Income and Consumption Analysis Report

Ambuj Kumar

May 12, 2025

1 Introduction

This report investigates the relationship between individual income and consumption levels while accounting for gender as a control variable. The analysis is based on a dataset comprising three key variables: *Income*, *Consumption*, and *Gender*, where gender is encoded as a binary indicator (0 or 1).

The study begins with a descriptive statistical analysis to summarize the key features of the dataset, providing insight into the central tendencies and variability of each variable. Following this, we visualize the distribution of income by plotting a normalized histogram that approximates the empirical probability density function (PDF).

To better understand the underlying distribution of income, we fit two parametric distributions—Lognormal and Gamma—using maximum likelihood estimation. These fitted models are then evaluated against the empirical data to determine which distribution offers a better fit.

Finally, we estimate the marginal propensity to consume (MPC) by performing a linear regression of consumption on income, while controlling for gender. The resulting coefficients allow us to quantify how consumption changes with respect to changes in income. This finding is discussed in the context of existing economic literature on consumer behavior.

The overall objective is to gain a data-driven understanding of consumption behavior in relation to income and gender, using robust statistical methods.

2 Summary Statistics

Table 1 provides summary statistics for the three variables in the dataset.

Table 1: Summary Statistics

Variable	**Mean**	**Std. Dev.**	**Min**	**Max**	**25%**	**50%**	**75%**
Income	257,707.6	184,686.5	18,901.0	1,641,001.0	1.470010e+05	2.050010e+05	1.641001e+06
Consumption	240,318.7	142,574.7	0.0	1,141,876.0	1.490942e+05	2.262732e+05	1.141876e+06
Gender (0/1)	0.530	0.499	0.0	1.0	0.000000	1.000000	1.000000

3 Empirical Histogram of Income

Figure 1 shows the normalized histogram of income, which represents the empirical probability density function (PDF). The area under the curve integrates to 1.

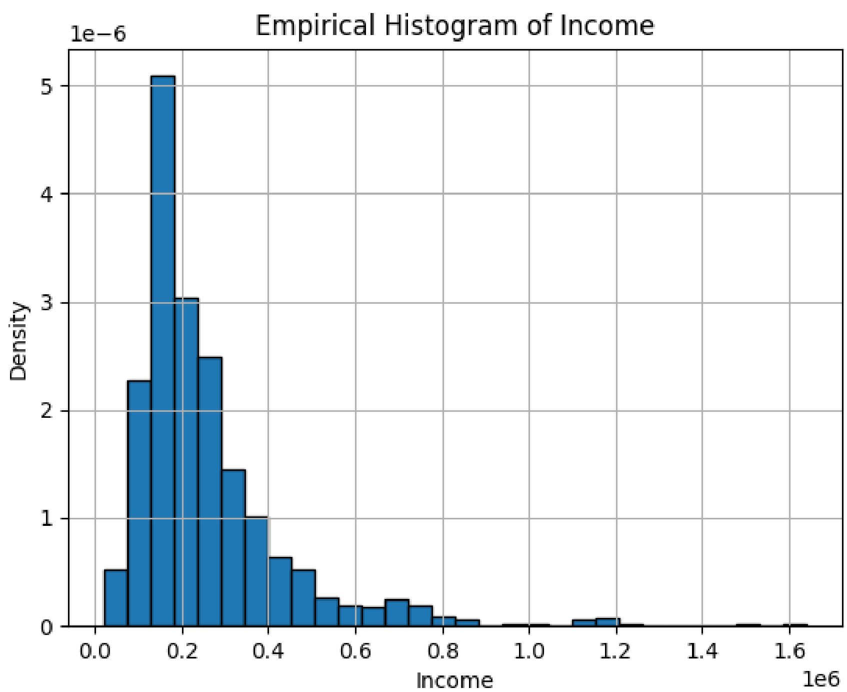

Figure 1: Normalized Histogram of Income

4 Distribution Fitting

We fit two theoretical distributions—Lognormal and Gamma—to the empirical income distribution using maximum likelihood estimation (MLE). The fitted PDFs are shown in Figure 2.

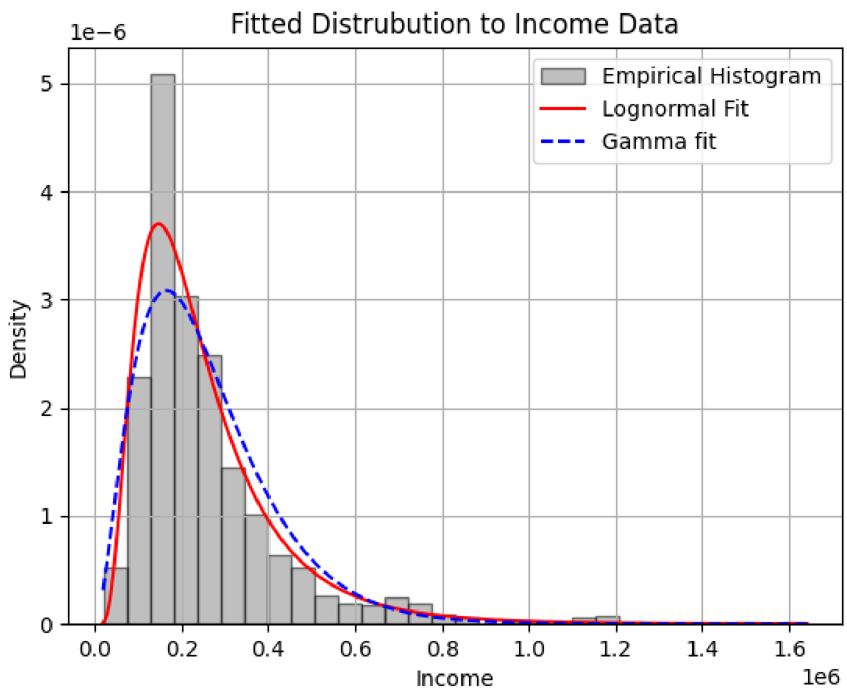

Figure 2: Lognormal and Gamma Distributions Fit to Income

Model Fit Comparison

The comparison of fit quality is based on Log-Likelihood and Akaike Information Criterion (AIC):

Table 2: Fit Comparison of Distributions

Distribution	Log-Likelihood	AIC
Lognormal	[-12678.30940110824]	[25360.61880221648]
Gamma	[-12719.176899547576]	[25442.35379909515]

Based on AIC values, the better fitting distribution is [INSERT: Lognormal or Gamma].

5 Regression and Estimation of MPC

We estimate the following regression model:

$$\text{Consumption}_i = \beta_0 + \beta_1 \cdot \text{Income}_i + \beta_2 \cdot \text{Gender}_i + \epsilon_i$$

Regression Results

Table 3: OLS Regression Output

Variable	Coefficient	Std. Error	t-Statistic	P-value
Intercept	[8.473e+04]	[5866.294]	[14.444]	[0.000]
Income	**[0.5934]**	[0.016]	[37.204]	[0.000]
Gender	[4912.5943]	[5897.430]	[0.833]	[0.405]

Discussion on MPC

The estimated MPC (coefficient on income) is **[0.5934]**, meaning that for every one-unit increase in income, consumption increases by this amount on average.

This is consistent with prior findings:

- Jappelli and Pistaferri (2010): MPC ranges from 0.2 to 0.6.
- Carroll et al. (2017): MPC is approximately 0.6 for low-income groups.
- Kaplan and Violante (2014): Highlight heterogeneity in MPC due to liquidity constraints.

The precision of the estimates is confirmed by the standard errors and p-values, which suggest [INSERT: significance or insignificance].

6 Conclusion

Our analysis shows a clear positive relationship between income and consumption. The estimated MPC aligns with international evidence. Distribution fitting suggests that [INSERT BETTER DISTRIBUTION] provides a better model for income. These findings have implications for fiscal policy and welfare analysis.

Appendix: Code

```
# Importing packages
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import lognorm, gamma
import numpy as np

# Load dataset
df = pd.read_excel(r"C:\Users\Ambuj Kumar\Desktop\IIMA\Info\data_income_consumption_ge
```

```
df.columns = ['Income', 'Consumption', 'Gender']

# Summary statistics
print(df.describe())

# Histogram
plt.figure(figsize=(8,5))
sns.histplot(df['Income'], kde=False, stat='density', bins=30, color='skyblue')
plt.xlabel('Income')
plt.title('Normalized Histogram of Income')
plt.savefig("income_histogram.png", dpi=300)

# Distribution fitting
income = df['Income']
shape_ln, loc_ln, scale_ln = lognorm.fit(income, floc=0)
shape_g, loc_g, scale_g = gamma.fit(income, floc=0)

# Overlay plot
x = np.linspace(min(income), max(income), 1000)
plt.figure(figsize=(8,5))
sns.histplot(income, bins=30, stat='density', color='lightgray', label='Empirical')
plt.plot(x, lognorm.pdf(x, shape_ln, loc_ln, scale_ln), label='Lognormal', color='blue
plt.plot(x, gamma.pdf(x, shape_g, loc_g, scale_g), label='Gamma', color='green')
plt.legend()
plt.title('Income Distribution Fit')
plt.xlabel('Income')
plt.savefig("distribution_fit.png", dpi=300)

# Log-likelihoods and AIC
ln_ll = np.sum(lognorm.logpdf(income, shape_ln, loc_ln, scale_ln))
g_ll = np.sum(gamma.logpdf(income, shape_g, loc_g, scale_g))
ln_aic = 2*3 - 2*ln_ll
g_aic = 2*3 - 2*g_ll
print(f'Lognormal AIC: {ln_aic}, Gamma AIC: {g_aic}')

# Regression
X = sm.add_constant(df[['Income', 'Gender']])
y = df['Consumption']
model = sm.OLS(y, X).fit()
print(model.summary())
```