperly specified. We can see in Table 7 that the predicted values are statistically significant at the 5 percent level while the squares of the predicted values are not, meeting the goals of the link test, and therefore, not showing model misspecification.

Table 7. Results of Link Test

	2017-2018	2019-2021
Characteristic	**OR (SE)**[1]	**OR (SE)**[1]
predicted	2.37* (0.404)	2.75** (0.314)
I(predicted^2)	0.96 (0.111)	1.00 (0.085)

[1] *p<0.05; **p<0.01; ***p<0.001

Abbreviations: OR = Odds Ratio, SE = Standard Error

6. Conclusions

The analysis presented several interesting results. First, there appears to be a shift in the willingness of women to ask for a raise that coincides to the timing of the #MeToo movement among other social changes. The results presented in Table 4 and Table 5 suggest that there was a statistically significant change in the willingness for some demographic groups to ask for a raise.

Second, this result is strongest for single women. Single women are likely more financially vulnerable than married women, as findings indicate that female-headed households have higher levels of financial vulnerability compared to male-headed households. This vulnerability has been increasing through 2016, and is highest for younger household heads, both of which are more prevalent among single women (Azzopardi, et al. 2019). The willingness to ask for a raise among single women was no longer statistically different than that of either single or married men during the 2019 – 2021 period. This is a similar finding to Cortés et al (2024), whose work focused on graduating students, who are primarily young and unmarried.

Third, because women were also more likely to be targets of sexual harassment (Hendrix, et al, 1998), and there is a relationship between sexual harassment and lower wages (Folke and Rickne (2022), there had been a negative economic impact of sexual harassment on women. While it is a challenge to measure the full financial impact, this analysis suggests the improvement in willingness to ask for a raise,

which is highly related to receiving a raise, may lead to a financial improvement, particularly for single women, who had been most impacted by sexual harassment and, thus, benefit the most from the social movements of the past several years.

There are several notable limitations to this analysis. First, as covered in the discussion of the data, there are potentially important variables not included in this analysis. Other household characteristics, job/industry characteristics and risk preferences, among others, may be correlated with both the willingness to ask for a raise and gender/marital status in a way that the included control variables insufficiently proxy for. Should these additional characteristics become available in a single dataset, a revision of the analysis would be recommended. Second, it has been several years since this analysis was initially conducted, and additional years of SHED data is now available. Conducting a similar analysis on the more recent data may reveal whether the impact was transient or lasting. Lastly, this analysis can be furthered by allowing for interaction between characteristics beyond gender and marital status, e.g., region and gender, where there may be a theoretical framework for greater differences in a gender gap in various parts of the country.

References

Azzopardi, D., Fareed, F., Lenain, P., & Sutherland, D. (2019). Assessing household financial vulnerability: Empirical evidence from the US using machine learning. *OECD economic survey of the United States: Key Research Findings*, 2019, 121-142.

Babcock, L., Laschever, S., Gelfand, M., & Small, D. (2003). Nice girls don't ask. *Harvard Business Review*, 81(10), 14-14.

Bertrand, M. (2011). New perspectives on gender. *In Handbook of Labor Economics,* 34, 1543–1590. Elsevier.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789-865. DOI: 10.1257/jel.20160995

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology, 100*(5), 1261–1293. https://doi.org/10.1086/230638

Cortés, P., French, J., Pan, J., & Zafar, B. (2024). Gender differences in negotiations and labor market outcomes: Evidence from an information intervention with college students (No. w32154). *National Bureau of Economic Research*.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448-474.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, 109, 573–598. http://dx.doi.org/10.1037/0033-295X.109.3.573

Folke, O., & Rickne, J. (2022). Sexual harassment and gender inequality in the labor market. *The Quarterly Journal of Economics*, 137(4), 2163-2212.

Giritli Nygren, K., Olofsson, A., Öhman, S. (2020). Conceptual Frames: Risk and Intersectionality. In: *A Framework of Intersectional Risk Theory in the Age of Ambivalence. Critical Studies in Risk and Uncertainty*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-030-33524-3_2

Hartig, F. (2024, October 17). DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. Theoretical Ecology, University of Regensburg. https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html

Hendrix, W. H., Rueb, J. D., & Steel, R. P. (1998). Sexual Harassment and Gender Differences. *Journal of Social Behavior and Personality*, 13(2), 235. http://proxy.library.nd.edu/login?url=https://www.proquest.com/scholarly-journals/sexual-harassment-gender-differences/docview/1292312054/se-2

Leibbrandt, A., & List, J. A. (2015). Do women avoid salary negotiations? Evidence from a large-scale natural field experiment. *Management Science*, *61*(9), 2016-2024.

O'brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual Quant* 41, 673–690 (2007). https://doi.org/10.1007/s11135-006-9018-6

Mazei, J., Hüffmeier, J., Freund, P. A., Stuhlmacher, A. F., Bilke, L., & Hertel, G. (2015). A meta-analysis on gender differences in negotiation outcomes and their moderators. *Psychological bulletin*, 141(1), 85.

Pampel, F. C. (2000). The logic of logistic regression. SAGE Publications, Inc., https://doi.org/10.4135/9781412984805

Small, D. A., Gelfand, M., Babcock, L., & Gettman, H. (2007). Who goes to the bargaining table? The influence of gender and framing on the initiation of negotiation. *Journal of personality and social psychology*, *93*(4), 600.

Appendix

Table A1: Summary of Demographic Variables by

	Year				
	2017	2018	2019	2020	2021
Asked for a Raise	7.1%	7.5%	10.1%	7.0%	8.9%
Female	50.3%	50.5%	48.3%	49.1%	49.8%
Age					
18 - 24	3.8%	4.2%	4.8%	6.9%	4.8%
25 - 34	14.4%	15.3%	18.0%	18.3%	16.4%
35 - 44	12.7%	14.3%	14.7%	14.8%	15.0%
45 - 54	14.8%	15.5%	14.4%	13.7%	14.4%
55 - 64	22.8%	22.8%	21.3%	20.5%	21.0%
65 - 74	21.7%	18.9%	18.2%	17.6%	19.1%
75+	9.8%	9.1%	8.6%	8.2%	9.3%
Race					
White	73.7%	69.5%	69.9%	69.7%	70.7%
Black	8.6%	10.7%	10.1%	10.0%	10.0%
Hispanic	12.1%	12.3%	12.0%	12.3%	11.8%
Asian	2.3%	3.6%	3.9%	4.3%	3.8%
Other	3.2%	3.9%	4.0%	3.8%	3.7%
Marital Status					
Married	57.0%	54.3%	57.3%	57.4%	59.8%
Widowed	6.3%	6.0%	4.7%	4.6%	4.9%
Divorced	13.0%	13.0%	10.9%	9.9%	10.0%
Separated	1.6%	1.9%	1.5%	1.7%	1.5%
Never Married	17.4%	19.1%	18.8%	20.0%	23.8%
Living w/ Partner	4.7%	5.7%	6.8%	6.4%	*[6]
Education					
Less Than HS	3.0%	3.6%	3.5%	4.5%	4.2%
HS or GED	21.4%	24.2%	22.3%	19.1%	19.5%
Some College	19.3%	19.1%	17.5%	19.8%	18.1%
Cert or Technical	5.8%	5.1%	5.3%	5.3%	5.1%
Associates	9.9%	9.8%	9.4%	8.6%	9.0%
Bachelor's	23.8%	22.0%	24.3%	24.1%	24.9%
Master's	10.6%	10.4%	11.0%	11.7%	12.2%
Professional	3.8%	3.5%	4.0%	4.4%	4.3%
Doctoral	2.5%	2.4%	2.2%	2.4%	2.6%

[6] The answer choices for 2021 Marital Status no longer included "Living with a Partner" as an option.

Income[7]	2017	2018	2019	2020	2021
$0	3.5%	4.2%	4.4%	5.1%	4.8%
$1 to $4,999	5.5%	5.3%	5.1%	5.2%	5.0%
$5,000 to $14,999	7.6%	7.8%	6.5%	7.0%	5.8%
$15,000 to	9.6%	9.6%	7.6%	6.9%	6.9%
$25,000 to	14.5%	14.6%	11.2%	10.5%	9.5%
$40,000 to	8.7%	8.4%	8.2%	7.6%	7.2%
$50,000 to	15.4%	14.1%	16.2%	14.6%	15.1%
$75,000 to	10.8%	10.8%	12.9%	12.1%	11.8%
$100,000 to	12.1%	12.9%	14.9%	15.5%	15.7%
$150,000 to	6.6%	6.7%	7.0%	8.0%	9.4%
$200,000 or higher	5.6%	5.5%	6.0%	7.5%	8.9%

[7] Income is the total income for the household.

Their "fair share" of taxes: Preferences, perceptions, and observations

Thomas T. Amlie
Pennsylvania State University at Harrisburg

Abstract

An argument which is central to much of the discussion of taxation is that certain groups of people do not pay enough in taxes, whilst others pay too much. Although there are innumerable references to people paying (or not paying) their "fair share", this "fair share" is never defined except as either being more than, or less than, some group is currently paying. Politicians make political points with voters by promising to make certain groups pay their fair share, and argue that increasing taxes on these groups will solve the fiscal problems facing the country.

While there is some literature, both academic and popular, that documents that people do believe the distribution of the tax burden is unfair, there is none that attempts to quantify public perceptions of "fair" and "unfair", or what the public believes the distribution of the burden to be. The purpose of this paper is to determine, based on survey data, what the popular perception of a "fair share" would be and the current beliefs about the distribution of the tax burden among income groups. These results are compared to IRS data to assess the degree to which the current distribution of the tax burden relates to the ideal and assumed distribution of the burden. It is found that there is a popular perception that individuals at the lower (higher) end of the income distribution bear too much (too little) of the tax burden. When compared to actual IRS data, it is found that individuals at lower (higher) income levels pay less than (more than) the "fair" amount.

This paper does not presume to venture an opinion as to what is "fair" or "unfair". It is simply an attempt to document and quantify public attitudes on a perennially contentious topic.

Introduction

Taxation, and the distribution of the tax burden, are topics of considerable and recurring debate. Tax revenues fund the activities of the state. Many people view these activities as necessary if not inherently desirable, so there must be enough tax revenues to fund them. When there are revenue shortfalls which threaten reductions in these expenditures, or when the government has to take on debt to fund them, there are accusations that these shortfalls arise because some taxpayers do not pay their "fair share". Some political figures make such claims a central part of their appeals to voters. Both the popular and academic literature are replete with writings which make these same claims. Unfortunately, this literature is limited to general questions of whether the tax system is fair or unfair, or whether different groups pay too much or too little. There is a dearth of research which quantifies exactly what tax burden distribution would be viewed as "fair", or the extent to which peoples' assumptions regarding the tax burden comport with reality.

This paper attempts to measure empirically public perceptions of "fairness" in the US income tax system. Survey data was collected which examined respondents' views on what constituted a "fair" distribution of the tax burden among various income groups, both in terms of the tax rate as well as the percentage of the total income tax burden. This was then compared to data from the Internal Revenue Service which shows the actual tax rates and shares of the tax burden for these same groups.

This paper addresses a gap in the literature by quantifying public perceptions of "fair" tax rates and burden distribution, an area underexplored in prior research. The use of IRS data to compare perceived and actual tax burdens provides a robust empirical foundation.

The paper opens with a review of some of the literature which appeals to the "fairness" (or the lack thereof) in the US income tax system, as well as some of the political positioning on these matters. The survey instrument and sampling procedures are then introduced, as well as the data available from the US Internal Revenue Service, including a discussion of some of the shortcomings of both the IRS

data as well as the necessarily-related survey measures. Survey results are then reported, and comparisons made. The "ideal" or "fair" distribution of the tax burden is compared to the respondents' beliefs about the distribution of the burden, which reveals that in the aggregate respondents believe that higher (lower) income groups bear too small (large) a portion of the tax burden. This is the genesis of the common claim that the tax burden is unfairly distributed in society. The ideal and assumed distributions are compared to the actual distribution based on information from the Internal Revenue Service, revealing that in the aggregate the respondents have inaccurate impressions of the distribution of the tax burden, and that the actual incidence of the tax burden does not comport with their perceptions of the unfairness of the system.

The paper closes with a discussion of the shortcomings of the data and survey instrument, and makes suggestions for further research.

"Tax fairness" in academic literature

There is a substantial body of literature in the fields of business ethics, public policy, accounting, and other areas which makes reference to taxpayers paying (or not paying) their "fair share of taxes". Unfortunately, the vast majority of this literature doesn't attempt to define "fairness": it simply asserts that a given tax burden or tax structure is "fair" or "unfair".

"Fairness", like beauty, is in the eye of the beholder. As McGee (2023) notes, "The 'fair share' argument is conveniently undefinable, and it makes a good political argument." In the business ethics literature, most references to "fairness" are made in the context of writings addressing tax avoidance or evasion. An exact specification or definition of this "fair share" is generally not provided. Bird and Davis-Nozmack (2018) use some variant of "fair" 25 times, without ever defining the term. Similarly, Payne and Raiborne (2018) cite "fair" or "fairness" with respect to taxation 34 times, again without definition.

There are many different avenues by which a tax burden can be deemed "fair" or "unfair". Some take a legalistic approach: the tax

system (whether created by the divine right of kings or the divine right of parliament) reflects an accepted social contract between the state and the taxpayer, and compliance with the enumerated code is prima facie evidence of "fairness". Some extend this by arguing that a tax code enacted by democratically elected legislators represents a reflection of the general will, and therefore embodies society's view of "fairness".

At the other extreme, some argue that income taxation is inherently unfair: that it represents the confiscation of property under threat of violence. Under a Lockean view of government, individuals delegate some of their rights to the state. Since an individual does not have the right to threaten violence to deprive a peaceful person of their legitimately obtained property, the individual cannot delegate that right to the state. No matter how many voters are in the majority that approves of it, those individuals cannot delegate that right since they do not possess it themselves.

Between these two extremes – the absolute right of the state and the absolute right of the individual – lies the vast gray area in which most people find their views. They do not believe that "whatever the state does is fair", nor do they believe that any tax burden is by definition unfair.

The tax code is inherently fair

In the literature on the ethics of tax avoidance, "unfairness" is held to result from avoiding legally prescribed taxes. To the extent that tax avoidance is held to result in taxpayers not paying their fair share, or that the failure to pay one's fair share is a result of avoidance, the implicit assumption would be that the tax system is "fair" as written, and the unfairness arises from not adhering to either the letter or the spirit of the tax code. Cruz et al. (2000) argue that "several concepts of justice permeate the federal income tax system, such as requiring all citizens to pay their 'fair share' of taxes."

There are several conceptual avenues – none entirely satisfactory – via which one might argue that the tax system is inherently fair, such that a "fair" share of taxes can be defined as that arrived at according to an impartial application of the tax code. The corollary

would be that any unfairness in the distribution of the tax burden must arise solely from tax evasion or avoidance activities. At the most simplistic level, one could assume the omniscience or moral infallibility of the legislators who create the tax system, arguing that they are both possessed of knowledge of what is "fair" and that they are solely motivated by concerns over fairness. These assumptions are never explicitly made, and can be passed over in silence.

The alternative is to argue that the tax code is created by democratically-elected legislators, and therefore reflects the "will of the people". In the context of taxation, Lenz (2020) refers to "democratically legitimized legislators" and "democratically-agreed legislation", while Sikka (2010) refers to "democratically agreed taxes". Mosquera Valderrama (2015) holds that by virtue of living in a society where legislators are elected, citizens participate in the design of the tax system: "the taxpayer represented by a democratically elected lawmaker (parliament/congress) will participate in the design of the tax system." This line of argument is itself subject to criticism, for a number of reasons. Legislators are generally elected by a small proportion of the voting-age adults in a given election district. A legislator who receives 60% of the vote (a landslide), in a district with 70% voter turnout, where only 80% of eligible adults are registered to vote, has the support of approximately 34% of the eligible voting base. Each legislator is elected based on numerous factors, of which their position on taxation is presumably only marginally important. Party affiliation, the success of campaign marketing efforts, personal charisma, views on abortion, education, immigration, or any of a number of other issues might take precedence in the voters' evaluation of candidates. To the extent that the candidate's stated views on taxation are known, it is an open question as to whether they extend beyond general sense of "less" versus "more", and whether the legislator's votes correspond with the position promulgated during their campaign. Finally, tax legislation is written by a small subset of the legislative assembly (e.g., the House Ways and Means Committee in the United States), and in the US at least is delivered as a fait accompli to the general legislative body, with inadequate time for legislators to read the legislation or offer amendments. To argue that the tax code which emerges from such a political process in any way

meaningfully reflects "the will of the people" seems overly optimistic.

The relationship and "contract" between the tax payer and taxing authority is not consensual in any meaningful sense of the term. Brennan and Buchanan (1980) argue that the state's authority to tax income often stems from majoritarian coercion, not genuine consent, violating democratic fairness.

Taxes are inherently unfair

Some argue that income taxes are by their very nature unfair. Taxes represent confiscation of honestly earned or held assets under threat of violence, which is tantamount to extortion. Income taxation has been criticized as a violation of individual property rights and self-ownership. John Locke's principle, articulated in Two Treatises of Government (1689), was that "every Man has a *Property* in his own *Person*. This no Body has any Right to but himself. The *Labour* of his Body, and the *Work* of his Hands, we may say, are properly his." Libertarian philosophers like Nozick (1974) argue that since individuals do not have the moral right to demand another's property, they cannot delegate this power to the state. Bastiat (1850) noted that "Since no individual acting separately can lawfully use force to destroy the rights of others, does it not logically follow that the same principle also applies to the common force that is nothing more than the organized combination of individual forces?" Rothbard (1982) in The Ethics of Liberty further contends that taxing income is morally arbitrary, as it targets labor's fruits without consent, undermining the ethical basis of property rights derived from one's efforts.

The uses to which the tax revenues are put, and the fact that those expenditures may benefit the taxpayer, do not alter the fundamental nature of the extraction. Lysander Spooner (1870), in a comparison of the highwayman and the tax collector, observed that the highwayman

> "… does not pretend that he has any rightful claim to your money, or that he intends to use it for your own benefit. … He has not acquired impudence enough to

> profess to be merely a "protector" … He does not persist in following you on the road, against your will; assuming to be your rightful "sovereign," on account of the "protection" he affords you. He does not keep "protecting" you, … by robbing you of more money as often as he finds it for his interest or pleasure to do so; and by branding you as a rebel, a traitor, and an enemy to your country, and shooting you down without mercy, if you dispute his authority, or resist his demands. … In short, he does not, in addition to robbing you, attempt to make you either his dupe or his slave." (Spooner, 1870)

Taxes can be fair, but the "fairness" is an elusive concept

Most people do not hold either of the above extreme views: that the distribution of the tax burden is inherently fair because it is based on a tax code created by democratically elected legislators, or that it is inherently unfair because taxation itself is morally indefensible. Instead, it is believed that there is some income tax structure which could be considered "fair". "Fairness" is an entirely subjective concept, which arises out of each individual's own moral compass and orientation: how each individual defines "fair".

Outside of the literature on tax avoidance, which generally tacitly accepts that the tax code as written is "fair", there are still innumerable references to some taxpayers not paying "their fair share". The Budget Lab at Yale (Yale University, 2024) argues that the tax code itself generates unfairness. This unfairness is apparently evidenced by different taxpayers with similar total incomes facing different tax burdens. This unfairness can be the result of differential treatment of different types of income (e.g., wages versus capital gains), differences in tax credits between individuals with similar income, or different family size (i.e., taxes relative to income-per-family-member). Calls to reform the tax code aspire to a tax code wherein all pay a "fair share". Oftentimes this includes calls to impose taxes on wealth or unrealized capital gains. Political campaigning and posturing often accuses some groups of "not

paying their fair share", with concomitant promises to rectify that injustice if elected. As is the case in the other literature referenced so far, a glaring omission in all of this is that a "fair share" of taxes is never defined, other than as being more than (or less than) is currently being paid.

The Gallup organization has conducted polls on taxation for decades. The results have been fairly stable over time. Over the past 24 years, the percentage of respondents who viewed their own tax burden as "fair" ranged from 46% to 64%, while the percentage who viewed it as unfair ranged from 33% to 51%. A similar survey question, over a slightly longer time frame, found that a small percentage (<4%) of respondents viewed their tax burden as too low, while 28% to 50% viewed it as "about right", while 45% – 68% viewed their burden as too high. When asked about the fairness of the tax burden imposed on different income groups, the general consensus of respondents is that lower income people (not defined) pay the right amount or too much, while higher income people (not defined) pay too little. Similarly, the Pew Research Center (2023) similarly finds that a plurality of Americans feel that some of the wealthy do not pay their "fair share". Unfortunately, as is common in such findings, neither "wealthy" nor "fair" is defined, and there is no indication that survey respondents were informed of the actual distribution of the tax burden. Such knowledge would seem to be a prerequisite for making an informed statement as to whether the distribution was "fair" or not.

"Fairness" in taxes is generally – although not universally – viewed through an "ability to bear" lens. This is the basis for progressive taxation, which the survey results presented here show to be widely although not universally supported. Those with higher incomes or wealth can better afford to pay more to underwrite the costs of the state: extracting a higher tax from the wealthy will still leave them with more after-tax income than that of a poorer person who pays less (or no) income tax. This idea is given a veneer of academic credibility by arguing that the "disutility" of a high tax payment to a wealthy party might be less than the disutility of a smaller tax payment to someone of lesser means. While declining marginal utility is a theoretically irrefutable concept, so is the subjective nature of utility and the impossibility of interpersonal

comparisons of utility. In the end, the arguments for progressive taxation reduce to "they can afford it".

Instead of focusing on the ability to bear taxes, some argue that taxes are fair based on "benefits received". Those who take a transactional view of the taxpayer/state relationship might argue that a "fair" tax burden would reflect the benefits received from the state – e.g., police protection, infrastructure, transfer payments, etc. Of course, this immediately gives rise to problems in assigning a value to those benefits, and how those values might vary from person to person and circumstance to circumstance: any attempt to place a value on a benefit received such as education or police protection or social welfare spending immediately gives rise to a host of arguments involving externalities.

The sobriquet "fair tax" has been arrogated by those who support a single national consumption or sales tax. Others of an egalitarian bent might argue that a flat per-capita tax (which would be regressive with respect to income) might be "fair", whilst others would argue for a flat tax rate applied to all incomes. Given all of the various ways in which "fair" might be defined, and since none of them has an irrefutable claim to be absolutely correct, the question might naturally arise as to how people on average define fairness. Since in a democracy or representative Republic the state is expected to reflect the will of the people, it seems reasonable to attempt to define and measure "fairness" based on the opinions of those subject to the rules of the state. With this in mind, those who argue that the tax code is "fair" because it reflects the "will of the people" as implemented by "democratically legitimized legislators" should be especially receptive to what "the people" deem to be fair levels of, and distribution of, tax burdens.

The survey instrument and available data

The survey used in this study (Appendix A) was intended to ascertain views on the "fair" level of taxation for different income groups as well as respondents' beliefs about the existing level of taxation. These observations would then be compared to the actual tax data from the Internal Revenue Service, and the agreement or discrepancy between the fair, assumed, and actual levels of taxation

explored. Taxation was measured both as the tax rate which the various different groups faced, as well as the percentage of the total tax burden borne by different groups.

There is no widely accepted extant survey research in this area on which to build, so the survey design was started from first principles. Aside from demographic information (age, education, employment status, etc.), interest centered on respondents' views on the fairness of the tax system in general, as well as their assumed and ideal distribution of the tax burden amongst the different income groups. An attempt was made to gauge respondents' general knowledge of Federal fiscal issues (i.e., overall spending, the current deficit, and the current Federal debt). The survey was tested with colleagues as well as with students in an undergraduate Advanced Financial Accounting class in order to determine whether any issues of clarity needed to be addressed.

In comparing public perceptions to actual tax information, there are several constraints on the creation of such a study. Since the survey asks respondents for their views on the tax burden as applied to different income groups, and that data will be compared to data from the Internal Revenue Service, it is critical that the income groupings or categories correspond to the income groupings reported by the Internal Revenue Service. The data on individual income taxes available from the Internal Revenue Service is categorized based on Adjusted Gross Income (AGI), e.g., tax returns from taxpayers with AGI from $15,000 to $40,000. For that reason the survey also asked about respondent views on taxes for groups constructed on the same measure, and in the same intervals. By the same token, questions about taxes and tax rates were based on AGI, even though taxes are generally thought of as being a function of taxable income. Although the IRS also reports the total taxable income for each group which is constructed based on AGI, survey questions which referenced both measures of income tended to be unwieldy and confusing, e.g., "For taxpayers with an AGI between $40,000 and $75,000, what is a fair tax rate as a percentage of taxable income?" In addition to making the questions themselves unwieldy, to ask questions regarding taxes as a percentage of taxable income, for groups based on AGI, would necessitate respondents' understanding of the difference between taxable and adjusted gross

income. Given that the income groups must be constructed based on AGI data, it was believed necessary to frame the survey questions on the same basis.

Another potential concern based on the data available from the Internal Revenue Service involves the size of the income intervals selected for the groupings. At lower income levels the intervals are fairly narrow (e.g., $5,000 intervals from zero to $30,000) while at higher income the income tiers are significantly broader (e.g., $200,000 to $500,000, or $5,000,000 to $10,000,000). Since presenting each IRS-provided interval as a survey question would involve 18 different intervals, and since several questions were asked about each interval, some aggregation was deemed necessary to keep the survey to a manageable size. Intervals were selected so that each interval would encompass 10 – 15% of the total adjusted gross income across all groups; this process was frustrated to an extent since the IRS-defined $100,000 - $200,000 and $200,000 - $500,000 groups accounted for 24.2% and 17.3% respectively of the total AGI across all groups. Those tax returns with an AGI below $15,000 were excluded from the study as the vast majority of those were not taxable returns.

To ensure content and construct validity, survey items were aligned with specific constructs of interest—namely, perceived fairness, ideal tax rates, and actual tax burden estimates—using income groupings based on IRS-defined adjusted gross income (AGI) categories. This alignment allowed for direct comparison between respondent perceptions and published tax data, thereby grounding the survey in relevant empirical benchmarks.

Face validity was addressed through pre-testing. The instrument was reviewed by colleagues and piloted with students in an upper-level accounting course to assess clarity and interpretability. Feedback from these test groups informed refinements to question wording and structure, helping to ensure that items were comprehensible and reasonably interpreted as intended.

Efforts were made to improve data quality and internal consistency during the data cleaning process. Incomplete responses, or those deemed unserious (i.e., those where “fair” tax rates appeared

to form a random pattern), were excluded from the final sample. In contrast, responses exhibiting flat, progressive, or regressive patterns were retained as legitimate reflections of consistent respondent viewpoints. Additionally, the survey incorporated procedural safeguards to reduce measurement confounds. Questions explicitly instructed participants to consider only federal income taxes, excluding state, local, and payroll taxes. For items requesting a breakdown of the tax burden across income groups, a normalization procedure ensured that responses summed to 100%, allowing for meaningful comparisons across participants and against IRS data.

Although the survey instrument was not subjected to formal psychometric validation, these steps provide support for its validity as a tool for examining public perceptions of tax fairness. The use of income categories matched against IRS data, pre-testing, structured exclusions, and consistency checks collectively reinforce the reliability and interpretive strength of the data collected.

Survey respondents

The survey instrument (appendix A) was made available via Survey Monkey, and participants were recruited from their pool of respondents. SurveyMonkey maintains an opt-in database of individuals willing to complete surveys. Individuals who complete a survey earn contributions ($.50 per completed survey) to the charitable organization of their choice. SurveyMonkey sent the URL of the survey to 221 participants randomly selected from its database. As many of the questions asked for responses which would involve some relatively careful thought, a significant proportion of respondents did not complete the survey, and it appears that some proportion of the respondents chose to put in seemingly random numbers when asked for tax rates. Responses with incomplete data were removed from the sample, as were responses which were deemed to be unserious. "Unserious" responses were judged based on the responses to the "what is a fair tax rate" for each of the income groups. There were apparently some "flat taxers" in the sample, as well as some who apparently thought that higher income taxpayers should be completely expropriated. These respondents were left in, as these are positions which some in society clearly hold. The responses deleted were those where the tax rate rose and fell randomly as the income levels increased; e.g., where

the $15,000 - $40,000 group's fair rate was 40%, while the $40,000 - $75,000 group was 5%, while the $75,000 to $100,000 group was 60%, with the $100,000 - $200,000 group at 13%. In short, any response with an unsystematic pattern was removed. Flat, monotonically increasing (progressive), and monotonically decreasing (regressive) patterns were deemed valid responses. The final sample was comprised of 86 usable complete survey responses.

The survey respondents who provided the 86 usable responses had the following characteristics. A majority (58%) were between the ages of 30 and 49; 16% were below 30, while 26% were above 49. The majority (68%) were college graduates. 60% were employed full time, while 18% were retired. Part-time workers and the unemployed comprised the majority of the remainder of the sample. 40% of the sample earned less than $75,000 per year; 47% earned between $75,000 and $150,000, while the remaining 13% earned more than $150,000 per year.

When asked to rate whether the tax system was "fair" on a 4-point scale (1 = very fair, 4 = very unfair) the survey average was 3 ("reasonably fair"), with no significant differences between income and age groups. When asked whether tax avoidance was a significant factor in the unfairness of the tax system (1 = a major factor, 3 = not a significant factor), the sample average was 1.73, with the highest income group ($2,000,000 +) holding the strongest views in that regard. Unfortunately, the small sample sizes for the various age and income categories preclude drawing any firm conclusions.

Survey respondents exhibited a general preference for a progressive tax structure. Several different measures of progressivity were used, all of which generated substantially the same results. The spread between the "fair" tax rates for the highest and lowest income groups, of the spread between the averages of the lowest two groups and highest two groups, or the ratio of the rates for the highest and lowest income groups are all based on the same data, so all yield essentially the same result. Interestingly, the greatest progressivity preferences were exhibited by the lowest income ($0 to $25,000) and highest income ($2,000,000 +) respondents, with the greatest progressivity preference among the

highest income group. Similarly, the greatest progressivity preference was among the 60 – 70+ year old age groups. An attempt was made to gauge the survey respondents' awareness of general fiscal issues by asking them their "best guess" as to the Federal debt, the current budget deficit, and Federal spending. An answer was deemed "correct" if it came within +/- 30% of the actual figure. Of the 86 survey respondents, 13 correctly estimated the total debt, while 12 correctly estimated the current deficit and spending.

IRS tax data

The IRS data used in this study was accessed at the irs.gov website, from the page entitled "SOI tax stats – Individual statistical tables by size of adjusted gross income". The data used was for tax year 2020, which was the most recent data available when this survey was conducted.

Survey questions

Survey respondents were asked for basic demographic information (age, education, employment status, and household income), how they prepared their taxes, their overall views on the fairness of the tax system and what might contribute to any perceived unfairness, and whether certain types of taxes are or would be "fair" or "unfair" (see appendix A).

In terms of areas of interest for this study, taxpayers were grouped into the following income categories. The percentage of total AGI which each group represents (excluding those with AGI below $15,000) was provided, and is indicated in table 1:

Table 1: Distribution of income among taxpayer groups

Group	Total AGI (in billion dollars)	As % of total AGI
AGI from $15,000 - $40,000	1,239.2	9.8%
AGI from $40,000 to $75,000	1,977.3	15.7%
AGI from $75,000 to $100,000	1,247.3	9.9%
AGI from $100,000 to $200,000	3,060.1	24.2%
AGI from $200,000 to $500,000	2,187.2	17.3%
AGI from $500,000 to $2,000,000	1,349.6	10.7%
AGI above $2,000,000	1,573.8	12.5%

For each of these groups, respondents were asked several questions. It was stressed that all answers should be based solely on

Federal income taxes, and should exclude state, local, and Social Security/Medicare taxes. First, they were asked to rate, on a 5-point Likert scale, whether they believed each group paid "much less than their fair share", "somewhat less than their fair share", "about their fair share", "somewhat more than their fair share", or "much more than their fair share". Respondents were then asked what they believed a fair tax rate (as a percentage of AGI) would be for the taxpayers in each group, what they believed each group's "fair share" of total income tax revenues would be, and what they believed each group currently paid as a percentage of total Federal income tax revenues.

Discussion of results

While there was a broad range of opinions, the mean responses indicated that people generally believe that taxpayers in lower income groups pay too much in taxes, whilst those in higher income categories pay too little. As can be noted from Table 2, a majority of respondents believe that taxpayers with an AGI less than $500,000 pay their fair share or more than their fair share. For taxpayers with an AGI above $500,000 a majority of respondents believe they pay somewhat or much less than their fair share. Those in the highest income category are thought by a majority to pay much less than their fair share: this is the only single category which a majority of the respondents agreed with.

Table 2: Perceived fairness of tax contributions by AGI group

Income category	Pays much less than their fair share	Pay somewhat less than their fair share	Pays about their fair share	Pays somewhat more than their fair share	Pays much more than their fair share
15k-40k AGI	7%	15%	39%	19%	21%
40k-75k AGI	2%	12%	48%	26%	13%
75k-100k AGI	0%	14%	38%	38%	9%
100k-200k AGI	5%	19%	41%	17%	19%
200k-500k AGI	14%	34%	21%	17%	14%
500k – 2,000k AGI	36%	24%	16%	6%	17%
2,000k + AGI	57%	7%	7%	10%	19%

These categorizations were based on participants' intuitive sense of "fairness", however they chose to define it, as well as their beliefs

about the distribution of the tax burden. To get a better understanding of what exactly would constitute a fair level of taxation, participants were next asked to specify what tax rate, as a percentage of AGI, would be a fair rate of taxation. The results, which clearly show a general preference for a progressive tax rate, are presented in table 3.

Table 3: Respondents choice of "fair" tax rate for each AGI group

Fair tax rate			
Income group	Mean	Standard Error	Median
AGI from $15,000 to $40,000	5.26	0.5	4
AGI from $40,000 to $75,000	7.58	0.77	5
AGI from $75,000 to $100,000	9.28	0.89	6.5
AGI from $100,000 to $200,000	11.76	1.15	10
AGI from $200,000 to $500,000	14.94	1.35	10
AGI from $500,000 to $2,000,000	19.2	1.7	15
AGI over $2,000,000	23.85	2.12	20

Now that we have some measure of what survey participants deem a "fair" rate of taxation for each group, we can compare it to actual effective rate based on Internal Revenue Service data. For this, the effective rate for each income group is calculated based on the total taxes paid by that group, divided by the total AGI of that group. The actual effective rates are as indicated in table 4:

Table 4: Actual tax rate as percentage of AGI

Income Group	Actual tax rate as percentage of AGI
15,000-40,000 AGI	2.78%
40,000-75,000 AGI	6.06%
75,000-100,000 AGI	8.11%
100,000 – 200,000 AGI	10.90%
200,000 – 500,000 AGI	16.77%
500,000 – 2,000,000 AGI	24.70%
2,000,000+ AGI	26.68%

Subtracting the actual effective rate from the survey participants' "fair" rate for each group yields the following measure of the taxes paid above (below) the fair amount, as shown in table 5.

Table 5: Actual tax rate minus "fair" tax rate

	Mean	Standard Error	t
AGI from $15,000 to $40,000	-2.473	0.503	-4.915
AGI from $40,000 to $75,000	-1.526	0.766	-1.992
AGI from $75,000 to $100,000	-1.17	0.892	-1.312
AGI from $100,000 to $200,000	-0.82	1.152	-0.712
AGI from $200,000 to $500,000	1.828	1.349	1.355
AGI from $500,000 to $2,000,000	5.504	1.697	3.243
AGI over $2,000,000	2.83	2.119	1.336

From this we can see that the actual effective rates paid by the lower (higher) income groups are lower (higher) than what survey participants deem "fair". Those taxpayers with an AGI between $15,000 and $75,000 pay a significantly lower tax rate than survey participants deem "fair", while those with an AGI from $500,000 to $2,000,000 pay significantly more than their "fair share".

Similar results obtain when examining views on the "fair share" of the total income tax burden. In the survey instrument respondents are informed what share of total AGI each group accounted for, and were asked what each group share of the total tax burden should be. Respondents were informed that if their "fair shares" did not sum to 100%, their answers would be scaled so that they would. For example, if a respondent's "fair shares" summed to 125%, each individual percentage would be multiplied by 0.8 so that the final answer summed to 100%. Information regarding the mean "fair share" for each group, as well as the actual share for each group, is presented in table 6.

The results in table 6 suggest that the group with an AGI below $100,000 bear less than their fair share of the total income tax

burden, while those with an AGI from $100,000 to $2,000,000 bear more than their fair share of the burden.

Table 6: Fair share of total tax revenue versus actual share

	Mean "fair share"	Actual share	Mean difference	Standard Error	Median
AGI from $15k to $40k: 9.8% of total income of all groups	6.65%	2.02%	4.63**	0.49	3.56
AGI from $40k to $75k. 15.7% of total income of all groups	9.59%	7.00%	2.59**	0.63	2.65
AGI from $75k to $100k:9.9% of total income of all groups	11.46%	5.91%	5.54**	0.46	4.96
AGI from $100k to $200k: 24.2% of total income of all groups	13.83%	19.57%	-5.74**	0.56	-5.28
AGI from $200k to $500k: 17.3% of total income of all groups	16.02%	21.45%	-5.43**	0.51	-5.32
AGI from $500k to $2,000k: 10.7% of total income of all groups	17.83%	19.50%	-1.67*	0.61	-0.6
AGI over $2,000k: 12.5% of total income of all groups	24.60%	24.55%	0.09	1.75	-2.57

The "share of total tax burden" measure itself (whether fair or actual) can be misleading due to the different proportion of income for each category. At first blush, it might appear odd that the "actual share" drops from 7% to 5.91% when moving from the $40k-$75k AGI category to the $75k – $100k AGI category. This result can easily be explained by the fact that the share of income in the latter category is only 63% of the share of income of the former.

To alleviate this concern, the shares of tax burden can be expressed relative to the share of income for each group. Dividing the share of taxes (whether the "fair share" or the "actual share") for each group by the share of income for each group yields the "share of taxes as a percentage of the share of income". If all groups had a value of 1.00, that would indicate that each group should have a tax

burden share equal to their income share. A value of less than 1.00 indicates that that group should (or does) pay an amount less than proportionate to their income share, while a value above 1.00 indicates that that group pays a greater than proportionate share of taxes, relative to their income share. Results for these revised measures are presented in table 7.

Table 7: Share of tax burden relative to share of income

	Fair	Actual	Difference	Standard Error
AGI $15,000-$40,000	0.68	0.21	0.47**	0.0498
AGI $40,000 - $75,000	0.61	0.45	0.16**	0.0402
AGI $75,000 - $100,000	1.16	0.6	0.56**	0.0469
AGI $100,000 - $200,000	0.57	0.81	-0.23**	0.0233
AGI $200,000 - $500,000	0.93	1.24	-0.31**	0.0292
AGI $500,000 - $2,000,000	1.67	1.82	-0.15**	0.0574
AGI Over $2,000,000	1.97	1.96	0.01	0.1404

These results paint the same picture as previously, namely that the proportion of the tax burden borne by taxpayers with an AGI below $100,000 is significantly less than their “fair share”, while the proportion of the tax burden borne by those with an AGI between $100,000 and $2,000,000 is significantly more than their fair share. Those with an AGI above $2,000,000 pay an amount which the surveyed sample deems a “fair” proportion of the total tax burden.

Tables 5 through 7 paint a consistent picture: The “fair tax rates”, and the “fair share of the tax burden”, are higher (lower) than the actual figures for groups comprised of lower (higher) AGI taxpayers. In contrast, table 2 shows that respondents believe the opposite: that lower-AGI taxpayers pay too much, whilst higher AGI taxpayers pay too little. The discrepancy is obviously attributable to a lack of information about the true distribution of the income tax burden in the United States. The final analysis therefore compares the beliefs about the relative tax burden for each group to the actual tax burden for each income group.

In table 8, information is presented indicating the assumed share of the tax burden borne by each income group, compared to the actual share of the tax burden.

Table 8: Assumed share minus actual share of total tax revenues

Income Group	Assumed share of tax burden	Actual share of tax burden	Difference	Standard Error
AGI from $15k to $40k: 9.8% of total income of all groups.	10.21	2.02	8.12**	1.23
AGI from $40k to $75k: 15.7% of total income of all groups.	12.82	7	5.82**	0.88
AGI from $75k to $100k: 9.9% of total income of all groups.	13.43	5.91	7.52**	0.61
AGI from $100k to $200k: 24.2% of total income of all groups.	15.16	19.57	-4.41**	0.61
AGI from $200k to $500k: 17.3% of total income of all groups.	15.33	21.45	-6.12**	0.52
AGI from $500k to $2,000k: 10.7% of total income of all groups.	15.41	19.5	-4.09**	0.69
AGI over $2,000k: 12.5% of total income of all groups.	17.63	24.55	-6.92**	1.3

The results in table 8 highlight the core problem with much of the public debate regarding whether certain groups pay more of less than their fair share of income taxes. On average, to the extent survey respondents are representative of the public, people are misinformed about the actual distribution of the tax burden. People overestimate the taxes paid by lower income taxpayers, and underestimate the taxes paid by higher income taxpayers. This apparent misunderstanding of the actual distribution of the tax burden appears to contribute to perceptions of tax unfairness, with that unfairness taking the form of lower (higher) income taxpayers paying too much (too little) in taxes. Given the results outlined in tables 5 through 7, it is apparent that when people are queried as to what they would deem a “fair” tax structure, their “fair” structure, when compared to actual tax data, suggests that lower income taxpayers instead pay too little, while higher-income taxpayers pay too much.

Other views on taxation

In addition to the above-discussed questions regarding fair tax rates and fair shares of the tax burden, survey respondents were also asked their views on what made the tax code fair or unfair, and whether certain specific tax measures would be fair or not. Results are indicated in table 9.

Table 9: Respondents' views on different aspects of the tax system's fairness

Percentage of respondents who answered 1 = "strongly disagree" through 5 = "strongly agree"

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	Mean
A major reason the tax system is unfair is because some taxpayers face a too-high or too-low tax rate as applied to their income					
6.08%	6.76%	33.11%	33.11%	20.95%	3.56
A major reason the tax system is unfair is because the complex tax code allows some people to legally avoid taxes through legal moves					
2.76%	5.52%	30.34%	28.28%	33.10%	3.83
A major reason the tax system is unfair is because the tax code is written to benefit special interests/political contributors/pressure groups					
2.78%	7.64%	28.47%	34.72%	26.39%	3.74
A flat tax system where everyone paid the same percentage of income (e.g., 15%) of all income above some minimum would be fair					
9.66%	14.48%	28.28%	32.41%	15.17%	3.29
When someone dies, the government should get part of their estate through estate taxes					
31.94%	14.58%	20.83%	21.53%	11.11%	2.65
If you save your money and accumulate wealth, the government should get a part of that wealth every year through a wealth tax					
20.28%	12.59%	30.77%	24.48%	11.89%	2.95

The first three questions indicate a tendency towards dissatisfaction with the way the tax code is written. We can observe again that people believe that tax rates are unfair, but as was noted previously there seems to be a general lack of understanding of the actual tax rates different income groups face. The tax system itself is suspect, either because its complexity allows some taxpayers to avoid taxes, or because it was written to benefit special interests. A flat tax enjoys insignificant support, while estate taxes, a perennial political issue, lean towards a lack of support. Respondents seemed neutral towards a wealth tax, although for both the estate and the wealth tax

there were a significant proportion who were strongly opposed to those.

Limitations and suggestions for future research

There are several limitations in both the survey instrument as well as the sample which should be addressed in subsequent research. In trying to assess respondents' views on taxation there are several factors which can affect the degree to which the survey accurately captures respondents' views.

There are numerous issues that make this a difficult topic to survey. The US income tax system is highly complex, where two taxpayers with identical gross income could have very different tax obligations based on their filing status, the types of income (i.e., wage versus business income versus capital gains), various tax credits, and so on. Similarly, two taxpayers with identical incomes could be thought to have differing "fair" tax burdens based on the number of dependents they have. Given these complexities, a single measure such as "the fair tax rate on AGI for taxpayers with an AGI between $75,000 and $100,000" is likely to lack some detail which would be important for a full understanding of views. As an alternative, a study such as this could be replicated through the use of an assortment of scenarios, with each scenario having a different level of gross income, different types of income, differing filing statuses, differing numbers of dependents, and differing levels of taxable income, with respondents asked to posit a "fair" income tax levy for that taxpayer. The results could be compared to the actual levy based on the current tax code.

Although it was explicitly stated in the survey instrument that the taxes under consideration were to be limited to Federal income taxes, and exclude payroll taxes and state income taxes, it is possible if not likely that when some respondents think of a taxpayer's tax burden they envision the total tax burden (i.e., including payroll and state taxes) rather than the narrower Federal income tax burden. Two respondents who answered that a hypothetical taxpayer's tax burden should be 20% might have very different intentions. One might be thinking of the total tax burden (e.g., 7.65% for payroll taxes, 3% for State income taxes, and 9.35% for Federal income

taxes, for a total of 20%), whilst the other might be properly contemplating a Federal income tax burden of 20%, for a total tax burden of 30.65% (20% Federal, 7.65% payroll taxes, and 3% State). Reframing the question in terms of the total tax burden (i.e., Federal, State, and payroll taxes) is unsatisfactory. Given different state income tax rates ranging from zero to 10% or more, the results would no longer be directly comparable to Federal tax rates, and comparisons with IRS tax data would no longer be appropriate. Given that Social Security taxes are different in nature from income taxes, including them in a measure of "fairness" would be inappropriate (Amlie and Gibney, 2023).

Using open-ended (i.e., fill in the blank) responses for the various numerical answers, as opposed to offering a menu of specific choices (e.g., 2%, 8%, 15%, 20%, 40%, etc.), may have impacted the results of the survey or how respondents interacted with it. While the open-ended format allows respondents to express their precise opinion, the use of a multiple-choice format might enable them to better appreciate the issue of progressive taxation. Additionally, since on-line survey are presumably often completed on mobile phones, the use of "radio buttons" for choices might encourage respondents to give an imprecise-yet-serious answer, rather than simply keying in random digits on a keypad.

The survey sample itself could be improved upon in future research. The use of the SurveyMonkey's opt-in participant panel, in a survey which required some degree of careful thought regarding numerical answers, led to a significant number of responses which could be characterized as unambiguously unserious. More than half of the responses had to be discarded either because they were incomplete or because the answers were clearly random. A smaller but more carefully targeted sample, where respondents have an existing interest in the topic, would likely be an improvement. When the survey was being developed it was completed by students in an upper-level accounting class with much cleaner results. However, using student responses in published research creates significant Institutional Review Board challenges. Unfortunately, a sample targeted at those who have an existing interest in the topic would likely be unrepresentative of the population at large.

Extensions of this research could be pursued along several avenues. A larger, more carefully targeted population of respondents would increase the reliance that could be placed on the results. It would be revealing to explore more fully why people deem the tax system to be unfair. In addition to the question of the inherent fairness of the tax code, the manner in which income is earned can also be a factor. In the literature on the drivers of tax compliance it has been noted by some (Alm and Gomez (2008), Bordignon (1993)) that people's compliance decisions are based in part on their views on the fairness of the tax system, and that fairness is in part a function of how just they perceive the tradeoff between taxation and benefits received, how just the distribution of income is, and the ways in which that income is earned (e.g., through governmental favoritism or subsidies).

Conclusion

A large body of literature, both academic and popular, as well as much political rhetoric, includes the claim that certain people or groups of people either pay too much or too little in income taxes. In the vast majority of cases, lower-income taxpayers are assumed to pay more than their fair share, whilst higher income taxpayers are assumed to pay less than their fair share. Literature in the areas of public policy, business ethics, sociology, and other fields makes this appeal to taxpayers paying "their fair share", and a common theme of politicians and their proxies in the media is that some groups don't "pay their fair share of taxes". Politicians court votes by promising to rectify this alleged injustice.

Unfortunately, this idea of a "fair share of taxes", while commonly appealed to, is seldom if ever defined. The closest that any of the literature comes to defining a "fair share" is the literature on tax avoidance, where the "fair share" is tacitly assumed to be that which is in accord with the spirit (however ascertained) of the tax code. If the tax code is held to represent a "fair" taxation system, then avoidance or evasion will by definition result in an unfair outcome.

This paper is based on a survey which attempted to determine popular views on what a "fair" distribution of the tax burden among various income groups would be, both in terms of the tax rate faced

by those groups as well as the share of the total income tax burden borne by those groups. The survey results were compared to IRS data and the following conclusions drawn:

a. People generally believe that lower-income taxpayers pay too much, whilst higher income taxpayers pay too little;
b. Actual IRS data suggests the opposite: that lower-income taxpayers pay a lower tax rate than survey participants deem fair, or pay a lower percentage of the total tax bill than survey participants deem fair;
c. Higher-income taxpayers pay a higher tax rate than survey participants would deem fair, while also paying a greater proportion of the total tax bill than survey participants would consider fair;
d. Survey participants significantly overestimated the taxes paid by lower-income taxpayers, and significantly underestimated the taxes paid by higher-income taxpayers.

The results presented in this paper suggest that popular concerns that certain taxpayers, or groups of taxpayers, do not pay "their fair share" are misplaced at best, and that that popular impressions of the "unfairness" of the US Income tax structure are due more to public misperceptions of the distribution of the tax burden than to the actual distribution of the tax burden. When respondents' beliefs on a "fair" distribution of the tax burden are compared to the actual distribution of the burden, it is seen that their views on the existing fairness or unfairness are in fact contrary to the actual state of affairs. Instead of lower income taxpayers paying "too much", they in fact pay too little relative to what survey respondents deem "fair". By the same token, higher income taxpayers pay more than what survey respondents would deem fair. Finally, when asked for their estimates of what share of the tax burden each income group bears, it is seen that their beliefs, while consistent with their belief that certain groups pay too much or too little, are inconsistent with the actual distribution of the tax burden. From this is appears that arguments which appeal to claims of fairness or unfairness of the tax burden should be largely discounted, as those claims are based on incorrect assumptions about the distribution of the tax burden. Given that a supermajority of the survey respondents were unable to identify the Federal debt, deficit, or budget within +/- 30% it should not be surprising that their knowledge of tax rates is similarly limited.

A relevant question raised by this research is what factors or biases lead to such systematic misestimation of the distribution of the tax burden among different income groups. There is general (although not universal) acceptance of the idea that taxes should be progressive: that higher income taxpayers should pay a higher percentage of their income as Federal income taxes. The misestimation results from the confluence of the respondents' general lack of knowledge of the actual distribution of the tax burden with the incessant claims in the political and media spheres that "the rich don't pay their fair share". It is taken by many as an article of faith that some people don't pay their "fair share", simply based on repeated assertions to that effect.

Given the frequency and vehemence with which these "fair share" claims are made in political campaigns and in the popular press, there is some likelihood that they affect peoples' voting decisions, and politicians' policy positions. It is therefore critically important that those to whom such claims are addressed be able to discern their accuracy or inaccuracy. Although individuals will invariably have widely divergent views of what a "fair" tax system looks like, unless they have knowledge of the true distribution of the tax burden, they will be unable to judge whether the current system comports with their idea of fairness. The first order of business in any discussion of "tax fairness" should be to delineate what a "fair" system looks like, and to educate people about the actual current state of affairs.

Appendix A
Survey instrument

The purpose of this survey is to try to understand how people view the fairness of the US Income tax system. "Fair" means different things to different people. Please use whatever personal definition of "fairness" you have in answering these questions. We appreciate your assistance in this study; your views will be reflected (Anonymously, of course) in any published research. You are of course free to discontinue participation at any time. Please restrict

your answers to the Federal income tax system, excluding state, local, and social security taxes.

1. What is your current age?
   - o 18-29
   - o 30-39
   - o 40-49
   - o 50-59
   - o 60-69
   - o 70-79
   - o 80-89
   - o 90 or older

2. What is the highest level of education you have completed?
   - o Did not graduate from high school
   - o Graduated from high school
   - o Some college
   - o 2-year college degree
   - o 4-year college degree
   - o Some graduate school
   - o Completed graduate degree
   - o Attended vocational/technical school

3. Which of the following categories best describes your employment status?
   - o Employed, working full-time
   - o Employed, working part-time
   - o Not employed, looking for work
   - o Not employed, NOT looking for work
   - o Retired
   - o Disabled, not able to work
   - o Student

4. What is your approximate average household income?
   - o $0-$24,999
   - o $25,000-$49,999
   - o $50,000-$74,999
   - o $75,000-$99,999
   - o $100,000-$149,999
   - o $150,000-$199,999
   - o $200,000 and up

5. How do you prepare your federal income tax return?
   - o I prepare it myself, either on-line or with paper/pencil.
   - o I use a volunteer community service such as VITA.
   - o I pay a preparer to do my taxes for me.
   - o I don't file a federal tax return.
   - o A friend or relative prepares my tax return

6. Do you think the current Federal income tax system is fair or not?
   - o Very unfair
   - o Somewhat unfair
   - o Reasonably fair
   - o Very fair

7. In discussing the fairness of the tax system, some people focus on the tax code itself, while others consider levels of tax evasion (people not paying taxes on all of their income as determined by the tax code). Leaving aside the fairness of the tax code itself, how big of a problem do you think tax evasion is in contributing to the unfairness of the US Federal income tax system? (select 1)

a. Tax evasion is a major factor contributing to the unfairness of the tax system. If there was no evasion the tax system would substantially fairer.
b. Tax evasion is a moderate factor contributing to the unfairness of the tax system. If there was no evasion the tax system would still be unfair, but less so.
c. Tax evasion is a minor factor contributing to the unfairness of the tax system. Tax evasion really doesn't add much to the unfairness of the tax system

8. The following groups are based on adjusted gross income (AGI) as reported on federal income tax returns. Adjusted gross income includes income from all sources. For each group, indicate whether you think that group pays less than or more than their fair share of federal income taxes:

AGI group	Pays much less than fair share	Pays somewh at less than fair share	Pays about their fair share	Pays somewhat more than their fair share	Pays much more than their fair share
15k – 40k	○	○	○	○	○
40k – 75k	○	○	○	○	○
75k – 100k	○	○	○	○	○
100k - 200k	○	○	○	○	○
200k – 500k	○	○	○	○	○
500k-2,000k	○	○	○	○	○
Over 2,000k	○	○	○	○	○

9. The following groups are based on adjusted gross income (AGI) as reported on Federal income tax returns. For each group, indicate the percentage of AGI you think would be fair as a federal income tax. Please indicate a value from zero to 100. For example, if you believe the 15,000 – 40,000 group should pay a federal tax rate of 4% of adjusted gross income write "4". Each answer should be between 0 and 100.

Income group	Fair tax rate
15k – 40k	
40k – 75k	
75k – 100k	
100k - 200k	
200k – 500k	
500k-2,000k	
Over 2,000k	

10. The following groups are based on adjusted gross income (AGI) as reported on Federal income tax returns. For each group, the percentage share of the total income that that group accounts for is indicated. For each group, indicate what their "fair share" of total Federal income tax revenues would be. For example, if you believe the 15,000 – 40,000 group should contribute 4% of Federal income tax revenues, write "4". Each answer should be between 0 and 100. Your answers will be scaled so that they sum to 100 (for example, if your answers add up to 80, they will all be multiplied by 1.25 so that they add up to 100). If your answers already add up to 100 no such adjustments will be needed.

Income group	Fair share of tax burden (%)
15k – 40k: 9.8% of income of all groups combined	
40k – 75k: 15.7% of income of all groups combined	
75k – 100k: 9.9% of income of all groups combined	
100k - 200k: 24.2% of income of all groups combined	
200k – 500k: 17.3% of income of all groups combined	
500k-2,000k: 10.7% of income of all groups combined	
Over 2,000k: 12.5% of income of all groups combined	

11. The following groups are based on adjusted gross income (AGI) as reported on Federal income tax returns. For each group, the percentage share of the total income that that group accounts for is indicated. For each group, indicate what you believe their current share of total Federal income tax revenues is. For example, if you believe the 15,000 – 40,000 group contributes 4% of Federal income tax revenues, write "4". Each answer should be between 0 and 100. Your answers will be scaled so that they sum to 100 (for example, if your answers add up to 80, they will all be multiplied by 1.25 so that they add up to 100). If your answers already add up to 100 no such adjustments will be needed.

Income group	Your belief regarding their actual share of tax burden (%)
15k – 40k: 9.8% of income of all groups combined	
40k – 75k: 15.7% of income of all groups combined	
75k – 100k: 9.9% of income of all groups combined	
100k - 200k: 24.2% of income of all groups combined	
200k – 500k: 17.3% of income of all groups combined	
500k-2,000k: 10.7% of income of all groups combined	
Over 2,000k: 12.5% of income of all groups combined	

12. For each of the following statements, indicate your level of agreement or disagreement (1 = Strongly Disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, or 5 = Strongly Agree):

Statement	Agreement
A major reason the tax system is unfair is because some taxpayers face a too-high or too-low rate as applied to their income	
A major reason the tax system is unfair is because the complex tax code allows some people to legally avoid taxes through legal moves	
A major reason the tax system is unfair is because the tax code is written to benefit special interests or political contributors or pressure groups.	
A flat tax system where everyone paid the same percentage of income (e.g., 15%) of all income above some minimum would be fair	
When someone dies, the government should get part of their estate through estate taxes	

If you save your money and accumulate wealth, the government should get a part of that wealth every year through a wealth tax	

13. Income from different sources often is taxed at different rates. For each type of income listed, please indicate what you think a "fair" rate of taxation would be. If you think the tax on wage income should be 20%, write "20". For perspective, assume the person makes $100,000 per year from all sources. Each answer should be between 0 and 100.

Type of income	Fair tax rate for that type of income
Wages/salary income	
Interest income (savings accounts, CD's, etc.)	
Dividend income (you own shares in a corporation or mutual fund which pays dividends)	
Capital gains (you sell an item at a profit; real estate, investments, a car, etc.)	
Gifts (for example, a relative gives you a substantial gift of cash or some other valuable item).	
Pension/retirement/social security income	

14. What would be your best estimate of the current Federal debt. Please write out your answer in a form such as "500 million dollars" or "600 billion dollars" or "2.6 trillion dollars.

15. What is your best estimate of the current year's Federal budget deficit (spending in excess of tax revenues)? Please write out your answer in a form such as "500 million dollars" or "600 billion dollars" or "2.6 trillion dollars".

16. What would you estimate as the current annual Federal spending? Please write out your answer in a form such as "500 million dollars" or "600 billion dollars" or "2.6 trillion dollars" (those aren't intended to be choices; simply illustrations of the desired answer format).

17. Open ended question: In what ways is the US personal Income tax system unfair? What changes would you recommend to increase fairness? Please restrict comments to the personal income tax, as opposed to corporate taxes.

References

Alm, J., and Gomez, J. L. (2008). "Social Capital and Tax Morale in Spain", Economic Analysis and Policy, Volume 38, Issue 1, 2008, Pages 73-87, ISSN 0313-5926, https://doi.org/10.1016/S0313-5926(08)50007-5.

Amlie T.T. and Gibney, R. F. (2023). Social Security and the Fairness of the US Tax System. Journal of Accounting, Ethics & Public Policy, JAEPP, 24 (1), 1-20.

Bastiat, Frédéric. *The Law (FEE ed.)*. Foundation for Economic Education, 1998.

Bird, R., and Davis-Nozemack, K. (2018). Tax avoidance as a sustainability problem: JBE. *Journal of Business Ethics, 151*(4), 1009-1025. doi:https://doi.org/10.1007/s10551-016-3162-2

Bordignon, M. (1993). A fairness approach to income tax evasion. Journal of Public Economics 52, 345-362.

Brennan, G., & Buchanan, J. M. (1980). The Power to Tax: Analytical Foundations of a Fiscal Constitution. Cambridge University Press.

Gallup "In depth: Topics A to Z. Taxes". https://news.gallup.com/poll/1714/taxes.aspx. Accessed 3/4/2025

Locke, J. (1689). *Two Treatises of Government*. Cambridge University Press (modern edition).

Lokanan, M. (2023). The morality and tax avoidance: A sentiment and position taking analysis. *PLoS ONE*; 2023; *18*, e0287327. [DOI: https://dx.doi.org/10.1371/journal.pone.0287327] [PubMed: https://www.ncbi.nlm.nih.gov/pubmed/37459305]

Lopez, K., and Specht, L. "The tax man cometh – will he be fair? A comparative ethical analysis of proposed tax reforms." *Southern Journal of Business and Ethics* 5 (2013): 108.

McGee, R. W. (1999). An ethical look at paying your "fair share" of taxes". Journal of Accounting, Ethics & Public Policy, 2(2). (Spring 1999) p 318-329.

McGee, R. W. (2023). How to Determine Your 'Fair Share' of Taxes: Ask Chatgpt (April 8, 2023). Working paper available at SSRN: https://ssrn.com/abstract=4413435 or http://dx.doi.org/10.2139/ssrn.4413435

Mosquera Valderrama, I. J. (2015). Legitimacy and the making of international tax law: The challenges of multilateralism. World Tax Journal, 7(3), 344–366.

Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.

Payne, D. M. and Raiborn, C. A. (2018) "Aggressive Tax Avoidance: A Conundrum for Stakeholders, Governments, and Morality" Journal of Business Ethics 147(3) p. 469-487

Pew Research Center (2023). Top tax frustrations for Americans: The feeling that some corporations, wealthy people don't pay fair share. https://www.pewresearch.org/short-reads/2023/04/07/top-tax-frustrations-for-americans-the-feeling-that-some-corporations-wealthy-people-dont-pay-fair-share/ Accessed 3/6/25.

Rothbard, Murray N. *The Ethics of Liberty*. Atlantic Highlands, NJ: Humanities Press, 1982.

Spooner, Lysander. *No Treason. No. VI. The Constitution of No Authority (1870)*. Lysander Spooner, 1870.

The Budget Lab at Yale (2024). Who is paying their fair share of taxes? A new analysis and interactive tool". The Budget Lab at Yale, Yale University, New Haven, CT. The Budget Lab Horizontal Equity Report 2024 rev1.pdf Accessed 3/4/2025

United States Internal Revenue Service. SOI tax stats – Individual statistical tables by size of adjusted gross income. https://www.irs.gov/statistics/soi-tax-stats-individual-statistical-tables-by-size-of-adjusted-gross-income

THE SOUTHERN BUSINESS AND ECONOMIC JOURNAL

College of Business
Auburn University at Montgomery
P.O. Box 244023
Montgomery, Alabama
36124-4023

AACSB ACCREDITED