Economic Trends in East Asia: GDP per Capita, Labor Force Participation, and Unemployment over Decades*

The Role of Demographics and Productivity in Explaining Regional Differences Across Asia

Ruiyang Wang

November 26, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

1.1 Structure of the Paper

The remainder of this paper is structured as follows:
- Section 2 describes the dataset and key variables by presenting graphs. - Section 3 presents the simple linear regression (SLR), multiple linear regression (MLR) models used to explore the effect of sample size on support percentage. - Section 4 presents the analytical results for dataset graphs and regression models

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) to analyze the dataset from The Economists (The Economist n.d.). This dataset provides annual economic indicators, including GDP per capita, unemployment rates, and hours worked per population for countries worldwide.

2.1.1 Why We Chose This Dataset

This dataset offers a comprehensive longitudinal perspective, covering trends from 1990 to 2023 across multiple economic and demographic variables. Its regional focus allows us to analyze East Asia in depth. It aligns with our goal to explore how economic and labor trends evolve and interact with demographic factors over decades.

2.2 Measurement

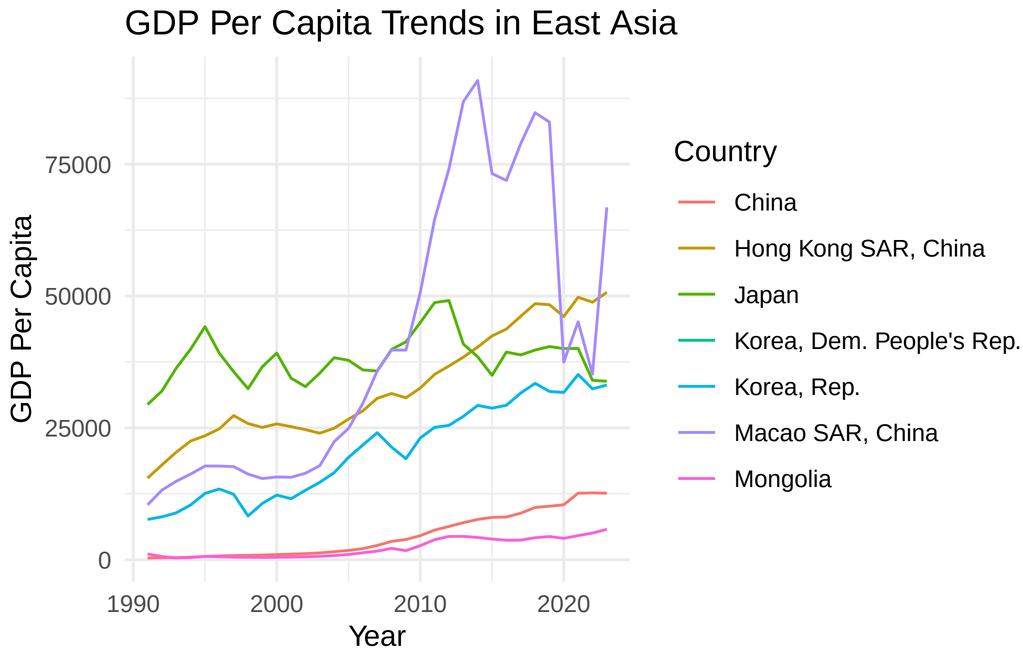
The dataset aggregates real-world economic and labor market indicators, converting them into structured data entries for analysis. Below, we explain the key measurements:

- **GDP per Capita:** Calculated as GDP divided by total population, capturing average economic output per person.
- **Unemployment Rate:** Percentage of the labor force that is unemployed, reflecting labor market health.
- **Hours Worked per Population:** Productivity indicator, calculated as total hours worked divided by population.

To ensure the consistency, most missing values were imputed, and countries outside of East Asia were excluded from the analysis. However,in order to make sure the intergrity of whole dataset, some NA values are not eliminated, since they only occur once or twice in a row, and the missing pattern has a pattern to follow, which can be avoided in later analysis

2.3 Outcome Variables

2.3.1 GDP per Capita



2.3.2 Unemployment Rate

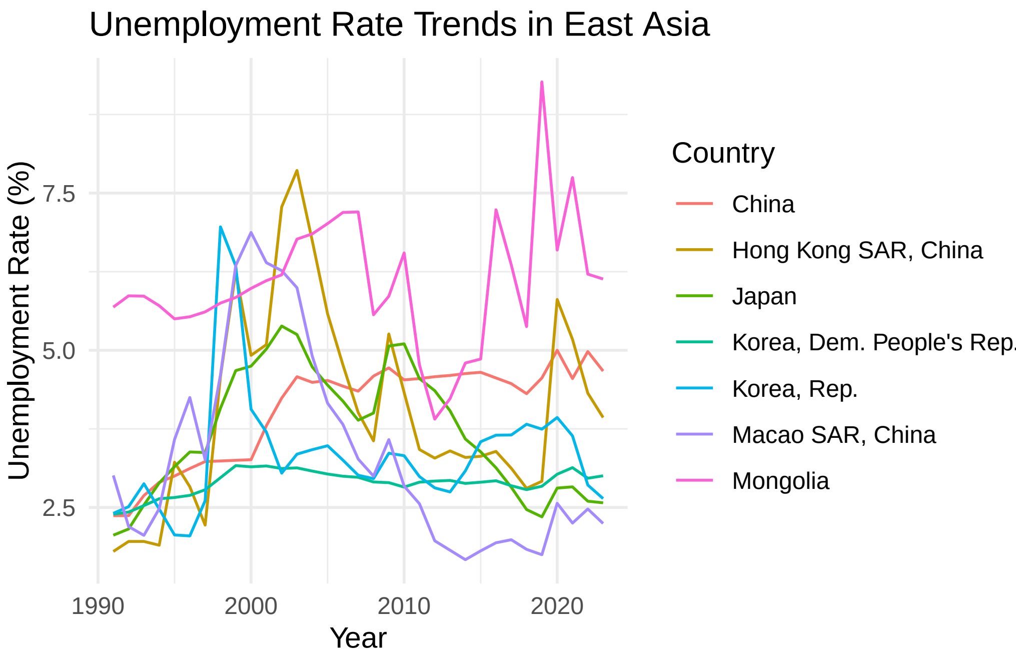


Figure 1: Unemployment Rate Trends in East Asia (1990-2023)

2.3.3 Hours Worked Per Population

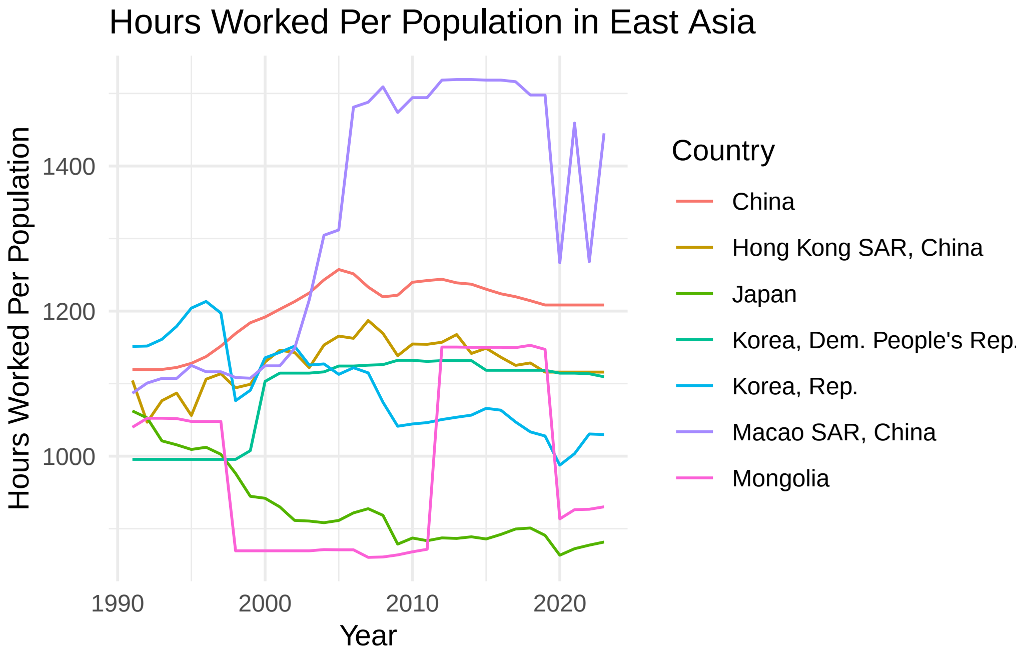


Figure 2: Hours Worked Per Population in East Asia (1990-2023)

2.4 Predictor variables

2.4.1 Population Aged 65 and Older

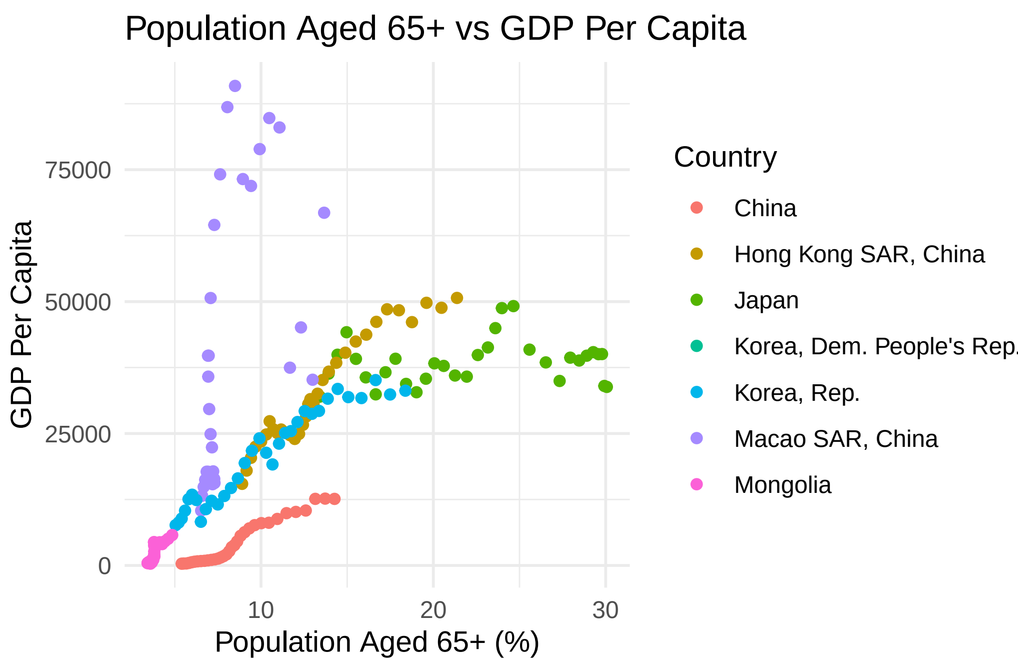


Figure 3: Population Aged 65+ vs GDP Per Capita in East Asia (1990-2023)

2.4.2 Youth Population

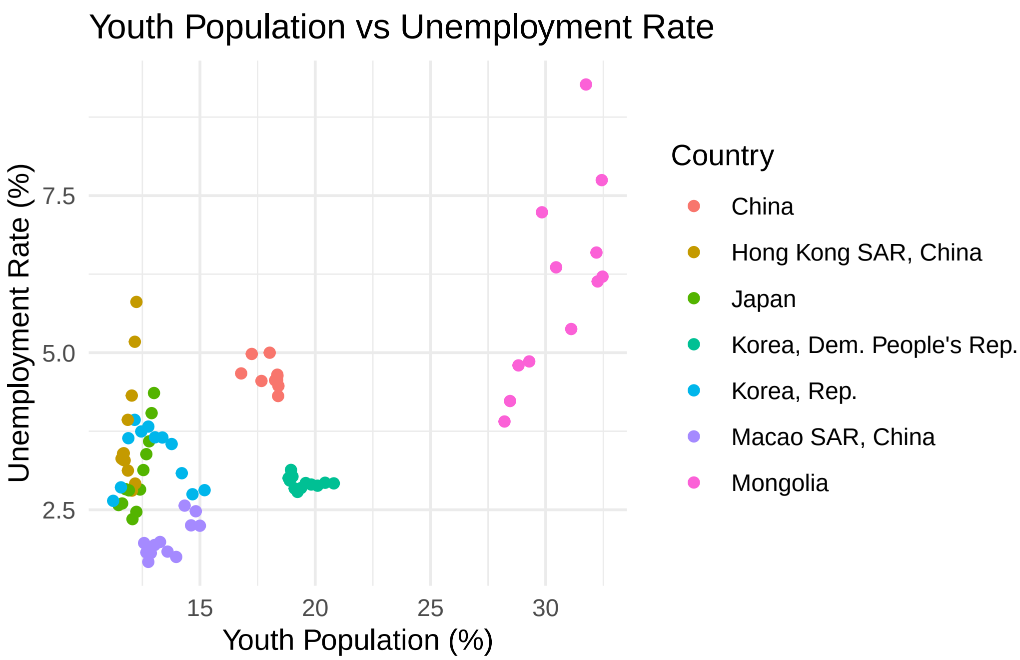


Figure 4: Youth Population vs Unemployment Rate in East Asia (1990-2023)

2.4.3 Labor Force Participation

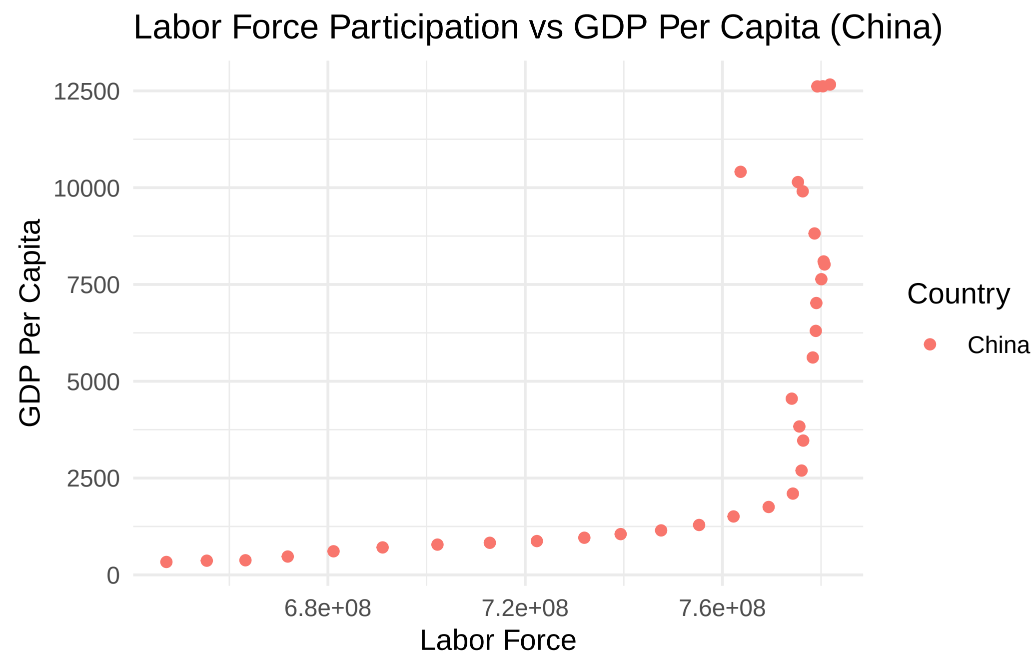


Figure 5: Labor Force Participation vs GDP Per Capita in East Asia (1990-2023)

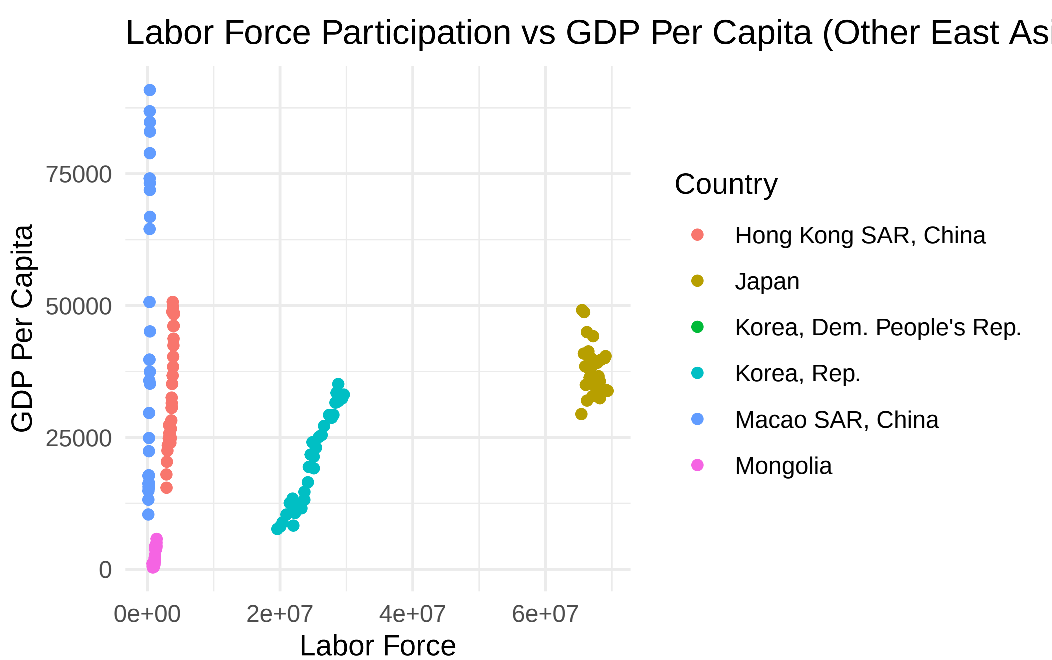


Figure 6: Labor Force Participation vs GDP Per Capita in East Asia (1990-2023)

3 Linear Regression Analysis

The goal of our regression analysis is twofold. First, we aim to capture the general trends in GDP per capita over time. Second, we explore how additional predictors, including unemployment, labor force participation, aging population, and productivity metrics, contribute to GDP per capita. Below, we outline our regression models and their justifications.

3.1 Model 1: Simple Linear Regression

Define (y_i) as the GDP per capita for country (i) in year (t). Let (Year_i) represent the year. The model is specified as:

[

$$y_i \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \beta_0 + \beta_1 \cdot \text{Year}_i \tag{2}$$

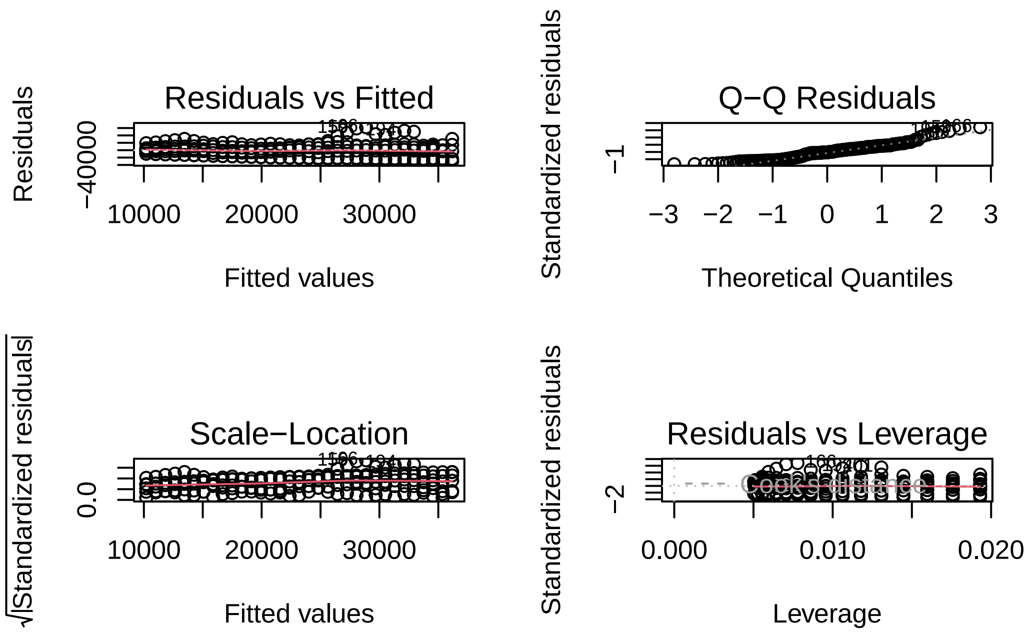
]

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.We fit this model in R using the `lm()` function:

```
Call:
lm(formula = gdp_over_pop ~ year, data = east_asia_data)

Residuals:
   Min     1Q Median     3Q    Max 
-30412 -15720   -868  10500  62002 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1606021.1   273836.1  -5.865 1.88e-08 ***
year             811.8      136.4   5.950 1.22e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18280 on 196 degrees of freedom
  (33 observations deleted due to missingness)
Multiple R-squared:  0.153, Adjusted R-squared:  0.1487 
F-statistic:  35.4 on 1 and 196 DF,  p-value: 1.217e-08
```

3.1.1 Diagnostics

To assess the model, we examine residuals for linearity, homoscedasticity, and normality:



3.2 Model 2: Multiple Linear Regression

We extend our analysis to include additional predictors: unemployment rate ((Unemployment_i)), labor force participation ((LaborForce_i)), aging population percentage ((AgingPop_i)), and hours worked ((HoursWorked_i)). Define (y_i) as the GDP per capita for country (i) in year (t):

[

$$y_i \sim \text{Normal}(\mu_i, \sigma) \tag{3}$$

$$\mu_i = \beta_0 + \beta_1 \cdot \text{Unemployment}_i + \beta_2 \cdot \text{LaborForce}_i + \beta_3 \cdot \text{AgingPop}_i + \beta_4 \cdot \text{HoursWorked}_i \tag{4}$$

]

We fit this model in R using the `lm()` function:

```
Call:
lm(formula = gdp_over_pop ~ unemployment_r + labor_force + pop_over_65 +
    hours_worked_over_pop_combined, data = east_asia_data)
```

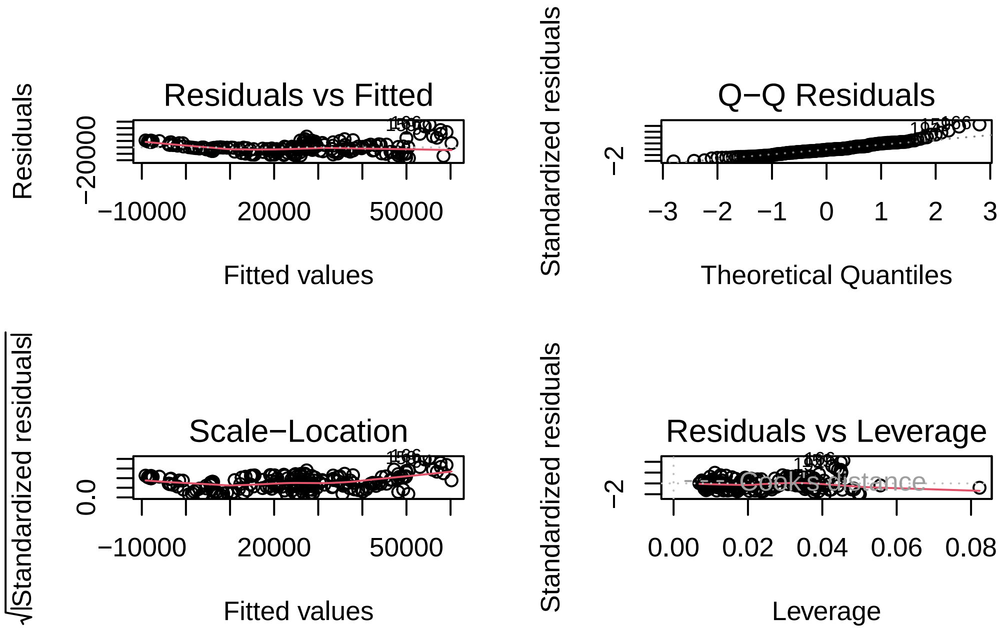
```
Residuals:
   Min     1Q Median     3Q    Max
-17558  -5163   -768   4279  35639

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                    -7.472e+04  6.983e+03 -10.701   <2e-16 ***
unemployment_r                 -9.284e+02  4.816e+02  -1.928   0.0554 .
labor_force                    -3.366e-05  2.302e-06 -14.622   <2e-16 ***
pop_over_65                     2.136e+03  1.089e+02  19.606   <2e-16 ***
hours_worked_over_pop_combined  7.463e+01  4.624e+00  16.140   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8604 on 193 degrees of freedom
  (33 observations deleted due to missingness)
Multiple R-squared:  0.8153,    Adjusted R-squared:  0.8114
F-statistic: 212.9 on 4 and 193 DF,  p-value: < 2.2e-16
```

3.2.1 Model Juestification

This multiple linear regression model enables us to explore the combined impact of key economic and demographic predictors on GDP per capita. By including these predictors, we aim to improve explanatory power and assess the relative contributions of each factor.

3.2.2 Diagnostics

To evaluate the assumptions of the MLR model, we examine residuals for: Linearity: The relationship between predictors and the outcome should be linear. Homoscedasticity: The residuals should have constant variance. Normality: Residuals should follow a normal distribution.



Multicollinearity occurs when predictors are highly correlated with each other, potentially inflating standard errors. We use the Variance Inflation Factor (VIF) to assess multicollinearity:

```
unemployment_r                  labor_force
      1.405317                     1.053640
   pop_over_65 hours_worked_over_pop_combined
      1.370456                     1.510884
```

The residuals should be independent. We test this assumption using the Durbin-Watson test:

```
	Durbin-Watson test

data:  multiple_model
DW = 2.532, p-value = 0.9999
alternative hypothesis: true autocorrelation is greater than 0
```

3.3 Model 3: Interaction Terms

To explore whether the effects of one predictor depend on another, we include interaction terms in the model. Specifically, we examine how the relationship between aging population percentage ((AgingPop_i)) and GDP per capita is influenced by labor force participation ((LaborForce_i)).

Define (y_i) as the GDP per capita for country (i) in year (t):

[

$$y_i \sim \text{Normal}(\mu_i, \sigma) \tag{5}$$
$$\mu_i = \beta_0 + \beta_1 \cdot \text{AgingPop}_i + \beta_2 \cdot \text{LaborForce}_i + \beta_3 \cdot (\text{AgingPop}_i \cdot \text{LaborForce}_i) \tag{6}$$

]

We fit this model in R using the `lm()` function:

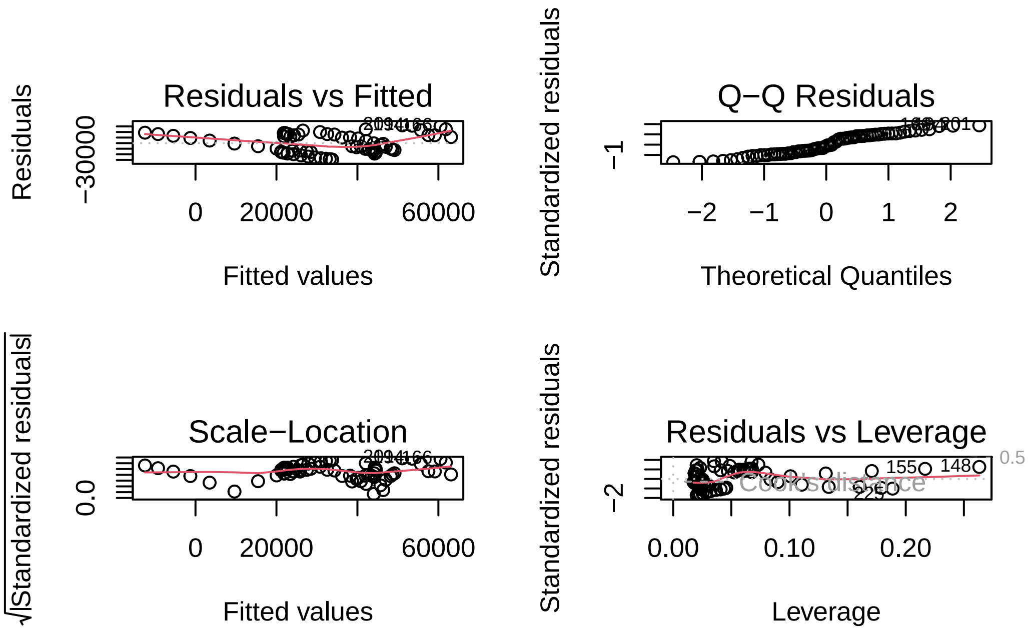
```
Call:
lm(formula = gdp_over_pop ~ pop_15_to_64 * labor_force, data = east_asia_data)

Residuals:
   Min     1Q Median     3Q    Max 
-29356 -14941  -2258  15447  31922 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)              -1.360e+05  2.967e+04  -4.582 2.02e-05 ***
pop_15_to_64              2.496e+05  4.243e+04   5.882 1.36e-07 ***
labor_force               1.191e-03  3.365e-04   3.540 0.000728 ***
pop_15_to_64:labor_force -1.742e-03  4.763e-04  -3.657 0.000499 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17340 on 68 degrees of freedom
  (159 observations deleted due to missingness)
Multiple R-squared:  0.4751,	Adjusted R-squared:  0.452 
F-statistic: 20.52 on 3 and 68 DF,  p-value: 1.412e-09
```

3.3.1 Model Juestification

This interaction model allows us to assess whether the relationship between the aging population percentage and GDP per capita changes depending on labor force participation. This is crucial for understanding how demographic and labor factors jointly influence economic performance.

3.3.2 Diagnostics

To evaluate the assumptions of the model, we examine residuals for: Linearity: The relationship between predictors (and interactions) and the outcome should be linear. Homoscedasticity: Residuals should have constant variance. Normality: Residuals should follow a normal distribution.



Multicollinearity Interaction terms can introduce multicollinearity. We check Variance Inflation Factors (VIFs) to ensure predictors remain interpretable:

```
      pop_15_to_64              labor_force pop_15_to_64:labor_force
          1.373217              2177.587572              2181.419713
```

Independence of Errors We test for autocorrelation in residuals using the Durbin-Watson test:

```

	Durbin-Watson test

data:  interaction_model
DW = 2.9521, p-value = 1
alternative hypothesis: true autocorrelation is greater than 0
```

4 Result

4.1 Data Visualization

The GDP per capita trends shown in Section 2.3.1 reveal significant variations among East Asian countries from 1990 to 2023. China demonstrates rapid growth, reflecting its economic transformation driven by industrialization, urbanization, and integration into global markets. In contrast, Japan maintains a steady GDP per capita, indicative of a mature and developed economy with limited growth potential. Macao SAR, China, consistently outperforms other countries in terms of GDP per capita due to its specialized economy, heavily reliant on tourism and gaming. However, countries like Mongolia and North Korea exhibit significantly lower GDP per capita, highlighting their developmental and structural economic challenges.

The unemployment rate trends shown inSection 2.3.2 display notable disparities across East Asia. Mongolia shows the highest volatility in unemployment rates, likely due to fluctuations in its resource-dependent economy and structural labor market issues. In contrast, Japan and Hong Kong SAR, China maintain relatively stable and low unemployment rates, underscoring their economic stability and efficient labor markets. China demonstrates a gradual decline in unemployment, reflecting improvements in job creation and labor force absorption as the economy expands.

Hours worked per population trends shown in Section 2.3.3 provide insight into productivity and labor market dynamics. Developed countries like Japan show a gradual decline in hours worked per population, reflecting shifts toward more balanced work-life policies and increased labor productivity. Macao SAR, China, exhibits fluctuations tied to its tourism-driven economy, where labor demand is sensitive to external economic conditions. In contrast, Mongolia experiences consistently low hours worked per population, indicative of inefficiencies in its labor market and economic structure.

The relationship between the percentage of the population aged 65 and older and GDP per capita highlights an interesting trend shown in Section 2.4.1. Countries with higher aging populations, such as Japan and Hong Kong SAR, China, tend to have higher GDP per capita. This reflects the economic maturity and wealth accumulation in these countries. However, countries with younger populations, like Mongolia, exhibit lower GDP per capita, which may indicate developmental challenges and a high dependency ratio.

The scatterplot of youth population percentage and unemployment rate shown in Section 2.4.2 reveals a clear pattern: higher youth populations are associated with higher unemployment rates. Mongolia stands out with a significant youth population and corresponding high unemployment, emphasizing the challenges of integrating young individuals into the labor market. In contrast, countries like Japan, with lower youth populations, maintain low unemployment rates, reflecting a more balanced labor market.

According to the two graphs of labor participation shown in Section 2.4.3, China's labor force participation shows a unique pattern due to its massive labor force. While GDP per capita increases with labor force participation, the relationship is nonlinear, suggesting diminishing returns or other contributing factors, such as productivity and industrial shifts. For other East Asian countries, smaller labor forces exhibit more clustered patterns, with countries like Japan and Hong Kong SAR, China showing high GDP per capita despite smaller workforces, highlighting the role of high productivity and economic efficiency.

These insights collectively highlight the diversity of economic and demographic trends across East Asia. While developed economies like Japan and Hong Kong SAR, China exhibit stability and high productivity, emerging economies like China showcase rapid growth and transformation. Structural challenges, such as unemployment and labor inefficiencies, remain prominent in countries like Mongolia, emphasizing the need for targeted policy interventions. The interplay between demographic factors, labor dynamics, and economic outcomes underscores the complexity of regional economic development.

4.2 Linear Regression Analysis

4.2.1 GDP per capita with time period

The results of the simple linear regression analysis provide insights into the relationship between GDP per capita and time in East Asia. The intercept, estimated at -1,606,021.07, represents the theoretical GDP per capita when the year is zero. While this value is not meaningful in a practical context, as it extrapolates to a time far before the dataset begins, it serves as a mathematical baseline for the regression model. Its primary importance lies in anchoring the regression line for interpreting changes over time.

The slope coefficient for the year variable is estimated at 811.76, indicating that GDP per capita increases by approximately 811.76 units per year on average. This positive trend is highly statistically significant (<0.001), suggesting that time is a critical factor driving economic growth in East Asia during the analyzed period (1990–2023). This result aligns with the broader understanding of the region's economic development, characterized by sustained growth and industrialization.

The model explains approximately 15.3% of the variability in GDP per capita (R square $=0.153$), with an adjusted R square of 0.1487, indicating that the year variable alone captures only a small portion of the economic variability across East Asia. This suggests that other factors, such as labor force dynamics, demographic shifts, and policy interventions, play a critical role in explaining GDP per capita trends.

Residual diagnostics show that while the model is statistically significant, the variability in residuals points to heteroscedasticity, and the unexplained variance hints at omitted variables. It can be observed that there are some discrete pattern in the tail of QQ plot, transformation can be applied to reduce the slight violation of aasumptions.

4.3 GDP per capita with multiple factors

The multiple linear regression model provides deeper insights into the factors influencing GDP per capita in East Asia by including additional predictors: unemployment rate, labor force participation, aging population percentage, and hours worked per population. The intercept, estimated at −74,720, represents the theoretical GDP per capita when all predictors are zero. While this is not directly interpretable in a meaningful economic context, it serves as a baseline for evaluating the effects of the predictors.

The coefficient for unemployment rate is estimated at −928.4, indicating that a one-unit increase in the unemployment rate is associated with a decrease in GDP per capita of approximately 928.4 units. However, with a p-value of 0.0554, this predictor is not statistically significant at the 5% significance level, suggesting that its impact on GDP per capita may be limited or influenced by other factors in the model.

Labor force participation exhibits a coefficient of -3.366×10^-5, indicating a negative relationship with GDP per capita. This suggests that increases in the labor force, without corresponding increases in productivity or economic capacity, may dilute GDP per capita. This predictor is highly significant ($p<0.001$).

The aging population percentage has a positive and significant impact on GDP per capita, with a coefficient of 2136.0. This implies that a one-unit increase in the percentage of the population aged 65 or older is associated with an increase in GDP per capita by approximately 2136 units. This finding reflects the economic structure of developed East Asian countries, where aging populations are often associated with higher productivity and income levels.

Finally, hours worked per population also shows a significant positive relationship with GDP per capita (=7.463, $p<0.001$), suggesting that higher average hours worked contribute to higher GDP per capita, likely reflecting labor productivity and economic effort.

The overall model explains approximately 81.5% of the variability in GDP per capita (R square =0.8153), with an adjusted R square of 0.8114. This indicates a substantial improvement in explanatory power compared to the simple linear regression model.

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

The Economist. n.d. "The Economist GDP per Hour Estimates." https://github.com/TheEconomist/the-economist-gdp-per-hour-estimates/blob/main/output-data/gdp_over_hours_worked_with_estimated_hours_worked.csv.