

Capital Formation and Economic Growth in India: A Baseline OLS and Robustness Analysis, 2000–2025

Kalyani Upadhyay

Department of Applied Economics, University of Lucknow
email: kalyani7054@gmail.com

Received: 05-07-2026; **Accepted:** 10-08-2026; **Published Online:** 7-09-2026

Cite as: Upadhyay, K. (2026). Capital Formation and Economic Growth in India: A Baseline OLS and Robustness Analysis, 2000–2025. *Asian Journal of Business Research & Innovation*, 2(3), 36 -58.

DOI: <https://doi.org/10.5281/zenodo.22694214>

Abstract

The paper reconsiders the contemporaneous relationship between real gross capital formation and real GDP growth in India by using 26 annual WDI measurements for the period 2000-2025. The bivariate OLS specification should be read as a baseline relationship, not as a growth regression in the spirit of a structural or causal growth model. The slope is positive ($\beta = 0.2628$) but not statistically significant at the usual 5% level ($p = 0.071$) and the model accounts for 12.9% of the variation in growth. Diagnostics indicate that 2020 is highly influential (Cook's $D = 0.916$) and results in significant non-normality of the residuals. Dropping 2020 or 2020-2021 in a lagged-GCF model with HC1, HC3 or Newey-West standard errors, the coefficient is positive but statistically insignificant. Adding a 2020 shock dummy brings the slope down to 0.1130 ($p=0.206$). The estimates for the sub-periods 2000-2007, 2008-2019 and 2020-2025 are highly volatile. When the sample is extended to 2025, the slope is a little steeper, but the message does not change: the relationship is positive but weak, and not a recognized benefit of investment-led growth.

Keywords

Capital formation, Economic growth, Indian economy, Investment-growth nexus

1. Introduction

Macroeconomic policies are primarily focused on maintaining growth in real output, which is the foundation for gains in employment, income, fiscal space and living standards. One of the main ways in which an economy can improve its productive capacity is through the creation of capital. Expenditure on equipment, buildings, roads, inventories and other types of created resources can not only increase current aggregate demand but also add to the stock of capital available for use in future output. For India, the investment-growth nexus is still politically and economically relevant, as investment in infrastructure and capacity building have been important features of the development approach.

The expectation of a positive correlation is well grounded but does not imply simple one-way causality. In the Boikos et al (2026) model, a higher saving investment rate increases the

capital per worker and the level of output in the transitory period to a higher steady state. The long-run per-capita growth is determined by the pace of technical innovation. Ouedraogo et al. (2026) contend that physical capital, human capital and population growth are the critical ingredients to explain why physical investment cannot explain all variation in production. Investment is part of aggregate expenditure (Koutsobinas, 2025) and thus Keynesian theory of investment has a short-run demand side story. Endogenous-growth models (Gamer, 2026) highlight knowledge, spillovers and the productivity of investment, rather than only its quantity.

The other channel is also important. Faster growth drives investment and impacts on the projected demand, profitability, internal finance and public income. Investment and growth can respond simultaneously to credit, government spending, trade, inflation, technology, labor supply and external shocks. Contemporaneous bivariate regression cannot disentangle these mechanisms. It can answer a more modest empirical question: was annual real GDP growth stronger or worse when India devoted more of its GDP to gross capital formation?

Differences for 2000-2025. The sample spans the pre-global financial crisis investment boom, the post-2008 slump, the financial and policy developments of the 2010s, the COVID-19 pandemic in 2020, and the 2021 rebound and growth and investment stability. The period is empirically useful but possibly harmful. In small samples, one or two observations can heavily influence the coefficient and standard error of a linear model.

A vast body of India related research has established a statistically meaningful relationship between capital formation and growth in a complex dynamic framework. Yet the estimated relationship is sensitive to the definition of investment, control variables, time horizon and econometric technique. More advanced time-series specifications are used by Bal et al. (2016), Sharma and Mittal (2019), Sarania (2021), Nazir and Mir (2023), Gupta et al. (2025), and Sinha (2025) to analyze long-term correlations, composition effects, complementary determinants and feedback. Recent studies have also emphasized the necessity of measuring fixed capital formation accurately and dividing it into public and private blocks (Preeti & Diwakar, 2026). The results suggest that a single variable OLS model is not sufficient to fully explain Indian growth.

The present study was therefore designed to be deliberately limited and explicit in its contribution. The first contribution is a replicable baseline estimate from one single international comparable data source until 2025. Second, we treat the impact of the pandemic as a real empirical result (not a footnote) by showing influence diagnostics, exclusion tests and a pandemic-shock dummy. Third, we address simultaneity problems by using lagged-GCF and exploratory one-lag predictive-causality tests, but do not claim causal identification. Fourth, we evaluate the samples 2000-2023, 2000-2024 and 2000-2025 to see if the latest observations considerably alter the estimated relationship. Sixth, we distinguish statistical evidence, economic magnitude, explanatory power, and causal interpretation. Not an investment multiplier. A positive coefficient. The main research question is: what does a simple but carefully diagnosed OLS benchmark reveal about the relationship between gross capital formation and annual real GDP growth in India over the period 2000-2025 and how robust is this finding to different inference, treatment of the pandemic, choice of lag structure

and sub-period? The formal empirical model remains simple by design, while the robustness analysis is used to determine how much confidence can reasonably be placed in that benchmark.

2. Literature Review and Theoretical Framework

2.1. Integrated theoretical framework

On the supply side, we start with the neoclassical production function. Bikos et al (2026) also argues that on the transition path capital accumulation can increase output per worker, but physical capital alone cannot by itself provide a permanent rise in output, because of diminishing returns. This constraint is clear in expanded model of Ouedraogo et.al (2026) by taking into consideration factor of human capital and population growth. The relevance for the present work is not that an investment ratio should tightly predict annual growth, but that even if a moderate R^2 is not inconsistent with a multi-factor growth process, a positive coefficient is not theoretically out of the question.

The Keynes effect works via aggregate demand. When resources are idle, investment expenditure can increase current income and output through the multiplier, and the investment products can be used later to expand capacity (Koutsobinas, 2025; Li et al; 2018). This gives a rationale for both contemporaneous association and for lagged effects. This stresses the non-arbitrariness of the lower down robustness checks with a one year lagged-GCF. This suggests there is a risk that it takes some time before investment projects pay off.

The theory of endogenous growth takes account of knowledge, learning and spillovers. Gamer (2026) shows that private investment in knowledge can have non-declining social dividends. In practice, the growth effect of a given investment ratio may be dependent on the technology, human capital, governance and project composition (Picasso, 2025; Bruns, S.B., & Ioannidis, 2020). This interpretation is consistent with the reviewer's caveat that GCF should not be used as a homogeneous or mechanically productive input at the aggregate level.

The primary endogeneity problem is the growth-investment channel. Companies will invest if they expect demand and profitability to grow, governments can finance higher capital spending from higher revenues and faster growth can help ease financial constraints. This allows to associate simultaneously the GCF and the GDP growth, even if the causal influence goes at least partly or mainly from growth to investment. So, a static OLS coefficient is a reduced form relationship. Lagged specifications and predictive-causality tests can be used to study timing, but identification is difficult with just 26 observations.

2.2. Empirical evidence and critical synthesis

A powerful cross-country benchmark is provided by Feng et al (2025), they explored the robustness of growth regressions to a wide variety of conditioning factors and found that the investment share is one of the most robust correlates of growth. The result provides support for the inclusion of investment as a baseline regressor, but it does not imply a one-line causal coefficient. More recent research also supports the hypothesis that the relationship is dependent on institution, composition and model structure. For instance, Kesar et al. (2023)

revealed that gross fixed capital formation has positive and significant effect in the BRICS countries when governance and population are included in the model.

The work on India is a different methodology. Bal et al. (2016) estimate the ARDL models for extended time series data for the period 1970-2012 and find that capital formation and growth are associated in the long run with trade openness, exchange rate, TFP and inflation. Mishra (2019) provides a better post-reform regression and shows a positive association between overall domestic capital formation and GDP. Sharma and Mittal (2019) application of nonlinear ARDL approach shows the favorable effect of gross domestic capital formation on while the economic circumstances are significant. The causality linkages between growth, openness and capital formation differ across the BRICS economies, as shown by Rani and Kumar (2019), which raises the danger of inferring direction from a static coefficient.

Composition is important too. Unnikrishnan and Kattookaran (2020) also find a positive and robust growth relation with both public and private infrastructure investment in India, although the influence of private investment is stronger in their specification. For instance, (Sarania 2021; Harb et al. 2024) examines the interaction between infrastructure, openness, FDI, fixed capital creation, labor and inflation and found considerable dynamic relationships. Nazir and Mir (2023) document a positive association between gross capital formation and growth in India for the period 1991-2020 adjusting for savings, government expenditure and inflation. The findings show that production of capital is entrenched in a more in the rest macro- economic system rather than being an autonomous engine of growth.

The recent literature has more comprehensive metrics and concerns about endogeneity. For example, in their panel analysis of growth and health outcomes in northern India, Para and Alam (2024) include gross fixed capital creation as one of the controls. Investment is usually seen as one of the factors of growth. The growth of capital formation is more essential than the static shares for West Balkans (Krasniqi and Krasniqi Markaj, 2024; Bianchi et al., 2023). This is an important result concerning the difference between levels and dynamics. Gupta et al (2025) draw differentiation between public and private investment in India and present short run and long run differing effects. Sinha (2025) is evidence for growth driven capital formation in post reform India. Preeti and Diwakar (2026) analyzed private GFCF and capital stock estimation for Punjab and found that measurement of aggregate investment may disguise significant variance at sector and ownership level.

Table 1. Comparative synthesis of selected literature

Study	Sample / capital measure	Method	Main finding	Relevance / remaining gap
Feng et al (2025)	Cross-country; investment share and growth	Sensitivity analysis	Investment share among more robust growth correlates	Supports investment as a baseline variable; not causal identification
Bal et al. (2016)	India, 1970–2012; capital formation + controls	ARDL/ECM	Positive short-run role; long-run relationship	Long sample and dynamic controls; ends before recent shocks
Sharma & Mittal (2019)	India, 1980/81–2016/17; GDCF, fiscal variables	NARDL	GDCF positively related to GDP; asymmetric fiscal effects	Shows conditioning by financing and nonlinear dynamics
Unnikrishnan & Kattookaran (2020)	India, 1961/62–2016/17; public/private infrastructure investment	Johansen/VECM	Both investment types matter; private stronger in their model	Aggregate GCF masks composition
Sarania (2021)	India, 1970–2018; infrastructure, FDI, openness, fixed capital, labour, inflation	ARDL + Toda–Yamamoto	Strong short- and long-run interactions	Supports multi-variable and feedback interpretation

Study	Sample / capital measure	Method	Main finding	Relevance / remaining gap
Para & Alam (2024)	Northern Indian states, 2002–2020; health + GFCF controls	Panel OLS/FE/RE	GFCF positive in broader state-growth model	Recent evidence that investment operates with other determinants
Gupta et al. (2025)	India, 1980/81–2019/20; public/private investment + macro controls	ARDL	Short-/long-run effects differ; crowding dynamics	Recent composition evidence; sample ends at pandemic onset
Sinha (2025)	India, 1990–2024; GFCF, growth, employment	Johansen/VECM	Finds long-run relation and growth-induced capital formation	Directly reinforces reverse-causality concern
Preeti & Diwakar (2026)	Punjab, 2011–2021; private GFCF and capital stock	Measurement/construction study	Official state estimates can understate private GFCF	Shows why aggregate/national GCF should be interpreted cautiously

Note: GDGCF = gross domestic capital formation; GFCF = gross fixed capital formation. The present paper uses the World Bank gross capital formation (GCF) ratio, a broader aggregate that includes fixed capital formation, inventory changes and valuables.

2.3. Research gap and contribution

The literature is well known, however, and there is no question that investment can be important for growth. For the purposes of this paper the main difference for a narrower scope is the behavior and credibility of a transparent current account benchmark. Nevertheless, a simple OLS estimate is of some use, if it is used as a diagnostic preliminary, rather than as a structural model. It gives us an interpretable base coefficient that we can compare to our dynamic or multivariate coefficient estimates and it tells us quickly how sensitive our outcome is to particularly influential obs.

What would matter for relevance is whether the extension to 2025 makes any difference to what one can learn. The paper does not claim to be innovative either, only two recent

observations have been incorporated. Instead, it provides correct estimates of the sample extension for 2000-2023, 2000-2024 and 2000-2025. This indicates the change of the sign, size or uncertainty of the coefficient with the new observations. To consider the pandemic as an empirical regime issue, the full sample estimate is reported, as well as the pre-pandemic, exclusion and dummy variable specifications. Sub-period estimates must be taken with utmost caution, due to the very small sample sizes. The GCF is expected to have a positive coefficient, but the formal hypothesis test is two-sided, to be compatible with journal-standard p-values published. The null is $H_0: \beta_1 = 0$ and the alternative is $H_1: \beta_1 \neq 0$ with $\beta_1 > 0$ as the theory-based sign expectation. This clarifies the issue of whether a one-sided substantive expectation exists in a two-sided statistical test.

3. Research Method

3.1. Research design, sample and data source

The paper is based on the study of the Indian macroeconomic time series data for the period 2000 to 2025 and thus covers 26 years of observations. The choice of an annual frequency is justified by the fact that the WDI GCF ratio is a continuous annual series on national accounts for the whole period and capital formation can be reasonably considered as an accumulation variable over a medium-term horizon. The quarterly Indian national accounts figure also report more disaggregated expenditure sectors, including GFCF, but both the investment concept and the data source would be changed. The annual technique thus privileges conceptual and source consistency at the cost of a small sample, a trade-off that is transparently reflected in the interpretation.

The two series were derived from the World Bank World Development Indicators (WDI) on 22 August 2026; The files have a last-updated date of 13 July 2026. Gross domestic product GDP growth indicator NY.GDP.MKTP.KD. ZG gross capital formation as a percentage of GDP indication indicator NE.GDI.TOTL. ZS no smoothing (interpolation or imputation) Full numbers by year are given in Appendix A. The replication is easy go to WDI/Databank and go down country= India indicators = the two codes given for 2000 – 2025 by year and get the data in excel or csv format. The sample is intentionally limited to the 21st century rather than going back to the maximum observations. This choice allows for the delineation of a somewhat coherent current policy period, and the investigation of the global financial crisis, the investment drought of the 2010s and the pandemic spike. The trade-off is decreased statistical power. With only 26 data points, confidence intervals are wide, unit-root tests have low power, and individual years can have a big impact.

3.2. Variable choice and measurement

Table 2. Variable definitions and measurement

Variable	Role	Measure	WDI code	Interpretation
----------	------	---------	----------	----------------

Variable	Role	Measure	WDI code	Interpretation
GDPG	Dependent	Real GDP growth, annual %	NY.GDP.MKTP.KD.ZG	Annual percentage change in constant-price GDP
GCF	Explanatory	Gross capital formation, % of GDP	NE.GDI.TOTL.ZS	GFCF + inventory changes + net acquisitions of valuables, divided by GDP

Source: World Bank (2026), *World Development Indicators*.

Economic growth (GDPG) is the annual percentage change in real GDP. The variable is a growth rate rather than a level and therefore does not mechanically trend upward with the size of the economy. Gross capital formation (GCF) is measured as a percentage of GDP. A one-unit change in GCF denotes a one-percentage-point change in the share of output devoted to capital formation.

GCF is retained rather than GFCF for three reasons. First, it is the exact broad investment series used in the original research question and provides a consistent WDI measure across the entire sample. Second, it captures total capital formation, including inventories, which can be relevant to short-run output fluctuations. Third, its breadth makes the analysis easy to reproduce internationally. The disadvantage is that GCF is less tightly linked to productive fixed assets than GFCF and does not separate public/private, infrastructure/non-infrastructure or sectoral investment. The coefficient must therefore be interpreted as an association with an aggregate capital-formation ratio, not as the return to fixed productive capital.

3.3. Baseline econometric specification and hypotheses

The baseline model is a contemporaneous bivariate OLS regression. It is explicitly treated as a reduced-form association model rather than a complete explanatory growth equation:

$$GDPG_t = \beta_0 + \beta_1 GCF_t + \varepsilon_t \quad (1)$$

where $GDPG_t$ is annual real GDP growth, GCF_t is gross capital formation as a percentage of GDP, β_0 is the intercept, β_1 is the contemporaneous association coefficient and ε_t contains omitted growth determinants and random shocks. The theory-based sign expectation is:

$$\beta_1 > 0 \quad (2)$$

The formal statistical hypotheses are $H_0: \beta_1 = 0$ and $H_1: \beta_1 \neq 0$. Two-sided p-values and 95% confidence intervals are reported throughout. A p-value below 0.05 is treated as statistically significant. Values between 0.05 and 0.10 are reported exactly and described, at most, as weak evidence; they are not treated as strong confirmation. This convention is especially important in a small sample.

3.4. OLS estimation and dynamic/endogeneity checks

OLS chooses the coefficient vector that minimizes the sum of squared residuals. In matrix notation:

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (3)$$

Contemporaneous OLS cannot resolve simultaneity. To assess whether the result survives a simple timing change, the paper estimates a one-year lagged-GCF model:

$$GDPG_t = \alpha_0 + \alpha_1 GCF_{t-1} + u_t \quad (4)$$

A small dynamic sensitivity model additionally includes lagged GDP growth:

$$GDPG_t = \gamma_0 + \gamma_1 GDPG_{t-1} + \gamma_2 GCF_{t-1} + v_t \quad (5)$$

Lastly, one-lag bivariate Granger predictive-causality tests are also run as exploratory tests in both directions. For a sample this small, failure to reject is not proof of no economic causality; the tests only examine whether one lag of one variable helps predict the other above and beyond its own lag. They are added to better address reverse causality and to avoid making an identification claim that the data do not support.

3.5. Diagnostics, influence analysis and robustness strategy

Residual variance, serial correlation, functional form and normalcy are examined by diagnostic tests. The Breusch-Pagan test for heteroskedasticity, Breusch-Godfrey LM test and Durbin-Watson statistic for first order serial correlation, Ramsey RESET test for a simple nonlinear misspecification indication, Jarque-Bera test to normality of the residuals. Because normality is particularly relevant for small-sample t and F inference, the effect of removing the most influential observation on the Jarque–Bera result is also reported.

Influence is evaluated using externally studentized residuals, leverage, Cook's distance, DFFITS and the DFBETA for the GCF coefficient. Practical screening thresholds are used only as diagnostics: Cook's $D > 4/n$, leverage $> 2p/n$, $|DFFITS| > 2\sqrt{(p/n)}$, and $|DFBETA| > 2/\sqrt{n}$, where $p = 2$ parameters in the baseline model. Exceeding a threshold does not justify automatic deletion; it indicates that the observation deserves economic interpretation and sensitivity analysis.

Robustness checks comprise HC1 and HC3 heteroskedasticity-consistent standard errors, Newey–West/HAC standard errors with one lag, exclusion of 2020 alone, exclusion of 2020–2021, a 2020 shock-dummy model, the lagged-GCF model, the dynamic lag model, pre-pandemic and reviewer-suggested sub-period estimates, and sample-extension estimates ending in 2023, 2024 and 2025. The purpose is not to search for a specification that produces significance; it is to reveal how dependent the baseline result is on reasonable alternative treatments.

3.6. Stationarity and rationale for level-ratio OLS

Time-series regressions can be spurious when non-stationary variables share trends. The dependent variable here is already a growth rate, and GCF is a bounded ratio rather than a macroeconomic level, but stationarity is nevertheless examined. The Augmented Dickey–

Fuller (ADF) test, with a constant and a parsimonious zero-lag specification suited to the short annual sample, strongly rejects a unit root for GDP growth but does not reject a unit root for GCF at 5%. The KPSS test reverses the null and does not reject level stationarity for either variable. Results for GCF are therefore mixed.

The mixed evidence is not surprising with only 26 observations, especially when the GCF ratio undergoes medium-run shifts rather than a smooth stochastic trend. The paper proceeds with level-ratio OLS for three reasons: GDPG is clearly stationary; KPSS does not reject stationarity for GCF; and the purpose is a contemporaneous descriptive association rather than estimation of a long-run cointegrating vector. No long-run equilibrium claim is made. The conflicting tests are treated as an additional reason for cautious inference, not as proof that the series is stationary.

4. Results and Discussion

4.1. Descriptive patterns and correlation

Table 3. Descriptive statistics and correlation (26 annual observations)

Statistic	GDP growth (%)	GCF (% of GDP)
Observations	26	26
Mean	6.254	34.171
Standard deviation	2.985	4.083
Minimum	-5.778	25.677
First quartile	5.295	30.875
Median	7.310	34.307
Third quartile	7.908	38.118
Maximum	9.690	41.951
Pearson correlation	0.360	—

Source: Author calculations from World Bank WDI data.

For the period 2000-2025, the average growth rate of real GDP was 6.25% per year, while GCF as a percentage of GDP was at an average value of 34.17%. The correlation is 0.360, hence there is a tiny positive linear relationship on the full dataset. The descriptive series also explain why a single full-sample coefficient might be misleading: the investment ratio goes up strongly before 2008, down during the 2010s, and partially recover, while growth has an unusual 2020 contraction and 2021 comeback.



Figure 1. India: real GDP growth, 2000–2025. Source: World Bank WDI; author presentation.

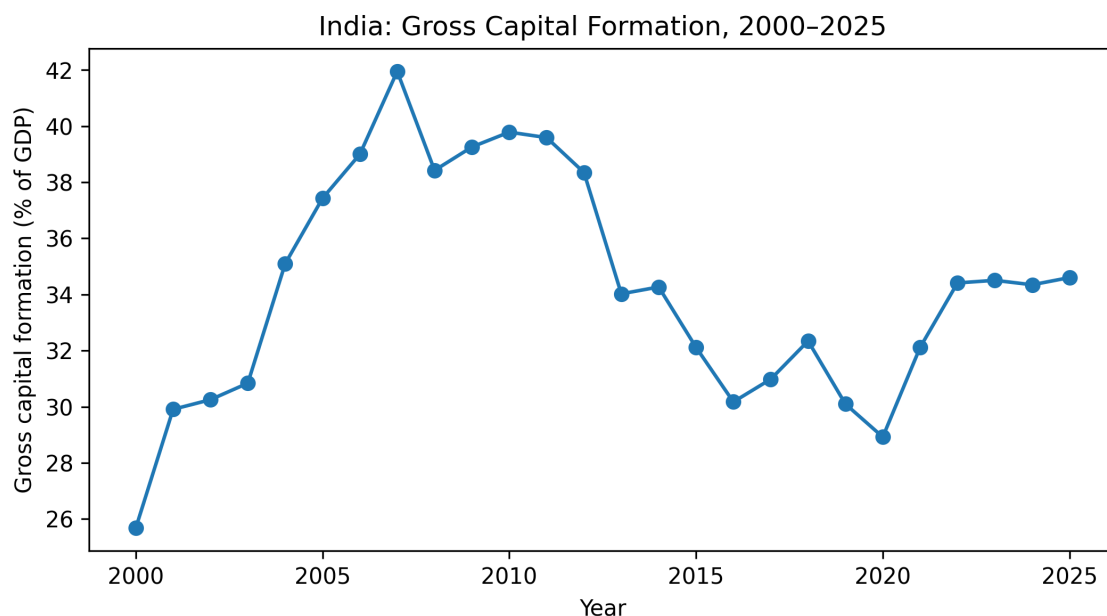


Figure 2. India: gross capital formation as a percentage of GDP, 2000–2025. Source: World Bank WDI; author presentation.

4.2. Baseline OLS estimate

Table 4. Baseline OLS regression: dependent variable = annual real GDP growth

Variable	Coefficient	SE	t	Two-sided p	95% CI

Variable	Coefficient	SE	t	Two-sided p	95% CI
Constant	-2.7265	4.7904	-0.569	0.575	[-12.6134, 7.1603]
Gross capital formation	0.2628	0.1392	1.888	0.071	[-0.0246, 0.5502]

$N = 26$; $R^2 = 0.129$; *adjusted* $R^2 = 0.093$; $F = 3.563$; $Prob(F) = 0.071$; *Durbin–Watson* = 1.998.

The fitted baseline equation is:

$$\hat{GDPG}_t = -2.7265 + 0.2628GCF_t \quad (6)$$

The coefficient of 0.2628 indicates that the fitted sample shows that a 1 percentage point increase in the proportion of GCF is associated with a 0.263 percentage point increase in annual real GDP growth. The important thing is the words “associated with”. The coefficient is not an identified causal multiplier because investment and growth are simultaneously determined and other drivers of growth are not included.

The two-sided p-value was 0.071, which is larger than the prespecified significance level of 5%. The 95% confidence interval was (-0.0246, 0.5502) and contained 0. Thus, the coefficient for the full sample is positive but not statistically significant. The p-value is below 0.10 but we do not take this as strong evidence (especially given the sample size and the influence results below).

R squared of 0.129 is a different animal than the coefficient. Thus, the simple linear correlation with GCF accounts for only 12.9% of the sample variation in annual GDP growth. That is not evidence that investment is economically irrelevant, just that it has low explanatory power. Other drivers of annual growth include consumption, public spending, productivity, trade, credit, inflation, weather and international conditions and shocks. A coefficient may be statistically significant, but the model may have low explanatory power. Neither is proof of causality.

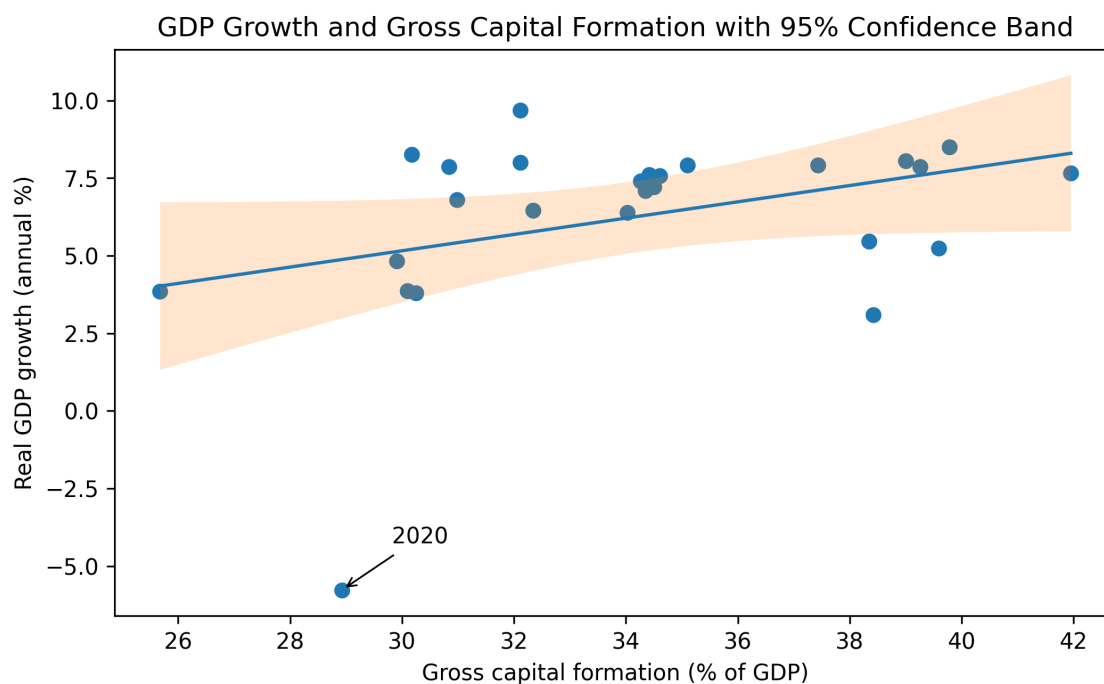


Figure 3. Gross capital formation and GDP growth with fitted OLS line and 95% confidence band. The 2020 observation is labelled because influence diagnostics identify it as exceptional.

4.3. Residual, influence and stationarity diagnostics

Table 5. Diagnostic and stationarity tests

Test	Null / purpose	Statistic	P-value	Interpretation
Breusch–Pagan LM	Homoskedastic errors	1.715	0.190	Do not reject heteroskedasticity null
Breusch–Godfrey LM(1)	No first-order serial correlation	0.0002	0.989	Do not reject
Durbin–Watson	Residual autocorrelation screening	1.998	—	Close to 2
Ramsey RESET (power 2)	No added quadratic fitted-value term	2.658	0.117	Do not reject
Jarque–Bera	Residual normality	66.789	<0.001	Reject normality

Test	Null / purpose	Statistic	p-value	Interpretation
Jarque–Bera excluding 2020	Residual normality	1.568	0.457	Do not reject after 2020 removal
ADF: GDP growth (0 lag, constant)	Unit-root null	-4.666	0.0001	Reject unit root
ADF: GCF (0 lag, constant)	Unit-root null	-2.312	0.168	Do not reject at 5%
KPSS: GDP growth (constant)	Stationarity null	0.064	>0.100	Do not reject stationarity
KPSS: GCF (constant)	Stationarity null	0.138	>0.100	Do not reject stationarity

ADF and KPSS results for GCF are not uniform. Given the 26-observation sample and different test nulls, the stationarity evidence is interpreted as mixed rather than definitive.

Table 6. Five observations with the largest Cook's distance

Year	Externally studentized residual	Leverage	Cook's D	DFBETS	DFBETA (GCF)
2020	-6.584	0.105	0.916	-2.250	1.789
2008	-1.626	0.082	0.110	-0.485	-0.353
2021	1.468	0.049	0.053	0.332	-0.152
2016	1.124	0.077	0.052	0.324	-0.229
2011	-0.905	0.109	0.050	-0.316	-0.254

Screening thresholds: Cook's $D > 4/n = 0.154$; leverage $> 2p/n = 0.154$; $|DFBETS| > 2\sqrt{(p/n)} = 0.555$; $|DFBETA| > 2/\sqrt{n} = 0.392$. Thresholds are diagnostic, not automatic deletion rules.

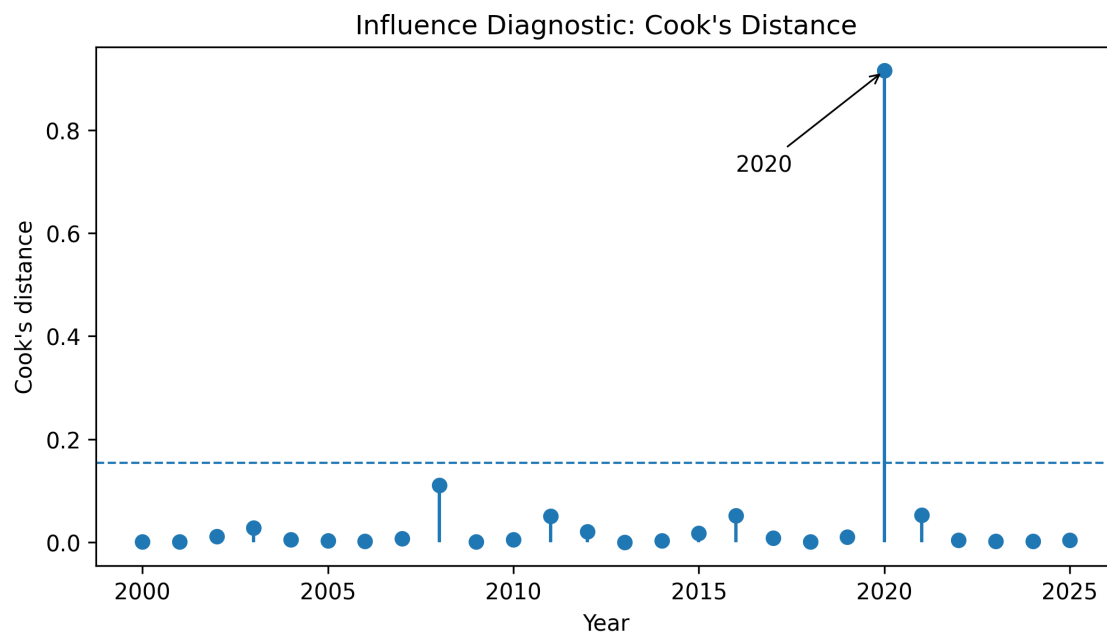


Figure 4. Cook's distance by year. The dashed line shows the common $4/n$ screening threshold.

The fundamental diagnostic issue is the 2020 observation, not heteroskedasticity or first-order serial correlation. Its externally studentized residual is -6.584, and the Cook's D is 0.916. The DFFITS is -2.250 and the DFBETA for GCF slope is 1.789. It is not very high, but it is above the influence cutoffs. That's because 2020 is having too much effect – 2020 is the pattern is 2020 is highly significant because the growth outcome is so extreme, not because its GCF value is some distance away from the investment-share range.

Residual non-normality is correspondingly severe in the full sample (Jarque–Bera $p < 0.001$), but after excluding 2020 the Jarque–Bera p -value rises to 0.457. This does not justify mechanically deleting a historically important year. Instead, it shows that conventional small-sample inference is highly sensitive to a one-off macroeconomic shock and motivates the robust and pandemic-specific specifications below.

The stationarity tests for GCF are deliberately reported as conflicting evidence. The ADF test does not reject a unit root at 5%, whereas KPSS does not reject stationarity. With only 26 annual observations, both tests have limited discriminating power and are sensitive to lag choice and medium-run shifts. The baseline regression is therefore interpreted as a sample-period contemporaneous association, not as evidence of a stable long-run equilibrium relationship.

4.4. Robustness, lag structure and pandemic sensitivity

Table 7. Robustness and alternative specifications

Specification	N	GCF coefficient	SE	Two- sided p	95% CI	R ²

Specification	N	GCF coeffie nt	SE	Two- sided p	95% CI	R ²
Baseline OLS	26	0.2628	0.1392	0.071	[-0.0246, 0.5502]	0.129
HC1 robust SE	26	0.2628	0.1608	0.115	[-0.0690, 0.5947]	0.129
HC3 robust SE	26	0.2628	0.1716	0.139	[-0.0914, 0.6170]	0.129
Newey– West/HAC(1)	26	0.2628	0.1611	0.116	[-0.0698, 0.5954]	0.129
Exclude 2020	25	0.1130	0.0868	0.206	[-0.0665, 0.2925]	0.069
Exclude 2020– 2021	24	0.1328	0.0818	0.119	[-0.0368, 0.3024]	0.107
2020 shock dummy	26	0.1130	0.0868	0.206	[-0.0665, 0.2925]	0.698
Lagged GCF (t–1)	25	0.0853	0.1493	0.573	[-0.2236, 0.3942]	0.014
Dynamic: GDPG(t–1) + GCF(t–1)	25	0.0867	0.1636	0.601	[-0.2525, 0.4260]	0.014

For HC1, HC3 and HAC rows, the OLS point estimate and R² are unchanged; only the covariance estimator changes. In the 2020-dummy model, the reported coefficient is the GCF slope while the 2020 indicator absorbs the extraordinary contraction.

Alternative inference materially weakens the apparent precision of the full-sample slope. The HC1 p-value is 0.115, HC3 is 0.139, and HAC (1) is 0.116; all corresponding 95% confidence intervals include zero. HC3 is particularly useful in small samples because it applies stronger leverage adjustments than HC1. These results reinforce the decision not to characterize the baseline $p = 0.071$ as strong evidence.

Pandemic treatment changes the magnitude as well as the precision. Excluding only 2020 reduces the slope from 0.2628 to 0.1130 ($p = 0.206$). Excluding 2020–2021 gives 0.1328 ($p = 0.119$). The 2020 shock-dummy model also shows a GCF coefficient of 0.1130 ($p = 0.206$), but the 2020 dummy itself is a large and highly significant -11.90 percentage points. The key

finding is that the positive slope in the full sample is more driven by the unusual pandemic contraction than a steady investment-growth link.

The lagged-GCF model does not show any evidence that last year's investment share has any predictive power for current year growth. Slope = 0.0853, $p = 0.573$. This result does not change when lagged GDP growth is added. Coefficient of lagged-GCF is 0.0867 ($p = 0.601$). The one-lag Granger tests are also not conclusive in rejecting the hypothesis of no predictive direction (GCF $\rightarrow$ GDPG: $p = 0.601$; GDPG $\rightarrow$ GCF: $p = 0.947$). The tests are underpowered ($n=26$) and the results do not prove lack of causation. But they do suggest that the positive contemporaneous association is not merely a one-year-ahead prediction trend.

4.5. Sub-period and 2025-extension analysis

Table 8. Reviewer-suggested sub-period estimates

Period	N	Mean growth	Mean GCF	Correlation	Slope	p-value	R ²
2000–2007	8	6.487	33.769	0.774	0.2763	0.024	0.600
2008–2019	12	6.443	34.950	-0.096	-0.0433	0.766	0.009
2020–2025	6	5.566	33.151	0.830	2.0547	0.041	0.688

Sub-period regressions are descriptive only. $N = 8$ and $N = 6$ in two segments, so coefficient and p-value instability is expected, and no structural-break claim is made.

The sub-period comparison shows pronounced instability. The 2000–2007 segment displays a strong positive association, the 2008–2019 segment displays essentially no linear association, and the 2020–2025 segment shows a very large positive slope because the 2020 contraction sits at a relatively low GCF ratio. The final segment contains only six observations and must not be interpreted as a stable post-pandemic parameter. Instead, the contrast demonstrates that the full-sample coefficient averages over different macroeconomic regimes and that the pandemic mechanically increases the apparent contemporaneous relationship.

Table 9. Does extending the sample through 2025 change the benchmark?

Sample	N	GCF slope	SE	p-value	95% CI	R ²
2000–2023	24	0.2611	0.1446	0.085	[-0.0388, 0.5610]	0.129
2000–2024	25	0.2615	0.1417	0.078	[-0.0316, 0.5547]	0.129

Sample	N	GCF slope	SE	p-value	95% CI	R ²
2000–2025	26	0.2628	0.1392	0.071	[-0.0246, 0.5502]	0.129

All p-values are two-sided conventional OLS p-values. The comparison isolates the incremental effect of the 2024 and 2025 observations on the same specification.

This is partly instructive, because it does not change anything in the result. 2000-2023 slope = 0.2611 (p=0.085) 2000-2024 slope = 0.2615 (p=0.078) 2000-2025 slope = 0.2628 (p=0.071) The additional observations raise the precision only modestly and do not alter the essential conclusion: the contemporaneous association is positive, but with a 95% confidence interval that still includes zero. Thus, the novelty of the 2025 extension is a contemporary validation of the benchmark and a demonstration of its stability to the latest normalization years, not a claim of a newly discovered investment-growth effect.

4.6. Interpreting statistical evidence, economic magnitude, explanatory power and causality

We distinguish four main kinds of evidence. First, the statistical evidence is about the coefficient's uncertainty. The p-value of the baseline model (p = 0.071) is not statistically significant at the 5% level as the 95% CI contains zero, but the robust and pandemic-adjusted p-values are larger. Second, economic significance is about the size of the point estimate: a coefficient around 0.26 would be economically important, if it were a causal effect, but the research does not show a causal effect. Third, explanatory power is about fit: R² = 0.129 suggests most of the variation in annual growth can be explained by something other than the bivariate model. Fourth, causal interpretation is about identification: we need a credible identification strategy to separate investment-to-growth effects from reverse causality, omitted variables and common shocks; the current architecture does not allow for such identification.

That matters for policy, too. The empirical evidence does not suggest that a 1 percent increase in the GCF/GDP ratio will lead to a 0.263 percentage point improvement in growth. They are only consistent with the more modest conclusion that years with larger shares of GCF were on average associated with higher fitted growth in the full 2000-2025 sample, though the association is statistically weak and sensitive to the pandemic treatment. The theoretical and empirical literature drives the larger policy suggestions on improving infrastructure quality, stimulating private investment, supporting technology adoption and enhancing human capital, not the baseline coefficient directly.

4.7. Discussion and policy implications

The full-sample results are consistent with the bulk of the investment-growth literature and India-specific literature that suggest the importance of capital formation in richer models (Bal et al., 2016; Nazir & Mir, 2023). But the robustness results are also compatible with literature construction, complementary determinants and feedback. Gupta et al. (2025) find that public

and private investment might have different short- and long-run consequences while Sinha (2025) stresses the relevance of growth driven capital accumulation. Therefore, the present result should be viewed within the framework of this conditional literature, rather than as evidence for a simple investment-led development process.

The other most informative outcome of the paper concerns the sensitivity to the epidemic. The presence of one remarkable year produces a large residual non-normality and a large rise of the GCF slope. Once 2020 is dropped or controlled for, the remaining normalcy improves and the investment coefficient becomes smaller and less precise. This is a case where, for empirical macroeconomic work with short annual samples, the shock diagnostics are not merely cosmetic: they can alter the economic significance of the primary coefficient.

Policy implications should be correspondingly modest. The paper supports continued attention to productive capital formation as part of a broader growth strategy, but it does not estimate the return to public capital expenditure or private fixed investment. Evidence from the wider literature suggests that investment composition, implementation quality, infrastructure bottlenecks, governance, technology, human capital and financing conditions influence the productivity of capital. Recommendations on these dimensions are literature-based implications rather than direct outputs of the bivariate OLS model.

The aggregate GCF measure also limits policy specificity. Inventories and valuables are included alongside fixed capital, and the series does not identify whether investment is public or private, infrastructure or equipment, or concentrated in high- or low-productivity sectors. Recent measurement work such as Preeti and Diwakar (2026) and composition studies such as Gupta et al. (2025) show why future research should move toward disaggregated investment measures when the objective is policy evaluation rather than a transparent baseline association.

4.8. Limitations

The principal limitation is the sample size. Twenty-six annual observations provide limited statistical power, make unit-root tests weak, widen confidence intervals and allow individual shocks to exert substantial influence. The second limitation is omitted-variable bias: growth depends on many factors not included in the one-regressor benchmark. Third, simultaneity and reverse causality remain unresolved. Lagged and Granger predictive checks add timing evidence but do not supply an instrument or quasi-experimental source of exogenous investment variation.

Fourth, GCF is an aggregate measure that combines fixed investment, inventories and valuables and does not distinguish public from private capital formation or asset types. Fifth, the pandemic produces an influential observation and possible regime change that a single linear equation cannot fully capture. Sixth, annual frequency limits observations; higher-frequency GFCF data could increase sample size but would change the measurement concept and require careful seasonal, structural-break and dynamic modelling. These limitations define the scope of the paper and are the reason the results are described as a baseline association rather than a structural growth model.

5. Conclusion

This paper's question is a narrow one. What would a simple, transparent and rigorously assessed OLS benchmark estimate reveal about the relation between gross capital formation and annual real GDP growth in India over 2000-2025? The coefficient for the full sample is positive at 0.2628, but the two-sided p-value is 0.071, and the 95% confidence interval contains zero. The bivariate model explains 12.9% of the variability in annual growth. Therefore, the sign of the association is consistent with theory, but the statistical precision and explanatory power are low. The robustness check qualifies the benchmark considerably. HC1, HC3 and HAC do not have significant slopes, dropping 2020 or 2020-2021 reduces the coefficient, a dummy for a shock in 2020 gives a slope of 0.1130, lagged-GCF, dynamic regressions do not point to any significant one-year-ahead predictions. Influence diagnostics show that the influential observation is 2020 and that removing it eliminates the extreme non-normality of residuals. The sub-sample estimates are quite unstable. The 2024-25 data do little to improve the precision of the full sample, without altering the implications. The novelty of the paper is therefore not that investment drives Indian growth, but a contemporaneous and reproducible benchmark for how sensitive that apparent relationship is to inference method, timing, and exceptional shocks. Future research should use longer or higher-frequency data where conceptually appropriate, disaggregate public and private fixed investment, incorporate theoretically selected controls, model structural breaks and, where possible, adopt identification strategies capable of addressing simultaneity and causal direction.

Acknowledgements

The author acknowledges the World Bank for providing open access to the World Development Indicators used in this study.

Funding Information

The author received no specific financial support for the research, authorship and/or publication of this article.

Conflict of Interest

The author declares no conflict of interest.

Data Availability and Reproducibility Statement

All observations are publicly available from the World Bank World Development Indicators. To reproduce the dataset, select India, years 2000–2025, indicator NY.GDP.MKTP.KD. ZG (GDP growth, annual %) and indicator NE.GDI.TOTL. ZS (gross capital formation, % of GDP) and export the series. The files used in this study were downloaded on 22 August 2026 and reported a World Bank last-updated date of 13 July 2026. No observations were imputed or transformed beyond the calculations reported in the paper. Appendix A reproduces the exact annual values to three decimal places.

Ethical Statement

The study uses publicly available aggregate macroeconomic data and does not involve human participants, personal data, animals or experimental interventions. Ethical approval was therefore not required.

References

- Bal, D. P., Dash, D. P., & Subhasish, B. (2016). The effects of capital formation on economic growth in India: Evidence from ARDL-bound testing approach. *Global Business Review*, 17(6), 1388–1400. <https://doi.org/10.1177/0972150916660403>
- Bianchi, C., Isabella, F., Picasso, S. (2023). Growth slowdowns at middle income levels: Identifying mechanisms of external constraints. *Metroeconomica*, 74 (2), 288–305
- Boikos, S., Panagiotidis, T., Serenidou, E., & Stengos, T. (2023). *Revisiting the Mankiw et al. (1992) growth regressions*. *Economics and Business Letters*, 12(3), 241–247.
- Bruns, S.B., & Ioannidis, J.P. (2020). Determinants of economic growth: Different time different answer? *Journal of Macroeconomics*, 63 , 103185
- Feng, G., Wang, C., & Kumbhakar, S.C. (2025). A Bayesian approach to modeling economic growth: Variable selection and cross-sectional dependence. *Journal of Econometrics*.
- Gamber, W. L. (2026). *Firm dynamics, inflation, and the transmission of monetary policy* (Finance and Economics Discussion Series 2026-003). Board of Governors of the Federal Reserve System. <https://doi.org/10.17016/FEDS.2026.003>
- Gupta, P., Rai, P., Pundir, V., & Bhawna. (2025). An empirical investigation into dynamics of growth of Indian economy impacted by public and private sector investment: Is there evidence of crowding out effect? *African Journal of Science, Technology, Innovation and Development*, 17(3), 360–375. <https://doi.org/10.1080/20421338.2025.2471605>
- Harb, G., Bassil, C., & El Sahli, Z. (2024). *Club convergence in productivity among European regions: Recent evidence and policy implications*. *The Journal of Economic Integration*, 39(3), 525–556. <https://doi.org/10.11130/jei.2024026>
- Kesar, A., Bandi, K., Jena, P. K., & Yadav, M. P. (2023). Dynamics of governance, gross capital formation, and growth: Evidence from Brazil, Russia, India, China, and South Africa. *Journal of Public Affairs*, 23(1), e2831. <https://doi.org/10.1002/pa.2831>
- Koutsobinas, T. (2025). *Keynesian behavioral macroeconomics*. Elsevier.
- Krasniqi, A., & Krasniqi Markaj, A. (2024). Analyzing the impact of final consumption expenditure and gross capital formation on GDP growth in the Western Balkans. *International Journal of Social Science and Human Research*, 7(7). <https://doi.org/10.47191/ijsshr/v7-i07-09>
- Li, F., Li, G., Qin, W., Qin, J., & Ma, H. (2018). *Identifying economic growth convergence clubs and their influencing factors in China*. *Sustainability*, 10(8), 2588. <https://doi.org/10.3390/su10082588>
- Mishra, M. (2019). Nexus between gross domestic capital formation and economic growth in India (2004–05 to 2018–2019). *International Journal of Advance and Innovative Research*, 6(1), 167–171.

- Nazir, A., & Mir, S. A. (2023). Influence of direct productive expenditure and inflation on economic growth in India. *Indian Journal of Economics and Development*, 19(4), 756–765. <https://doi.org/10.35716/IJED-23039>
- Ouedraogo, I., Ramde, F., & Bakouan, P. (2026). *Heterogeneous effects of human capital on income in developing countries: A gender-based approach*. *Cogent Economics & Finance*, 14, 2700059. <https://doi.org/10.1080/23322039.2026.2700059>
- Para, I. A., & Alam, T. (2024). Dynamics of economic growth and health outcomes among the northern states of India: A panel data analysis. *Global Business Review*. Advance online publication. <https://doi.org/10.1177/09721509241230811>
- Picasso, S. (2025). *Beyond β , σ and clubs: A unified approach to measure economic convergence*. Research Square. <https://doi.org/10.21203/rs.3.rs-7924994/v1>
- Preeti, K., & Diwakar, B. (2026). Constructing private gross fixed capital formation and capital stock for Punjab (India): New evidence (2011–2021). *Millennial Asia*. Advance online publication. <https://doi.org/10.1177/09763996261423451>
- Rani, R., & Kumar, N. (2019). On the causal dynamics between economic growth, trade openness and gross capital formation: Evidence from BRICS countries. *Global Business Review*, 20(3), 795–812. <https://doi.org/10.1177/0972150919837079>
- Sarania, R. (2021). Interactions among infrastructure, trade openness, foreign direct investments and economic growth in India. *Journal of Infrastructure Development*, 13(1), 21–43. <https://doi.org/10.1177/09749306211023621>
- Sharma, V., & Mittal, A. (2019). Fiscal deficit, capital formation, and economic growth in India: A nonlinear ARDL model. *DECISION*, 46(4), 353–363. <https://doi.org/10.1007/s40622-019-00223-8>
- Sinha, J. K. (2025). Strategic investment and inclusive growth in India: Unpacking the interlinkages between capital formation and employment generation. *Journal of Behavioral Economics and Policy*, 1(1), 1–19. <https://doi.org/10.55121/jbep.v1i1.535>
- Unnikrishnan, N., & Kattookaran, T. P. (2020). Impact of public and private infrastructure investment on economic growth: Evidence from India. *Journal of Infrastructure Development*, 12(2), 119–138. <https://doi.org/10.1177/0974930620961477>
- World Bank. (2026). *World Development Indicators* [Data set]. Indicators NY.GDP.MKTP.KD.ZG and NE.GDI.TOTL.ZS. <https://data.worldbank.org/>

Appendix A. Annual data used in the regression

Table A1 reproduces the exact WDI observations used in the analysis to three decimal places. Calculations used the downloaded numerical values before display rounding.

Table A1. India: real GDP growth and gross capital formation, 2000–2025

Year	GDP growth (%)	GCF (% of GDP)
2000	3.841	25.677
2001	4.824	29.908
2002	3.804	30.251

Year	GDP growth (%)	GCF (% of GDP)
2003	7.860	30.840
2004	7.923	35.096
2005	7.923	37.428
2006	8.061	38.999
2007	7.661	41.951
2008	3.087	38.422
2009	7.862	39.257
2010	8.498	39.786
2011	5.241	39.590
2012	5.456	38.347
2013	6.386	34.023
2014	7.410	34.268
2015	7.996	32.117
2016	8.256	30.173
2017	6.795	30.982
2018	6.454	32.343
2019	3.871	30.096
2020	-5.778	28.922
2021	9.690	32.116
2022	7.609	34.412
2023	7.210	34.505
2024	7.099	34.346
2025	7.567	34.603

Source: World Bank (2026), World Development Indicators. Files downloaded 22 August 2026; source files report last updated 13 July 2026.