Indian Firms In International Trade: A Firm Level Analysis

Group-17: Anjali (210147), Harsh (210415), Mihir (210607), Pulak (210791), Yatendra (211202)
ECO412 (International Economics And Finance): Research Paper
Submitted To Professor Somesh K. Mathur

Abstract

This paper investigates the role of Indian firms in international trade by integrating firm-level heterogeneity into traditional trade models, challenging country- and industry-level approaches that treat firms as homogeneous. We demonstrate how firm-specific characteristics, particularly productivity, influence trade patterns through adjustments in both the extensive (number of exporting firms) and intensive (volume exported per firm) margins. Using firm-level data, we explore how proximity to infrastructure, like the Golden Quadrilateral, affects trade flows. This research is highly relevant for policymakers, as it provides insights into how infrastructure investments and targeted interventions can improve trade participation and productivity, particularly for firms in less connected regions, ultimately contributing to more inclusive economic growth.

Introduction

In this study, we examine the role of Indian firms in international trade by challenging traditional country- and industry-level trade theories through the integration of firm-level heterogeneity into the analysis. Conventional models often assume that firms within the same industry are homogeneous, failing to account for the substantial differences across firms in terms of productivity, size, and export behavior. As a result, they are unable to explain why only a small fraction of firms participate in exporting activities. Heterogeneous-firm models, in contrast, provide a more nuanced understanding of how variations in firm productivity shape trade patterns and drive resource allocation, ultimately leading to improvements in aggregate productivity. This paper aims to demonstrate how firm-level heterogeneity transforms our understanding of international trade dynamics, particularly by highlighting the central role of the extensive margin—the number of firms that engage in trade—in shaping overall trade patterns. Additionally, we look to emphasize on the importance of transaction-level data, which offers new insights into the adjustments that occur along both the intensive margin (the volume exported by each firm) and the extensive margin. Finally, we analyze how the distance of firms from key infrastructure, such as the Golden Quadrilateral Road Network, affects their trade flows, contributing to the broader discourse on the relationship between infrastructure proximity and export performance.

Literature Review

In recent years, the study of international trade has shifted focus from traditional models based on country-level analyses to firm-level heterogeneity, revolutionizing our understanding of trade dynamics. This shift is prominently illustrated in the works of Bernard et al. (2007) and Chaney (2008), which introduce significant innovations to trade theory by emphasizing firm-level characteristics and the roles of both intensive and extensive margins of trade.

Bernard et al. 's (2007) seminal paper, “Firms in International Trade”, provides a comprehensive empirical analysis of firm participation in global trade. Their study highlights that only a small fraction of firms engage in exporting, and these firms tend to be systematically larger, more productive, capital-intensive, and offer higher wages than those solely serving domestic markets. This self-selection process is driven by the significant fixed costs associated with exporting, resulting in only the most productive firms entering international markets. Bernard et al. argue that international trade leads to the reallocation of resources within industries, wherein high-productivity firms expand, and less productive firms exit, thereby driving aggregate welfare and productivity gains at the macroeconomic level. This reallocation mechanism introduces heterogeneous-firm models, a key development that departs from traditional trade models which treated firms as homogeneous units. The authors underscore the critical role of the extensive margin (number of firms engaged in exporting) in shaping trade flows, a concept previously un derexplored in classical trade theories.

On the theoretical side, Chaney’s (2008) work, “Distorted Gravity: The Intensive and Extensive Margins of International Trade”, introduces firm heterogeneity into the widely-used gravity model of trade, building on earlier models by Krugman (1980) and Melitz (2003). Chaney’s model provides a more nuanced understanding of how trade flows are shaped not only by economic size and distance but also by firm-level productivity distributions. A key contribution of this work is the differentiation between the intensive margin (the volume of exports per firm) and the extensive margin, which responds differently to changes in trade costs. Higher trade costs decrease the number of firms that export (extensive margin), while increasing the volume exported by firms already engaged in trade (intensive margin). Importantly, Chaney finds that the elasticity of substitution plays a pivotal role: higher elasticity amplifies the impact on the intensive margin but dampens the response of the extensive margin. His model, based on a Pareto distribution of firm productivity, predicts that the elasticity of trade with respect to trade barriers is larger when accounting for firm heterogeneity, thereby offering more realistic insights into global trade patterns.

Bernard and Chaney’s models have been further validated by empirical studies such as Eaton, Kortum, and Kramarz (2007), which examined French data to confirm the significance of firm size and productivity in determining both the number of foreign markets a firm enters and the intensity of its exports. These findings reinforce the notion that the extensive margin of trade is particularly sensitive to trade barriers, especially in low-competition sectors.

When contextualized in the Indian setting, the insights from these works take on additional layers of complexity. Mukim (2011) provides evidence that, for Indian firms, productivity remains the key determinant of export participation, mirroring trends in developed economies. However, Indian firms exhibit a higher sensitivity to the fixed costs of exporting due to regulatory hurdles and logistical challenges, highlighting the unique constraints faced by firms in emerging markets. Nataraj (2011) further supports the assertion that larger and more productive firms disproportionately benefit from trade liberalization, aligning with Chaney's predictions regarding firm productivity as a critical factor in determining market entry and export volumes.

Despite the robust findings of Chaney's (2008) model, its application in the Indian context remains limited. Traditional gravity models, which focus on country-level factors such as economic size, distance, and trade barriers, still dominate the literature on Indian trade. These models, while useful, overlook the significant role of firm-level heterogeneity, suggesting a need for further research that incorporates firm-specific variables into the analysis of India's trade patterns. Studies examining the relationship between firm characteristics, market entry costs, and trade volumes could yield more accurate predictions of how Indian firms might respond to changes in global trade policy.

In conclusion, the shift towards firm-level analysis in international trade, spearheaded by Bernard et al. (2007) and Chaney (2008), has provided valuable insights into how trade barriers, firm productivity, and heterogeneity shape global trade patterns. While their models offer a strong foundation, further research, particularly in the Indian context, is necessary to fully understand the dynamics at play in emerging markets. By integrating firm-level data into traditional gravity models, researchers can uncover deeper insights into the mechanisms driving trade flows and firm performance, offering more targeted recommendations for trade policy and market regulation.

Research Gap

While models of firm-level heterogeneity and trade margins, such as those developed by Bernard et al. (2007) and Chaney (2008), have been extensively applied to developed economies like the U.S. and France, their application in emerging markets like India remains underexplored. This gap is particularly significant given the substantial differences in market structures, trade costs, and firm characteristics between developed and developing countries. For instance, developed economies often benefit from more sophisticated infrastructure, fewer regulatory barriers, and greater institutional support, all of which reduce the costs associated with exporting. In contrast, firms in developing countries like India face unique constraints that are not fully captured by models tailored to advanced economies.

India's market presents several distinctive challenges. Trade liberalization following the 1991 reforms led to significant growth in the extensive margin, allowing more firms to enter export markets and diversify their product lines. However, high fixed costs associated with market entry—such as regulatory barriers, infrastructural inadequacies, and complex tariff structures—continue to constrain export growth, particularly in sectors characterized by fierce competition and low product differentiation. These challenges suggest that existing models may need to be adapted to account for the specific trade barriers faced by firms in emerging markets.

Furthermore, while studies on Indian trade, such as those by Mukim (2011) and Nataraj (2011), have confirmed that larger, more productive firms benefit disproportionately from trade liberalization, there remains a lack of in-depth analysis on how firm heterogeneity influences the extensive and intensive margins of trade in India. The interaction between firm size, productivity, and export performance, particularly in relation to infrastructure development, has not been comprehensively examined. This research gap is even more evident when considering the role of key infrastructure projects like the Golden Quadrilateral (GQ) road system in shaping firm performance. Incorporating the geographical distance of firms from major infrastructure such as the GQ into the analysis of firm performance provides an additional layer of complexity. It is crucial to understand how proximity to such infrastructure influences a firm's ability to overcome trade barriers and improve export performance. This is especially relevant in India, where infrastructural gaps remain a major constraint to trade growth. The unique economic landscape of India presents an ideal context to test whether the predictions of models developed for advanced economies hold true in a developing country setting. Exploring these dimensions can offer critical insights into trade policy design and the role of infrastructure in facilitating trade expansion in emerging markets.

In summary, while existing models have provided valuable insights into firm-level trade dynamics in developed economies, their applicability to emerging markets like India has not been sufficiently explored. This creates an opportunity for research that adapts and extends these models to account for the specific challenges and opportunities presented by India's unique trade environment, including infrastructural factors such as proximity to the Golden Quadrilateral road system.

Objectives And Hypothesis

The primary objective of this study is to explore how firm heterogeneity influences international trade in the Indian context, building on models proposed by Bernard et al. (2007) and Chaney (2008). By applying these models to India, we aim to assess how factors like firm size, productivity, and proximity to infrastructure shape trade dynamics in an emerging market.

Objectives:

- To analyze firm-level heterogeneity in India's trade patterns, particularly focusing on the differences between exporters and non-exporters.
- To evaluate the role of proximity to key infrastructure projects, that is, the Golden Quadrilateral, EDFC, WDFC in improving firm performance and overcoming trade barriers.
- To investigate how Distance and Total Factor Productivity (TFP) of firms impacts trade flows in India, particularly in sectors with lower product differentiation and higher competition.
- To assess the extensive and intensive margins of trade in India, focusing on how firm entry and trade volumes are influenced by factors like fixed costs and market barriers.

Hypotheses:

- <u>Hypothesis-1</u>: Only a small fraction of Indian firms engage in international trade, and a limited number of those firms dominate the trade volume. This hypothesis seeks to confirm whether the patterns observed in developed economies, where a few firms account for a large share of exports, hold true in India.

- Hypothesis-2: Even within the same industry, exporter firms in India are larger, more productive, more skill- and capital-intensive, and tend to pay higher wages than non-exporters. This will test whether the export premia observed in other countries are present in India, suggesting that exporting firms tend to be more advanced and competitive.
- Hypothesis-3: The self-selection hypothesis posits that Indian firms engaging in exporting already possess higher productivity levels before they enter export markets. This implies that firms need to surpass a productivity threshold to cover the fixed costs associated with exporting.
- Hypothesis-4: Adjustments along the extensive margin help explain the "gravity model" of trade in India. Specifically, greater distances between trading partners dampen trade flows, and firms that trade fewer products or with fewer destinations are more sensitive to distance and other trade costs.
- Hypothesis-5: Proximity to major infrastructure projects, such as the Golden Quadrilateral Road Network, Delhi-Meerut Expressway, EDFC, and WDFC, significantly improves firm performance. This hypothesis will examine how access to well-developed transport infrastructure influences firms' ability to engage in trade (international trade - exports), particularly by reducing trade costs and logistical obstacles.

By testing these hypotheses, the study aims to provide deeper insights into how international trade operates in the Indian context, offering policy-relevant conclusions about the role of firm heterogeneity and infrastructure in shaping trade outcomes.

Model And Methodology

Here we explain the methodology followed by us to test each hypothesis. Later in the Results section we show the results of our exercises and the running STATA/Python codes.

Hypothesis I: Firm Exporting Is Relatively Rare

In the context of international trade, we propose that engaging in exporting activities is a relatively rare event for firms:

1. To conduct this test, we begin by summarizing the proportion of manufacturing firms within each industry category that engage in exporting. These categories are defined using the NIC codes, which classify industries into different sectors, as illustrated in the table below.
2. We investigate whether firm exporting is relatively uncommon, particularly across various manufacturing sectors. To focus on active participants in international trade, we exclude data from non-exporting firms, ensuring that the mean export percentage accurately reflects firms engaged in exporting.
3. For each industry category, we calculate the mean exports as a percentage of total shipments.
4. While traditional trade theory explains net imports and exports at the industry level, it does not address firm-level decisions to export or how those decisions interact with comparative advantage.

This exercise was done in excel, and following is an image showcasing our calculations.

d2/Current Total Annual Sales	n3//Last Year Total Annual Sales	d3a	d3b	d3c	Export Shipments	Total Exports
50720000000	45280000000	35	25	40	32968000000	65
86627000000	76323000000	70	0	30	25988100000	30
60000000000	0	60	0	40	24000000000	40
90000000000	86800000000	75	0	25	22500000000	25
50000000000	40000000000	70	0	30	15000000000	30
15000000000	13600000000	5	0	95	14250000000	95
40455030000	42576550000	65	0	35	14159260500	35
35000000000	32000000000	60	0	40	14000000000	40
15000000000	20000000000	15	0	85	12750000000	85
12000000000	12000000000	0	0	100	12000000000	100
20000000000	0	40	0	60	12000000000	60
12407850000	16496485000	10	0	90	11167065000	90
25000000000	18000000000	60	0	40	10000000000	40
9000000000	7500000000	0	0	100	9000000000	100
8000000000	6800000000	0	0	100	8000000000	100
10830000000	8500000000	30	20	50	7581000000	70
7000000000	3000000000	0	0	100	7000000000	100
8000000000	0	30	0	70	5600000000	70
6000000000	0	10	0	90	5400000000	90
20060000000	0	75	0	25	5015000000	25
9090000000	0	45	0	55	4999500000	55
9090000000	0	45	0	55	4999500000	55
8102500000	7211100000	40	0	60	4861500000	60
5000000000	4600000000	4	0	96	4800000000	96
6000000000	6250000000	20	0	80	4800000000	80
8500000000	7500000000	45	0	55	4675000000	55
5000000000	4500000000	10	0	90	4500000000	90
45000000000	48000000000	90	0	10	4500000000	10
4951800000	5029100000	10	0	90	4456620000	90
8000000000	7000000000	45	25	30	4400000000	55

Total Shipments	1117888419417
Top 15% Shipments	872173470321
Account For	78%

Hypothesis II: Exporter Firms Differ From Importer Firms

1. Exporters also tend to be more capital-intensive, meaning they invest more in machinery and equipment, which allows them to produce higher-quality goods. The ability to pay higher wages reflects the higher productivity of exporting firms, which can afford to attract and retain more skilled workers.
2. We test this hypothesis using our dataset of Indian firms, emphasizing the role of trade liberalization in enhancing productivity via resource reallocation across firms.
3. We calculate the export premia of exporting firms by comparing the mentioned characteristics. Estimation of Export Premia is done by regressing each characteristic over an exporter dummy:

$$Y_i = \beta_0 + \beta_1 \text{Export}_i + \epsilon_i$$

By incorporating firm size and industry fixed effects into the analysis, we ensure that the estimated impact of exporting on key variables like productivity, wages, and capital intensity is not driven by these other factors:

$$Y_i = \beta_0 + \beta_i \text{Export}_i + \gamma_j \text{Industry}_j + \epsilon_i$$

$$Y_i = \beta_0 + \beta_i \text{Export}_i + \Sigma\gamma_j \text{Industry}_j + \gamma_e \text{Log(employment)} + \epsilon_i$$

4. Since the dependent variable data are in logarithms, we can simply interpret the estimated coefficients as percentages while drawing conclusions. The variables used on the right-hand side of our analysis are listed in the table below:

5. Following is the **Python code for testing this hypothesis. It runs OLS following the mentioned three equations for each of the variables:**

```
import pandas as pd
import statsmodels.api as sm
import numpy as np
import argparse

	def parse_arguments():
		parser = argparse.ArgumentParser(description='Run regression analysis with various
	specifications')
	parser.add_argument('--file', type=str, default='classfn_4064_final.csv',
			help='Path to the CSV file (default: classfn_4064.csv)')
	parser.add_argument('--industry-dummies', action='store_true',
			help='Include industry dummies in regression')
	parser.add_argument('--employment-control', action='store_true',
			help='Include log of employment as control variable')
	parser.add_argument('--output', type=str, default='regression_results.csv',
			help='Path to save regression results (default: regression_results.csv)')
	return parser.parse_args()

	def run_regression(df, columns, include_industry_dummies=False, include_employment=False):
	results_summary = {}

	# If employment is included as control, create log employment variable
	if include_employment:
		df['log_employment'] =
	np.log(df['Current_Total_Employment'].replace({0:np.nan}).fillna(1e-10))
	# Remove employment from dependent variables if it's being used as control
	columns = [col for col in columns if col != "Current_Total_Employment"]

	for col in columns:
	print(f"\nProcessing column: {col}")

	try:
		# Convert column to numeric and handle missing values
		df[col] = pd.to_numeric(df[col], errors='coerce')

		# Add small constant before taking log to handle zeros
		df[f"log_{col}"] = np.log(df[col].replace({0: np.nan}).fillna(1e-10))

		# Dependent variable
		y = df[f"log_{col}"].astype(float)  # Ensure float type

		# Initialize X with Export_Dummy
		if "Export_Dummy" in df.columns:
			df["Export_Dummy"] = pd.to_numeric(df["Export_Dummy"],
	errors='coerce').fillna(0).astype(float)
		X = df[["Export_Dummy"]]
		else:
		raise ValueError("Export_Dummy column is missing.")

		# Add employment control if requested
		if include_employment:
		X = pd.concat([X, df[['log_employment']]], axis=1)

		# Add industry dummies if requested
		if include_industry_dummies:
		if "NIC Classification Code" in df.columns:
```

```
            df["NIC Classification Code"] = df["NIC Classification Code"].astype(str)
            industry_dummies = pd.get_dummies(df["NIC Classification Code"],
                                              prefix="Industry",
                                              drop_first=True)
            X = pd.concat([X, industry_dummies], axis=1)
        else:
            raise ValueError("NIC Classification Code column is missing.")

        # Add intercept
        X = sm.add_constant(X)

        # Convert all X columns to float
        X = X.astype(float)

        # Drop rows with NaN values
        valid_mask = ~(X.isna().any(axis=1) | y.isna())
        X = X[valid_mask]
        y = y[valid_mask]

        print(f"Number of valid observations: {len(y)}")
        print(f"Number of independent variables: {X.shape[1]}")

        # Verify data types before regression
        print(f"X dtypes:\n{X.dtypes}")
        print(f"y dtype: {y.dtype}")

        # Check for any remaining non-finite values
        print(f"Non-finite values in X: {np.sum(~np.isfinite(X.values))}")
        print(f"Non-finite values in y: {np.sum(~np.isfinite(y.values))}")

        # Perform regression
        model = sm.OLS(y, X)
        results = model.fit()

        # Store basic results
        result_dict = {
        "Dependent_Variable": col,
        "Constant_(Intercept)": results.params['const'],
        "Exporter_Dummy_Coefficient": results.params.get('Export_Dummy', np.nan),
        "Exporter_Dummy_P_Value": results.pvalues.get('Export_Dummy', np.nan),
        "R_squared": results.rsquared,
        "Adjusted_R_squared": results.rsquared_adj,
        "Number_of_observations": len(y),
        "Number_of_variables": X.shape[1],
        "F_statistic": results.fvalue,
        "F_pvalue": results.f_pvalue
        }

        # Add employment coefficient if included
        if include_employment:
            result_dict["Log_Employment_Coefficient"] =
results.params.get('log_employment', np.nan)
        result_dict["Log_Employment_P_Value"] = results.pvalues.get('log_employment', np.nan)

        # Add industry dummy coefficients if included
        if include_industry_dummies:
        industry_cols = [col for col in X.columns if col.startswith('Industry_')]
```

```
        for ind_col in industry_cols:
            result_dict[f"{ind_col}_Coefficient"] = results.params.get(ind_col, np.nan)
            result_dict[f"{ind_col}_P_Value"] = results.pvalues.get(ind_col, np.nan)

        results_summary[col] = result_dict

    except Exception as e:
        print(f"Error in regression for {col}: {str(e)}")
        continue

return results_summary

def save_results(results_summary, output_file):
results_df = pd.DataFrame.from_dict(results_summary, orient='index')
results_df.to_csv(output_file)
print(f"\nResults saved to {output_file}")

def main():
args = parse_arguments()

# List of columns to perform regression on
columns = [
"Current_Total_Employment",
"Current_Total_Annual_Sales",
"Value_Added_per_Worker",
"tfp_lp",
"Wage per Worker",
"Capital_per_Worker",
"Skill_Per_Worker",
"min_distance_to_GQ",
"distance_to_Delhi_Meerut",
"distance_to_WDFC",
"distance_to_EDFC"
]

print(f"Loading data from {args.file}")
try:
df = pd.read_csv(args.file)
except FileNotFoundError:
print(f"Error: Could not find file '{args.file}'")
return
except Exception as e:
print(f"Error loading file: {str(e)}")
return

print(f"Running regressions with specifications:")
print(f"- Industry dummies: {'Yes' if args.industry_dummies else 'No'}")
print(f"- Log employment control: {'Yes' if args.employment_control else 'No'}")

results = run_regression(df, columns,
                    include_industry_dummies=args.industry_dummies,
                    include_employment=args.employment_control)

# Display results
for col, summary in results.items():
print(f"\nRegression Results for {col}:")
print(f"Constant (Intercept): {summary['Constant_(Intercept)']:.4f}")
```

```
    print(f"Exporter Dummy Coefficient: {summary['Exporter_Dummy_Coefficient']:.4f}")
    print(f"Exporter Dummy P-Value: {summary['Exporter_Dummy_P_Value']:.4f}")
    if args.employment_control:
        print(f"Log Employment Coefficient: {summary['Log_Employment_Coefficient']:.4f}")
        print(f"Log Employment P-Value: {summary['Log_Employment_P_Value']:.4f}")
    print(f"R-squared: {summary['R_squared']:.4f}")
    print(f"Adjusted R-squared: {summary['Adjusted_R_squared']:.4f}")
    print(f"Number of observations: {summary['Number_of_observations']}")
    print(f"F-statistic: {summary['F_statistic']:.4f}")
    print(f"F-statistic p-value: {summary['F_pvalue']:.4f}")
    print("=" * 80)

    # Save results to CSV
    save_results(results, args.output)

if __name__ == "__main__":
    main()
```

One can run the Python code as follows:

a. Run `python lm.py` command to get results for regression results for each variable without any fixed effects into account
b. Run `python lm.py --industry-dummies` command to get results for regression results for each variable with intra-industry fixed effects
c. Run `python lm.py --industry-dummies --employment-control` command to get results for regression results for each variable with control for intra-industry and firm-size fixed effects. Total employment has been used as a proxy for firm-size.

Calculation of TFP:

We also calculate TFP as one of the factors, for each of the firms. We also estimate the productivity threshold by analyzing Total Factor Productivity (TFP) data.

TFP measures the efficiency of all inputs used in production. It captures the portion of output not explained by traditional inputs like labor and capital.

Levinsohn And Petrin (2003) propose using a commonly observable variable (intermediate input) to control for unobserved productivity. This method has advantages over plain OLS method:

a. Addresses the issue of endogeneity in production function estimation.
b. Utilizes intermediate inputs such as materials & energy as a proxy for unobservable productivity shocks.

With Levinsohn & Petrin, after the estimation procedure we get input elasticities of the inputs fed. Once we have the input elasticities, we can use the following formulation to get TFP (we assume Cobb-Douglas production function):

$$TFP = \frac{Output}{Labour^{\beta} \cdot Capital^{\gamma} \cdot Intermediate^{\delta}}$$

We ran Levinsohn & Petrin on our dataset using STATA codes using the ***prodtest*** function which is readily available. Following is the **complete STATA code used for this exercise:**

```
* Load the cleaned CSV file
import delimited "C:\\classfn_4064_final.csv", clear

* Create separate observations for current and previous years
preserve
gen time_var = 1  // Current year
tempfile current
save `current'

* Restore and adjust for previous year data without redefining time_var
restore
replace current_total_annual_sales = last_year_total_annual_sales
replace current_total_employment = previous_total_employment
replace inter_in = prev_inter_in
replace capital = previouscapital
gen time_var = 0   // Set time_var for previous year only

* Append current year data to previous year data
append using `current`

* Generate log transformations
gen lnoutput = ln(current_total_annual_sales)
gen lnlabor = ln(current_total_employment)
gen lncapital = ln(capital)
gen lnintermediate = ln(inter_in)

* Set the panel structure using idstd and time_var
xtset idstd time_var

* Estimate TFP using the Levinsohn-Petrin method in prodest
prodest lnoutput, method(lp) free(lnlabor) proxy(lnintermediate) state(lncapital) poly(3)
valueadded reps(250)
```

P.T.O

This is how we classified the firms into their respective industries.

NIC Code	NIC Industry
10	Manufacture of food products
11	Manufacture of beverages
12	Manufacture of tobacco products
13	Manufacture of textiles
14	Manufacture of wearing apparel
15	Manufacture of leather and related products
16	Manufacture of wood and products of wood and cork
17	Manufacture of paper and paper products
18	Printing and reproduction of recorded media
19	Manufacture of coke and refined petroleum products
20	Manufacture of chemicals and chemical products
21	Manufacture of pharmaceuticals, medicinal chemical, and botanical products
22	Manufacture of rubber and plastics products
23	Manufacture of other non-metallic mineral products
24	Manufacture of basic metal
25	Manufacture of fabricated metal products, except machinery and equipment
26	Manufacture of computer, electronic and optical products
27	Manufacture of electrical equipment
28	Manufacture of machinery and equipment n.e.c
29	Manufacture of motor vehicles, trailers, and semi-trailers
30	Manufacture of other transport equipment
31	Manufacture of furniture
32	Other manufacturing
33	Repair and installation of machinery and equipment

Python Code for classifying the dataset is as follows:

```
import pandas as pd
df = df.read_csv('classified_classfn.csv’)

# Define the classification dictionary based on the list
classification_dict = {
       "10": "Manufacture of food products",
       "11": "Manufacture of beverages",
       "12": "Manufacture of tobacco products",
       "13": "Manufacture of textiles",
       "14": "Manufacture of wearing apparel",
       "15": "Manufacture of leather and related products",
       "16": "Manufacture of wood and products of wood and cork",
       "17": "Manufacture of paper and paper products",
       "18": "Printing and reproduction of recorded media",
       "19": "Manufacture of coke and refined petroleum products",
       "20": "Manufacture of chemicals and chemical products",
       "21": "Manufacture of pharmaceuticals, medicinal chemical, and botanical products",
       "22": "Manufacture of rubber and plastics products",
       "23": "Manufacture of other non-metallic mineral products",
       "24": "Manufacture of basic metal",
```

```
    "25": "Manufacture of fabricated metal products, except machinery and equipment",
    "26": "Manufacture of computer, electronic and optical products",
    "27": "Manufacture of electrical equipment",
    "28": "Manufacture of machinery and equipment n.e.c",
    "29": "Manufacture of motor vehicles, trailers, and semi-trailers",
    "30": "Manufacture of other transport equipment",
    "31": "Manufacture of furniture",
    "32": "Other manufacturing",
    "33": "Repair and installation of machinery and equipment"
}

# Function to classify each entry based on keywords in the activity description
def classify_activity(activity):
    activity = activity.lower()
    if "food" in activity or "bakery" in activity:
    return "10"
    elif "beverage" in activity or "brewery" in activity:
    return "11"
    elif "tobacco" in activity:
    return "12"
    elif "textile" in activity or "cotton" in activity or "garment" in activity:
    return "13"
    elif "wearing apparel" in activity or "clothing" in activity or "suit" in activity:
    return "14"
    elif "leather" in activity or "footwear" in activity:
    return "15"
    elif "wood" in activity or "timber" in activity:
    return "16"
    elif "paper" in activity or "print" in activity:
    return "17"
    elif "recorded media" in activity:
    return "18"
    elif "coke" in activity or "petroleum" in activity:
    return "19"
    elif "chemical" in activity:
    return "20"
    elif "pharmaceutical" in activity or "medicine" in activity:
    return "21"
    elif "plastic" in activity or "rubber" in activity:
    return "22"
    elif "non-metallic mineral" in activity:
    return "23"
    elif "metal" in activity and "basic" in activity:
    return "24"
    elif "fabricated metal" in activity:
    return "25"
    elif "computer" in activity or "electronic" in activity:
    return "26"
    elif "electrical equipment" in activity:
    return "27"
    elif "machinery" in activity or "equipment" in activity:
    return "28"
    elif "motor vehicle" in activity or "trailer" in activity:
    return "29"
    elif "transport equipment" in activity:
    return "30"
    elif "furniture" in activity:
```

```
        return "31"
        elif "manufacturing" in activity:
        return "32"
        elif "repair" in activity or "installation" in activity:
        return "33"
        else:
        return "Unclassified"

# Apply classification to each entry in the dataframe
df['Classification Code'] = df['Main Manufacturing Activity'].apply(classify_activity)

# Map classification code to description
df['Classification Description'] = df['Classification Code'].map(classification_dict)

# Save the classified data to a new CSV file
classified_file_path = 'classified_classfn.csv'
df.to_csv(classified_file_path, index=False)
```

Hypothesis III: Self-Selection of Exporting Firms

1. Roberts and Tybout (1997) introduced the concept of sunk entry costs for export markets, highlighting that firms must incur fixed costs upfront to enter international trade. These costs include establishing distribution networks, complying with foreign regulations, and marketing their products abroad. Only the most productive firms can generate enough profits to cover these fixed costs and remain competitive in foreign markets.

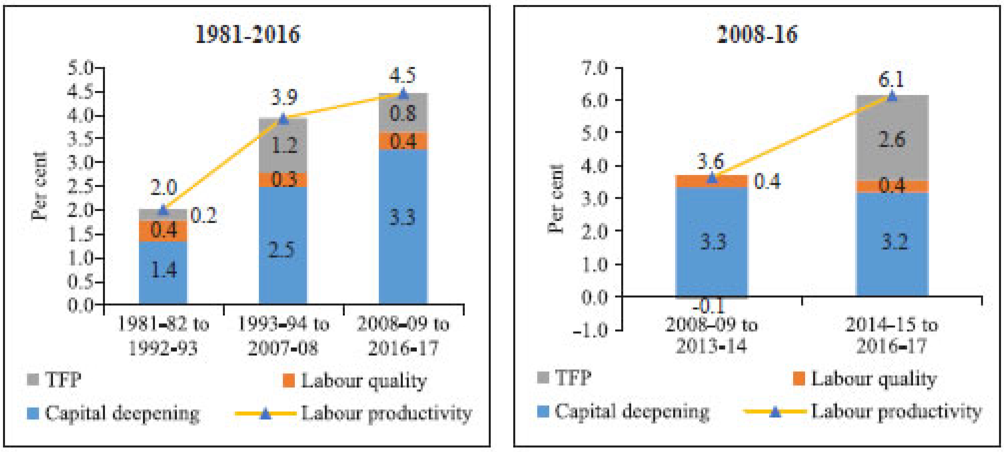
2. Further, Chaney (2008) extends traditional trade models like Krugman (1980) by incorporating firm heterogeneity—i.e., differences in productivity across firms—and the concept of fixed export costs. In this model, firms face not only variable costs (such as tariffs and transportation costs) but also fixed costs, which create a threshold productivity level that firms must reach to engage in exporting. Therefore, leaving only a subset of firms as exporters, Less productive firms are not able to generate enough profits abroad to cover the fixed cost of entering foreign markets.

The profits firm with productivity φ earns when exporting from i to j are

$$\pi_{ij}(\varphi) = \mu/\sigma Y_j[\sigma/(\sigma\text{-}1)(w_i\tau_{ij}\varphi/P_j)^{1-\sigma} - f_{ij}$$

3. We estimate this productivity threshold by analyzing Total Factor Productivity data.

Chart 1: Components-wise Labour Productivity Growth



Source: India KLEMS, RBI.

The Total Factor Productivity (TFP) values we calculated using the Levinsohn-Petrin (LP) method on our original dataset align closely with the Labor Productivity Growth values illustrated in the diagram above. According to the KLEMS dataset from RBI, theoretical TFP values for the period 2014-15 to 2016-17 were approximately 6.1. **Our calculated values were consistent with these results with sound theoretical backing**, ranging between 5.9 and 6.1. Further details are provided in the *results section of Hypothesis III.*

Hypothesis IV: Firm Heterogeneity in Gravity Models

1. In traditional trade models, such as Krugman's (1980) model, firms are assumed to be homogeneous, and trade is driven by differences between countries rather than firms. However, Chaney (2008) demonstrates that when firm heterogeneity (differences in productivity across firms) and fixed export costs are introduced, the predictions of the Krugman model are overturned, and subsequently estimated coefficient's sign changes.
2. When firms face both variable trade costs (such as tariffs or transportation costs) and fixed costs of entering a foreign market, the model shows that the distribution of firm productivity interacts with these trade barriers differently depending on the elasticity of substitution.
    a. Low Elasticity of Substitution: When the elasticity of substitution is low, consumers are less willing to switch between products, meaning that firms must be very productive to compete in foreign markets. In this scenario, only the most productive firms will export, and changes in trade barriers primarily affect the extensive margin (the number of firms exporting).
    b. High Elasticity of Substitution: When the elasticity of substitution is high, more firms can export, and the intensive margin (the volume of exports by existing firms) becomes more sensitive to changes in trade barriers. Thus, as trade barriers fall, existing exporters will expand their trade more rapidly compared to a low elasticity scenario.

3. Chaney's model introduces two margins of adjustment for trade flows:
    a. Intensive Margin: Refers to the size of trade for each existing exporting firm. When trade barriers fall, firms already engaged in trade will typically increase the volume of their exports. The extent of this increase depends on the elasticity of substitution, with higher elasticity leading to more pronounced changes in the intensive margin.
    b. Extensive Margin: Refers to the number of firms that start exporting as trade barriers decrease. With fixed export costs, not all firms will find it profitable to export. As these barriers fall, more firms will be able to cover their fixed export costs and enter foreign markets, increasing the extensive margin. In the presence of firm heterogeneity, this margin is strongly influenced by the distribution of firm productivity.

4. The impact of trade barriers (such as distance) on trade flows is dampened by the elasticity of substitution, and not magnified:
    a. A higher elasticity makes the intensive margin more sensitive to changes in trade barriers, whereas it makes the extensive margin less sensitive
    b. With assumption of Pareto distribution of firm productivity the effect on the extensive margin dominates

Chaney's proposed model is of the following form:

$$\text{Exports}_{AB} = \text{Constant} * \text{GDP}_A * \text{GDP}_B / (\text{Trade Barriers}_{AB})^{\varepsilon'(\sigma)} \qquad \text{with } \varepsilon'(\sigma) < 0$$

5. The complete equation as derived in Chaney (2008) is as follows. Total exports (f.o.b.) Xhij in sector h from country i to country j are given by:

$$X_{ij}^{\ h} = \mu_h \cdot (Y_i . Y_j) / (Y) . \ (w_i \tau_{ij} / \theta_j^h)^{-\gamma_h} . (f_{ij}^h)^{(\gamma_h / (\sigma_h - 1) - 1)}$$

where, exports are a function of country sizes (Yi and Yj), workers' productivity (wi), the bilateral trade costs, variable (τ_ij) and fixed (f_hij), and the measure of j's remoteness from the rest of the world (θ_hj).

We were not able to derive results for this hypothesis due to unavailability of fine data of countries traded with individual firms for the Indian firms.

Hypothesis V: Impact Of Infrastructure On Firm Performance

We compare the performance of firms located near major infrastructure projects (e.g., Golden Quadrilateral, Delhi-Meerut Expressway, EDFC, WDFC) with those further away. We basically check if the exporting firms are located closer to key infrastructure as compared to non-exporting firms.

We also use reverse geocoding to identify firm locations (and plot them on the map of India) and then analyze their performance relative to transportation barriers and proximity to infrastructure.

The nearest distance between each firm/city and the polyline (for e.g. GQ highway points) is calculated using the Haversine formula, which computes the great-circle distance between two points on Earth's surface.

The Haversine formula calculates the distance between two points on Earth, accounting for its curvature. The formula is:

$$d = 2R \cdot arcsin\left(\sqrt{sin^2\left(\frac{\Delta\phi}{2}\right) + cos\left(\phi_1\right) \cdot cos\left(\phi_2\right) \cdot sin^2\left(\frac{\Delta\lambda}{2}\right)}\right)$$

Where,

- d: Distance between two points
- R: Earth's radius (approximately 6371 km).
- Δϕ: Difference in latitude
- Δλ: Difference in longitude

For example the process for calculating the minimum distance of each firm from Golden Quadrilateral would be -

1. Represent GQ as a polyline of geographic coordinates. The GQ is represented by the coordinates of 19 major cities along its path. Each city is represented by its latitude and longitude.
2. Extract locations of the firms (latitude and longitude) from the dataset.
3. For each firm in the dataset, the code would calculate the Haversine distance between the firm's location and each GQ city.
4. After calculating distances to all cities, the code records the minimum distance to the nearest GQ city and identifies which city that is.

After obtaining the minimum distances of each firm from each of the 4 key infrastructure that we've considered (GQ, Delhi-Meerut, EDFC, WDFC), we regress the log of these distances on their exporter dummy variable, also later incorporating firm size and industry fixed effects into the analysis (similarly to what was done in Hypothesis 2).

$$\log(\text{min\_dist\_GQ}_i) = \beta_0 + \beta_1 \text{ExportDummy}_i + \epsilon_i$$

$$\log(\text{min\_dist\_GQ}_i) = \beta_0 + \beta_i \text{ExportDummy}_i + \gamma_j \text{Industry}_j + \epsilon_i$$

$$\log(\text{min\_dist\_GQ}_i) = \beta_0 + \beta_i \text{ExportDummy}_i + \Sigma\gamma_j \text{Industry}_j + \gamma_e \text{Log(employment)} + \epsilon_i$$

It yields a table of coefficients similar to what we obtained in the 2nd hypothesis. It is shown in the results section where it is interpreted too.

Python code for plotting the firms on map of India:

```
from matplotlib.pyplot import plt
import pandas as pd

# Load the dataset
From matplofile_path = 'classfn_4064_final.csv'
data = pd.read_csv(file_path)
# Plotting the latitude and longitude points directly without a base map
plt.figure(figsize=(10, 10))
plt.scatter(data['Longitude'], data['Latitude'], color='red', s=10, label="Locations")
# Setting latitude and longitude boundaries typical for India's geographic area
plt.xlim(68, 98)  # India's approximate longitude range
plt.ylim(6, 38)   # India's approximate latitude range
# Customizing the plot -- for plotting the firms on map
plt.title("Geographical Plot of Locations within India")
plt.xlabel("Longitude")
plt.ylabel("Latitude")
plt.legend()
plt.show()
```

Python code for calculating distance using Haversine formula

```
import pandas as pd
import numpy as np

# Load the dataset
file_path = 'classfn_4064_final.csv'
```

```
data = pd.read_csv(file_path)

# Define city coordinates
cities = {
       "Delhi": (28.6139, 77.2090),
       "Mumbai": (19.0760, 72.8777),
       "Kolkata": (22.5726, 88.3639),
       "Chennai": (13.0827, 80.2707),
       "Ahmedabad": (23.0225, 72.5714),
       "Pune": (18.5204, 73.8567),
       "Surat": (21.1702, 72.8311),
       "Vadodara": (22.3072, 73.1812),
       "Jaipur": (26.9124, 75.7873),
       "Udaipur": (24.5854, 73.7125),
       "Nagpur": (21.1458, 79.0882),
       "Varanasi": (25.3176, 82.9739),
       "Allahabad": (25.4358, 81.8463),
       "Kanpur": (26.4499, 80.3319),
       "Agra": (27.1767, 78.0081),
       "Gwalior": (26.2183, 78.1828),
       "Ranchi": (23.3441, 85.3096),
       "Bhubaneswar": (20.2961, 85.8245),
       "Vishakhapatnam": (17.6868, 83.2185)
}

# Step 1: Calculate minimum distance to GQ for each firm
firm_distances = []
nearest_cities = []
for index, row in data.iterrows():
       firm_lat = row['Latitude']
       firm_lon = row['Longitude']
       min_distance = float('inf')
       nearest_city = None

       for city, (city_lat, city_lon) in cities.items():
       R = 6371
       phi1 = np.radians(firm_lat)
       phi2 = np.radians(city_lat)
       delta_phi = np.radians(city_lat - firm_lat)
       delta_lambda = np.radians(city_lon - firm_lon)

       a = np.sin(delta_phi / 2) ** 2 + np.cos(phi1) * np.cos(phi2) * np.sin(delta_lambda / 2) ** 2
       c = 2 * np.arctan2(np.sqrt(a), np.sqrt(1 - a))
       distance = R * c

       if distance < min_distance:
              min_distance = distance
              nearest_city = city

       firm_distances.append(min_distance)
       nearest_cities.append(nearest_city)

data['min_distance_to_GQ'] = firm_distances
data['nearest_city_to_GQ'] = nearest_cities

# Step 2: Calculate distance to Delhi-Meerut Expressway
delhi_meerut_points = [
```

```
    (28.6139, 77.2090),
    (28.6270, 77.2773),
    (28.7440, 77.4995),
    (28.9845, 77.7064),
    (29.0832, 77.7109)
]

delhi_meerut_distances = []
for index, row in data.iterrows():
    firm_lat = row['Latitude']
    firm_lon = row['Longitude']
    min_distance = float('inf')

    for point_lat, point_lon in delhi_meerut_points:
    R = 6371
    phi1 = np.radians(firm_lat)
    phi2 = np.radians(point_lat)
    delta_phi = np.radians(point_lat - firm_lat)
    delta_lambda = np.radians(point_lon - firm_lon)

    a = np.sin(delta_phi / 2) ** 2 + np.cos(phi1) * np.cos(phi2) * np.sin(delta_lambda / 2) ** 2
    c = 2 * np.arctan2(np.sqrt(a), np.sqrt(1 - a))
    distance = R * c

    if distance < min_distance:
        min_distance = distance

    delhi_meerut_distances.append(min_distance)

data['distance_to_Delhi_Meerut'] = delhi_meerut_distances

# Step 3: Calculate distance to WDFC
wdfc_points = [
    (28.7041, 77.1025),
    (27.0238, 74.2179),
    (26.9124, 75.7873),
    (25.4358, 78.5685),
    (22.7196, 75.8577),
    (23.2599, 77.4126),
    (21.1702, 72.8311),
    (19.0760, 72.8777)
]

wdfc_distances = []
for index, row in data.iterrows():
    firm_lat = row['Latitude']
    firm_lon = row['Longitude']
    min_distance = float('inf')

    for point_lat, point_lon in wdfc_points:
    R = 6371
    phi1 = np.radians(firm_lat)
    phi2 = np.radians(point_lat)
    delta_phi = np.radians(point_lat - firm_lat)
    delta_lambda = np.radians(point_lon - firm_lon)

    a = np.sin(delta_phi / 2) ** 2 + np.cos(phi1) * np.cos(phi2) * np.sin(delta_lambda / 2) ** 2
```

```
        c = 2 * np.arctan2(np.sqrt(a), np.sqrt(1 - a))
        distance = R * c

        if distance < min_distance:
            min_distance = distance

        wdfc_distances.append(min_distance)

data['distance_to_WDFC'] = wdfc_distances

# Step 4: Calculate distance to EDFC
edfc_points = [
        (30.9000, 75.8573),   # Ludhiana
        (26.4499, 80.3319),   # Kanpur
        (24.9807, 84.0374)    # Sonnagar
]

edfc_distances = []
for index, row in data.iterrows():
        firm_lat = row['Latitude']
        firm_lon = row['Longitude']
        min_distance = float('inf')

        for point_lat, point_lon in edfc_points:
        R = 6371
        phi1 = np.radians(firm_lat)
        phi2 = np.radians(point_lat)
        delta_phi = np.radians(point_lat - firm_lat)
        delta_lambda = np.radians(point_lon - firm_lon)

        a = np.sin(delta_phi / 2) ** 2 + np.cos(phi1) * np.cos(phi2) * np.sin(delta_lambda / 2) ** 2
        c = 2 * np.arctan2(np.sqrt(a), np.sqrt(1 - a))
        distance = R * c

        if distance < min_distance:
            min_distance = distance

        edfc_distances.append(min_distance)

data['distance_to_EDFC'] = edfc_distances

# Save the updated DataFrame to a new CSV file
output_file_path = 'proximity.csv'
data.to_csv(output_file_path, index=False)
```

Results:

Hypothesis-I Results

The smaller set of Indian firms active in industries that are more predisposed to exporting - like those in the manufacturing that produce tradable goods - only 13.87% were exporters in 2022.
2719/18621 out of the firms are exporting firms, that is, **14.60% are exporting firms**.
The value of total shipments is INR 11,17,888.42 million.

The **top 15% exporting firms** account for the shipment value of INR 8,72,173.47 million. Thus they **account for 78% of the total shipments.**

Hypothesis-II Results

	(1)	(2)	(3)
Log Employment	0.7489*	0.7477**	
Log Shipments	1.0618**	1.0753*	0.1565**
Log Value-Added Per Worker	0.3129**	0.3276**	0.1565**
Log TFP	0.2297**	0.2411**	0.1672*
Log Wage per Worker	0.1075**	0.0958*	0.0511
Log Capital Per Worker	0.7176**	0.7229**	1.1875**
Log Skill Per Worker	0.2879**	0.2706**	0.3097*
Additional Covariates	None	Industry Fixed Effects	Industry Fixed Effects Log Employment

Firms that export look very different from non exporters along a number of dimensions. We highlight these differences by reporting Indian manufacturing exporters' "export premia" for 2022 in the above Table.

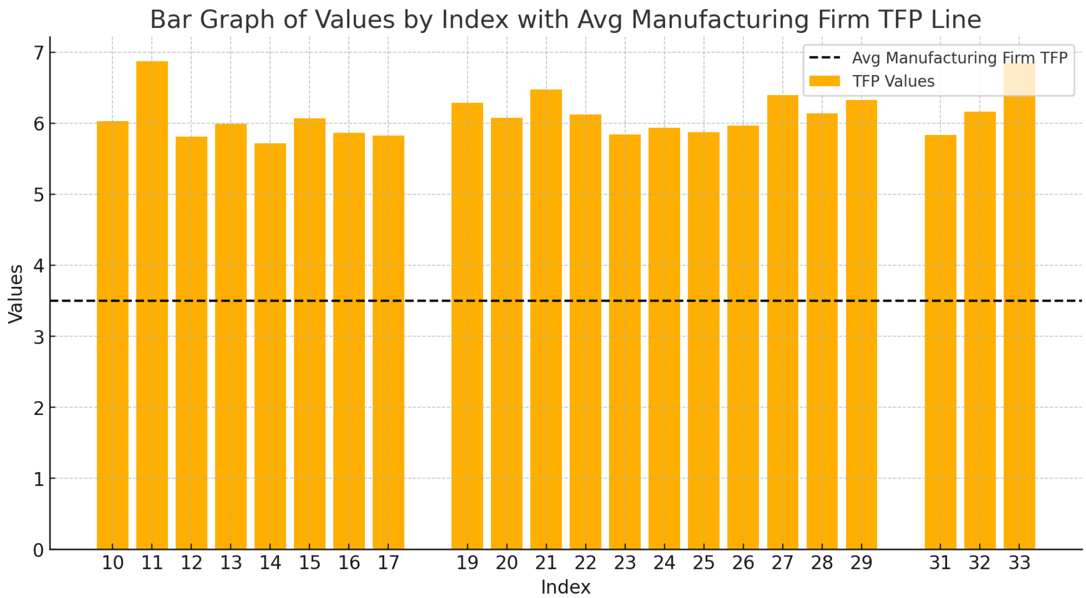
Each row of the table summarizes the average percent difference between exporters and non exporters for a particular firm characteristic. For example, the first column of the table reports the results of a series of bivariate ordinary least squares regressions. The dependent variables are employment, shipments, value-added per worker, and the other variables noted in the first column, all in logs. The explanatory variable is a dummy variable indicating whether the firm is involved in exporting or not.

Since the dependent variable data are in logarithms, the coefficients can be interpreted as percentages. In other words, exporting firms have 74 percent more employment, 106 percent higher shipments, 31 percent higher value-added per worker, and so on.

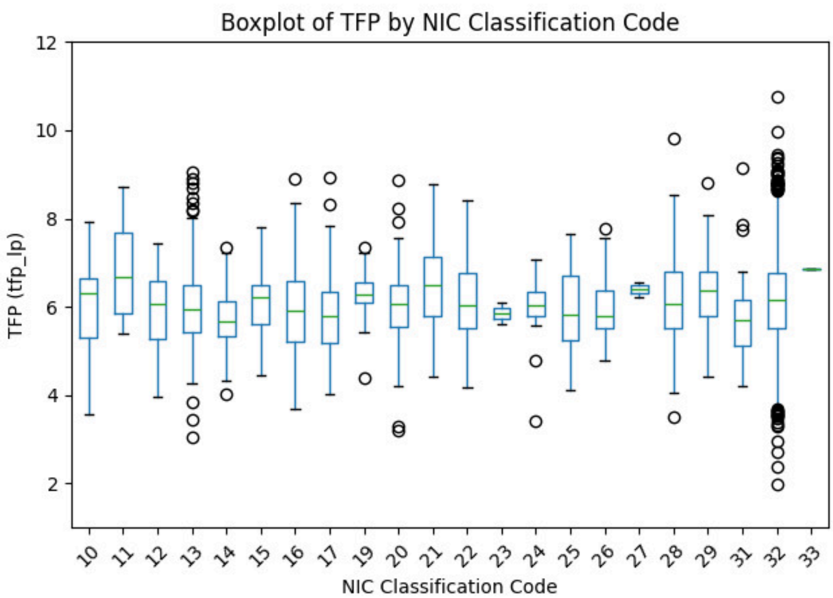
The second column repeats these regressions but now includes industry fixed effects in the explanatory variables to control for differences in firm characteristics across industries. Here too, exporters remain different from non exporters even in the same detailed industry. Exporters are significantly larger than non exporters, by approximately 74 percent for employment and 107 percent for shipments; they are more productive by roughly 32 percent for value-added per worker and 24 percent for total factor productivity; they also pay higher wages by around 9.5 percent. Finally, exporters are relatively more capital- and skill-intensive than non exporters by approximately 72 and 27 percent, respectively.

The observed differences between exporters and non exporters are not driven solely by size. When we control for firm size as measured by log employment as well as industry effects in column 3, the differences between exporters and non exporters within the same industry on all other economic outcomes continue to be statistically significant.

Hypothesis-III Results

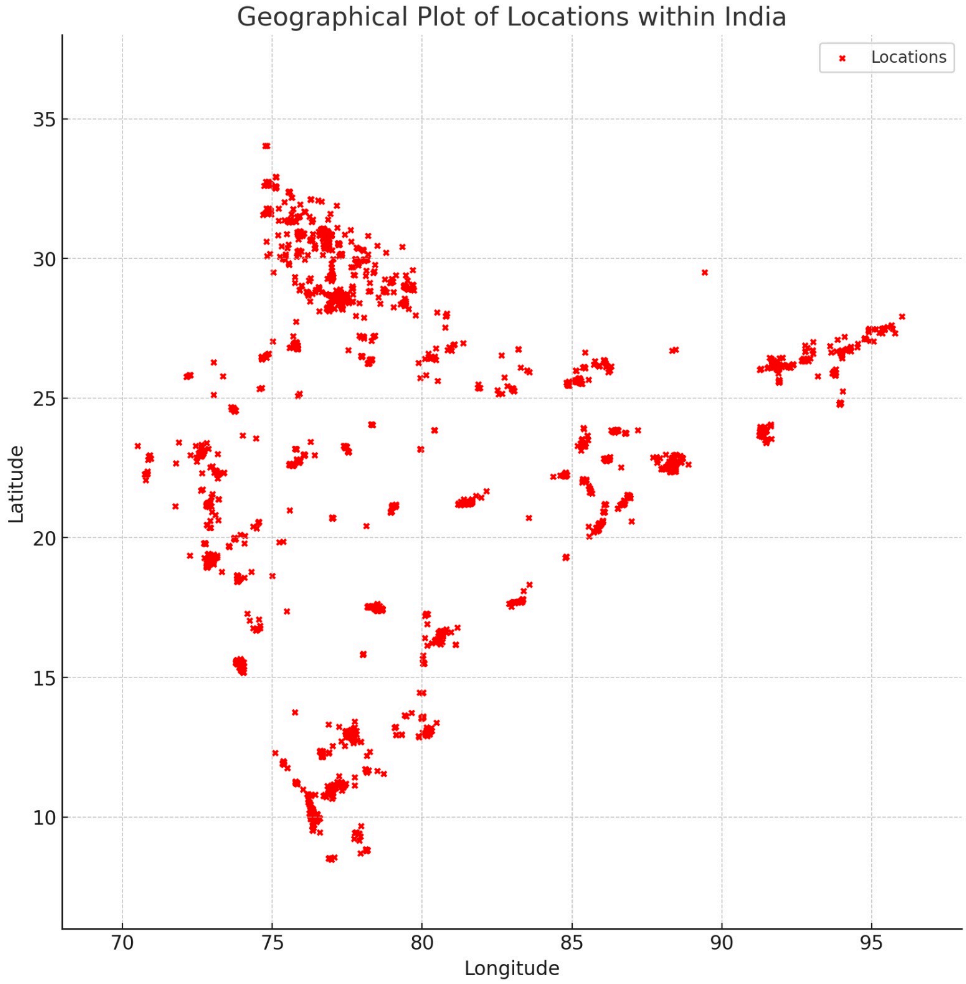


These are the industry-wise Total Factor Productivity values. All the industries were classified into the 24 National Industrial Classification Criteria & the values of TFP for each industry were calculated.
The dashed line represents the Total Factor Productivity of all manufacturing firms. While the bars represent the TFP values of only the exporting firms for each individual classification of industries. The graph clearly shows on average the TFP of exporting firms were much higher than the national average, showing ***self-selection*** in exporting.



Within the same industry classification itself there is firm-level heterogeneity as is evident from the box-plot given above. Large lengths of interquartile distance lines and abundance of outliers present away from the median value, provides empirical evidence that ***firm-level heterogeneity*** is considerably high even within the same industries.

Hypothesis-V Results



	(1)	(2)	(3)
Log Distance To EDFC	-0.2762**	-0.2764**	-0.2553**
Log Distance To WDFC	-0.3872**	-0.3594**	-0.3988**
Log Distance To Delhi Meerut Expressway	-0.8196**	-0.8106**	-0.7924*
Log Min Distance To GQ	-0.2044*	-0.1725*	-0.3642*
Additional Covariates	None	Industry Fixed Effects	Industry Fixed Effects Log Employment

The interpretation for these results is similar to what was done for the table shown in hypothesis 2. For example, we can see in the first column that exporter firms are 27 percent closer to EDFC, 38 percent closer to WDFC, 81 percent closer to Delhi-Meerut Expressway and 20 percent closer to the Golden Quadrilateral.

The second column repeats these regressions but now includes industry fixed effects in the explanatory variables to control for differences in firm characteristics across industries. Here too, exporters remain closer to key infrastructure as compared to non exporters even in the same detailed industry. Exporters are significantly closer to infrastructure than non exporters.

The observed differences between exporters and non exporters are not driven solely by size. When we control for firm size as measured by log employment as well as industry effects in column 3, the differences between exporters and non exporters within the same industry on all other economic outcomes continue to be statistically significant.

Conclusions

Our conclusions revolve around the 5 hypotheses which we stated earlier. The final conclusions are -

Conclusion-1

Only a small fraction of firms engage in international trade and within that group a small number of firms dominate the trade volume. Even within the same industry, exporter firms are larger, more productive, more skill and capital intensive firms and tend to pay higher wages than non-exporting firms. Traditional trade theories, such as the Ricardian and Heckscher-Ohlin models, emphasize comparative advantage at the country or industry level to explain trade patterns. However, we argue that firm-level differences play a crucial role in determining who trades and how much trade occurs. In the U.S. in 2000, only 4% of firms engaged in exporting, and the top 10% of these firms accounted for 96% of total U.S. exports. Within the smaller set of Indian firms active in industries more predisposed to exporting - like those in the manufacturing that produce tradable goods - only 13.87 percent were exporters in 2022. 2719/18621 out of the firms are exporting firms that is 14.60% are exporting firms. The value of total shipments is INR 11,17,888.42 million. The top 15% exporting firms account for the shipment value of INR 8,72,173.47 million. Thus they account for 78% of the total shipments.

Conclusion-2

Exporting firms tend to be larger, more productive and more capital-intensive and skill-intensive than the non-exporting firms. These firms also pay higher wages and have greater employment levels, creating a significant difference between exporters and non-exporters within the same industry. Exporting is more prevalent and export intensity is higher in skill-intensive industries such as computer manufacturing compared to labor-intensive industries such as apparel.

Conclusion-3

This conclusion deals with the idea of self-selection and heterogeneity, that is, the firms that engage in exporting already possess higher productivity levels before they start exporting. Firms need to reach a certain productivity threshold before they can export. Do exporters belong to a distinct group, self-selecting into exporting or is there an increase in productivity stemming from the learning opportunities associated with exporting, as has been observed in many developing countries? We emphasize the idea of self-selection, where only the most productive firms enter export markets. This self-selection occurs because exporting requires overcoming significant fixed costs that are the costs of marketing, distribution and complying with foreign regulations. As a result, only the most efficient firms can profitably export, which explains why productivity differences exist before firms even start exporting.

Productivity Advantage Before Exporting: We present evidence showing that the higher productivity of exporters isn't caused by exporting itself, but rather that firms need to reach a certain productivity threshold before they can export. This is in line with the "Melitz model" of international trade (Melitz, 2003), where only the most productive firms can profitably enter foreign markets.

Entry Costs And Barriers To Exporting: The cost of entering export markets (called "sunk costs") is high, which is why many firms do not export. For instance, only 15% of firms in manufacturing, mining, and agriculture export, despite these industries being traditionally more predisposed to international trade. Even when firms do export, many sell only a small portion of their output abroad (on average, about 14% of total sales).

Implication: The selection mechanism explains why aggregate productivity rises as trade barriers are lowered—because more productive firms expand, while less productive firms shrink or exit the market altogether. Traditional trade models assume all firms within an industry are alike, but the reality, as this paper intends to highlight, is that there is significant heterogeneity (variation) among firms, even within the same industry. This is where newer models of trade (such as the Melitz model and others discussed in the paper) come in. These models build on the idea that firms differ in their productivity, size, and capital intensity, and these differences have major implications for trade patterns.

Old vs New Models of Trade

Old trade models (like the Ricardian model) explain trade based on comparative advantage due to differences in opportunity costs, assuming all firms within a country are identical. They also predict that trade flows occur mainly between industries (inter-industry trade). – New trade models (Krugman, 1980) introduced economies of scale and consumer preferences for product variety, explaining intra-industry trade, where similar countries exchange goods within the same industry (like Germany and the U.S. trading cars). However, these models also assume all firms are identical within industries. – Heterogeneous Firms Model: The Melitz (2003) model integrates firm-level heterogeneity, recognizing that not all firms in an industry will export. Instead, only the most productive firms do so, creating a dynamic where trade liberalization benefits the more efficient firms disproportionately and raises aggregate productivity by reallocating resources to these firms

<u>Conclusion-4</u>

Adjustments along the extensive margins help understand the gravity model of trade - greater distance between trading partners dampens trade flows. The fewer products and destinations traded, the more sensitive firms are to distance. We highlight two important dimensions of trade: the extensive margin (the number of products and countries firms trade with) and the intensive margin (the value of trade per product per country).

Transaction-level data shows that the extensive margins—how many products firms export and how many countries they trade with—are crucial in understanding trade patterns. For example, most exporting firms in India export to only one country, but firms that export to many countries dominate the trade value.

Adjustments along the extensive margins are particularly important for understanding the "gravity model" of trade, which shows that distance between trading partners dampens trade flows. The fewer products and destinations traded, the more sensitive firms are to distance. Implications for Trade Liberalization: Trade liberalization (e.g., lowering trade barriers) disproportionately benefits the most productive firms, allowing them to grow faster. At the same time, it puts pressure on less productive firms, leading to their contraction or exit from the market. This process of reallocation improves aggregate productivity and raises welfare, which is a source of welfare gains overlooked by traditional trade models. – In countries like Chile, trade liberalization led to a significant increase in industry productivity, largely because of the reallocation of resources to more productive firms.

Conclusion-5

We examined the impact of proximity to major infrastructure projects such as Golden Quadrilateral Road Network, Delhi-Meerut Expressway, EDFC (Eastern Dedicated Freight Corridor) and WDFC (Western DFC) on firm performance. The conclusion for this is that proximity to key infrastructure like the Golden Quadrilateral, Delhi-Meerut Expressway and the Western Dedicated Freight Corridor significantly enhances firm sales/exports and reduces transportation obstacles.

7. Data Sources

1. World Bank Enterprises Survey Data (For Firm-Level Data)
2. CMIE (Centre For Monitoring Indian Economy Private Limited)

idstd	Current_T	Last_Yea	National	Indirect_	Direct_Ex	Export_Sh	Total_%	Current_T	Professio	Semi_Ski	Skill_Per	Previous_	Wages
869371	3E+08	2.3E+08	0	0	100	3E+08	100	110	110	0	1	60	2.8E+07
864133	2.5E+07	2.7E+07	0	0	100	2.5E+07	100	8	8	0	1	8	1728000
867038	6.3E+07	7.6E+07	0	0	100	6.3E+07	100	72	68	4	0.94	72	1.4E+07
870186	7.5E+08	8E+08	0	0	100	7.5E+08	100	1400	1300	100	0.93	1310	1.6E+08
869969	8000000	7000000	0	20	80	8000000	100	50	45	5	0.9	60	2500000
867101	1E+08	9.5E+07	0	60	40	1E+08	100	140	125	15	0.89	140	8000000
867493	1.1E+09	1E+09	0	0	100	1.1E+09	100	350	310	40	0.89	310	4.6E+07
870363	2E+07	1.8E+07	0	50	50	2E+07	100	16	14	2	0.88	20	1800000
864766	1.5E+08	1.4E+08	0	60	40	1.5E+08	100	80	70	10	0.88	80	3.8E+07
865804	4E+08	3.9E+08	0	60	40	4E+08	100	400	350	50	0.88	400	6E+07
865304	1.6E+09	1.2E+09	0	50	50	1.6E+09	100	320	280	40	0.88	340	5E+07
865252	2.5E+09	2.5E+09	0	0	100	2.5E+09	100	240	210	30	0.88	245	4.4E+07
871052	1.6E+08	1.3E+08	0	30	70	1.6E+08	100	80	70	10	0.88	80	2E+07
863668	1.4E+07	1E+07	0	0	100	1.4E+07	100	23	20	3	0.87	22	3500000
865682	2.3E+07	1.9E+07	0	0	100	2.3E+07	100	30	26	4	0.87	27	5200000
863704	2E+07	1.7E+07	0	60	40	2E+07	100	35	30	5	0.86	35	3000000
863954	4000000	3800000	0	0	100	4000000	100	14	12	2	0.86	20	1500000
869798	3000000	2900000	0	0	100	3000000	100	7	6	1	0.86	7	800000
868329	1E+08	9.5E+07	0	0	100	1E+08	100	65	55	10	0.85	75	3E+07
872645	3E+07	2.8E+07	0	30	70	3E+07	100	18	15	3	0.83	20	2500000
864094	5000000	4700000	0	0	100	5000000	100	18	15	3	0.83	18	1575000
871173	7.7E+07	8.1E+07	0	0	100	7.7E+07	100	42	35	7	0.83	73	8668800
870179	7E+08	6E+08	0	30	70	7E+08	100	60	50	10	0.83	80	1E+07
866483	4E+07	3E+07	0	50	50	4E+07	100	12	10	2	0.83	12	5000000
864888	1.2E+08	1.1E+08	0	50	50	1.2E+08	100	60	50	10	0.83	60	5000000
869392	8500000	6000000	0	25	75	8500000	100	18	15	3	0.83	18	3200000
866505	1.3E+07	1.2E+07	0	60	40	1.3E+07	100	12	10	2	0.83	12	1875000
863685	9500000	7500000	0	0	100	9500000	100	18	15	3	0.83	14	3240000
871055	1.8E+07	1.3E+07	0	75	25	1.8E+07	100	12	10	2	0.83	10	3200000
864582	7.8E+08	6.5E+08	0	20	80	7.8E+08	100	220	180	40	0.82	190	1.8E+07
864890	5E+07	4.6E+07	0	50	50	5E+07	100	55	45	10	0.82	55	5000000
863774	8500000	7500000	0	0	100	8500000	100	16	13	3	0.81	13	2300000
865471	8.6E+07	7.2E+07	0	50	50	8.6E+07	100	62	50	12	0.81	58	2.2E+07
863671	2E+08	1E+08	0	0	100	2E+08	100	150	120	30	0.8	130	4.5E+07
863748	1E+08	8.5E+07	0	20	80	1E+08	100	50	40	10	0.8	50	1.5E+07
865055	1E+07	6500000	0	40	60	1E+07	100	15	12	3	0.8	15	2500000
866654	4E+07	3.7E+07	0	60	40	4E+07	100	15	12	3	0.8	20	1800000
866067	9500000	8500000	0	0	100	9500000	100	25	20	5	0.8	2	3500000
869813	1.4E+08	1.6E+08	0	0	100	1.4E+08	100	150	120	30	0.8	160	3.2E+07
867310	7500000	7000000	0	60	40	7500000	100	15	12	3	0.8	15	2000000
863801	5E+07	5E+07	0	60	40	5E+07	100	50	40	10	0.8	50	9500000

idstd	Main_Ma	NIC Class	Wage pe	Latitude	Longitud	NIC Class	Classifica	Value_Ad	Capital	Previous	Capital_p
869371	manufacture of co		254545	17.3922	78.4763	32	Other ma	2727273	4.7E+07	4.3E+07	502843
864133	Manufacturer of La		216000	28.5184	77.2939	13	Manufac	3125000	2.8E+07	2.5E+07	4032606
867038	Manufacturer of h		189267	28.5083	77.2735	13	Manufac	878503	3E+07	2.6E+07	492602
870186	Manufacturing of (		113729	28.502	77.3011	13	Manufac	536487	8E+07	6.7E+07	66517.8
869969	manufacturing of r		50000	17.5486	78.2377	32	Other ma	160000	2.6E+07	2.1E+07	616716
867101	manufacturing of p		57142.9	17.4849	78.4156	32	Other ma	714286	3.3E+07	2.7E+07	275341
867493	Manufacturer of Ci		131429	28.5173	77.0764	13	Manufac	3142857	1E+08	8.4E+07	349637
870363	Manufacture of Al		112500	17.4847	78.4606	32	Other ma	1250000	2.7E+07	2.2E+07	1990107
864766	Manufactures of M		468750	17.5648	78.2004	32	Other ma	1875000	7500000	5957820	109375
865804	Manufacturing of b		150000	17.5418	78.4531	10	Manufac	1000000	5.5E+07	4.3E+07	159238
865304	Manufacturing of b		156250	17.4471	78.5396	10	Manufac	5000000	1.4E+08	1.1E+08	513392
865252	Manufacturing of L		185400	28.5555	77.2535	32	Other ma	1E+07	1.9E+07	1.5E+07	93519.7
871052	Sale and manufac		250000	17.4175	78.5142	32	Other ma	2000000	3.7E+07	3E+07	544716
863668	manufacturing and		152174	16.5342	80.5978	32	Other ma	608696	2.7E+07	2.1E+07	1362555
865682	manufacturing gra		173333	15.5033	80.0284	32	Other ma	766667	2.8E+07	2.2E+07	1069773
863704	Manufacturer of H		85714.3	17.5266	78.471	32	Other ma	571429	2.7E+07	2.1E+07	909763
863954	Manufacturer of su		107143	17.4652	78.4145	31	Manufac	285714	2.6E+07	2E+07	2178608
869798	manufacturing of c		114286	17.4444	78.4928	16	Manufac	428571	2.6E+07	2E+07	4345241
868329	manufacturing of i		461538	17.5528	78.3327	32	Other ma	1538462	3.3E+07	2.5E+07	593042
872645	Manufacture of Re		138889	17.5011	78.4458	20	Manufac	1666667	2.8E+07	2.1E+07	1815554
864094	Manufacturer of cc		87500	17.4664	78.4345	28	Manufac	277778	2.6E+07	2E+07	1699130
871173	Manufacturer of H		206400	28.5094	77.2663	32	Other ma	1824148	3.1E+07	2.3E+07	871129
870179	Manufacturer of In		166667	17.5356	78.4577	32	Other ma	1.2E+07	7.6E+07	5.7E+07	1480715
866483	manufacturers of e		416667	17.4807	78.46	32	Other ma	3333333	2.1E+07	1.6E+07	2041667
864888	manufacturing of a		83333.3	17.4423	78.5434	32	Other ma	2000000	9000000	6743169	175000
869392	Manufacturing of g		177778	17.2746	80.1838	32	Other ma	472222	1.2E+07	8990892	777778
866505	manufacturing of g		156250	17.2848	80.1262	32	Other ma	1041667	5100000	3821129	495833
863685	manufacturing of i		180000	16.6052	80.5235	32	Other ma	527778	2.7E+07	2E+07	1720086
871055	motor vehicle part		266667	17.4164	78.5034	29	Manufac	1458333	6825000	5113570	663542
864582	Manufacture of La		81818.2	28.3809	76.9013	13	Manufac	3552273	8.2E+07	6.1E+07	434885
864890	Manufacture of me		90909.1	17.5002	78.4031	32	Other ma	909091	2.9E+07	2.2E+07	624663
863774	manufacturer of a		143750	16.52	80.6588	32	Other ma	531250	2.6E+07	1.9E+07	1929858
865471	manufacturing of u		354839	17.4601	78.5071	15	Manufac	1393548	3.2E+07	2.3E+07	603350
863671	battery manufactu		300000	18.0936	83.3855	32	Other ma	1333333	4E+07	2.9E+07	312868
863748	Manufacture of we		300000	9.9263	76.2696	32	Other ma	2000000	3.3E+07	2.4E+07	770955
865055	Manufacturer of Ai		166667	17.441	78.5395	32	Other ma	666667	7875000	5687010	612500
866654	Manufacturer of ai		120000	17.4686	78.4155	32	Other ma	2666667	2.9E+07	2.1E+07	2234548
866067	Manufacturer of Al		140000	15.8131	78.0233	32	Other ma	380000	2.7E+07	1.9E+07	1238462
869813	Manufacturer of cu		210000	28.5266	77.2578	13	Manufac	961809	1944600	1404312	15124.7
867310	Manufacturer of In		133333	17.4619	78.4099	32	Other ma	500000	2.6E+07	1.9E+07	2052927
863801	manufacturer of p		190000	17.4527	78.6305	32	Other ma	1000000	7.8E+07	5.6E+07	1820000

idstd	Export_D	ln_outpu	ln_capita	ln_labou	ln_tfp_lp	tfp_lp	min_dist	nearest_	distance	distance	distance_to_EDI
869371	1	19.5193	17.6744	4.70048	6.29952	544.311	422.288	Nagpur	1254.48	620.162	1021.61
864133	1	17.0344	17.1352	2.07944	6.53696	690.188	13.4703	Delhi	12.1822	27.8448	298.954
867038	1	17.9626	17.23	4.27667	5.36133	213.007	13.325	Delhi	13.2031	27.4352	299.049
870186	1	20.437	18.1953	7.24423	4.57634	97.1579	15.3568	Delhi	14.0966	29.6831	300.903
869969	1	15.895	17.09	3.91202	3.70478	40.641	409.819	Nagpur	1234.88	590.722	1012.99
867101	1	18.4207	17.3133	4.94164	5.15539	173.363	413.146	Nagpur	1243.6	610.846	1016.2
867493	1	20.8186	18.4684	5.85793	6.11916	454.482	16.8207	Delhi	16.8207	20.9223	289.917
870363	1	16.8112	17.1221	2.77259	5.67156	290.487	412.387	Nagpur	1244.09	615.395	1013.97
864766	1	18.8261	15.8304	4.38203	6.82464	920.247	408.934	Nagpur	1232.76	586.407	1012.05
865804	1	19.807	17.8155	5.99146	5.30842	202.031	406.247	Nagpur	1237.69	612.757	1009.14
865304	1	21.1933	18.763	5.76832	6.43075	620.638	415.284	Nagpur	1249.1	624.654	1012.82
865252	1	21.6194	16.7724	5.48064	8.11946	3359.19	7.80698	Delhi	7.80698	22.1347	293.478
871052	1	18.8907	17.4359	4.38203	6.08799	440.535	418.921	Nagpur	1252.09	623.102	1017.05
863668	1	16.4546	17.1062	3.13549	4.98315	145.933	306.574	Vishakha	1386.96	818.556	1004.83
865682	1	16.951	17.13	3.4012	5.21902	184.753	270.416	Chennai	1486.32	856.676	1133.56
863704	1	16.8112	17.1221	3.55535	4.9389	139.616	407.609	Nagpur	1239.56	615.058	1009.48
863954	1	15.2018	17.0791	2.63906	4.20858	67.2611	415.325	Nagpur	1245.77	611.401	1018.38
869798	1	14.9141	17.0764	1.94591	4.57106	96.6465	416.295	Nagpur	1248.89	620.013	1015.81
868329	1	18.4207	17.3133	4.17439	5.87354	355.504	407.305	Nagpur	1235.27	600.196	1010.49
872645	1	17.2167	17.1481	2.89037	5.95381	385.219	410.839	Nagpur	1242.12	613.357	1013.33
864094	1	15.4249	17.0818	2.89037	4.19513	66.3621	414.834	Nagpur	1245.84	613.38	1017.19
871173	1	18.1543	17.2611	3.73767	6.04196	420.718	12.9029	Delhi	12.9029	26.9205	298.628
870179	1	20.3666	18.1482	4.09434	7.47769	1768.15	406.852	Nagpur	1238.43	613.423	1009.44
866483	1	17.5044	16.86	2.48491	6.76478	866.772	412.838	Nagpur	1244.53	615.473	1014.37
864888	1	18.603	16.0127	4.09434	6.77978	879.879	415.756	Nagpur	1249.67	625.195	1013.05
869392	1	15.9556	16.3004	2.89037	5.11572	166.62	325.107	Vishakha	1296.76	725.526	945.331
866505	1	16.3412	15.4448	2.48491	6.3079	548.891	330.99	Vishakha	1294.32	722.104	946.83
863685	1	16.0668	17.0941	2.89037	4.83086	125.319	310.581	Vishakha	1377.44	808.169	1000.22
871055	1	16.6777	15.7361	2.48491	6.49898	664.462	419.209	Nagpur	1252.1	622.048	1017.79
864582	1	20.4767	18.2223	5.39363	6.33472	563.812	39.6892	Delhi	39.6892	40.9605	297.726
864890	1	17.7275	17.1982	4.00733	5.39419	220.125	411.695	Nagpur	1241.78	609.062	1014.78
863774	1	15.9556	17.0914	2.77259	4.83123	125.366	301.386	Vishakha	1390.04	822.613	1004.08
865471	1	18.2745	17.2832	4.12713	5.78656	325.891	414.342	Nagpur	1247.3	620.924	1013.52
863671	1	19.1138	17.51	5.01064	5.68576	294.642	48.5576	Vishakha	1326.23	845.993	768.773
863748	1	18.4207	17.3133	3.91202	6.11911	454.46	559.638	Chennai	2080.28	1080.78	1864.39
865055	1	16.1181	15.8792	2.70805	5.65909	286.887	415.959	Nagpur	1249.77	624.844	1013.4
866654	1	17.5044	17.1734	2.70805	6.39951	601.549	414.934	Nagpur	1245.4	611.383	1017.99
866067	1	16.0668	17.0941	3.21888	4.52338	92.1467	388.234	Chennai	1425.83	655.336	1196.17
869813	1	18.7872	14.4806	5.01064	6.87094	963.858	10.8105	Delhi	10.8105	24.8842	296.537
867310	1	15.8304	17.0887	2.70805	4.76784	117.664	415.764	Nagpur	1246.09	611.047	1018.82
863801	1	17.7275	18.1722	3.91202	4.99731	148.015	413.447	Nagpur	1249.52	633.659	1007.04

8. References

[1] BERNARD, A.B., JENSEN, J.B., REDDING, S.J., SCHOTT, P.K. - Firms in International Trade, Journal of Economic Perspectives, 21 (3), 105-130, 2007.

[2] CHANEY, T. - Distorted Gravity: The Intensive and Extensive Margins of International Trade, The American Economic Review, 98 (4), 1707-1721, 2008.

[3] SRIVASTAVA, A., MATHUR, S.K., CHAUHAN, M., DE, P. - The Comparative Study of the Impact of the Indian Transportation Infrastructure on Firm Level Performance Using Difference In Difference (DID) with Focus on Inland Waterways.