



Chapter 5

E-Commerce Expansion in India: Emerging Trends, Growth Drivers, and a Regression-Based Empirical Analysis (2016–2023)

Shamit Ghai¹, Yuganshu Tickoo²

¹BBA (International business) Student, Sharda School of Business Studies, Sharda University, Greater Noida, Email Id:

2023356487.shamit@ug.sharda.ac.in

²Assistant Professor, Sharda School of Business Studies, Sharda University, Greater Noida, Email Id: yuganshu.tickoo@sharda.ac.in

Abstract: The e-commerce industry in India has experienced a transformational growth in 2016-2023, which has been triggered by the explosive adoption of smartphones, the widespread adoption of digital payment infrastructure and a post-pandemic seismic change in consumer buying behavior. In this research, the authors examine the scale and the factors of e-commerce development in India throughout this eight-year sample based on secondary panel data in IBEF, NASSCOM, RBI, and Statista. Using Ordinary Least Squares (OLS) multiple regression analysis and Compound Annual Growth Rate (CAGR) computation and

Pearson correlation matrix, the study measures the relative impact of internet users, smartphone penetration, digital payment volumes, and the density of logistics network on gross e-commerce revenue. Findings indicate that the total CAGR of the study period is 27.4% and that the strongest positive determinants of sectoral revenue are digital payments and smartphone penetration. The null hypotheses are both rejected at the 1% level and confirm the strength of the empirical results. The article also breaks down market share by category - fashion, electronics, groceries, financial services, and travel - in order to shed light on structural changes. Discussed are policy implications of regulators, logistics providers and digital marketers.

Keywords: E-Commerce, Digital Payments, OLS Regression, CAGR, Smartphone Penetration, India, Online Retail.

1. Introduction

E-commerce has emerged as one of the fastest growing entities in the Indian economy. The industry has blossomed into a USD 112 billion ecosystem, the third most significant e-commerce market in the world behind China and the United States, out of a nascent market of USD 16 billion in 2016. This phenomenal growth is not an accident; it is the result of converging structural, technological and behavioural forces that have radically transformed the way goods and services are purchased and sold in the country. These growth enablers are the Jio-led telecom revolution of 2016 which democratised access to affordable internet, the Unified Payments Interface (UPI) which made peer-to-peer and merchant digital payments simpler, and the COVID-19 pandemic of 2020/21 which increased the rate of digital adoption among all demographic and geographic groups. By 2023, India has recorded above 900 million internet users and above 750 million smartphone users - demographic dividends which perpetually populate an ever-growing consumer base to online platforms. Although earlier research concentrated on the initial development stage (2009/2013)



based on descriptive trend analysis, little of the existing empirical research has been conducted with the 2016-2023 frame and multivariate regression analysis. This paper fills this gap by estimating the independent impacts of four macroeconomic and infrastructural factors on e-commerce revenues and conditioned on the time dynamics over the eight year panel. The paper is organised as follows: Section 2 is a review of recent literature; Section 3 is the hypothesis; Section 4 describes the methodology; Section 5 is the statistical analysis and results; Section 6 is the discussion of the findings and Section 7 is the conclusion with policy recommendations.

2. Literature Review

Scholarly research on Indian e-commerce has undergone significant changes since the early 2010s, as more information has shifted towards descriptions towards empirical studies which have become more rigorous. Arora and Aggarwal (2018) examined the role of mobile internet in the adoption of online retail and discovered that an increase in smartphone penetration by 10% was associated with a 7.3 percent increase in the volumes of e-commerce transactions. They also confined their analysis to urban tier-I cities and failed to take into consideration the digital divide that was released in the rural areas after 2016. Mehta and Srivastava (2019) surveyed consumers at the consumer level in five metropolitan cities and found that trust in digital payments, which is measured by the levels of UPI adoption, is the most important predictor of repeat online purchases. They suggested that government-sponsored payment assurance plans had the potential to lead to additional development in tier-II and tier-III markets. The logistics-commerce nexus was examined by Kumar and Pandey (2020), who showed that the density of warehousing and the connectivity of the last-mile delivery system had a significant impact on the e-commerce sales per capita in states with a higher

warehousing density and better connectivity to the last-mile delivery. This observation highlights the importance of physical infrastructure and digital infrastructure. Rath, Jain and Choudhary (2021) examined the structural change in the consumer behaviour that occurred due to the pandemic, reporting that the number of first-time online buyers increased by 37 percent in the period between March 2020 and December 2020. Their difference-in-differences estimation revealed that 28 per cent of new users registered in rural areas in this time frame indicating a geographical diversification of the consumer base. Desai and Murthy (2022) looked at the supply side and the way platform-level investments in vernacular-language user interfaces and voice commerce increased market access to non-English speaking populations. In their quasi-experimental study, rollouts of vernacular features boosted monthly active users on sampled platforms by an average of 14%. Globally, Turban et al. (2018) reported that cross-country differences in the penetration rates of e-commerce depend on the interaction between regulatory environment, quality of logistics and digital payment infrastructure. They give theoretical support to the determinants studied in the current research. Extant literature therefore narrows down to four main determinants: (1) internet and smartphone penetration, (2) adoption of digital payment, (3) logistics infrastructure, and (4) consumer trust and behaviour. The current research project summarizes these threads into one multivariate construct, which is used to analyse the latest and most extensive data on India.

2.1. Research Objectives

1. To determine the annual growth trend of e-commerce in India between 2016 and 2023.



2. To determine and measure the relative impact of the important macroeconomic and infrastructural factors on e-commerce revenue with OLS multiple regression.
3. To break down market development by product/service lines and gauge the structural changes in consumer expenditure.
4. To determine the strength of association between e-commerce revenue and the determinants of e-commerce revenue using Pearson correlation matrix.
5. To make policy implications to industry stakeholders, regulators and digital infrastructure providers.

3. Research Hypotheses

According to the research goals and the literature reviewed, the following null and alternative hypotheses are developed:

H_{01} : There is no statistically significant growth in the Indian e-commerce market revenue during 2016–2023.

H_{11} : There is a statistically significant growth in the Indian e-commerce market revenue during 2016–2023.

H_{02} : Determinants (internet users, smartphone penetration, digital payment volume, logistics density) identified do not have a significant effect on e-commerce revenue.

H_{12} : The identified determinants significantly and positively influence e-commerce revenue.

4. Research Methodology

4.1. Research Design

The study adopts an explanatory and analytical research design. The explanatory component defines the growth path and structural makeup of the e-commerce market and the analytical

component uses the inferential statistics to test the hypothesised relationship among the variables. This two-fold study design is suitable considering the objective of the study to describe macro-level patterns and measure causal relations.

4.2. Data Sources and Period

The research is based solely on secondary data (eight annual observations) within 2016-2023. India Brand Equity Foundation (IBEF) reports, NASSCOM Strategic Review publications, Statista Digital Market Outlook (India) and Reserve Bank of India (RBI) Annual Reports were the sources of data on e-commerce revenue and market segmentation. Internet penetration, smartphone subscribers, and the number of digital payments were obtained in the TRAI Annual Reports and NPCI (National Payments Corporation of India) statistical releases. The figures of the logistics density were based on the report DPIIT Logistics Ease Across Different States (LEADS).

4.3. Variables Dependent

Variable: Annual e-commerce gross revenue of India (USD Billion). Independent Variables: X₁ - Internet users (millions). X₂ - Smartphone users (millions) X₃ — Digital payment transaction volume (billion transactions per year) X₄ — Density index of logistics network (composite index, 0100)

4.4. Analytical Tools

- i. Compound Annual Growth Rate (CAGR): $CAGR = [(End\ Value / Start\ Value)^{(1/n)} - 1] \times 100$, where n = number of years.
- ii. Pearson Correlation Matrix: Used to examine bivariate linear relationships between e-commerce revenue and each independent variable.



- iii. OLS Multiple Regression: The regression model is specified as: $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \varepsilon$, where Y is annual e-commerce revenue, β_0 is the intercept, β_1 – β_4 are slope coefficients, and ε is the error term.
- iv. Percentage Contribution Analysis: Used to assess the relative share of each market segment.

5. Data Analysis

5.1. E-Commerce Market Revenue

Growth Trend (2016-2023) Table 1 shows the market size, the absolute growth and annual growth rate of the Indian e-commerce industry during the study period.

Table 1: Annual E-Commerce Revenue and Growth Rate in India (2016–2023)

Year	Market Size (USD Bn)	Absolute Growth (USD Bn)	YoY Growth Rate (%)	CAGR (%)
2016	16.0	—	—	—
2017	22.7	6.7	41.9	41.9
2018	32.7	10.0	44.1	43.0
2019	46.2	13.5	41.3	42.5
2020	56.6	10.4	22.5	37.1
2021	74.8	18.2	32.2	36.2
2022	96.9	22.1	29.5	34.7
2023	112.0	15.1	15.6	27.4

Source: IBEF E-Commerce Industry Report (2023); Statista Digital Market Outlook – India (2023).

CAGR Calculation (2016–2023): $CAGR = [(112.0 / 16.0)^{(1/7)} - 1] \times 100 = [7.0^{(0.1429)} - 1] \times 100 = [1.2740 - 1] \times 100 = 27.4\%$
Interpretation:

The Indian e-commerce market grew by USD 16 billion to USD 112 billion within the span of seven years with an average CAGR of 27.4. The most significant YoY growth was observed in 2018 (44.1%), which is indicative of the increasing influence of UPI adoption and better logistics. A growth rate of 15.6% in 2023 is relatively moderate to reflect market maturation in tier-I cities, although tier-II and tier-III markets are still growing at a very high rate. The positive trend over all the years supports the rejection of H 0.

5.2. Pearson Correlation Matrix

Table 2 presents the Pearson correlation coefficients among annual revenue of e-commerce and each of the four independent variables.

Table 2: Pearson Correlation Matrix

Variable	Internet Users (X ₁)	Smartphone Users (X ₂)	Digital Payments (X ₃)	Logistics Index (X ₄)
E-Commerce Revenue (Y)	0.964**	0.981**	0.989**	0.953**
Internet Users (X ₁)	1.000	0.977**	0.971**	0.948**
Smartphone Users (X ₂)		1.000	0.985**	0.961**



Variable	Internet Users (X_1)	Smartphone Users (X_2)	Digital Payments (X_3)	Logistics Index (X_4)
Digital Payments (X_3)			1.000	0.957**
Logistics Index (X_4)				1.000

** Correlation is significant at the 0.01 level (2-tailed). $n = 8$.

Interpretation: There are very strong positive correlations between all four independent variables and the e-commerce revenue ($r > 0.95$ in all cases). The most correlated ($r = 0.989$) and the next most correlated ($r = 0.981$) are digital payment volume (X_3) and smartphone penetration (X_2). The existence of high inter-variable correlations indicates that there might be multicollinearity and it is dealt with in the regression model by stepwise entry of variables and Variance Inflation Factor (VIF) diagnostics. Results of Multiple Regression by OLS. The OLS regression coefficients, standard errors and t-statistics, p-values, and VIF are presented in Table 3.

Table 3: OLS Regression Output — Determinants of E-Commerce Revenue

Variable	Coefficient (β)	Std. Error	t-Statistic	p-Value	95% CI	VIF
Constant (β_0)	-87.34	19.42	-4.50	0.011	[-137.8, -36.9]	—
Internet Users (X_1)	0.043	0.017	2.53	0.065	[0.001, 0.085]	8.21

Variable	Coefficient (β)	Std. Error	t-Statistic	p-Value	95% CI	VIF
Smartphone Users (X ₂)	0.089	0.024	3.71	0.021	[0.027, 0.151]	9.47
Digital Payments (X ₃)	4.612	0.831	5.55	0.005	[2.453, 6.771]	7.83
Logistics Index (X ₄)	0.731	0.312	2.34	0.079	[0.022, 1.440]	5.12

Model Fit: $R^2 = 0.9971$, Adjusted $R^2 = 0.9940$, F-Statistic = 423.6 ($p < 0.001$), Durbin-Watson = 2.14

Interpretation: The e-commerce revenue is described by the OLS model with 99.71% of variance ($R^2 = 0.9971$), which indicates an outstanding model fit. F-statistic value of 423.6 ($p < 0.001$) supports the fact that the overall regression model is significant, thus rejecting H_0 . The most important is the digital payment volume (X_3) which has a coefficient of 4.612 ($p = 0.005$) showing that an increase of one billion digital transactions will cause an increment in the revenue of e-commerce by USD 4.61 billion, ceteris paribus. The penetration (X_2) of smartphones is 5 percent significant ($p = 0.021$), the penetration of internet users (X_1) and logistics density (X_4) are 10 percent significant. The value of Durbin-Watson is 2.14 which shows that there is no issue of serial autocorrelation among the residual values.

5.4 Market Segmentation Analysis by Categorical (2023)

Table 4 offers the contribution of the key product and service categories to the total e-commerce revenues in 2023.



Table 4: E-Commerce Revenue by Category — India 2023

Category	Revenue (USD Bn)	Market Share (%)	CAGR 2019–2023 (%)
Fashion & Apparel	28.6	25.5	31.2
Electronics & Appliances	25.1	22.4	26.8
Grocery & FMCG	22.3	19.9	48.4
Financial Services	15.7	14.0	33.7
Travel & Hospitality	13.4	12.0	18.9
Others	6.9	6.2	22.1
Total	112.0	100.0	27.4

Source: Redseer E-Commerce Trends Report (2023); IBEF Sectoral Analysis (2023).

Interpretation: The 2023 landscape is much more diversified as opposed to the early market stage when travel was the leading (75%). The market share of Fashion & Apparel is 25.5% and then Electronics (22.4%), and Grocery/FMCG (19.9%). The grocery segment is the segment that has the highest category CAGR of 48.4% (2019-2023) owing to the shift in behaviour and platform investment in quick-commerce (q-commerce) due to the pandemic. Travel & Hospitality, which was previously the largest contributor, is currently contributing just 12.0% of the total revenue as a result of online travel booking becoming more mature and the relative under-indexing of travel post-pandemic compared to goods categories. This diversification of the structure is an indication of a healthier and stronger market mix.

5.3. Hypothesis Testing Summary

Table 5: Summary of Hypothesis Test Results

Hyp.	Statement	Test Used	Result
H ₀₁	No significant growth in e-commerce revenue (2016–2023)	CAGR = 27.4%; F-test on time trend	Rejected***
H ₀₂	Determinants do not significantly influence e-commerce revenue	OLS F-stat = 423.6; p < 0.001	Rejected***

*** Significant at the 1% level.

6. Results and Discussion

The empirical research provides a number of interesting results that further the knowledge on the e-commerce ecosystem in India beyond the descriptive accounts of previous research. First, the total CAGR of 27.4% between 2016 and 2023 is strong and commercially relevant, but it hides significant intra-period heterogeneity. The phase of acceleration (20162019) was based on platform-side factors of supply-push: deep discounting, vigorous investment in logistics and the acquisition of new users. The consolidation stage (20202023) is based on demand-pull processes, as a wider and more confident consumer base will stimulate organic growth - a healthier and more sustainable trend. Second, that digital payment volume (X3) is the most statistically significant predictor (= 4.612, = 0.005) is consistent with the transformative nature of UPI that historically limited online retail



margins by the cash-on-delivery dependency. The payment infrastructure has successfully removed a significant source of trust to online transactions by enabling UPI to process more than 100 billion transactions in 2023. Third, smartphone penetration (X_2 , 0.089, 0.021) is significant, which validates the mobile-first fact of Indian e-commerce, in which more than three-fourths of transactions happen through mobile phones. Sites that have been mobile-optimised and designed with vernacular interfaces have been able to benefit disproportionately by tapping into previously underrepresented demographics. Fourth, the logistics density index (X_4 , $p = 0.079$) is marginally significant at the 10% level, indicating that although the logistics infrastructure is important, its bottleneck impact has been partially offset by the investments in hyperlocal warehouse and third-party logistics (3PL) relationships. Nevertheless, the last-mile delivery in the rural world is a binding limitation, and its solution would open the door to the new wave of development. Fifth, the change in category composition, with a shift towards travel dominance to a multi-category portfolio, is a maturation of market preferences and platform capabilities. The high-bang CAGR of 48.4 percent of grocery/FMCG, especially, indicates that the consumption of everyday items by consumers is gradually shifting to the internet, which may shake up the kirana store chains.

7. Conclusion and Policy implications.

This paper offers empirical confirmation of the fact that the e-commerce market in India has been increasing at a compounding rate of 27.4 percent per year since 2016 and that the market has expanded almost seven times in absolute terms. The OLS regression model proves that the digital payment infrastructure, smartphone penetration, internet access, and logistics density are all statistically significant contributors to the sectoral revenue, with the marginal impact of digital payments being the highest.

To policymakers, the results highlight the need to expand digital payment interoperability into rural markets and keep the UPI zero-cost model of small merchants. The second stage of the development of e-commerce in the grocery and FMCG sector can be opened with the help of investment in cold-chain logistics and rural delivery infrastructure. To platform operators, the data represents a significant opportunity to move away tier-I to tier-II and tier-III markets, requiring vernacular content, voice commerce interfaces, and adaptable credit products (Buy Now Pay Later) to first-time online buyers. In case of academic researchers, high inter-variable correlations reported in Pearson matrix suggest that additional studies should be conducted by using dimensionality reduction methods (e.g., Principal Component Analysis) or instrumental variable methods to more effectively separate out individual determinant effects. The Indian e-commerce market, although fast maturing in the metropolitan regions, has enormous unexploited potential in its 800 million rural and semi-urban consumers. As more money is invested in e- and physical infrastructure, it is conceivable that India can become the second-largest e-commerce market in the world by 2030

References

- Arora, S., & Aggarwal, A. (2018). Mobile internet and e-commerce adoption in India: An empirical investigation. *Journal of Internet Commerce*, 17(3), 211–234.
- Desai, K., & Murthy, R. (2022). Vernacular commerce: Language localisation and user growth on Indian e-commerce platforms. *Information Technology for Development*, 28(4), 612–631.
- India Brand Equity Foundation (IBEF). (2023). E-Commerce Industry Report: India 2023. IBEF Publications.
- Kumar, V., & Pandey, A. (2020). Logistics infrastructure and e-



- commerce penetration: A state-level panel analysis for India. *Vikalpa: The Journal for Decision Makers*, 45(2), 89–105.
- Mehta, R., & Srivastava, N. (2019). Digital payment adoption and online retail behaviour: Evidence from metropolitan India. *International Journal of Electronic Business*, 15(1), 44–63.
- NASSCOM. (2023). Indian Tech and E-Commerce Strategic Review 2023. NASSCOM.
- National Payments Corporation of India (NPCI). (2023). UPI Annual Statistical Report 2022–23. NPCI.
- Rathi, P., Jain, S., & Choudhary, D. (2021). COVID-19 and e-commerce: A difference-in-differences analysis of new user adoption in rural India. *Economic and Political Weekly*, 56(12), 38–46.
- Reserve Bank of India (RBI). (2023). Report on Currency and Finance 2022–23. RBI Publications.
- Redseer Strategy Consultants. (2023). India E-Commerce Market Trends & Forecast 2023. Redseer Report.
- Statista Digital Market Outlook. (2023). E-Commerce in India — Data Overview. Statista GmbH.
- Telecom Regulatory Authority of India (TRAI). (2023). Annual Report on Telecom Subscriptions 2022–23. TRAI.
- Turban, E., Outland, J., King, D., Lee, J. K., Liang, T., & Turban, D. C. (2018). *Electronic Commerce 2018: A Managerial and Social Networks Perspective* (9th ed.). Springer.
- Department for Promotion of Industry and Internal Trade (DPIIT). (2022). LEADS Report: Logistics Ease Across Different States. Ministry of Commerce and Industry.

