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 28, 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

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. Section 5 discusses further discussions based on paper discoveries.

*Code and data are available at: https://github.com/Ruiyang-Wang/STA304-Final-East-Asia-GDP-Analysis.git.

2 Data

2.1 Overview

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

This dataset offers a comprehensive longitudinal perspective, covering trends from 1990 to 2023 across multiple economic and demographic variables. Its global 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

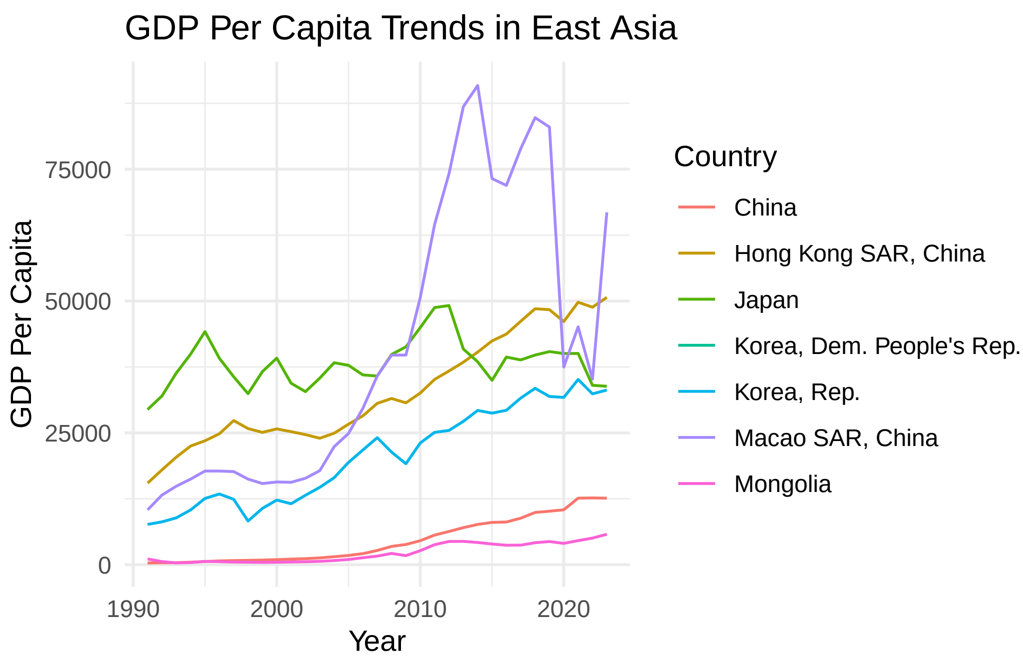
The dataset aggregates real-world economic and labor market indicators, converting them into structured data entries for analysis. Below, we explain the key measurements:

- **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 by utilizing R functions in combination of dplyr and lahman packages (Friendly et al. 2020). 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 all missing values has a pattern to follow, which can be avoided in later graphing and analyzing.

2.3 Outcome Variables

2.3.1 GDP per Capita



2.3.2 Unemployment Rate

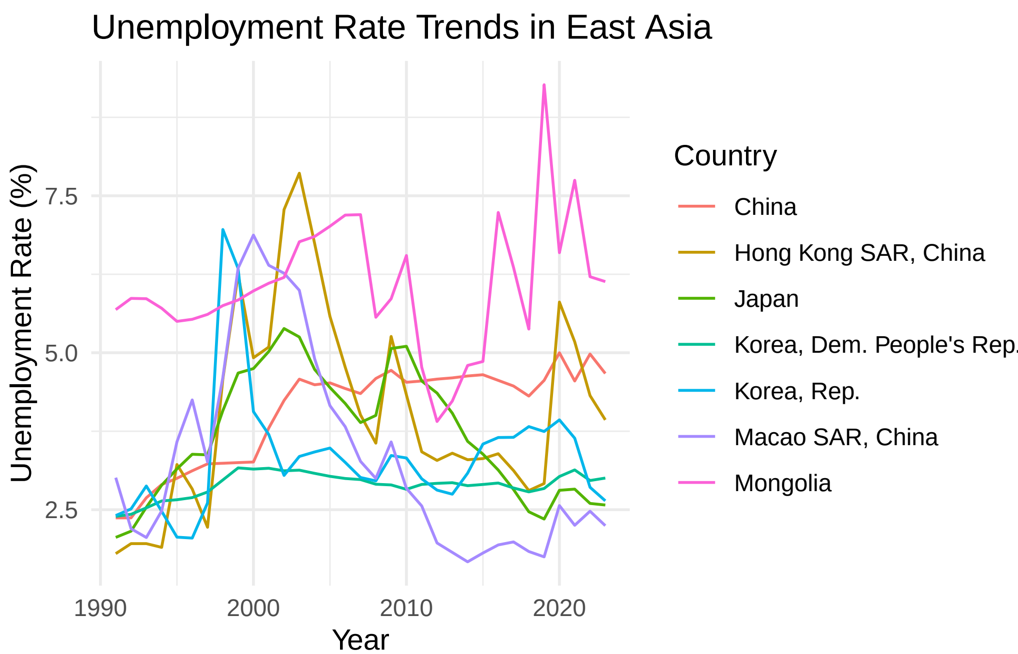


Figure 1: Unemployment Rate Trends in East Asia (1990-2023)

2.3.3 Hours Worked Per Population

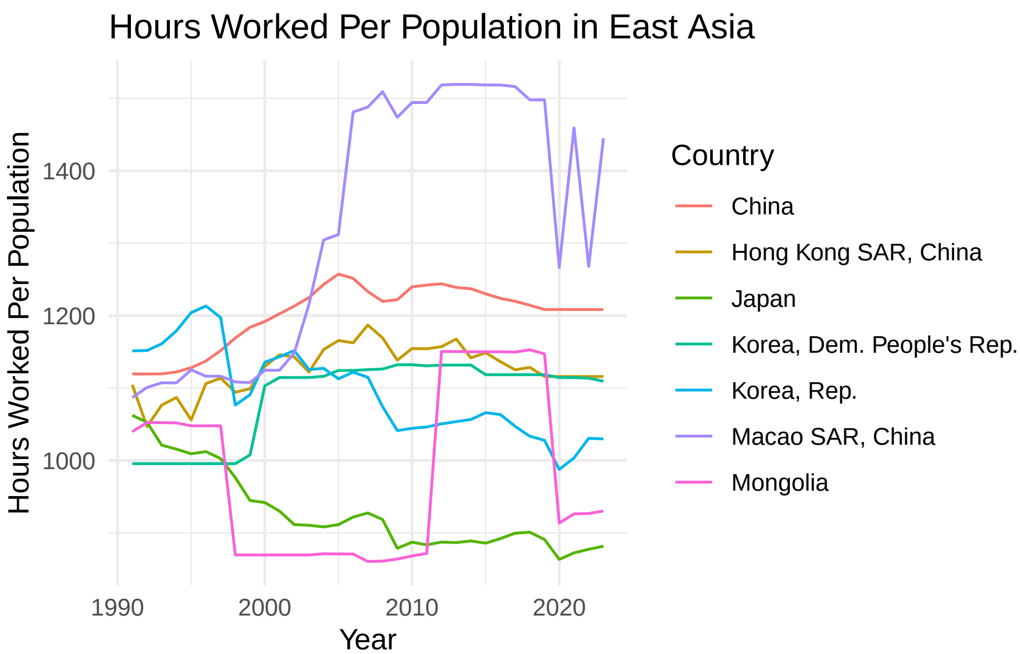


Figure 2: Hours Worked Per Population in East Asia (1990-2023)

2.4 Predictor variables

2.4.1 Population Aged 65 and Older

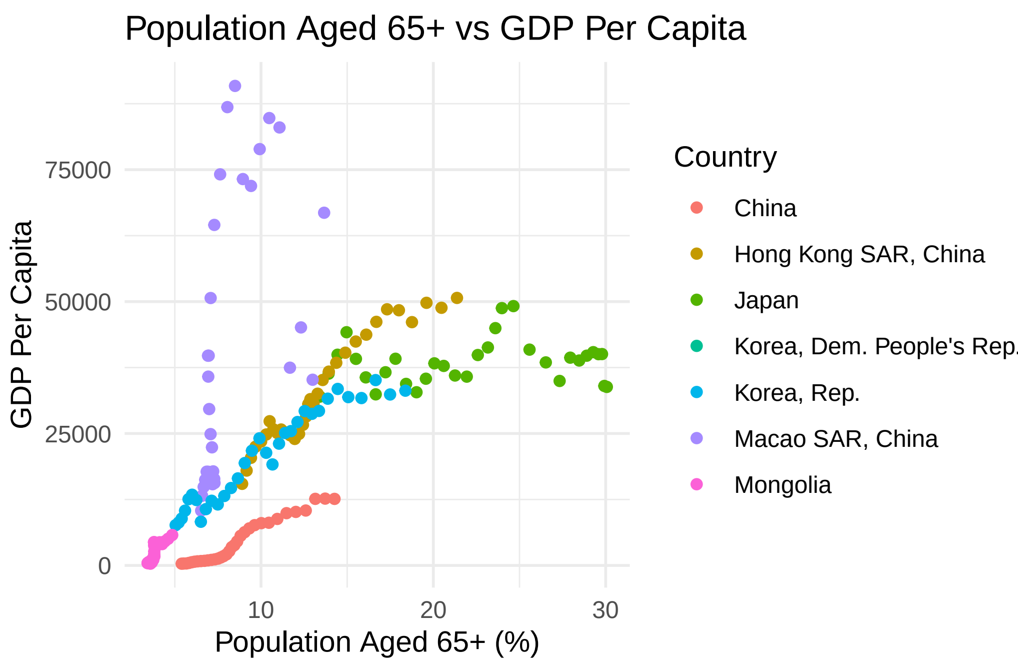


Figure 3: Population Aged 65+ vs GDP Per Capita in East Asia (1990-2023)

2.4.2 Youth Population

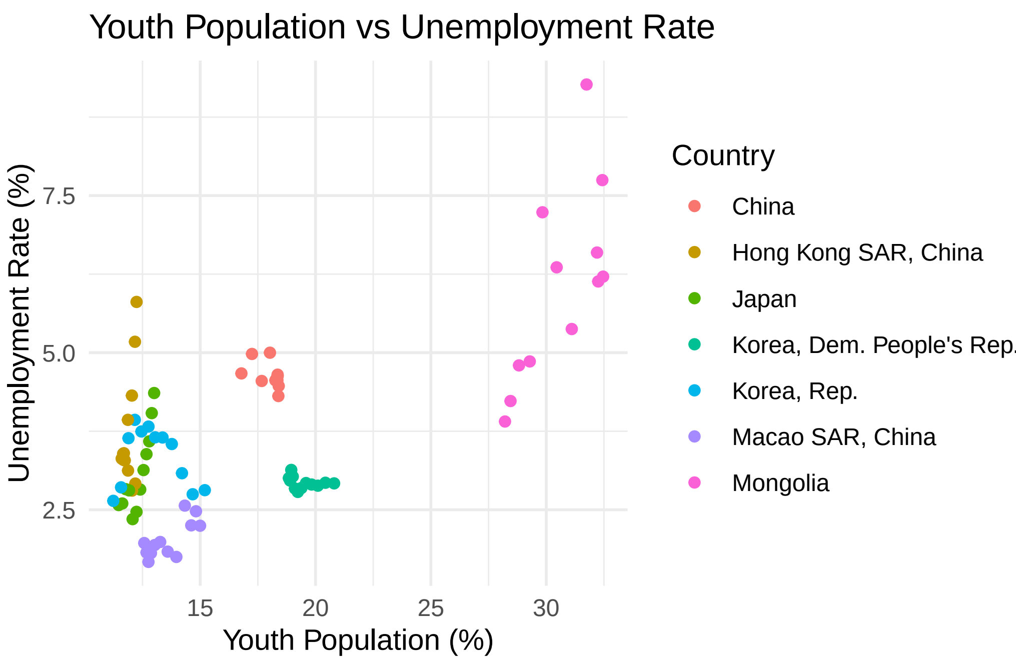


Figure 4: Youth Population vs Unemployment Rate in East Asia (1990-2023)

2.4.3 Labor Force Participation

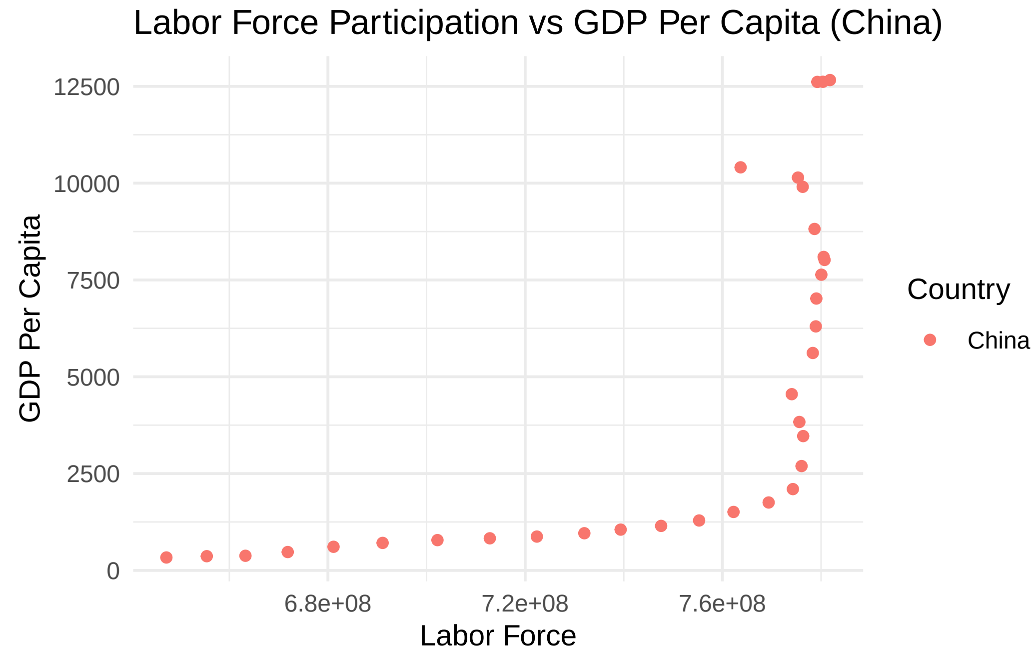


Figure 5: Labor Force Participation vs GDP Per Capita in East Asia (1990-2023)

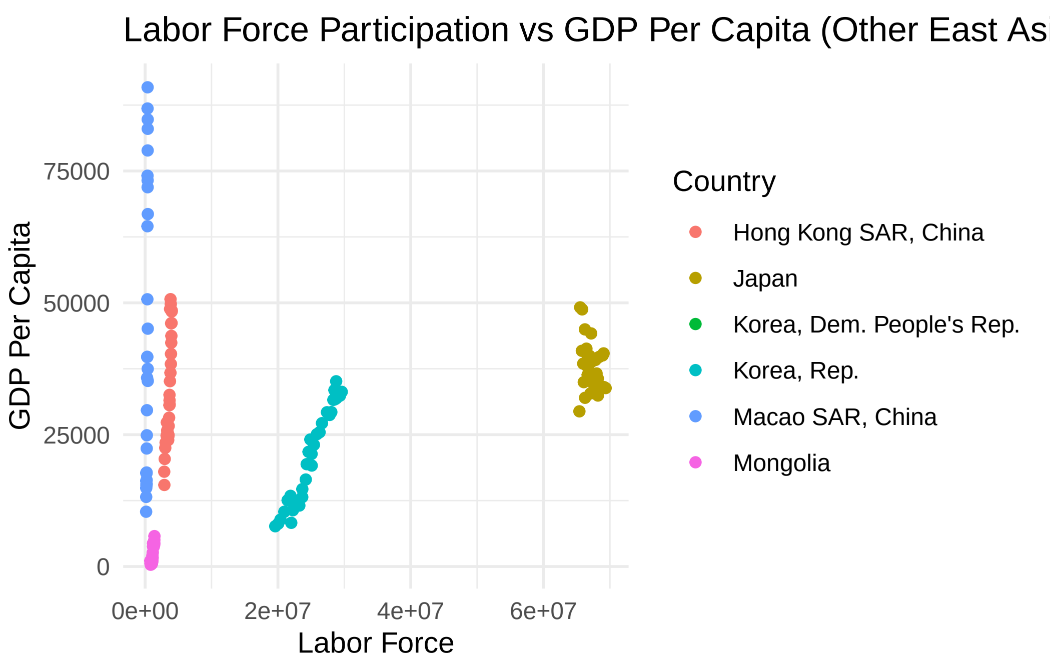


Figure 6: Labor Force Participation vs GDP Per Capita in East Asia (1990-2023)

3 Linear Regression Analysis

We did our Linear regression analysis based on the `tidyverse` package. The goal of our regression analysis is divided into three steps. First, we aim to capture the general trends in GDP per capita over time. Second, we explore how interaction terms affect GDP per capita. Third, we want to find out how additional predictors, including unemployment, labor force participation, aging population, and working hours, contribute to GDP per capita.

	(1)
(Intercept)	−1 606 021.070
	(273 836.070)
year	811.764
	(136.439)
Num.Obs.	198
R2	0.153
R2 Adj.	0.149
AIC	4452.1
BIC	4461.9
Log.Lik.	−2223.039
F	35.398
RMSE	18 188.22

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:

[

$$y_i \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$y_i = \beta_0 + \beta_1 \cdot \text{Year}_i \tag{2}$$

]

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, and the result is shown above.

This linear regression assumption check is located in Appendix at Section 6.1.1.

	(1)
(Intercept)	−135 971.649
	(29 673.340)
pop_15_to_64	249 604.222
	(42 432.871)
labor_force	0.001
	(0.000)
pop_15_to_64 × labor_force	−0.002
	(0.000)
Num.Obs.	72
R2	0.475
R2 Adj.	0.452
AIC	1615.7
BIC	1627.1
Log.Lik.	−802.869
RMSE	16 848.56

3.2 Model 2: 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):

[

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

$$\mu_i = \beta_0 + \beta_1 \cdot \text{AgingPop}_i + \beta_2 \cdot \text{LaborForce}_i + \beta_3 \cdot (\text{AgingPop}_i \cdot \text{LaborForce}_i) \tag{4}$$

]

We run the model in R (R Core Team 2023). We fit this model in R using the `lm()` function and result is shown above.

This linear regression assumption checks are located in Appendix at Section 6.1.3.

	(1)
(Intercept)	−74 721.033
	(6982.654)
unemployment_r	−928.355
	(481.623)
labor_force	0.000
	(0.000)
pop_over_65	2135.567
	(108.922)
hours_worked_over_pop_combined	74.633
	(4.624)
Num.Obs.	198
R2	0.815
R2 Adj.	0.811
AIC	4156.6
BIC	4176.3
Log.Lik.	−2072.283
RMSE	8494.18

3.3 Model 3: 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):

[

$$y_i \sim \text{Normal}(\mu_i, \sigma) \tag{5}$$

$$\mu_i = \beta_0 + \beta_1 \cdot \text{Unemployment}_i + \beta_2 \cdot \text{LaborForce}_i + \beta_3 \cdot \text{AgingPop}_i + \beta_4 \cdot \text{HoursWorked}_i \tag{6}$$

]

We run the model in R (R Core Team 2023). We fit this model in R using the `lm()` function and result is shown above.

This linear regression assumption checks are located in Appendix at Section 6.1.2.

4 Result

4.1 Data Visualization

The per capita GDP trends shown in Section 2.3.1 show significant differences between East Asian countries from 1990 to 2023. China's rapid economic growth reflects the economic transformation driven by industrialization, urbanization and integration into the global market. In contrast, Japan's GDP per capita remained stable, indicating a mature and developed economy with limited growth potential. The per capita GDP of the Chinese Macao SAR has consistently exceeded that of other countries due to its specialized economy, which relies heavily on tourism and gambling industries (Greenwood and Dwyer (2016)). However, countries such as Mongolia and North Korea have significantly lower GDP per capita, highlighting the challenges they face in terms of development and structural economies.

The unemployment rate trend shown in Section 2.3.2 displays significant differences in East Asia. Mongolia's unemployment rate fluctuated the most, likely due to volatility in its resource-dependent economy and structural labor market issues. In contrast, Japan and China's Hong Kong Special Administrative Region have maintained relatively stable and low unemployment rates, emphasizing their stable economies and efficient labor markets. China's unemployment rate is fluctuating, displays instability.

Section 2.3.3 shows trends in per capita hours worked, providing insight into productivity and labor market dynamics. In developed countries such as Japan, the number of hours worked per person has gradually declined, reflecting policies towards a greater work-life balance and rising labor productivity. Labor demand in the Macao Special Administrative Region of China is sensitive to external economic conditions, and its fluctuations are related to the tourism-driven economy. Mongolia, by contrast, has a consistently low number of hours worked per capita, which suggests that its labor market and economic structure are inefficient.

The relationship between the proportion of people aged 65 and over and GDP per capita highlights an interesting trend shown by Section 2.4.1. Countries with older populations, such as Japan and China's Hong Kong Special Administrative Region, tend to have higher GDP per capita. This reflects the maturity of their economies and the accumulation of wealth. However, countries with younger populations, such as Mongolia, have lower GDP per capita, which may indicate development challenges and a high dependency ratio.

The scatterplot of youth population and unemployment rate shown in Section 2.4.2 shows a clear pattern: the larger the youth population, the higher the unemployment rate. Mongolia is distinguished by its large youth population and correspondingly high unemployment rate, which underscores the challenge of integrating young people into the Labour market. By contrast, countries with smaller young populations, such as Japan, maintain lower unemployment rates, reflecting a more balanced labor market.

As can be seen from the two labor participation rate charts shown in Section 2.4.3, China's labor force is huge, and the labor participation rate presents a unique pattern. While GDP per capita increases as labor force participation increases, the relationship is non-linear, indicating diminishing returns or other influencing factors such as productivity and industrial relocation. For other East Asian countries, smaller workforces exhibit more of a cluster pattern, with countries like Japan and China's Hong Kong Special Administrative Region having high GDP per capita despite a smaller workforce, highlighting the role of high productivity and economic efficiency.

4.2 Linear Regression Analysis

4.2.1 GDP per capita with time period

The results of a simple linear regression analysis Section 3.1 provide insights into the relationship between GDP per capita and time in East Asia.

The slope coefficient of the annual variable is estimated to be 811.76, indicating that GDP per capita increases by about 811.76 units per year on average. This positive trend is statistically significant ($p < 0.001$), indicating that time is the key factor driving East Asian economic growth over the period analyzed (1990-2023). This result is in line with a broader understanding of the region's economic development, which is characterized by continued growth and industrialization.

The model explains about 15.3% of the change in GDP per capita (R squared = 0.153), with an adjusted R squared of 0.1487, suggesting that the year variable reflects only a small part of the change in East Asian economies. This suggests that other factors, such as labor force dynamics, demographic changes, and policy interventions, play a key role in explaining trends in GDP per capita.

4.3 GDP per capita with Working-Age Population and Labor Force Participation

The interaction model Section 3.2 introduces a key dimension to the analysis by looking at how the relationship between the share of the population aged 15-64 and GDP per capita depends on the labor force participation rate.

The coefficient for the percentage of the population aged 15 to 64 was estimated at 249,600, indicating a strong positive correlation with GDP per capita. Every one unit increase in this percentage increases GDP per capita by about 249,600 units. This result is highly statistically significant ($p<0.001$), highlighting the importance of this demographic in driving economic performance.

The positive coefficient of the labor force participation rate is 1.191×10^-3, which means that a higher labor force participation rate contributes positively to GDP per capita. This result ($p=0.000728$) reflects the role of labor force participation in economic productivity.

The interaction term between the proportion of the population aged 15 to 64 and the labor force participation rate is a negative coefficient (-1.742 $\times 10\hat{}-3$), indicating that the positive impact of the working-age population on GDP per capita decreases as the labor force participation rate increases. This finding (p=0.000499) highlights the complexity of the relationship between demographic composition and economic outcomes. It may reflect diminishing returns to labor, or the challenge of integrating a growing workforce into productive economic activity.

The overall model explains 47.51% of the change in GDP per capita (R squared =0.4751), with an adjusted R squared of 0.452. Although this is a lower explanatory power than a multiple linear regression model, the inclusion of interaction terms provides valuable insights into the interactions between population and labor factors.

4.4 GDP per capita with multiple factors

The multiple linear regression model Section 3.3 is introduced to increase the overall explanatory power of the model.

The unemployment rate coefficient is estimated at -928.4, indicating that every unit increase in the unemployment rate reduces GDP per capita by about 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 effect on GDP per capita may be limited or influenced by other factors in the model.

The coefficient of the labor force participation rate is -3.366 $\times 10\hat{}$-5, indicating that it is negatively correlated with GDP per capita. This suggests that an increase in the labor force could dilute GDP per capita without a corresponding increase in productivity or economic capacity. This prediction is highly significant ($p<0.001$).

The proportion of population aging has a significant positive impact on per capita GDP, and its coefficient is 2136.0. This means that for every unit increase in the proportion of people over 65, GDP per capita increases by about 2,136 units. This finding reflects the economic structure of developed East Asian countries, where aging populations tend to be associated with higher levels of productivity and income.

Finally, hours worked per capita were also significantly positively correlated with GDP per capita (=7.463, $p<0.001$), suggesting that higher average hours worked contributed to an increase in GDP per capita, which may reflect labor productivity and economic effort.

The overall model explains about 81.5% of the variation in GDP per capita (R squared =0.8153), with an adjusted R squared of 0.8114. This shows a substantial improvement in explanatory power compared to simple linear regression models.

5 Discussion

5.1 Full model or reduced model?

In the multiple regression analysis, the summary table shows that the unemployment rate is not as influential as other predictors, an ANOVA test is required to compare the explanatory power between full model and reduced model.

```
Analysis of Variance Table

Model 1: gdp_over_pop ~ unemployment_r + labor_force + pop_over_65 + hours_worked_over_pop_co
Model 2: gdp_over_pop ~ labor_force + pop_over_65 + hours_worked_over_pop_combined
  Res.Df        RSS Df  Sum of Sq      F  Pr(>F)
1    193 1.4286e+10
2    194 1.4561e+10 -1 -275020737 3.7155 0.05538 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA test results show that when we compare the full model (which includes the predictor unemployment_r) to the reduced model (which does not include unemployment_r), the p-value is 0.05538. This indicates that the difference in model fit is not statistically significant at the usual 0.05 threshold. In other words, removing the unemployment_r predictor from the full model does not lead to a substantial loss in explanatory power, which implies we should keep the predictor for the overall data integrality.

5.2 Diminishing Impact of the Demographic Dividend

As (Mason 1997) says in his book "Demography and the Asian Economic Miracle", an increase in the working-age population ratio would lead to an increase in income levels. That rapid increase in the working-age population due to a low birth rate would produce a "demographic dividend." However, demographic dividend occur less as less as time goes on.

In recent years, major East Asian countries discussed in this paper have faced economic slowdowns. Countries like Japan, South Korea, and China, which once benefited from a growing working-age population, are now encountering challenges related to aging populations and declining birth rates.

Japan, has suffered severe economic stagnation in recent decades (Hoshi and Kashyap (2004)). Although Japan was one of the first countries to experience a "demographic dividend" during the post-World War II economic boom, its working-age population is shrinking due to low fertility rates and an aging society.According to Takeo Hoshi, Japan's GDP growth has slowed, and the country faces increasing social welfare costs for its elderly population. Against this backdrop, the "demographic dividend" that fueled Japan's economic rise appears to have largely dissipated, replaced by challenges such as labor shortages and falling consumer demand.

In China, the one-child policy (which lasted from 1979 to 2015) reduced the birth rate, resulting in a growing elderly population. While China benefited from a demographic dividend during its rapid industrialization and economic growth shown in Section 2.3.1, the country is likely to face a "demographic crisis" in the future since its working-age population begins to shrink. Despite the effect of COVID-19, part of China's slowdown can be attributed to a shrinking workforce and the growing fiscal burden in order to feed the aging population.

In conclusion, while the demographic dividend once provided a huge boost to East Asia's economies, the region's major economies are now facing the negative effects of aging populations and declining birth rates. Over time, the demographic shift from a young, expanding workforce to an aging, shrinking population is becoming more evident in the economic performance of these countries, signaling that the era of demographic dividends may be coming to an end, leaving an significant macroeconomic problem to the government.

5.3 "Impact of Model Complexity on R-Squared

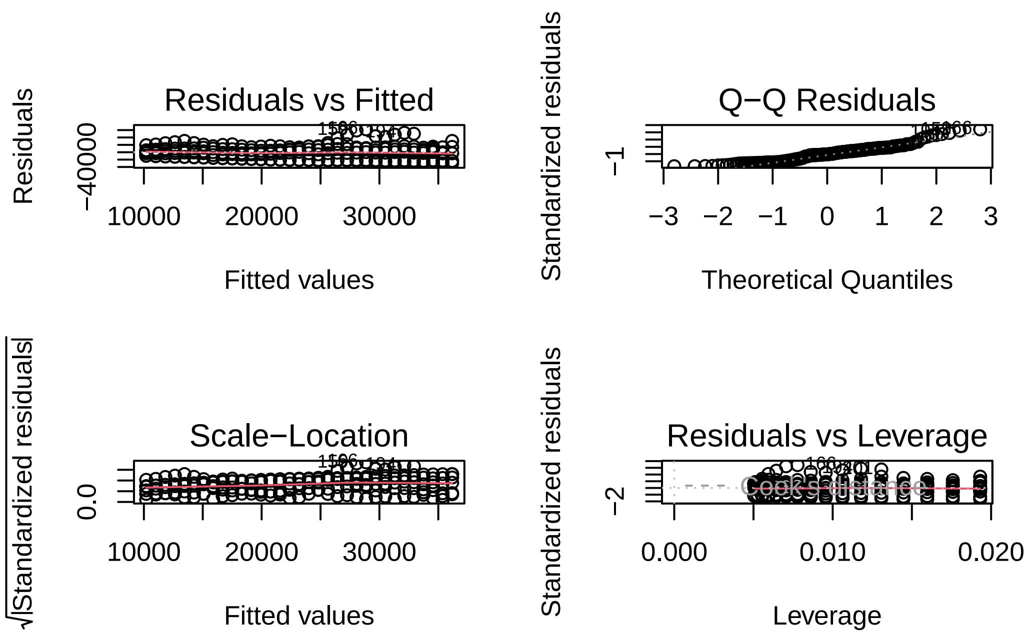
In this paper, we observe that both simple linear regression and linear regression with interaction terms have R-squared values below 0.5, suggesting that these models can only explain less than half of the variance in the outcome variables, which is, GDP per capita. This suggests that either the model specification is missing important predictors, or the relationship between the variables is more complex than what is already captured. A multiple linear regression model, however, with four variables in this paper, showed a relatively higher R-squared value of more than 0.8, indicating a much better fit to the data because including additional variables help explain a large portion of the variance. While a higher R-square generally indicates a stronger model, it also emphasizes the importance of selecting relevant variables; Too many predictors can lead to overfitting, with models capturing noise rather than the true underlying pattern.

6 Appendix

6.1 Diagnostics

6.1.1 Simple Linear Regression

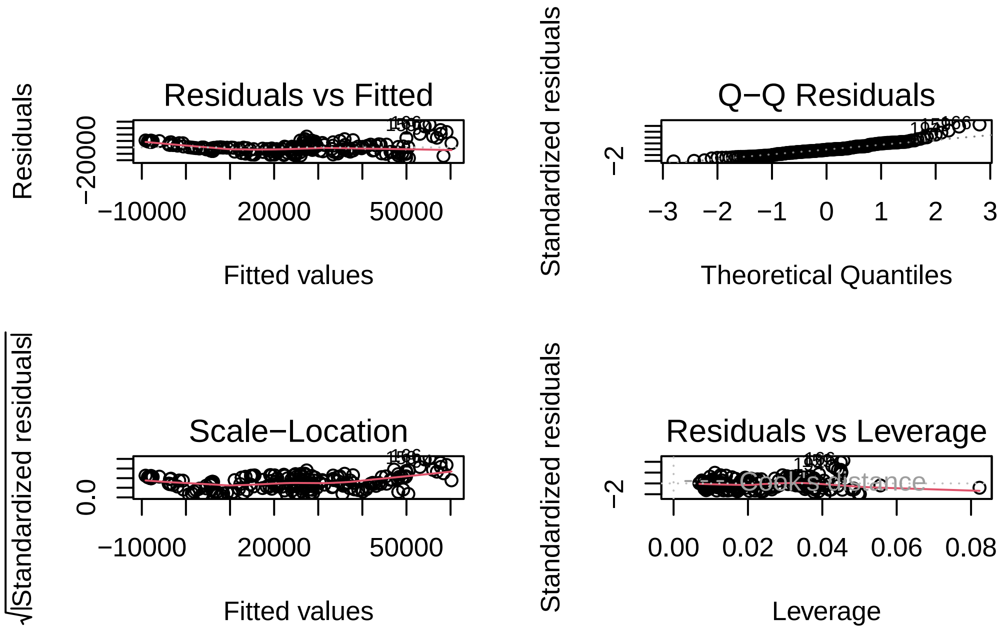
To assess the simple regression model, we examine residuals for linearity, homoscedasticity, and normality,and uncorrelated errors.



The residual diagnosis shows that while the model is statistically significant, the variability of the residual points to heteroscedasticity, while the unexplained variance suggests missing variables. It can be observed that there are slightly discrete patterns in the tail of the QQ graph, which does not affect to overall accuarcy.

6.1.2 Multiple Linear Regression

To evaluate the assumptions of the MLR model, we examine residuals for linearity, homoscedasticity, and normality,and uncorrelated errors.



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
```

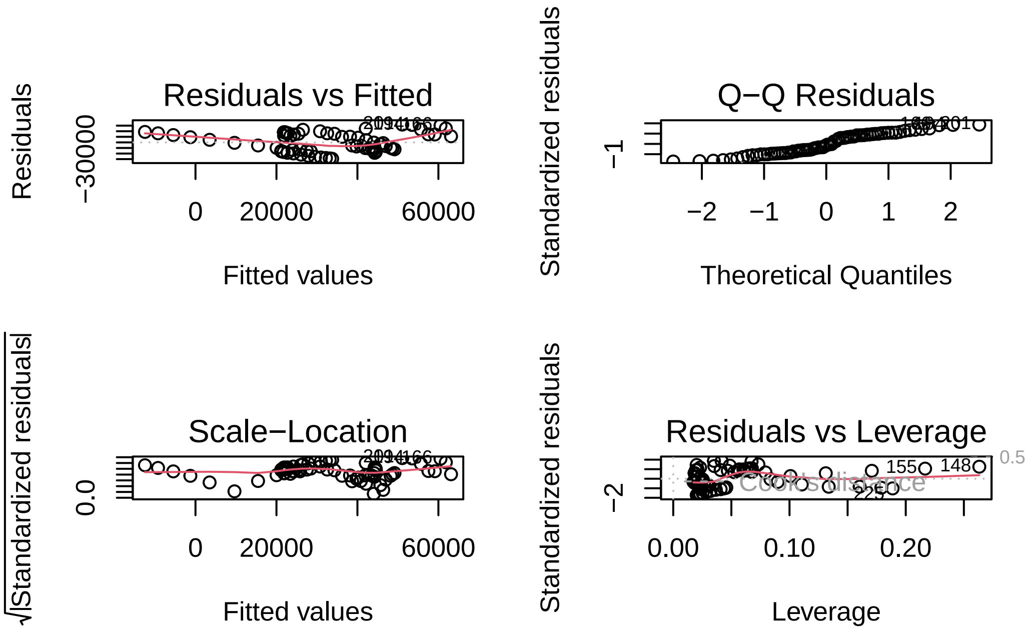
The residuals should be independent. We test this assumption using the Durbin-Watson test:

```
	Durbin-Watson test

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

6.1.3 Multiple Linear Regression (Interaction term included)

To evaluate the assumptions of the model, we examine residuals for linearity, homoscedasticity, and normality,and uncorrelated errors.



Interaction terms can introduce multicollinearity. As a result, 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
```

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
```

References

Friendly, Michael, Chris Dalzell, Martin Monkman, and Dennis Murphy. 2020. *Lahman: Sean "Lahman" Baseball Database*. https://CRAN.R-project.org/package=Lahman.

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

Greenwood, Verity Anne, and Larry Dwyer. 2016. "Reinventing Macau Tourism: Gambling on Creativity?" *Tourism Geographies* 18 (5): 580–602. https://doi.org/10.1080/13683500.2016.1187585.

Hoshi, Takeo, and Anil K. Kashyap. 2004. "Japan's Financial Crisis and Economic Stagnation." *Journal of Economic Perspectives* 18 (1): 3–26. https://doi.org/https://doi.org/10.1257/089533004773563412.

Mason, Andrew. 1997. "Population and the Asian Economic Miracle." *Asia-Pacific Population & Policy* 43: 1–4.

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. 2023. "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.