Economic Development Indicators: A Comparative Analysis of China, India, and the United States

Nicole Chen, Zihan Liang, Phoebe Pan, Mingke Tian

Table of contents

1 Introduction

This project investigates macroeconomic development indicators across China, India, and the United States from 2000 to 2023, focusing on GDP growth, GDP per capita, and the employment-to-population ratio. These indicators provide insights into both the output and distributional aspects of national economies. Using data from the World Bank World Development Indicators (WDI) database, the project applies econometric and machine learning methods to analyze the interrelationships between these variables. By combining exploratory visualizations, regression models, and predictive tools, this report aims to uncover structural economic trends and policy-relevant patterns among the three countries.

Our central research question is: **How do GDP growth, per capita income, and employment evolve differently across China, India, and the United States?**
We summarize that China's economic growth has been rapid but labor-light, India's development has been volatile and employment-sensitive, and the United States demonstrates steady, mature economic behavior.

2 Data Description

The dataset was retrieved from the World Bank API and included annual observations for the selected indicators. The countries under study—China, India, and the United States—offer a diverse range of economic trajectories. China's rapid transformation, India's emerging economy, and the United States' mature market economy serve as contrasting case studies. The original dataset was saved in wide format and included 216 rows and 6 columns, covering the years 2000 through 2023.

2.1 Data Merging Steps

To prepare the dataset, we merged indicator series into a unified panel structure. The original wide-format WDI table (one year per column) was transformed into a long-format dataset, enabling time-series and panel regressions.

2.2 Data Cleaning and Pre-processing

During cleaning, we removed entries with missing values and standardized variable names. Using DuckDB's SQL UNPIVOT operation, we ensured consistent time series formatting across all countries and indicators, allowing reliable longitudinal analysis.

The final cleaned dataset was saved as `wdi_long_clean.csv` for downstream statistical modeling.

2.3 Summary of Key Variables

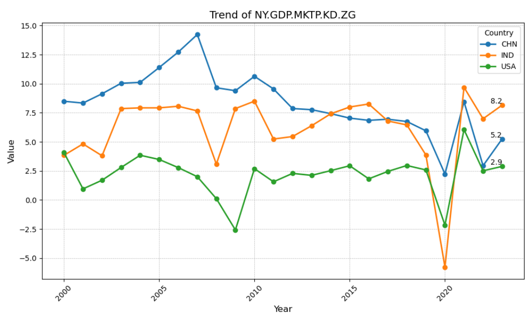
The three indicators analyzed are:

- **Annual GDP Growth Rate** (`NY.GDP.MKTP.KD.ZG`): Measures year-over-year economic expansion.
- **GDP per Capita** (`NY.GDP.PCAP.KD`): Reflects average individual income in constant 2015 USD.
- **Employment-to-Population Ratio** (`SL.EMP.TOTL.SP.ZS`): Indicates labor market participation relative to the working-age population.

3 Exploratory Analysis

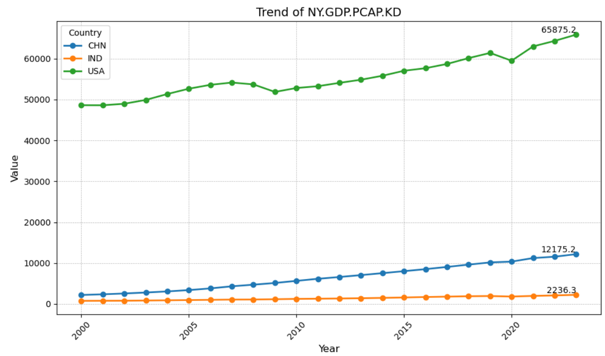
3.1 Visual Summary of Core Insights

GDP Growth (2000–2023)



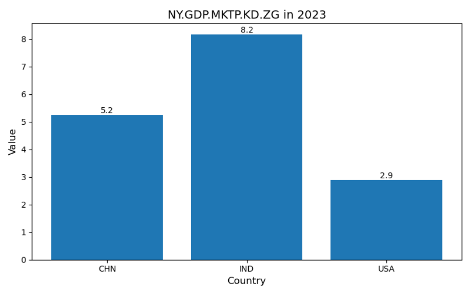
China's growth peaked before 2008. India remained volatile. U.S. dipped in 2008 & 2020.

GDP per Capita (2000–2023)



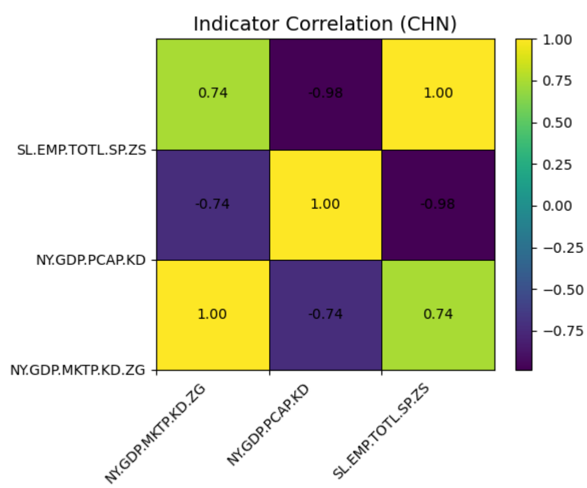
China's income converged. U.S. dominates. India grows slowly.

2023 GDP Growth



India leads post-COVID. China follows. U.S. shows slower recovery.

China Correlation Matrix



Employment decreases as income rises — indicating productivity-driven growth.

These visual comparisons reinforce the distinct developmental trajectories of China, India, and the United States, highlighting how patterns in growth, income, and employment reflect deeper structural and policy-driven differences across the three economies.

4 Data Analysis

4.1 Linear Regression Results

To explore the relationships among GDP growth, GDP per capita, and employment, we conducted a series of pairwise Ordinary Least Squares (OLS) regressions using the full panel dataset.

Dependent	Predictor	Coef	R^2	p-value
GDP_Growth	GDP/Capita	-0.0001	0.42	0.000
GDP_Growth	Employment Ratio	0.1532	0.08	0.013
GDP/Capita	GDP_Growth	-4462.63	0.42	0.000
GDP/Capita	Employment Ratio	374.24	0.01	0.384

Linear regression reveals that GDP growth declines as GDP per capita rises (coef = -0.0001, $p < 0.001$), reflecting diminishing returns in wealthier economies. Growth is positively linked to employment (coef = 0.1532, $p = 0.013$), though with limited explanatory power ($R^2 = 0.08$), suggesting labor contributes to output but isn't the main driver.

In reverse, higher growth slightly reduces GDP per capita (coef = -4462.63, $R^2 = 0.42$), indicating transitional dynamics in developing economies. Employment shows no significant effect on income ($p = 0.384$), implying that income levels are shaped more by structural factors than labor participation.

4.2 Panel Regression Results

To better account for unobserved heterogeneity and country-specific effects, we estimated panel regression models for each target variable: GDP growth, GDP per capita, and the employment-to-population ratio.

Target	Model	R^2
GDP Growth	Basic	0.5555
	Fixed Effects	0.5795
	+ Time Trend	0.5959
GDP/Capita	Basic	0.5194
	Fixed Effects	0.9869
	+ Time Trend	0.9952
Employment Ratio	Basic	0.2361
	Fixed Effects	0.9249
	+ Time Trend	0.9731

Panel regressions highlight the importance of country-specific factors. For GDP growth, explanatory power rises modestly from 55.6% to 59.6% with fixed effects and a time trend, suggesting partial influence from national context and temporal dynamics.

GDP per capita shows stronger path dependence, with R^2 increasing from 0.52 to 0.99 across models—driven by persistent structural traits like institutions and infrastructure. Employment follows a similar pattern, with model fit improving from 0.24 to 0.97, indicating stable, country-driven labor trends.

These results affirm that macroeconomic indicators are largely shaped by national trajectories, with development marked by gradual, historically grounded change.

5 Modeling and Forecasting

5.1 Time Series Forecasting (ARIMA)

We fit ARIMA models per country-indicator:

- **China**: ARIMA(1,1,0) for GDP growth; ARIMA(0,2,2) for GDP per capita.
- **India**: GDP growth fit ARIMA(0,0,0), indicating volatility.
- **United States**: ARIMA(0,1,0) and ARIMA(1,1,0) models fit various indicators.

These results align with the prior economic narrative: China's rapid transformation, India's volatile path, and the U.S.'s stable evolution.

5.2 Machine Learning Predictions

To evaluate predictive accuracy and complement our statistical analysis, we implemented machine learning models using GDP growth as the response variable. We compared three approaches: standard linear regression, elastic net regularization, and random forest.

Model	RMSE	R^2
Linear	2.2841	0.5834
Elastic Net	2.2806	0.5847
Random Forest	1.8247	0.7341

The linear model served as a baseline, with an RMSE of 2.2841 and R^2 of 0.5834. Elastic net slightly improved the fit (RMSE = 2.2806, $R^2 = 0.5847$), showing limited gains from penalizing overfitting. Both models captured broad patterns, but struggled with nonlinear relationships.

The random forest model significantly outperformed the others, achieving an RMSE of 1.8247 and R^2 of 0.7341. As an ensemble method that averages over many decision trees, it effectively captures nonlinear interactions and variable importance without strong parametric assumptions. Its superior performance suggests that economic indicators interact in complex, non-additive ways, particularly across heterogeneous countries.

6 Discussion

The results of this study highlight key developmental contrasts between the three economies. China's rapid GDP growth is well-documented, but its declining employment ratio post-2015 may reflect industrial upgrading, automation, or demographic aging. India, despite consistent GDP growth, exhibits volatile employment trends, perhaps due to informal labor market structures and inconsistent policy implementation. The U.S. stands out for its consistent high income and labor stability, even with moderate GDP growth—underscoring the role of advanced institutions and infrastructure.

One limitation of our analysis is that we focused on three countries and three indicators. Future work could disaggregate sectors, include external shocks (e.g., COVID-19), or expand to other emerging economies for broader validation.

7 Conclusion

This report provides a multi-faceted, reproducible analysis of economic development across China, India, and the United States. It shows how output growth, income convergence, and labor dynamics differ substantially across national contexts.

Understanding not just the levels of these indicators, but their dynamic behavior across countries and over time, is critical for designing adaptive, effective economic policies.

8 Further Reading

- World Bank WDI Data
- World Bank API Docs
- Indicator Metadata
- ARIMA Models Guide