Very Preliminary Draft

Domestic tax reform and imported inputs: Evidence from India *

Maxime Berrou †

July 22, 2025

Abstract

This paper investigates the impact of India's 2017 Goods and Services Tax (GST) reform, which replaced a fragmented commodity taxation system with a nationwide Value-Added Tax (VAT) and eliminated cascading taxes on cross-state transactions. Using a panel of manufacturing firms from 2012 to 2020, and exploiting heterogeneity in firms' eligibility for input tax credits, I employ a difference-in-differences approach to compare treated firms—those able to claim input tax credits—with untreated firms. The results show that treated firms increase input sourcing post-reform, particularly through newly acquired inputs. The findings indicate that reductions in input costs significantly influence firms' sourcing and production decisions, with some firms adjusting their product mix to gain access to tax credits. Additionally, the reform appears to have facilitated greater international sourcing of inputs, especially for treated firms in inland states, and there is evidence of complementarity between lower input tariffs and GST exposure.

Keywords: input sourcing, tax reform, India, plant panel.

*I would like to thank my supervisor Pamela Bombarda, Pierre-Louis Vézina, Maria Bas, Zaki Wahhaj, participants at ThEMA internal seminar (CYU) and at Econ Reading Groups (KCL) for helpful comments and suggestions. I would also like to thank Johannes Boehm for providing information the Annual Survey of Industries (ASI), and Mohammad Hoseini for clarifications regarding the construction of the aggregated firm data. I am also grateful to the Ministry of Statistics and Programme Implementation (MOSPI) staff for their support with data access. All errors are my own.

†CY Université, THEMA (UMR CNRS 8184), France. maxime.berrou@cyu.fr

1 Introduction

Value-added taxation (VAT) has been implemented in many developing countries across the world in the last forty years (?). It has become a favored instrument for policymakers and economists alike due to its ability to increase tax revenue and to promote formality without distorting production decisions. This article is about how deviations from textbook VAT design may impact sourcing decisions.

Can the design of VAT impact plants' decision to source inputs domestically or internationally? Because it only taxes value added and leaves relative prices unchanged, VAT implementation should be neutral, and therefore trade-neutral (?). This neutrality of VAT has been studied in theory, but there has been little empirical evidence that it actually holds. A notable exception is the work by ?, who show significant but very small trade effects of VAT changes. Their focus is on European countries with well-functioning VAT systems. A natural question is whether those observations still hold in developing economies, where VAT design and implementation may deviate significantly from the typical "textbook" VAT. A tax on consumption, a uniform single-rate VAT design is inherently regressive[1]. Because of that, VAT rates are often higher for intermediate goods used in firm's production processes, especially for complex and high-tech inputs, which tend to be imported. Additionally, to simplify compliance to the tax system, large parts of the economy are exempted from VAT (?). The effects of those two features of VAT design on international sourcing decisions is what this paper explores.

My contribution is twofold. First, I document long-run trends and stylized facts about commodity taxation in India. Most previous work focuses on short-run approaches. Second, I provide causal evidence that tax design matters for sourcing decisions. To my knowledge, this paper is the first to exploit the implementation of a nationwide VAT. I exploit micro-level information about plants' pre-reform input mix, as well as their exemption status, to highlight a phenomenon theorized by ?: that plants exempt from VAT are differentially impacted following the implementation of VAT, since the inputs they purchase are taxed but they cannot claim input tax credit (ITC).

To study the effects of tax reforms on sourcing decisions, I take advantage of an indirect taxation overhaul in India - the Goods and Services Tax (GST) implementation in 2017. The GST replaced a complex system characterized by multiple state-level tax systems. Before 2017, India had over 30 state-level commercial tax systems (?). The implementation of state-level VAT in the early 2000s was an initial simplification of the commercial tax system, but many inefficiencies remained. Among those imperfections, perhaps the most salient is the lack of a well-functioning input tax credit scheme between states. The ability for firms to claim taxes paid on inputs is a crucial feature of VAT systems. That system made firm-to-firm transactions costly. This issue was particularly pronounced for transactions between states. The implementation of GST was

[1] Like all taxes on consumption in general. Tariffs are not commercial taxes, but are also generally regressive as shown by ?

meant to simplify the system and unify the Indian market to boost domestic trade.

The empirical analysis relies on two salient features of the Indian GST reform: (i) some products are exempt, based on their classification code, and (ii) non-exempt products are not taxed uniformally. Those two characteristics allow me to construct plant-level exposure to the reform, and to compare those relatively more exposed to the reform to those relatively less exposed. More specifically, the empirical strategy compares GST-eligible plants, which are able to claim refunds on taxes paid on inputs, to GST-exempt plants, to which the input tax credit scheme does not apply. To address identification concerns, I provide evidence that GST implementation decreased tax burden at the plant level, but only for those plants able to claim a refund on the taxes paid on inputs. The higher the tax paid on inputs, the greater the difference in taxes paid between eligible and non-eligible plants.

Having shown that the reform provides plausibly exogenous variation in taxes paid by plants, and therefore their production cost, I then investigate whether those changing costs may have impacted plants' sourcing decisions. I hypothesize that, relative to exempt plants, plants producing taxable goods and able to claim tax refunds should be more likely to source high-taxed intermediates because they may then claim ITC and not pay the input VAT. I find empirical evidence that to ITC-eligible plants tend to import relatively more following the reform than ITC-ineligible plants. In my preferred specification relying on a double-difference estimation, I show that ITC-eligible plants with a high input tax have import-to-domestic sourcing growth rates 15% higher than similar plants which cannot claim input VAT. My results thus suggest that that the implementation of GST is not trade-neutral, *i.e.*, plants changed their international sourcing decisions following the reform, and that those results are quantitatively meaningful.

Related literature. This study relates to three strands of literature. First, it relates to recent empirical works highlighting the effects of value-added taxation. In theory, VAT should not distort firm's sourcing decisions because it should leave their relative prices unchanged. Replacing sales taxes or turnover taxes by VAT reduces distortions along the supply chain and increases overall efficiency (**?** for China, **?** for India; **?** for Brazil). The assumptions on which the theory is based are not necessarily met in practice. **?** and **?** develop a simple theoretical framework highlighting the assumptions under which the non-distortionary effect of VAT holds. Those assumptions are rarely met in developing economies. First, VAT can still distort trade between firms when informality exists, as shown by**?**. In addition, tax design matters and deviations from this "ideal" neutral VAT may induce changes in (international) production decisions (**?**). **?** finds that VAT implementation can hurt export performance if the input tax credit works imperfectly. My contribution is to focus on the production side and to provide micro-level evidence of those effects on sourcing decisions when VAT has multiple rates and large parts of the economy are excluded , and to show that it matters for sourcing decisions.

Second, this research adds to a literature on access to imported inputs in developing economies. New international inputs leads to productivity gains (see evidence by **?** for Hungary, **?** for Chile and ? for Indonesia). This is especially the case if they are of higher quality (?). It is of particular importance in developing economies because it permits technology upgrading (?). The extensive margin of imported inputs, import switching, is a prevalent phenomenon in both developed and developing economies (see ? for India, ? for Colombia). The ability of plants to add and drop inputs also provides a significant margin of adjustment for firms facing a shock. ? show that in Spain, firms facing a negative demand shock switched from imported to domestic inputs, resulting in productivity losses. The main focus of the paper is finding evidence of import switching - specifically, to switch back from international to domestic sourcing following a domestic shock.

Finally, this study adds to a growing literature investigating the effects policy heterogeneity within countries. Fiscal federalism XXX ?. In particular, different tax systems within a country lead to spatial misallocation. ? shows that different state taxes induces spatial misallocation in the United States. Recent research has highlighted the importance of those policies in fragmenting developing economies. ? develops a quantitative two-sector model à la (?) to quantify trade barriers between Indian states. Those barriers are high, and in significant part they are due to the tax system. Still in India, ? show that differences in court quality impacts buyer-seller transactions. While this study does not study spatial misallocation *per se*, it investigates the effects of the integration of the Indian market and tax policies under a single framework.

Section 2 presents the conceptual framework, and section 3 then presents the data used. Section 4 describes the institutional background and key empirical facts. Sections 5 presents the construction of the main explanatory variable, the identification strategy, and the econometric models used. Section ?? presents the results and robustness checks, and section 9 concludes.

2 Conceptual framework

An indirect taxation scheme such as the 2017 Goods and Services Tax corresponds to the implementation of a nationwide VAT. This section motivates the research question by providing theoretical insights on the effects of VAT on production decisions of firms, and in particular, how it might distort trade flows.

A VAT is designed to tax the value added created at each stage of the production chain. Firms making transactions declare the VAT collected when selling their output ("output VAT"), and the VAT they paid when purchasing inputs ("input VAT"). Firms can then credit the input VAT paid to suppliers against the output VAT on their supplies (?). In addition, VAT follows the destination principle: imports are taxed but exports are zero-rated, meaning exporters can still credit input VAT against the (zero) output VAT.

Under some "textbook" assumptions, VAT leaves production decisions unchanged relative to a world without tax. It does not modify relative prices because all sectors of the economy are taxed; an increase in the tax rate should increase all prices by the same amount and thus leave relative prices unchanged. In addition, it is considered to be trade neutral, despite the fact that imports are taxed and exports are zero-rated. To show the (trade-)neutrality of VAT, **?** provide a simple framework describing the effects of a VAT on the economy. They develop a model of VAT in which all sectors are taxed at the same unique rate. They argue that VAT is trade-neutral even in that case, and that the destination-principle is actually necessary in order to not tax domestic producers relatively more.[2]

? make explicit that their key result, the trade-neutrality of VATs, relies on specific assumptions on tax design. More specifically,in their model, VAT is implemented following a "textbook", or "idealized", design. This design refers to VAT systems which have a single rate applied on all goods, and which covers the entire economy (**?**). This means that VAT is trade neutral if it all firms pay the same rate of output VAT and can perfectly claim input VAT.

The key question is then whether textbook VATs are actually implemented. This is not a mere theoretical curiosity; in fact, it may be of great empirical relevance. In practice, few VAT systems follow a "textbook" design. VAT almost universally has at least two rates, sometimes more, even in developed economies.[3] In that case, output VAT varies between sectors. Moreover, the practice of exempting sectors is also widespread, especially in developing economies, leaving entire parts of the economy out of the possibility to claim input VAT (**?**). In that case, the tax will not purely tax value added and does not leave relative prices unchanged.

?, commenting on **?**, stressed that "the treatment of intermediate inputs" was "a vital aspect of a VAT", especially with multiple rates and/or exempt sectors. He developed a framework in which the domestic economy produces two goods: a traded good subject to VAT, and a nontraded good exempt from VAT. The model shows that VAT is not trade neutral because of the way intermediate inputs are taxed: (i) VAT raises the producer price of the nontraded good, (ii) it raises the cost of nontraded inputs for the traded good sector, and (iii) it raises the cost of traded inputs for the nontraded good sector. This last channel is the one that this paper will seek to provide evidence of: that plants which do not pay input VAT (through ITC) pay less input cost than exempt firms, and that exempt firms react by switching from low-taxed, domestically produced inputs.

We expect imports to increase because importing is costly (**?**). So the ...

It turns out that **?**'s model is be a good approximation of the Indian economy. Indeed, India's GST features exempt sectors (for instance, textiles), and has higher rates on goods which tend to be imported (like high-tech intermediates). Further evidence will be presented in section 4 about the institutional background.

[2]Their result assumes full pass-through of VAT from seller to buyer. This may not actually be the case under ceratin market conditions (*e.g.*, market concentration). In this review, I focus on papers which assume full VAT pass-through to restrict the analysis to the effects of VAT design itself.

[3]The decision to tax at a regular rate, a reduced rate, or to exempt from VAT matters for producers and is a debated issue in most countries (see, for instance, reactions to proposed changes in VAT thresholds in France in 2025).

3 Data

Data on manufacturing plants comes from the Annual Survey of Industries (ASI), which is conducted by the Ministry of Statistics and Programme Implementation of the Government. The ASI covers manufacturing establishments registered under the Factories Act. It is meant to be representative of establishments above 10 workers and using power or above 20 workers and not using power. Because it only follows registered establishments, it does not take into account the informal sector.[4]

I restrict the data to the 2012-2013 wave of the survey to 2019-2020 wave. Due to the inclusion of non-producing establishments, I also restrict the sample to establishments which were open and engaged in economic activity (*i.e.*, with total gross sale value strictly greater than zero).[5] Out of those open establishments, I follow the data cleaning procedure of ASI by **?** and drop a small number for which some variables take implausible values (specifically, the year of initial production and total labor employed). In the final unbalanced sample, each year has about 40,000 observations.[6]

The ASI is an ideal dataset to study establishments' production decisions (**?**). Each observation corresponds to a manufacturing plant, not firm. This implies that unlike firm-level data, in which the values reported may be aggregates from multiple production sites, the ASI permits to precisely connect inputs used and output produced in a specific unit of production. In a limited number of cases, when the owner owns multiple establishments in the same state, this is not possible because information is reported jointly by the owner. This case is uncommon; in 2016, 93% of observations had information reported for one plant, and 97% for two plants or less.[7] The ASI contains detailed information on intermediate inputs and products manufactured by plants (value, quantity and unit price), and distinguishes between imported and domestically-sourced inputs. It also contains typical firm-level information from firms' balance sheets (labor, capital, location, industry, etc.).

I also gather data on Indian tax and trade policy. Data on GST rates come from official publications by the Gazette of India. Those publications provide information on whether each product is exempt from GST, as well as the GST rate to be applied to the (non-exempt) product. The products are both described in words and are assigned a product code following the Indian Trade Classification (ITCHS), which is similar to the HS classification. I then matched the tax rates to

[4]This is not an issue for the purpose our this paper, which studies how a tax reform distorts sourcing decisions of tax-paying producers, but not their decision to select into the tax scheme.

[5]The ASI surveys establishments which may be non-producing. **?** document that establishments often become "dormant" for some time before actually closing, which they argue is due to high regulatory barriers to exit.

[6]I keep the sample unbalanced throughout the analysis. While the ASI provides a representative sample of all registered manufacturing establishments in India, it only covers large establishments every year, with smaller establishments covered on a sampling basis (**?**). Therefore, keeping a balanced sample would select only large establishments.

[7]In cases when the one observation corresponds to multiple plants, I divide reported values by the number of factories and consider that all factories for that observation contributed equally.

the Central Product Classification (CPC) Version 2.1 at the 5-digit level, which corresponds to the classification used in ASI up to five digits.[8] I then merge the tax data with the within-plant manufactured products and intermediate inputs used. In case of no matching, I consider the product (input) to be non-exempt and to be taxable at a rate of 18%, following official regulations.[9] Data on tariffs come from WITS, downloaded at the 6-digit Harmonized System (HS) classification, which are then merged to a 5-digit CPC Version 2.1. I gather data on Indian MFN tariffs and on Indian preferential tariffs with respect to ASEAN partners.

4 Institutional background

4.1 Taxation in India

Before 2017. Before GST was implemented, India had a fragmented commercial tax system. It was characterized by multiple, and often overlapping, state-level and central-level commercial taxes (?).[10] On average, tax rates were high, and varied by type of good and location. The pre-GST regime involved significant inefficiencies because of cascading taxation along two dimensions: across space, and across production stages. Firms seeking to source inputs or to sell products in other states had to pay high taxes (with often multiple taxation for the same transaction), and those taxes added up along supply chains, increasing the price of products at each stage of production, thus distorting relative input prices and decreasing production efficiency (?).

There had been attempts to introduce a more effective tax system. The implementation of a state-level VAT in the early 2000s marked an step towards a simplification of the tax structure by reducing cascading effects within states. Firms could now claim input tax credit within their state. An ITC mechanism for inter-state transactions, the Central Value-Added Tax (CENVAT), was also introduced in 2004. Despite this, significant inefficiencies remained, especially on inter-state transactions. Before 2017, states still implemented high entry taxes on goods coming from other states, and the ITC for inter-state supplies did not work well because CENVAT and state-level VAT could not be credited against each other (?). Prior to GST implementation, it is estimated that internal trade barriers represented approximately 40% of total Indian trade barriers, and that the tax system was a major reason for those high internal trade costs (?). For those reasons, GST implementation is the first time a nationwide well-functioning and unifying commercial taxation scheme was introduced in India.

Some facts.

[8]Concordance tables were obtained through the website of the Statistical Division of the United Nations.

[9]The Gazette of India states that "goods which are not specified in Schedule I, II, IV, V or VI" should be assigned the basic rate of 18%.

[10]Central taxes included central value-added tax (CENVAT), central excise duty tax, service tax, central sales tax, countervailing duties and special additional duties of customs. State taxes also had their own value-added tax, sales tax, as well as an entry tax, a luxury tax and an entertainment tax (?).

The Goods and Services Tax (2017). On July 1st, 2017, India introduced a nationwide value-added tax (**?**). It has been referred to as the most important tax reform since Independence in 1947. For the first time, India had a functioning nationwide ITC scheme, rationalizing commercial taxes across production stages and location.

There are two main characteristics of the Indian GST as it has been designed. First, the supply of some goods is within the scope of GST, thus activating possibility to use the ITC scheme, whereas other goods are exempt, meaning that the plant cannot claim refunds on inputs. Second, for goods within the scope of GST, there are multiple tax rates. Tax slabs are reported in table 1. The basic rate is 18%, which is the category in which the most products were classified. Products are relatively evenly spread out between slabs; many of them belong to the 5%, 12% and 28% and exempt categories. On the other hand, very few products belong to the 0,25% and 3% slabs: those categories are only for precious stones and jewelry. The main tax categories are therefore exempt goods, and goods taxed at 5%, 12%, 18% and 28%.

Table 1: Number of products by GST slab

Tax rate (%) (1)	Number of products (2)
0% (exempt)	1,091
0.25%	1
3%	50
5%	610
12%	763
18%	1150
28%	768

Notes: This table presents the repartition of the main tax slabs in the basic GST scheme. Column (2) reports the number of products belonging a a particular slab (5-digit CPC Rev.2.1).

To better understand the scope of this paper, it is useful explain other secondary features of GST. First, as is often the case in for VAT systems, plants may opt out of paying taxes if they operate below a certain turnover. For instance, in most states (except in the North-East), plants operating below a turnover of 4 million INR are exempt from GST, whatever good they produce. In addition, plants between a turnover of 4 million INR and 15 million INR may opt for a “composition scheme” in which the supply is taxed at a low rate of 1%, but with an inability to claim ITC. Second, exports and supplies to Special Economic Zones are zero-rated, with the ability to claim ITC. The empirical analysis will address those specificities to make sure they do not drive the results.

Academic research as well as anecdotical evidence suggest that the ability to claim refunds on taxes paid on inputs matters a great deal for firms. In pre-GST West Bengal, **?** document that firms with turnover below mandatory VAT registration thresholds still opt into the tax scheme because the lower taxes paid on inputs more than compensates lower tax rates on output. After 2017, multiple court rulings had to clarify which supplies were exempt or not[11], and some firms have been found guilty of producing fake invoices to claim ITC.

Summary statistics of characteristics of the reform further showcase the relevance of focusing on the input-side effects of commercial taxation made by **?**. Table 2 presents, for two-digit industries, the average tax rate in that industry, the share of exempted products and the share of imports. Three striking facts emerge. First, for taxed goods, the higher the GST rate of taxed industries, the more technological that industry. For example, column (1) shows that the category "Products of agriculture, horticulture and market gardening" (average tax rate of 5.0%), and "Meat, fish, fruits, vegetables, oils and fats" (8.4%) face among the lowest tax, whereas "Medical appliances, precision and optical instruments, watches and clocks" and "Special purpose machinery" are taxed higher rates (20.1% and 18.1%). Those findings are logical: VAT-like systems typically at least two rates, with lower rates for food or agricultural products (**?**). Consumption taxes tends to be lower on final, consumption goods, and higher on intermediates and luxury goods, to mitigate their inherent regressivity.

Second, and importantly for this study, large parts of production are exempted from GST. Column (2) of table 2 reveals large heterogeneity between industries: the share of exempted products ranges from 91.9% (agriculture) and 2.6% (special-purpose machinery). The average exemption share is lower but still substantial in industries which make up most of Indian manufacturing output, such as textiles (24.1%). Technological goods tend to be non-exempt from GST.

Third, column (3) shows the average import share for each two-digit industry.[12] It is unsurprising that industries that produce more complex goods, which tend to require specific components that India does not make, or not in sufficient quantity, tend to import more. Some industries rely much more on imports than others; with agricultural and food products importing very little, and capital goods or technological goods importing much more.

[11]For instance, the Supreme Court ruled on October 1st 2024 that companies are eligible for ITC on construction costs for commercial buildings for rental purposes (civil appeal no. 2948 of 2023).

[12]Note that column (3) is compiled using ASI and is an aggregate of direct plant imports. It does not account for imports made further up value chains when the plants' suppliers (or the suppliers of its suppliers) when they are not in the same two-digit industry.

Table 2: Description of GST by two-digit industries

Industry	Average GST rate (%) (1)	Share of exempted goods (%) (2)	Share of imports (%) (3)
Products of agriculture, horticulture and market gardening	5.0	91.9	4.4
Live animals and animal products (excluding meat)	9.1	81.4	0.3
Forestry and logging products	13.2	55.6	16.1
Fish and other fishing products	5.0	96.8	0.7
Other minerals	4.6	12.5	45.3
Electricity, town gas, steam and hot water	5.0	33.3	3.9
Meat, fish, fruits, vegetables, oils and fats	8.4	41.5	35.0
Dairy products and egg products	11.6	30.8	0.1
Grain mill products, starches and starch products; other food products	13.3	44.9	7.4
Beverages	21.3	25.0	9.4
Yarn and thread; woven and tufted textile fabrics	8.9	7.5	5.7
Textile articles other than apparel	11.6	24.1	11.4
Knitted or crocheted fabrics; wearing apparel	8.5	3.3	7.4
Products of wood, cork, straw and plaiting materials	19.7	3.0	12.1
Pulp, paper and paper products; printed matter and related articles	13.4	23.9	23.5
Basic chemicals	11.9	4.4	27.9
Other chemical products; man-made fibres	16.7	8.8	18.5
Rubber and plastics products	16.8	6.9	17.7
Glass and glass products and other non-metallic products n.e.c.	20.3	3.9	10.7
Furniture; other transportable goods n.e.c.	16.1	10.9	54.2
Wastes or scraps	12.2	30.6	17.3
Fabricated metal products, except machinery and equipment	18.5	2.8	14.0
Special-purpose machinery	18.1	2.6	26.4
Medical appliances, precision and optical instruments, watches and clocks	20.1	3.6	36.8
Transport equipment	14.5	4.2	15.0

Notes: This table presents a simple average of Goods and Services Tax within two-digit CPC (version 2.1). Column Average GST rate (%) is the simple average of GST rate of non-exempt goods belonging to a two-digit industry. Share of exempted goods (%) is the share of the number of exempt goods over the total number of goods belonging to a two-digit industry. This table presents all two-digits industries for which the share of exempted goods is non-zero.

In short, table 2 shows that low-complexity consumption goods are taxed a low rate, are more likely to be exempt from GST and are generally not reliant on imports. On the contrary, high-complexity intermediate goods are taxed a high rate, are generally not exempt and rely much on imported inputs. In order to see the implications of those statistics, a (fictional) example may be useful. Consider two manufacturing plants operating in the textile industry. One plant (plant A) produces an exempt good, the other plant (plant B) produces a non-exempt good for which input-tax credit can be claimed. Therefore, plant A effectively pays taxes on its inputs, unlike plant B. In that setting, plant A total cost will be high relative to plants B, despite being operating in the same sector. In order to gain competitiveness, plant A will find ways to reduce the amount of taxes paid by modifying its input mix towards less taxed inputs, which tend to be produced domestically. In that case, plant A would be more likely do import switching than plant

B, depending on the elasticity of substitution between inputs.

The goal of the empirical analysis will be to find evidence of import switching. A key question, then, is what have been the patterns of Indian plants, and particularly importers, in the years before and after the reform.

4.2 Indian manufacturing and global trade

This section documents some facts about formal manufacturing plants in India. Table 3 reports summary statistics of interest in 2015, prior to the reform, for different types of plants. Columns (1)-(3) present those statistics for all plants, whereas columns (4)-(6) restrict the analysis to importing plants only. Within each sample, results are further decomposed by the plant's ability to claim input tax under the forthcoming GST.

Unexpectedly, there are big differences between plants engaging in international trade and those which do not. Only 23% of Indian manufacturing plants are importers, and 15 % are exporters. Importing plants are larger, both in terms or sales and number of employees. They are more capital-intensive, reflecting greater reliance on capital inputs and presumably foreign technology. Logically, since they rely less on labor, importing plants have lower labor costs (lower value of Wages per sales_i). A key finding for the purpose of this study is the prevalence of adding and dropping inputs. Relative to the previous year, 28 % of all plants added at least one input, whereas for dropped inputs the figure is 26%.[13]. This phenomenon is stronger for importers as 39% and 37% of them add and drop inputs, respectively. Those results show that many plants add and drop products at a given year, a common result in empirical studies. For instance, **?** note the importance of the extensive margin for Indian firms in the 1990s, as firms start sourcing previously unavailable inputs. For Colombian importers, **?** find that approximately 60% of plants add inputs per year, the higher figure potentially due to the level of disaggregation at which their analysis is carried out (10-digit HS).

Another point worth noting is that while differences between the full sample and the sample restricted to importers are important, there does not appears to be many differences within those samples between plants able to claim input taxes under GST, and those ineligible. Note that this observation is reassuring for the identification strategy, as it illustrates no large differences prior to the reform between groups of plants which are subsequently differently impacted.[14]

Now turning to dynamic trends in plants' sourcing decisions, I find that in the past decade, the trend in India has been to source relatively less inputs internationally. This is true across both the extensive and the intensive margins. To see this, I follow a similar approach to **?** and **?**. Figure 1 plots the year dummies resulting from the estimation equation 1.

The time dummy coeffients are plotted relative to the year 2012. Panel (a) focuses on the

[13]Adding an input means sourcing a 5-digit CPC Rev. 2.1 input for the first time in 2015, and dropping an input means that 2015 is the last year the plant sourced that 5-digit CPC Rev. 2.1 input.

[14]Of course, they may still have differing trends, a concern which the identification strategy addresses.

Table 3: Plant characteristics in 2015, by ITC eligibility and import status

	All plants			Importing plants		
	All (1)	ITC (2)	Non-ITC (3)	All (4)	ITC (5)	Non-ITC (6)
Age_i	20.29	19.13	23.65	20.50	20.48	20.56
	(18.68)	(16.02)	(24.51)	(16.36)	(16.36)	(16.36)
Labor_i	221.64	226.84	206.63	463.21	475.17	423.11
	(726.23)	(787.44)	(509.54)	(937.64)	(924.54)	(979.60)
$\ln \text{Sales}_i$	18.55	18.44	18.87	20.37	20.41	20.25
	(2.33)	(2.37)	(2.16)	(1.67)	(1.67)	(1.66)
Capital per labor_i	12.62	12.53	12.87	13.30	13.34	13.17
	(1.89)	(1.96)	(1.65)	(1.50)	(1.48)	(1.55)
Wages per sales_i	0.30	0.35	0.18	0.14	0.15	0.14
	(9.97)	(11.54)	(1.60)	(2.05)	(2.31)	(0.60)
Importer_i	0.23	0.24	0.20	1.00	1.00	1.00
	(0.42)	(0.43)	(0.40)	(0.00)	(0.00)	(0.00)
Exporter_i	0.14	0.14	0.13	0.33	0.33	0.33
	(0.35)	(0.35)	(0.34)	(0.47)	(0.47)	(0.47)
Adds inputs_i	0.29	0.28	0.29	0.39	0.39	0.38
	(0.45)	(0.45)	(0.45)	(0.49)	(0.49)	(0.49)
Drops inputs_i	0.26	0.26	0.26	0.37	0.38	0.36
	(0.44)	(0.44)	(0.44)	(0.48)	(0.48)	(0.48)
Observations	27,205	20,207	6,998	6,215	4,788	1,427

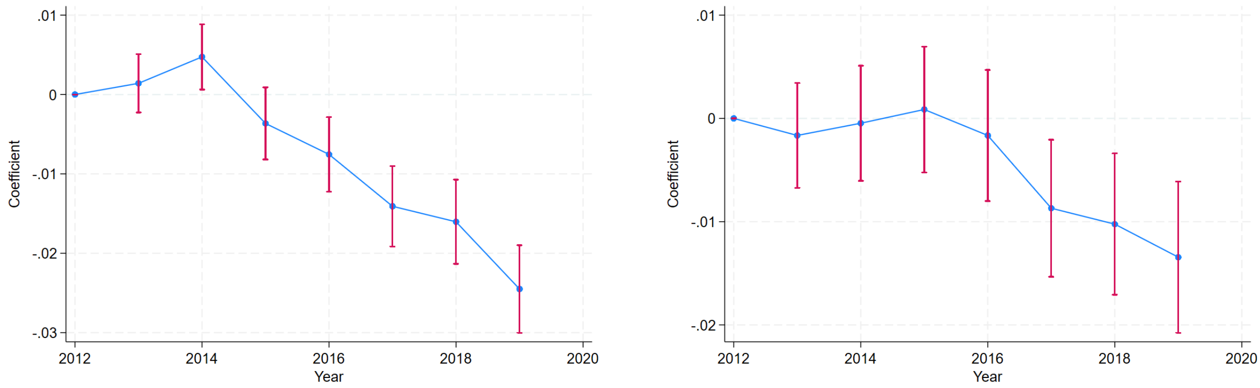
Notes: This table plots the main characteristics of manufacturing plants in the final sample for estimation. It reports means and standard deviations (in parentheses). *Importing plants* refers to plants with non-zero import value of intermediates in 2015. *ITC* takes value 1 if the plant produces a good which is taxed under GST and thus eligible for input tax credit, and zero otherwise. *Non-ITC* takes value 0 if *ITC* is equal to 1, and value 1 if *ITC* is equal to 0. Data from the Annual Survey of Industries, covering only plants registered under the Factories Act, 1947.

extensive margin of importing. It shows that on average, Indian manufacturing plants are less likely to import over time since 2014. This is an illustration of what has been characterized as India's "inward return" (**?**). Indeed, after 2014, in a shift from the policies of the 1990s and the 2000s, the country has implemented policies to promote domestic trade, rather than international trade.[15] Panel (b), on the other hand, focuses on the intensive margin, and therefore only focuses on importers. It shows that importers too import less on average relative to 2012, but the change only becomes significant in 2017 and subsequent years.

This section has documented three important findings. First, in 2017, one of the most significant reforms in the past decades in India has been the implementation of a nationwide indirect taxation scheme to unify the domestic market. That reform has made the sourcing of inputs less

[15]**?** note that this inward trend is exacerbated by the common belief in India that the Indian domestic market is large enough to sustain economic growth, and that the growth opportunities provided by international trade are limited.

Figure 1: Plants' average importing decisions



(a) Whether imports

(b) Import share if imports > 0

Notes: Panels (a) and (b) plot the year dummy coefficients on the following regression, following **?** and **?**:

$$\mathrm{M}_{it} = \alpha_0 + \beta_i + \gamma_t + X_{it} + \varepsilon_{it} \tag{1}$$

where the outcome variable, M_{it}, is a plant-level measure of engagement in international sourcing of inputs. α_0 is the intercept, β_i is a plant dummy, γ_t is a year dummy, X_{it} is a vector of time-varying plant-level characteristics which are relevant to the decision to source inputs. It contains wages-to-sales ratio (to account for changing labor costs) and capital-to-labor ratio (to account for capital deepening). For panel (a), M_{it} is a binary variable taking value one if the plant imports any input, and zero otherwise. For panel (b), M_{it} variable is the plant-level share of imported intermediate inputs over total intermediate inputs, and the sample is restricted to importing plants. Base year is 2012. Only manufacturing plants from the ASI dataset.

costly across states for plants eligible to tax refunds on inputs. This is not the case for ineligible plants, which still have to pay taxes on inputs, and thus have greater input costs relative to eligible plants. Second, import shares of importers has decreased in 2017 and after, relative to 2012. Third, adding and dropping inputs is a common phenomenon, and therefore is a potential margin of adjustment for plants when a shock occurs. The remainder of this paper will seek to establish a causal link between those three facts to show that plants in exempted sectors and with high input taxes switched from high- to low-taxed inputs after the reform, and, since imported technology tend to be taxed at a higher rate, from imported to domestic inputs.

5 Empirical strategy

This section details the identification strategy and the construction of the plant-level exposure to the policy.

5.1 Measuring plant exposure to the reform

A common challenge in empirical works seeking to estimate the effect of a policy which is implemented nationwide is to find a control and a treated group. In this section, I detail how I construct three types of treatment variables, based on two features of the reform. What those three types of treatment variables have in common is that they rely on the fact that some plants can claim refunds on the taxes paid on inputs, and some cannot. In order to distinguish between those two types of plants, I rely on two features of the design of GST.

The first feature I exploit is the fact that, depending on the type of product manufactured by the plant, the plant will be able to claim input tax credit or not. In other words, the 5-digit code the good manufactured by the plant belongs to determines the treatment status of the plant. If the plant produces a good within the scope of GST, then it is able to claim ITC on the tax paid on its inputs, and I consider it treated. If the plan produces a good outside the scope of GST (an exempt good), then I consider it as a control. In most cases, the plant's main product is the only product it makes, making this classification process straightforward. However, roughly one third of plants are multiproduct. For those plants, I consider the product which has the greater ex-factory value as the main product.

$$\text{ITC}_j^{GST} = \begin{cases} 0, & \text{if product } j \text{ is on the list of exempt products} \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

ITC_j^{GST} is the first type of explanatory variable that I use throughout the analysis. It is the most simple way to distinguish between a treated group and a control group: some plants producing a good eligible for ITC are treated, since the scheme did not exist prior to the reform, and plants producing exempt goods are controls, since for them the situation is as before (they pay taxes on inputs). Note that the source of identification when using this treatment variable only has two dimensions: (i) time and (ii) ability to claim ITC.

I refine measurement of exposure to GST by exploiting an additional feature of the GST reform: its multi-rate structure. That second feature allows me to compute the plant's average GST rate, which is a weighted average of the GST rate paid on each of the inputs which are entering in the plant's input mix. Previous works have shown the large degree of heterogeneity in the sourcing of inputs, even within narrowly-defined industries (**?**). The ASI provides information about plant-level production processes. Using this information, I construct this exposure measure as an input tariff, in the spirit of **?**. Intuitively, plants which were sourcing inputs that end up being assigned a higher tax level will have a higher input-weighted GST rate. The plant-level input-weighted GST rate is given by:

$$\text{InputTax}_i^{GST} = \sum_k \frac{\text{Value}_{ki}^T}{\sum_k \text{Value}_{ki}^T} \times \text{Rate}_k^{GST} \tag{3}$$

In equation 3, Value_{ki}^{T} is the average purchase value of input k sourced by plant i at a given initial period T. Input value is a simple average of the purchase value for that input during the initial period. In most regressions, T corresponds the first two years a plant appears in the dataset. For instance, if the plant appears first in 2012 and then in 2013, the variable Value_{ki}^{T} is simply $\frac{1}{2}\left(\text{Value}_{ki}^{2012} + \text{Value}_{ki}^{2013}\right)$.[16] The variable Rate_k is the GST rate applied on the sale of input k from 2017 onwards. Equation 3 defines a continuous variable. In addition, I discretize the exposure measures based on the relative value of InputTax_i^{GST} (see appendix B).

Combining the treatment variable ITC_i^{GST} with information about plant's average GST paid on inputs allows me to construct more refined treatment and control variables. In order to distinguish the basic treatment variable ITC_j^{GST} depending only on the output produced by the plant, I call those more complex treatment variables Exposure_i^{GST} and $\text{Exposure}_i^{GST,\ell}$. Those variables are defined as follows:

$$\text{Exposure}_{ij}^{GST} = \text{InputTax}_i^{GST} \times \text{ITC}_j^{GST} \tag{4}$$

$$\text{Exposure}_{ij}^{GST,\ell} = \text{InputTax}_i^{GST,\ell} \times \text{ITC}_j^{GST} \quad \forall \ell \in \{HIGH, MID, LOW\} \tag{5}$$

Note that those two additional exposure measures based on the plant's inclusion in the ITC scheme and its initial input mix allow me to exploit an additional source of variation. Here, identification relies on variation in (i) time, (ii) ability to claim ITC and (iii) plant-level input-weighted tax. Therefore, estimation in which $\text{Exposure}_{ij}^{GST}$ enters as explanatory variables introduce much more variation for identification.

It is important to well understand the similarities and differences between the treatment variables defined by 2, 4 and 5. They all define a measure of exposure to the that is based on ability or not to claim taxes paid on inputs. In all cases, those variables are equal to 0 for plants which are producing exempted goods. Those plants are the control group in this setting: before and after the reform, internal trade within India remains costly due to the different taxes and inability to claim taxes on inputs. Where those variables differ, however, is in the value assigned to plants producing non-exempted goods and which can to claim ITC. For variable $\text{Exposure}_{ij}^{GST}$, unlike typical difference-in-differences settings when units are assigned a binary outcome, treatment value is continuous (that is, treated units obtain a certain "dose" of treatment). That is, there will be (in theory) as many differnent values of treatment as there are plants because the input-weighted GST is unique to a plant. For $\text{Exposure}_{ij}^{GST,\ell}$, categories are created based on the relative value of $\text{InputTax}_{ij}^{GST}$, meaning that the variable can only take values 0 or 1. However, in that case, only a subset of plants for which input taxes can be claimed will be defined as treated (depending on the relative level of the input-weighted tax). Throughout the analysis, the main explanatory variable

[16] I choose those initial years because it is unlikely that plants made sourcing decisions in anticipation of a potential GST implementation. Before 2014, GST implementation appeared unlikely due to political divisions in the Lok Sabha, India's lower chamber of Parliament. There remained uncertainty about the rates as late as January 2017 (**?**). Additionally, I pick several years instead of one to mitigate potential year-specific volatility that might impact plants' input choices.

will either be ITC_j^{GST}, $\text{Exposure}_{ij}^{GST}$ or $\text{Exposure}_{ij}^{GST,\ell}$, depending on the type of empirical model.

5.2 Empirical frameworks

This section presents the three empirical frameworks used in this study to provide causal evidence of import switching at the plant-level. All frameworks rely on a difference-in-differences methodology. I start by investigating whether plants relatively more exposed to the reform have different sourcing decisions after 2017 than plants relatively less exposed. The first econometric model is carried out at the plant level and takes the following form:

$$\begin{aligned} y_{it}^{IMP} = \beta_0 + \beta_1 \text{Post}_t^{2017} + \beta_2 \text{ITC}_j + \beta_3 \text{Post}_t^{2017} \times \text{ITC}_j \\ + \gamma_i + \delta_t + X_{it} + \text{Trends}_{it}^{2012} + \varepsilon_{it} \end{aligned} \tag{6}$$

In most regressions, the outcome variable y_{it}^{IMP} is a measure of imported input sourcing in plant i at time t. It is regressed on Post_t^{2017} is a post-2017 dummy variable, ITC_j is the first measure of exposure to the reform: whether it is eligible for input tax refunds or not. In many specifications, we decompose the $\text{Post}_t^{2017} \times \text{ITC}_j$ in three categories, depending on the relative level of input-weighted GST paid by the plant. The vector X_{it} contains time-varying plant-level characteristics relevant in the import switching literature: capital per labor (which accounts for capital deepening), wages per sales (which accounts for changes in labor costs), and whether the plant exports, since importers also tend to be exporters (**?**). To check whether the results are not contaminated by pre-trends, a vector of initial characteristics are sometimes included, and we also estimate a dynamic panel DiD specification, which is detailed in equation B.4 of appendix B.3.

Despite providing with a useful first step, as well as a way to visually test for pre-trends, there are two identification issues that models 6 and its dynamic version do not fully address. The first issue has to do with the nature of the research question. Sourcing decisions tend to be highly serially correlated: past values influence current values and future values. In this setting, this means that a shock on imports in the previous period, such as a change in trade policy, may have modified plants' international sourcing decisions in the current and in future periods. That issue can be partially addressed by the inclusion of pre-trends of key variables of interest, such as past imports, by including as many time-varying plant-level and policy controls as possible. Another approach is to pool the regression between pre- and post-reform periods, as recommended by **?**. The second issue that model 6 may not fully address is the possibility of unobserved time-varying heterogeneity at the plant-level which may be correlated with the explanatory variable (selection into ITC) or the dependent variable (import trends): it is not possible to ensure that 6 accounts for those.

To address those two issues, I adopt an additional econometric model used in the literature on import switching, recently by **?** and **?**. The baseline is estimated in first-differences between pre and post-reform averages: it compares changes in average outcome Y_i before and after the reform, and thus addresses the issue of serial autocorrelation raised by **?**.

$$\Delta Y_i^{IMP} = \beta_0 + \beta_1 \text{ITC}_j + \beta_2 X_{i,pre} + \delta_{js} + \delta_d + \varepsilon_i \tag{7}$$

For plant i, ΔY_i^{IMP} is the log change of an outcome variable, typically imports. $X_{i,pre}$ includes pre-reform plant characteristics similar to those in specification 6. The inclusion of δ_{js} accounts for changes in 2d-product-state characteristics, thereby capturing difference in product-state aggregate trends (which could be an industrial policy decided by a particular state, for instance). Similarly, δ_d captures local trends, such as local labor market conditions or a climate shock. Note that a first-difference model such as model 7 cancels out time-invariant plant-level characteristics, as did model 6 did with the inclusion of a fixed effect. The value of model 7 becomes more obvious relevant when defining another outcome variable. The final econometric model is thus the following:

$$\Delta Y_i^{IMP} - \Delta Y_i^{IND} = \beta_0 + \beta_1 \text{ITC}_j + \beta_2 X_{i,pre} + \delta_{js} + \delta_d + \varepsilon_i \tag{8}$$

Model 8 is a double-difference framework: the outcome variable consists in the log change between two log changes. It has been used in the trade literature to assess the effect of single-even, binary policies (**?**). On the left-hand side, ΔY_i^{IMP} is the log change in imports by plant i between the pre- and post-reform periods averages, and ΔY_i^{IND} is a similar change for inputs sourced domestically. The outcome variable should therefore be interpreted as a change in plant imports relative to plant domestic sourcing. Model 8 is the most demanding specification in this study because it controls for a greater part of the time-varying unobserved heterogeneity at the plant-level, which was not the case under models 6 and 7, limiting the risk of omitted variable that would bias the estimates. Note that the specification here controls not only for plant time-invariant characteristics, but also time-varying: it accounts for different total sourcing (imports plus domestic inputs) trends between plants. The identifying assumption here is that, conditional on the controls, plants had similar trends in imports and domestic sourcing.

Finally, in all regressions, standard errors are clustered at the 5-digit main product which the plant manufactures, because it is the level of variation of the main explanatory variable, ITC_j.

5.3 Identification

I now discuss the identifying assumptions on which the empirical analysis is based. The multiplicity of tax rates and the existence of exempt products allow me to introduce heterogeneity in exposure to GST implementation. Identification thus relies on variation across three plant-level dimensions: (i) time, (ii) input-weighted tax rates, and (iii) product exemption. This approach assumes a quasi-natural experiment setting, providing a plausibly exogeneous shock to the Indian economy.

All plants over India - located in different areas, engaged in different activities, and at different stages of development - were impacted. The inclusion of the multiplicity of rates and of exempt products should not threaten identification. Of course, the existence of those features may itself be endogeneous: plants assigned higher GST rates This is unlikely to be the case. The reform took place not only suddenly (leaving no possibility for anticipatory effects), but also uniformly across space, making it hard for plants and sectors to negociate based on their own local interests with the central government.

Nonetheless, in the remainder of this section, i check that the reform truly is plausibly exogenous to plant. First, I investigate whether there are differing trends in the tax rate paid by exposed plants relative to non-exposed plants, before and after the reform. If it truly is capturing the implementation of GST, the coefficient attached to the exposure variable should not be statistically significant prior to the reform, and it should have stronger effects the higher the taxes paid on inputs are. I thus estimate the following equation, which is based on equation 6:

$$\text{TaxRate}_{it} = \beta_1 \text{Post}_t^{2017} + \beta_2 \text{ITC}_j + \beta_3 \text{Post}_t^{2017} \times \text{ITC}_j + \gamma_i + \delta_t + X_{it} + \text{Trends}_{it}^{2012} + \varepsilon_{it} \quad (9)$$

The outcome variable of equation 9, TaxRate_{it} is a measure of the average tax rate paid by plant i at time t. It is the share of total taxes reported by the plant on its supplies, over total sales. Table 4 reports the results. In columns (1) to (7), the dependent variable Tax per NS_{it} is the share of total taxes paid to total ex-factory sale value. In column (1), the coefficient associated to the $Post_t^{2017}$ dummy is negative, suggesting that on average all plants paid lower tax rates after the reform (as average tax rates decreased). On the contrary, ITC_j is associated to a positive coefficient, which is to be expected since the supply of exempt products is not taxed. The coefficient on the interaction between the two, $\text{Post}_t^{2017} \times \text{ITC}_j$, is the difference-in-differences estimator. It is negative and significant: after the reform, those plants which were producing a non-exempt good and could therefore claim input tax credit, paid relatively less taxes. Column (2) estimate a similar regression (the inclusion of plant and year fixed effects drops the coefficients for $Post_t^{2017}$ and ITC_{ij}). The interaction coefficient is still negative and significant. Columns (3) to (6) go further by decomposing the ITC_j variable in three categories based on the value of the plant-level input-weighted GST rate. In those columns, the results show that there is no statistically significant difference between plants which can and cannot claim input tax credit when plants use inputs that are taxed a low rate. For medium and high exposure measures, as the input-weighted tax increases, the difference between exempt and non-exempt plants matters much more: as plants' inputs are taxed a higher rate, plants which can claim refunds relatively pay less taxes. Those results are robust to further controls. In column (4), I check that results still hold when using an alternative measure of plant-level average tax rate.[17] In columns (5) and (6),

[17]That alternative measure is, for each plant, the sum of the input-weighted average GST rate and the output-weighted average GST rate. This alternative measure tests that the results are not driven by the taxes that plants pay based on their supply of goods, therefore ensuring that what matters triggers the lower tax rates here is on

I control for state-year FE and district-year FE to control for local shocks in the economy besides GST implementation.[18] Finally, in column (7), the outcome variable becomes Tax per GS_{it}, the share of total taxes paid by plant i over gross sale value. The robustness of the results with both net and gross sale value in the denominator prove that the results are not due to changes in distributive expanses other than taxes (for instance, greater transportation costs in the context of greater internal trade following 2017). Overall, this table suggest that GST implementation did impact taxes paid for treated plants, just as expected.

Table 4: The effects of GST implementation on average tax rate paid by plants

	Tax per NS_{it}						Tax per GS_{it}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post_t^{2017}	-0.004***						
	(0.001)						
ITC_j	0.018***						
	(0.003)						
$\text{Post}_t^{2017} \times \text{ITC}_i$	-0.007***	-0.011***					
	(0.002)	(0.003)					
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH}$			-0.017***	-0.019***	-0.019***	-0.012***	-0.015***
			(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID}$			-0.009***	-0.006**	-0.006**	-0.005***	-0.008***
			(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW}$			-0.003	-0.002	-0.002	-0.002	-0.002
			(0.004)	(0.005)	(0.005)	(0.003)	(0.003)
Plant FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes	No	No	Yes
State × year FE	No	No	No	No	Yes	No	No
District × year FE	No	No	No	No	No	Yes	No
X_{it}	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,586	186,918	186,918	186,918	186,918	117,201	186,918
R-squared	0.034	0.656	0.657	0.658	0.658	0.695	0.656

Notes: OLS estimation. Only manufacturing. Tax per NS_{it} is the plant-level total tax paid reported by plant i at time t over total sale value net of distributive expenses. Tax per NS_{it} is the plant-level total tax paid reported by plant i at time t over total gross sale value. X_{it} is a vector of time-varying plant characteristics containing log capital per employee, log wage per gross sales and a binary variable for export status. $Post_t^{2017}$ is a binary variable taking value 1 if the year is 2017 or after, zero otherwise. ITC_j is a binary variable taking value 1 if the plant manufactures a product that is part of the input tax credit scheme, and zero otherwise. $\text{Exposure}_{ij}^{GST,HIGH}$, $\text{Exposure}_{ij}^{GST,MID}$ and $\text{Exposure}_{ij}^{GST,LOW}$ are three-category discrete versions of the continous exposure variable defined in equation 3 (see appendix B). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

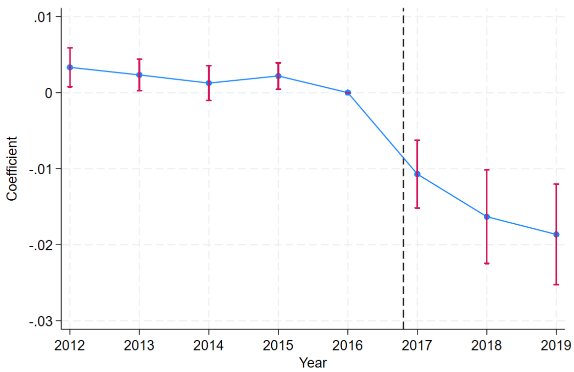
To interpret table 4, first note that the taxes reported by plants excludes taxes paid on inputs, which is crucial for this test. A key assumption for identification is that prior to the reform, there were no differing trends between treated and untreated plants in the tax rates paid. Table 4 addresses this issues in several ways. First, the inclusion of state-year and district-year FE controls

the input-side, not the output-side.

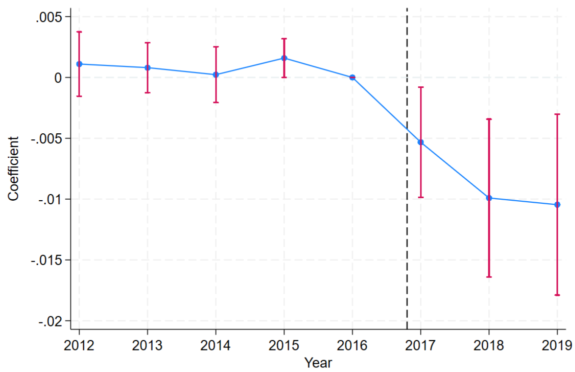
[18]The reduced number of observations in column (6) is due to the inclusion of district controls. The panel version of ASI never reports districts, but merging between the survey and the panel versions on a sufficient number of variables permitted to recover district information before 2010. Unfortunately, the ASI stopped reporting which district plants are located in after a law was passed to prevent plants from being identified. As a result, plants entering the panel after 2010 are not assigned a district, and are therefore dropped from estimation.

for policies changes, such as changes in tax rates, at the local level. In addition, the inclusion of the ITC_j dummy accounts for the fact that some products may have specificities which may increase their probability of selection in the input tax credit scheme However, those controls only partially accounts for the risk of pre-trends. Indeed, there may still be time-varying shocks at the product level, and at the local-product level, impacting taxes paid. For example, a single state may have decreased tax rates for a specific product in the years prior to the reform to promote domestic production for certain goods. This possibility is likely, but not dealt with by the local time-varying FE. To check the absence of pre-trends, I estimate a dynamic version of equation 9. As in column (3), which the main explanatory variable is decomposed by tax level as defined in 5 and thus and takes three forms, $Post_t^{2017} \times Exposure_{ij}^{GST,HIGH}$, $Post_t^{2017} \times Exposure_{ij}^{GST,MID}$ and $Post_t^{2017} \times Exposure_{ij}^{GST,LOW}$. The results are plotted in figure 2.

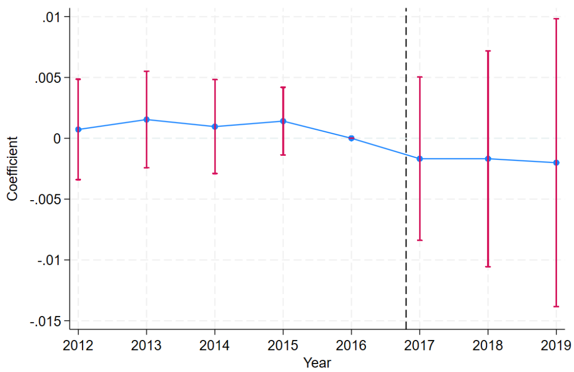
Figure 2: Plants' exposure to reform and average tax rate



(a) High



(b) Medium



(c) Low

Notes: Panels (a), (b) and (c) plot the $\text{Exposure}_{ij}^{GST,\ell}$ for $\ell \in \{HIGH, MID, LOW\}$ defined in equation 5 interacted with year dummies. Reference year is 2016 in all panels. The outcome variable is defined as the total of plant-level taxes paid over ex-factory value (net of distributive expenses) of total sales.

Figure 2 permits to rule out the existence of pre-trends between plants able to claim input tax credit under GST and plants unable to do so. Barring a small pre-trend in panel (a), the coefficient is not statistically significant from zero before the reform. After the reform, the coefficient is negative and statistically significant for panels (a) and (b), meaning that plants which have a high and medium input-weighted GST and which can claim the input tax credit pay relatively less taxes per sales than plants with a similar input-weighted GST. This is not the case in panel (c), again likely because when taxes are low the ability to claim tax refunds does not matter as much. An alternative results of the dynamic estimation of 6 is presented in appendix C, where I show that the overall effect (*i.e.*, not separating by relative exposure) is negative and significant only after 2017.

This section has shown that the implementation of GST appears to have been an exogeneous shock on the Indian economy. Some plants were impacted relatively more than others due to their inclusion or exclusion from the input tax credit scheme and their initial input mix, but that impcat was not due to other policies or time-varying shocks on products. In other words, GST implementation provides an arguably exogeneous source of variation in input costs for plants. The next section presents the main results of the empirical analysis: the effects of the reform on international input sourcing.

6 Effects on the sourcing of inputs

6.1 Baseline results

I now turn to the core part of the analysis, which is to provide causal evidence of import switching due to differential exposure of plants to the reform. I start with estimating the econometric model defined by equation 6 in which the outcome is a measure of international sourcing of the plant. Specifically, in column (1), the outcome is an importer dummy, and in column (2) it is the log of imports of intermediate inputs (so columns (2)-(6) exclude non-importers of intermediates from the estimating sample). On the right-hand side, the explanatory variable of interest is the same as columns (2)-(7) of table 4: a post-2017 dummy interacted with three categories of exposure variables as defined in 5. All columns include time-varying plant-level controls, captured by the vector X_{it}. I follow the literature on import switching and include controls that may have modified the relative need for imported products. X_{it} thus includes log capital per labor (capturing capital deepening) and wages per sales (reflecting changes in labor costs, which may be substitutable or complementary with imported inputs). Because the outcome variable is in log, X_{it} also contains a measure of plant size, the log of gross sale value. Finally, $\text{Trends}_{it}^{2012}$ accounts for pre-existing plant characteristics in 2012 interacted with year dummies to control for heterogeneous trends. In table 5, it is the 2012 value of total imports. All regressions include plant level FE, year-district FE and 2-digit product FE, to account for unobserved heterogeneity and time-specific shocks.

Table 5 establishes a link between eligibility to input tax refunds and imports for plants which face a relatively high level of GST on their inputs, but only in some cases. In column (1), the reform does not have a significant impact on the probability of plants to import, showing that in the context of internal trade barriers reductions, adjustments are not made on the extensive margin. Columns (2)-(5) focus on the intensive margin only by only keeping importers in the sample. In column (2), the effect of the reform is positive and significant, which is expected because plants which are treated, and can claim input tax credit, will be able to source high-taxed goods, many of which are complex and technological goods, at a relatively lower cost. More puzzling, this positive effect is only for low-taxed plants. Column (3) checks that this effect is not driven by changes in import competition, access to cheaper foreign inputs and changing demand for Indian goods by controlling for output tariff τ_{it}^{O}, τ_{it}^{I} and whether the plant is an exporter. Next in column (4), I control for total inputs sourced to focus on reallocations between imported and domestic inputs. The results are robust. Finally, in column (6), I include a plant-level trend, the log of imports in the initial year of the sample. The inclusion of that trend changes the results: treated plants still import relatively more, but the effect is driven by plants with high input taxes. This change in results may be explained by the fact that imports, especially for imports, is a highly serially-autocorrelated variable (past import habits determine greatly current imports). Controlling for past imports takes away some of that serial correlation.

Table 5: The effects of GST implementation on plant-level imports

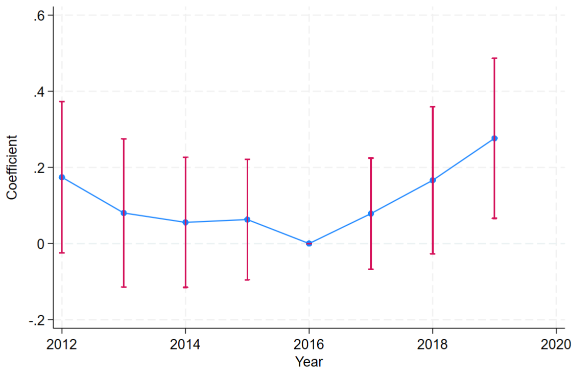
	Whether imports_{it}	$\ln \text{imports}_{it}$ if $\text{imports}_{it} > 0$			
	(1)	(2)	(3)	(4)	(5)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH}$	0.002	-0.058	0.067	0.055	0.268**
	(0.008)	(0.086)	(0.113)	(0.109)	(0.120)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID}$	-0.006	0.001	0.070	0.026	0.109
	(0.007)	(0.071)	(0.088)	(0.086)	(0.094)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW}$	0.010	0.247*	0.402**	0.341**	0.248
	(0.008)	(0.135)	(0.156)	(0.153)	(0.159)
τ_{it}^{I}			-0.482	-0.608	-1.729*
			(0.833)	(0.927)	(1.037)
τ_{it}^{O}			-0.092	-0.259	0.573
			(0.658)	(0.648)	(0.709)
Whether exports_{it}			0.040	0.032	0.040
			(0.042)	(0.042)	(0.042)
$\ln \text{Inputs}_{it}$				1.191***	1.144***
				(0.088)	(0.107)
ln Trend: $\text{Imports}_{it}^{2012}$					-0.037***
					(0.005)
X_{it}	Yes	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes
2d product × Year FE	Yes	Yes	Yes	Yes	Yes
District × Year FE	Yes	Yes	Yes	Yes	Yes
Observations	146,405	41,592	28,817	28,817	21,089
R-squared	0.768	0.735	0.757	0.763	0.758

Notes: OLS estimation. Only manufacturing. Whether imports_{it} is a binary variable taking value 1 if plant i imports any intermediate inputs at time t. $\ln \text{imports}_{it}$ if $\text{imports}_{it} > 0$, is the log of plant imports of intermediates by plant i at time t, and is only defined if plant imports are non-zero. $\text{Exposure}_{ij}^{GST,HIGH}$, $\text{Exposure}_{ij}^{GST,MID}$ and $\text{Exposure}_{ij}^{GST,LOW}$ are three-category discrete versions of the continous exposure variable defined in equation 3 (see appendix B). X_{it} is a vector of time-varying plant characteristics containing log capital per labor, wages per sales, log gross sale value and whether the plant is an exporter. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

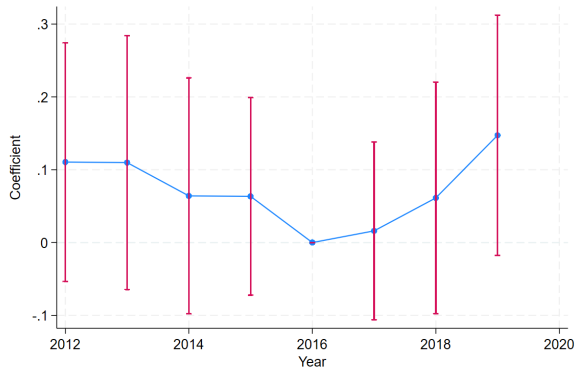
The results of column (5) of table 5 suggest that relative to plants which cannot claim input tax credit, plants which can tend to import more, especially if the taxes paid on inputs are high. If this results is taken causally, it provides an illustration of theoretical insight by (?) discussed above that the cost of traded inputs for the exempt sector would rise due to the inability to claim input tax credit.

A natural follow-up question is what drives this differential effect on imports: do credit-eligible plants source relatively more of the same inputs, or do they source more types of inputs? I estimate a dynamic version of 6 to produce 3. Each panel plots, by relative input-tax exposure, the effect of being eligible for input tax credit. The outcome variable is the number of directly imported products.[19]

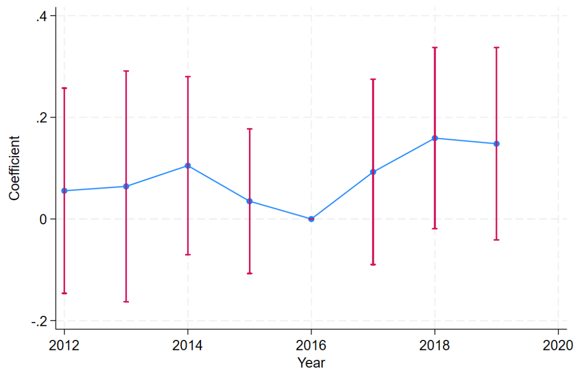
Figure 3: Plants' exposure to reform and average number of imported inputs



(a) High



(b) Medium



(c) Low

Notes: Panels (a), (b) and (c) plot the $\text{Exposure}_{ij}^{GST}$ defined in equation 5 interacted with year dummies. Reference year is 2016 in all panels. The outcome variable is defined as the number of imported products defined at the 5-digit level (CPC Version 2.1). Controls include input tariff, output tariff, export dummy, log capital per labor, wages per sales, log of imports, plant FE and state-year FE.

Again, the effect is positive and significant for highly-exposed plants in panel (a). Plants which can claim input tax credit import a relatively higher number of imported inputs. The effect is not significant for panels (b) and (c). Note that the results do not say that treated plants source more abroad after the reform. It is possible that imports increased, or decreased, for all types of plants, treated or not. The empirical approach simply allows us to say that there is only a differential effect between eligible and non-eligible plants after the 2017 reform. A possible explanation is that treated plants do not face increased costs like ineligible plants, so they may source foreign inputs as before. However, plants which cannot claim input tax credit now face greater input costs tif they source high-taxed goods. Since all plants within a narrowly-defined industry produce the

[19]Note that the Annual Survey of Industries does not report the origin country of plant imports. Therefore, it is impossible to know whether the policy had an impact on input variety.

same type of good (exempt or non-exempt), competition may not so much be between plants which can claim input tax credit and those which cannot, but rather sourcing decisions become a margin of adjustment for plants to cut costs to remain competitive. Since the GST rates were announced in 2017 and remained relatively unchanged since then, this means that ineligible plants had time to switch their sourcing decisions towards low-GST goods.

I also study how the reform modified the probability to drop inputs. To do so, I use information on inputs sourced by the plant over time and define added and dropped inputs. Adding and dropping is measured at the 5-digit. An input is added the first year it enters a plants' input mix (except if this is the year the plant enters the panel). Similarly, an input is dropped the last year it enters a plant's input mix except except if it is the last year the plant enters the panel. From that, I create plant-level dummy variables taking value 1 if a plant has added/dropped an input that year, and zero otherwise. I do so for total inputs and for categories of inputs. Categories are high-taxed and low-taxed goods, differentiated and homogeneous goods (following the classification by **?**), and capital, intermediate and consumption goods (using SNA classifications).For instance, suppose a plant has dropped only one 5-digit input: a highly-specific car component. With that, I gather information that the plant has dropped an input (dropped dummy takes value 1), a differentiated good, an intermediate input but not a consumption good (dropped consumption good dummy takes value 0).

Table 6: The effects of GST implementation on the probability to drop inputs

	Total	Tax rate		(?)		End-use		
VARIABLES		High	High	Differentiated	Homogeneous	Capital	Intermediate	Consumption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH}$	-0.060***	-0.062***	-0.060***	-0.064***	-0.017**	-0.047***	-0.069***	0.019*
	(0.018)	(0.010)	(0.010)	(0.013)	(0.007)	(0.010)	(0.016)	(0.010)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID}$	-0.053***	-0.008	-0.007	-0.032**	-0.015**	-0.008	-0.057***	-0.009
	(0.017)	(0.009)	(0.008)	(0.012)	(0.006)	(0.008)	(0.016)	(0.011)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW}$	0.008	0.003	0.001	0.016	-0.017**	0.002	0.019	0.002
	(0.033)	(0.015)	(0.015)	(0.016)	(0.008)	(0.009)	(0.031)	(0.014)
$\text{Add}_{it}^{GST,HIGH}$			0.121***	-0.004	0.000	-0.004	0.003	-0.008
			(0.008)	(0.008)	(0.004)	(0.005)	(0.009)	(0.005)
τ_{it}^I	0.464***	0.101***	0.101**	0.123*	0.072	0.046*	0.316***	0.155*
	(0.130)	(0.038)	(0.039)	(0.066)	(0.056)	(0.028)	(0.097)	(0.082)
τ_{it}^O	-0.080	-0.029	-0.017	-0.053	0.033	-0.021	-0.015	-0.132
	(0.200)	(0.060)	(0.062)	(0.127)	(0.038)	(0.084)	(0.170)	(0.104)
X_{it}	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	42,224	42,224	42,224	42,224	42,224	42,224	42,224	42,224
R-squared	0.367	0.276	0.286	0.308	0.238	0.353	0.349	0.279

Notes: OLS estimation. Only manufacturing plants that import. All outcome variables are binary variables taking value 1 if the plant drops a product belonging to the described category, and zero otherwise. $\text{Exposure}_{ij}^{GST,HIGH}$, $\text{Exposure}_{ij}^{GST,MID}$ and $\text{Exposure}_{ij}^{GST,LOW}$ are three-category discrete versions of the continous exposure variable defined in equation 3 (see appendix B). $\text{Add}_{it}^{GST,HIGH}$ takes value 1 if plant i added a high-taxed good in the period. X_{it} is a vector of time-varying plant characteristics containing log capital per labor, wages per sales, log gross sale value and whether the plant is an exporter. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Results of those regressions are presented in table 6. Column (1) shows that plants which are eligible for input tax credit are less likely than ineligible plants to drop inputs following the reform if their input mix is high, as is expected since they do not pay taxes on the sourcing of this input.

In column (2), I include input and output tariff controls. Results are very similar, highlighting the fact that the results are not driven by changes in trade policy. Next, in column (3), I regress a regression similar to column (2) except that it contains a $\text{AddedInputs}_{it}^{GST,HIGH}$ dummy variable, which takes value 1 if the plant has added a high-taxed GST that year. It ensures that the results are not just picking up the fact that plants add and drop high-GST inputs as a part of the typical switching process highlighted by ?. Then, in columns (4)-(5), I introduce heterogeneity in the type of goods sourced, focusing on the distinction between homogeneous goods and differentiated inputs. Differentiated inputs are not traded on organized exchanges not reference-priced, and are therefore less substitutatble with other inputs. Comparing columns (4) and column (5) indicates a substantial difference in the value of the coefficient attached to $\text{Exposure}_{ij}^{GST,HIGH}$: plants eligible for input tax credit are much less likely to drop differentiated inputs (relative to ineligible plants), than to drop homogeneous goods. This could be reflect the fact that eligible plants can afford not dropping an input for which they cannot find a substitute easily, whereas they will more easily drop homoegeneous inputs since they can be easily replaced. On the other hand, ineligible plants face short-term pressure to decrease cost and may have to drop inputs, even differentiated ones. In column (6)-(8), I study the probability to drop a product by end-use. Eligible plants are less likely to drop capital goods and intermediate goods, which tend to be high-taxed under GST. However, they are slightly more likely to drop consumption goods than plants ineligible for tax refunds. This result highlights that compared to eligible plants, uneligible plants do not drop low-taxed (consumption) goods that enter their production process, since alternatives may be taxed a higher rate and ultimately be more costly.

6.2 Robustness and sensitivity

Alternative specification.

Different initial exposure. Industry share of intermediates, input-weighted GST.

Similar initial sourcing decisions.

Contemporaneous policies.

The previous subsection has permitted to uncover that plants are differentially impacted by the tax reform, and in particular that they adapt their sourcing decisions accordingly. All those results have been based on econometric model 6. While this model presents a number of strengths, it cannot address some identification concerns as argued in 5. I now move on to test the robustness of those results to other empirial frameworks, defined by models 7 and 8, which control for greater unobserved heterogeneity as well as serial autocorrelation in the outcome variable.

The results are presented in table C.2. All results are for plants importing at least once prior to the reform and once after the reform. Columns (1)-(3) keep the total sample of importers. Column (1) shows that plants eligible to ITC are not more or less likely to import less in the post-period relative to the pre-period. However, in column(2), plants which can get tax credits on high taxes paid on inputs experience greater increases (or lesser decreases) in imports relative to ineligible plants. In column (3), the results are preserved and similar in size when controlling for the contemporaneous change in domestic sourcing. The results in column (3) suggest that inclusion in the ITC-scheme increased the imported-to-domestic sourcing growth rate by 15.2% for plants with a high input-weighted GST, relative to similar plants excluded from ITC. The remainder of the columns split the sample between plants which import less in the post-reform period (columns (4)-(5)) and those which import more (columns (6)-(7)). The results are only significant, and stronger, for plants which import less between both periods. It means that the ability to claim input tax credit is only associated to relatively higher imports growth when plants source less imported inputs overall. One way to interpret this result could be that the margin which I study here, differences in is really for plants which are struggling and already are on a downward part of their life cycle, looking to decrease costs. Plants which are increasing imports over the period may be in the process of expansion, looking for new international suppliers, may care less about taxes paid on inputs relative to the gains offered by greater international sourcing. Input tax credit may only play a big role for plants which are already costc-constrained. This suggests that the results I am picking up are not so much increases in imports for ITC-eligible plants, but really decreases in imports for ITC-non-eligible plants.

6.2.1 Alternative mechanisms

Selection into tax-paying. The results could be driven by the fact that some low-productivity plants operating informally, or at least outside the scope of tax authorities, may decide to start paying taxes following the reform. This is plausible for plants facing high taxes on inputs and which may prefer to pay taxes on their supplies but be able to claim input tax refunds. **?** show that VAT participation is costly, taking West Bengal's VAT system as an example. While they do not study directly GST implementation, they argue that it likely reduced the cost of VAT-registration due to greater opportunities to source inputs from other states. In that case, some firms may have decided to register under the new GST scheme. Table C.2 provides evidence of that intuition by plotting the density of plant turnover by year and by category of states (regular and special-category states have different tax registration thresholds). Note that in many cases GST implementation did not modify mandatory registration thresholds, but aligned with the state-level existing VAT-threshold. In both panels (a) and (b), we can see that there is less bunching just below registration thresholds: plants after GST implementation appear to have been more likely to operate at a level of sales which would make them become taxable.

My empirical analysis takes GST implementation as an exogeneous shock to the plant. To rule out reverse causality (the plant choosing its GST participation), all regressions are restricted to plants which operated above mandatory GST thresholds before and after the reform. That way, the final sample does not contain plants switching tax registration status. Moreover, the parts of the analysis which reduce the sample to importers are even less likely to be contaminated by this potential effect as importers tend to be very large and cannot escape tax authorities (if they do not pay taxes, it will be because of legal exemptions, such as zero-rating for exporters, rather than indicating informal status).

Selection into ITC-eligible products. Some plants, anticipating or after the reform, may opt for producing non-exempt products to be able to claim input tax credit and minimize costs. Empirical works have shown that in India, plants add products in response to changes in input sourcing conditions (**?**). In particular, **?** have shown that the decision to produce goods depends on plants' initial input mix. Their analysis, which does not focus on tax policy, nonetheless suggests that plants with a given input mix and producing an exempt good may decide to enter the production of exempt goods, thus creating an endogeneity issue by which plants opt into the input-tax credit scheme. I address this issue in two ways. First, in all results, the exposure variable is fixed in the initial period way before the 2017 reform. This is true for the input mix, but also for the choice of good produced. This approach effectively rules out the possibility of endogeneity. A specific issue remains in the case of multi-product plants. In that case, the plant is eligible for input tax credit if the main product it makes in terms of value is in the scheme. As a robustness check, I drop all multi-products plants from the analysis, since they are more likely to add and drop products. Results are presented in table C.3. The main result that plants which are eligible for input tax credit are more likely to import if their input mix is biased towards are taxed goods still holds.

Subsequent amendments to GST. Following the reform, India announced several modifications to the 2017 policy by periodically announcing new rates and/or exemptions. I do not take into account those amendments, because they may emanate from plants or industries negotiating lower rates or seeking (non-)exemption. Omitting this time-varying nature of GST is unlikely to be an issue here. First, because many amendments actually are for services, not goods, and as a result should not impact directly manufacturing plants. In addition, most other changes are announcements concerning supply to specific organizations, such as the state or charities; thefore, they do not apply to specific inputs in the plant's input mix or to specific products made by plants. Rather, they apply across all plants in the panel that serve that type of client. Finally, gathering data about amendments for goods reveals that a very small number of goods have changed rates and/or exemption status. At the plant-level, out of 30,000 plants, only a few dozens change exemption status overall, for instance. Out of those changes, most did not take place before January 1st 2019 and October 1st 2019, at the very end of the period of analysis, making those changes unlikely to influence the results.

Contemporaneous shocks. Two policies were implemented in the years 2012-2020 which may have had a confounding effect on the results: demonetization, and the India-ASEAN trade agreement.

Demonetization. In 2016, financial conditions changed suddenly when the Indian governement announced that it would demonetize INR 500 and INR 1000 banknotes of the Mahatma Gandhi series. The goal was to fight against the shadow economy and to promote cashless transactions. Financial conditions conditions matter for import sourcing decisions. **?** carry out robustness checks to rule out the possibility that their results are driven by financial frictions in the midst of the Great Financial Crisis.[20] In addition, **?** show that this shock increased tax reporting because, in the context of limited cash, firms used electronic forms of payment which are more difficult to hide from authorities. In my research setting, I argue that this is not an issue because I focus on formal manufacturing plants which were registered, were above a relatively high size threshold, and have likely not been as impacted as informal firms by demonetization.

Trade agreements. In most regressions, I control for the average world average MFN input and output tariff constructed at the plant-level, based on the plants' input and output mix. During the period 2012-2020, India maintained a relatively protectionist trade policy and did not modify its average world MFN output tariffs, as shown in appendix A. However, there is one exception which MFN tariffs do not capture: the India-ASEAN trade agreement, signed in the previous decade, which was implemented in the 2010s. There could be a concern that more integrated trade with Asian partners increased competition with countries producing similar goods as India, and that may have forced plants to reduce costs by import switching. There is also a possibility that some Indian plants may have decided to source new inputs from those countries. It is likely not the case, given the general inward trend in India discussed above, but also because India's trade agreements tend to be quite shallow and ultimately do not integrate domestic firms in global value chains **?**. Nonetheless, I test the robustness of table 6 to the inclusion of the plant-level exposure to ASEAN preferential tariff change. Results are presented in C.4, and are virtually unchanged.

6.3 Heterogeneity

I also check whether there are heterogeneous effects underlying the ability of ITC. There could be several factors which may drive a plants' ability to switch inputs.

Institutional quality.

Sourcing of international inputs.

First, research has shown that court quality, which varies by state, may impact sourcing decisions (**?**). Indeed, seller-buyer relationship may be less easy to forge in states with low court quality, for instance where there is corruption or where disputes take years to be settled. In those states, we would expect a smaller effect of the ability to claim ITC because plants able to claim

[20] They focus on Spanish firms in the late around in the late 2000s-early 2010s. Access to credit, especially in Southern Europe, is likely to have been impacted by the crisis.

taxes on inputs may be reluctant to switch from current suppliers to new potential suppliers because of a potential dispute (*e.g.*, about invoice permitting ITC). Another source of heterogeneity could be the type of output produced. Plants producing highly-specific goods may have greater market power, since their product is differentiated. Their clients cannot easily find substitutes, and so they may not negociate lower prices (**?**). Therefore, we would expect that the difference between ITC-eligible and non-eligible plants would be greater for differentiated goods, since ITC-eligible plants in those sectors face less demand-side constraints to reduce costs and can source more expensive international inputs relatively. Finally, we would expect greater adjustments following GST in districts bordering other states. The removal of inter-state barriers and the establishment of a well-functioning cross-state ITC system should benefit them particularly, since firms in other states are closer.

Results are presented in C.1 in the appendix. Columns (1) and (2) show that the gains of ITC-eligible plants relative to non-eligible plants are smaller in states with low court quality, as expected. In columns (3) and (4), the type of output produced does not generate statistically significant differences for the number of imported inputs, but ITC-eligible plants producing a specific input appear to be even less likely to drop high-taxed goods after the reform, as expected. Finally, columns (5) and (6) do not suggest any differential response by location of the firm next to a state border.

7 The role of the reform on production

8 Import switching, quality and productivity

In this section, I do two things. First, I estimate quality of imported goods as done by **?**. First, I structurally estimate the productivity losses associated by import switching **?**. Second, I estimate quality of imported goods as done by **?**.

9 Conclusion

The design of tax policy has implications in terms of domestic and international trade flows. The main message from this paper is that complex VAT designs featuring tax-exempt goods and multiple rates can impact plants differently. Plants which cannot claim refunds on input VAT are more likely to switch between high-taxed goods to low-taxed goods, and by doing so, to import less, as higher-taxed goods tend to be sourced internationally. The results are robust to alternative explanations and specifications. Those findings have policy implications. For countries seeking to implement a nationwide VAT, granting exempt status to some industries can be a tempting option, both because it is easy to implement and because it means an output VAT of zero.

Both output VAT and input VAT should be considered when deciding VAT exemptions for a given sector. Indeed, exemption disrupts the input tax credit chain, potentially undermining the reform's intended efficiency gains by reintroducing cascading taxes, and distorting production decisions. Further empirical research is needed to uncover the potential productivity losses and misallocation induced by imperfect VAT systems.

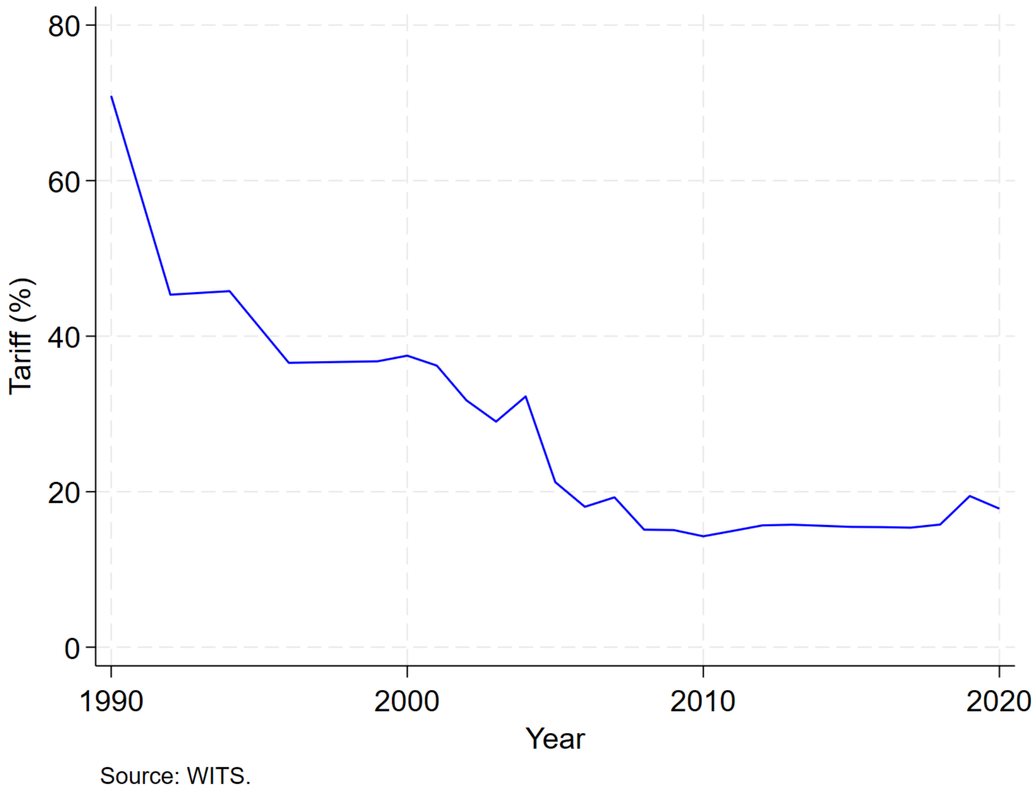
References

Appendices

A Institutional background

This section provides background information the institutional background of the reform.

Figure A.1: Average applied tariff imposed by India on European and US goods, 1990-2020



B Data and descriptive statistics

This section provides additional details about the datasets used and how variables were created.

B.1 Datasets

Table B.1: Summary of data

Dataset	Source	Description	Years
Annual Survey of Industries (ASI)	MOSPI	Survey data on formal manufacturing plants' production processes (inputs and products). Also contains information on employment, revenue, expenditures, and capital stock.	2012-2013 to 2019-2020
GST rates	Gazette of India	Official announcement of the Integrated GST (IGST) that is levied on inter-state supplies of goods and services. Non-exempt products and their rates are listed in the document "Notification No.1/2017", and exempt producs are listed in the document "Notification No.2/2017". Changes in rates and exempt products are compiled using subsequent notifications.	2017-2019
Tariffs	WITS	Average world MFN tariff; preferential tariff for the India-ASEAN trade agreement (2010)	2012-2019
Good specificity	**?**	Classification of SITC 4-digit industries in three categories (products traded organized exchanges, reference priced products, and differentiated products). Merged to CPC Rev.2.1 using concordance tables from UNSTATS.	N/A
Court quality	Daksh India	Data on pending High Court cases retrived from **?**	2002

B.2 Variable construction

Relative GST exposure. Plant-level exposure the the reform is constructed following equation 3. It defines a continuous variable. From this, I create categorical variables with either two, three, four or five categories. The classification process is based on the quantile of the plant's input-weighted tax rate. For example, for two categories (high exposure and low exposure):

$$\text{InputTax}_i^{GST,HIGH} = \begin{cases} 1, & \text{if } \text{InputTax}_i^{GST} > \text{median}\left[\text{InputTax}_i^{GST}\right] \\ 0, & \text{otherwise} \end{cases} \tag{B.1}$$

$$\text{InputTax}_i^{GST,LOW} = \begin{cases} 1, & \text{if } \text{InputTax}_i^{GST} \leq \text{median}\left[\text{InputTax}_i^{GST}\right] \\ 0, & \text{otherwise} \end{cases} \tag{B.2}$$

The process is similar when there are more categories. With three categories, a plant has low exposure when its $\text{InputTax}_i^{GST,HIGH}$ is below or equal to the first tercile; medium exposure when it is strictly above the first tercile but below the second tercile; and high exposure when it is strictly above the second tercile.

Court congestion. Data on state-level court congestion was compiled by Daksh India and retried from **?**. They define a continuous measure of state-level congestion of courts based on the number of pending cases in the state's main courts in 2002. The states for which they have information are Andhra Pradsh, Bihar, Chandigarh, Delhi, Gujarat, Haryana, Himachal Pradesh, Jammu and Kashmir, Karnataka, Kerala, Maharashtra and Punjab. I define a dummy variable Cong^s for relative high level of congestion, which is defined as:

$$\text{Cong}^s = \begin{cases} 1, & \text{if } \text{CongIndex}_s > \text{median}\left[\text{CongIndex}_s\right] \\ 0, & \text{otherwise} \end{cases} \tag{B.3}$$

Similarly, the dummy for non-congested states is defined as $\text{NotCong}^s \equiv 1 - \text{Cong}^s$.

Industry specificity. Industries are classified as specific if they are coded as "differentiated" in the classification by **?**, that is, if they are assigned the letter "n", and as non-specific otherwise.

B.3 Dynamic panel DiD specification

For plant i at time t, the dynamic panel difference-in-differences specification is given by:

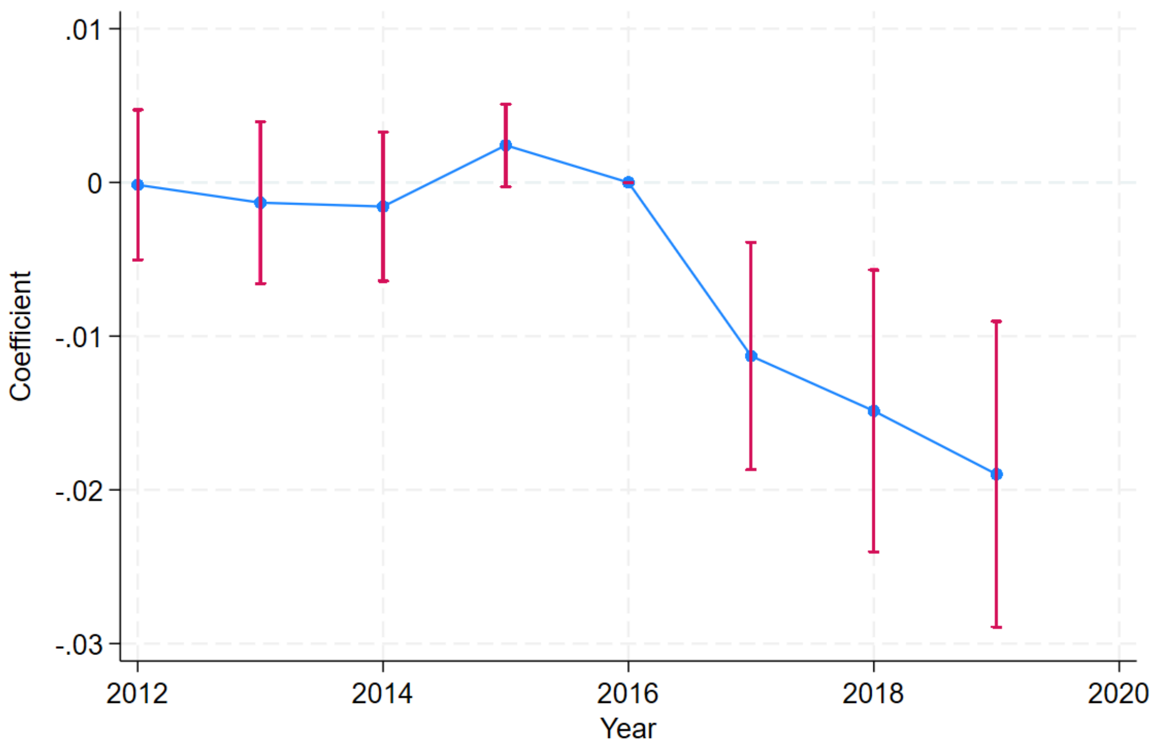
$$y_{it} = \beta_0 + \beta_s \sum_{\substack{s=2012 \\ s \neq 2016}}^{2019} (\mathbb{1}_{s=t}) \times \text{ITC}_i + \gamma_i + \delta_t + X_{it} + \text{Trends}_{it}^{2012} + \varepsilon_{it} \tag{B.4}$$

Variables in equation B.4 are defined similarly as in equation 6, and $(\mathbb{1}_{s=t})$ is a dummy variable taking value 1 if the indexed year s is the same as the observed year t, and zero otherwise.

C Robustness

This section provides additional regression results and robustness checks.

Figure C.1: Plants' exposure to reform and average tax rate



Notes: This figure plots the ITC_i^{GST} defined in equation 4 interacted with year dummies. Reference year is 2016 in all panels. The outcome variable is defined as the total of plant-level taxes paid over ex-factory value (net of distributive expenses) of total sales.

Table C.1: The heterogeneous effects of GST implementation on plant sourcing

	Court congestion		Output specificity		State border	
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	#Imp. inputs$_{it}$	Drop$_{it}^{GST}$	#Imp. inputs$_{it}$	Drop$_{it}^{GST}$	#Imp. inputs$_{it}$	Drop$_{it}^{GST}$
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH} \times \text{NotCong}_s$	0.170***	-0.071***				
	(0.066)	(0.019)				
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH} \times \text{Cong}_s$	0.009	-0.057***				
	(0.039)	(0.011)				
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID} \times \text{NotCong}_s$	0.049	-0.018				
	(0.058)	(0.015)				
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID} \times \text{Cong}_s$	-0.033	-0.005				
	(0.035)	(0.009)				
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW} \times \text{NotCong}_s$	0.182	0.017				
	(0.125)	(0.024)				
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW} \times \text{Cong}_s$	0.014	-0.002				
	(0.047)	(0.016)				
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH} \times \text{NotSpec}_j$			0.012	-0.053***		
			(0.038)	(0.012)		
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH} \times \text{Spec}_j$			0.086	-0.071***		
			(0.062)	(0.014)		
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID} \times \text{NotSpec}_j$			0.019	-0.015		
			(0.036)	(0.009)		
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID} \times \text{Spec}_j$			-0.066	0.002		
			(0.041)	(0.010)		
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW} \times \text{NotSpec}_j$			0.059	0.008		
			(0.055)	(0.021)		
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW} \times \text{Spec}_j$			0.022	-0.008		
			(0.069)	(0.015)		
$\text{Post}_t^{2017} \times \text{Exposure}_{i}^{GST,HIGH} \times \text{NotInternal}_d$					0.087	-0.065***
					(0.071)	(0.020)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH} \times \text{Internal}_d$					0.002	-0.066***
					(0.052)	(0.013)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID} \times \text{NotInternal}_d$					0.027	-0.004
					(0.073)	(0.013)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID} \times \text{Internal}_d$					-0.071*	-0.017*
					(0.041)	(0.011)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW} \times \text{NotInternal}_d$					-0.000	0.028
					(0.081)	(0.028)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW} \times \text{Internal}_d$					0.064	0.004
					(0.063)	(0.019)
$\text{Add}_{it}^{GST,HIGH}$		0.121***		0.121***		0.121***
		(0.008)		(0.008)		(0.010)
Plant FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
X_{it}	Yes	Yes	Yes	Yes	Yes	Yes
τ_{it}	Yes	Yes	Yes	Yes	Yes	Yes
Observations	42,218	42,224	42,218	42,224	29,653	29,653
R-squared	0.732	0.286	0.732	0.286	0.730	0.277

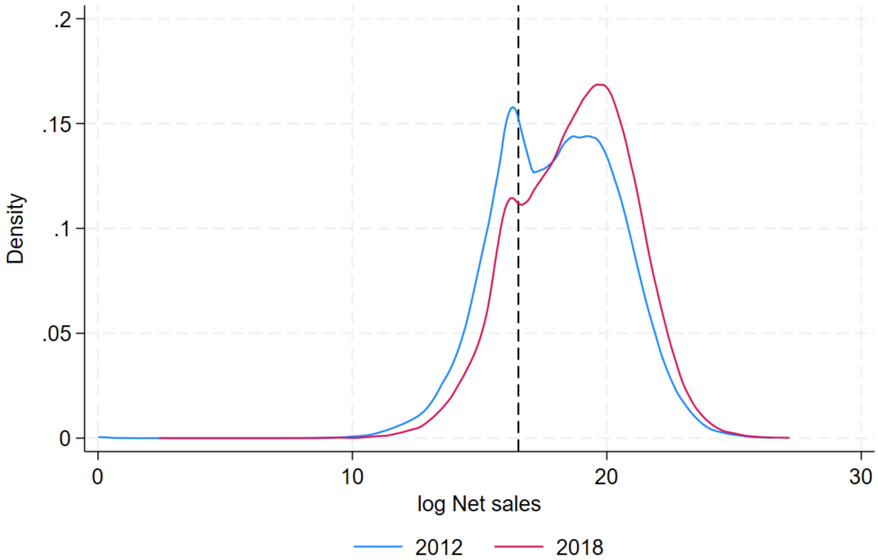
Notes: OLS estimation. Only manufacturing plants that import. Outcome variables are #Imp. inputs$_{it}$ (the number of inputs imported by plant i at time t) and Drop$_{it}^{GST}$ (a dummy variable taking value 1 if plant i has dropped a high-GST-taxed inputs at time t, and 0 otherwise). Exposure$_i^{GST,HIGH}$, Exposure$_i^{GST,MID}$ and Exposure$_i^{GST,LOW}$ are three-category discrete versions of the continous exposure variable defined in equation 3 (see appendix B). Cong$_s$ is a dummy variable taking value 1 if state s has above median congestion value and 0 otherwise. Spec$_j$ is a dummy variable taking value 1 if industr yj is classifed as being differentiated and 0 otherwise. Internal$_d$ is a dummy variable taking value 1 if district d is borders another state and 0 otherwise. X_{it} is a vector of time-varying plant characteristics containing log capital per labor, wages per sales, and whether the plant is an exporter. τ_{it} are input and output tariff constructed using the plant's initial input mix. Add$_{it}^{GST,HIGH}$ is a dummy variable taking value 1 if plant i has added high-taxed inputs in year t. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.2: The effects of GST implementation on plant-level imports, alternative specification

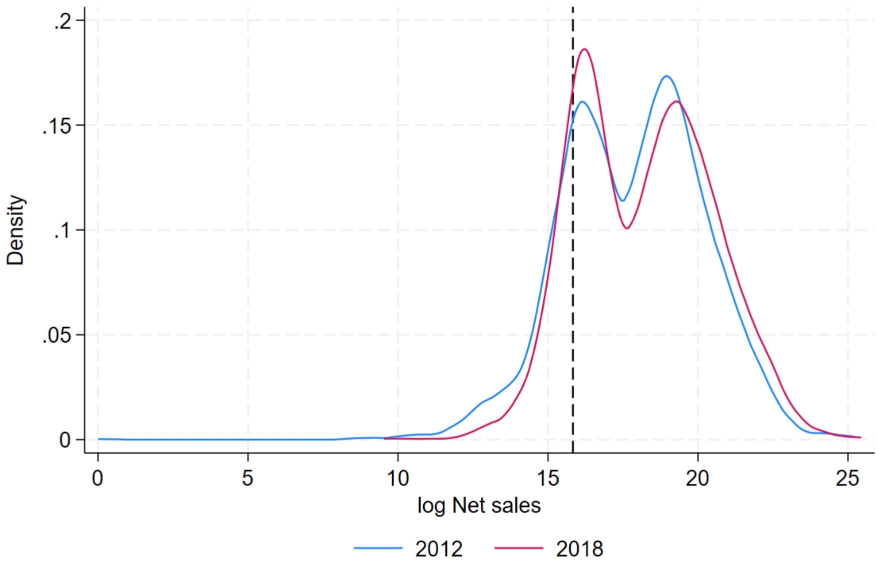
VARIABLES	All importing plants $\mathbb{1}\left[\Delta Y_i^{IMP} < 0\right]$ (1)	ΔY_i^{IMP} (2)	$\Delta Y_i^{IMP} - \Delta Y_i^{IND}$ (3)	Plants with $\Delta Y_i^{IMP} < 0$ ΔY_i^{IMP} (4)	$\Delta Y_i^{IMP} - \Delta Y_i^{IND}$ (5)	Plants with $\Delta Y_i^{IMP} > 0$ ΔY_i^{IMP} (6)	$\Delta Y_i^{IMP} - \Delta Y_i^{IND}$ (7)
$\text{Exposure}_{ij}^{GST,HIGH}$	-0.028	0.163**	0.152*	0.327**	0.380**	0.060	0.036
	(0.036)	(0.082)	(0.086)	(0.130)	(0.155)	(0.111)	(0.111)
$\text{Exposure}_{ij}^{GST,MID}$	-0.014	0.058	0.026	0.218*	0.161	-0.052	-0.044
	(0.031)	(0.074)	(0.076)	(0.114)	(0.142)	(0.099)	(0.099)
$\text{Exposure}_{ij}^{GST,LOW}$	0.022	0.141	0.062	0.283	0.175	-0.029	-0.116
	(0.048)	(0.125)	(0.138)	(0.233)	(0.281)	(0.156)	(0.175)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × 2d-product FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,541	3,455	3,415	1,325	1,305	1,919	1,903
R-squared	0.160	0.176	0.158	0.294	0.306	0.251	0.254

Notes: OLS estimation. Only manufacturing single-product plants. $\mathbb{1}\left[\Delta Y_i^{IMP}\right]$ is a dummy variable taking value 1 if the plant imports less in the post-reform period, and zero otherwise. ΔY_i^{IMP} is the log change of average imports before and after the period. ΔY_i^{IND} is the log change of average domestic input sourcing before and after the period. $\text{Exposure}_{ij}^{GST,HIGH}$, $\text{Exposure}_{ij}^{GST,MID}$ and $\text{Exposure}_{ij}^{GST,LOW}$ are three-category discrete versions of the continous exposure variable defined in equation 3 (see appendix B). X_{it} is a vector of time-varying plant characteristics containing log capital per labor, wages per sales, log gross sale value and whether the plant is an exporter. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure C.2: Plants' self-selection into taxpaying



(a) Regular states



(b) Special states

Notes: Turnover is proxied with ln net sale vlue of the plant. Dotted line is the turnover threshold in the state. Special states are Arunachal Pradesh, Assam, Himachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura, Uttarakhand, and Telangana. Regular states are all remaining states.

Table C.3: The effects of GST implementation on plant-level imports, excluding multi-product plants

	Whether imports_{it}	$\ln \text{imports}_{it}$ if $\text{imports}_{it} > 0$			
	(1)	(2)	(3)	(4)	(5)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,HIGH}$	0.005	-0.067	0.121	0.084	0.352*
	(0.012)	(0.154)	(0.170)	(0.166)	(0.209)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,MID}$	0.000	-0.047	0.163	0.107	0.301
	(0.011)	(0.138)	(0.157)	(0.156)	(0.197)
$\text{Post}_t^{2017} \times \text{Exposure}_{ij}^{GST,LOW}$	0.004	0.216	0.355	0.298	0.153
	(0.010)	(0.200)	(0.246)	(0.243)	(0.247)
τ_{it}^{I}			-0.454	-0.596	-0.384
			(1.145)	(1.276)	(1.322)
τ_{it}^{O}			-1.427	-1.441	0.190
			(0.987)	(0.884)	(0.730)
Whether exports_{it}			0.082	0.080	0.070
			(0.062)	(0.061)	(0.062)
$\ln \text{Inputs}_{it}$				1.085***	0.984***
				(0.124)	(0.136)
ln Trend: Imports_i^{2012}					-0.035***
					(0.008)
2d product × Year FE	Yes	Yes	Yes	Yes	Yes
District × year FE	Yes	Yes	Yes	Yes	Yes
Observations	80,976	20,906	13,396	13,396	9,721
R-squared	0.780	0.748	0.782	0.788	0.791
Plant FE	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimation. Only manufacturing single-product plants. Whether imports_{it} is a binary variable taking value 1 if plant i imports any intermediate inputs at time t. $\ln \text{imports}_{it}$ if $\text{imports}_{it} > 0$, is the log of plant imports of intermediates by plant i at time t, and is only defined if plant imports are non-zero. $\text{Exposure}_{ij}^{GST,HIGH}$, $\text{Exposure}_{ij}^{GST,MID}$ and $\text{Exposure}_{ij}^{GST,LOW}$ are three-category discrete versions of the continous exposure variable defined in equation 3 (see appendix B). X_{it} is a vector of time-varying plant characteristics containing log capital per labor, wages per sales, log gross sale value and whether the plant is an exporter. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: The effects of GST implementation on plant-level imports, accounting for India-ASEAN trade agreement

	All importing plants			Plants with $\Delta Y_i^{IMP} < 0$		Plants with $\Delta Y_i^{IMP} > 0$	
VARIABLES	$\mathbb{1}\left[\Delta Y_i^{IMP} < 0\right]$	ΔY_i^{IMP}	$\Delta Y_i^{IMP} - \Delta Y_i^{IND}$	ΔY_i^{IMP}	$\Delta Y_i^{IMP} - \Delta Y_i^{IND}$	ΔY_i^{IMP}	$\Delta Y_i^{IMP} - \Delta Y_i^{IND}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\text{Exposure}_{ij}^{GST,HIGH}$	-0.036	0.169**	0.137	0.327**	0.404**	0.086	0.071
	(0.037)	(0.081)	(0.084)	(0.132)	(0.158)	(0.114)	(0.114)
$\text{Exposure}_{ij}^{GST,MID}$	-0.011	0.032	0.027	0.181	0.142	-0.062	-0.027
	(0.033)	(0.075)	(0.074)	(0.111)	(0.142)	(0.105)	(0.104)
$\text{Exposure}_{ij}^{GST,LOW}$	-0.010	0.155	0.082	0.242	0.116	-0.011	-0.076
	(0.050)	(0.131)	(0.147)	(0.249)	(0.297)	(0.173)	(0.197)
$\ln \text{Capital per labor}_{i,pre}$	-0.011	0.028	0.050*	0.022	0.048	0.050	0.044
	(0.011)	(0.025)	(0.028)	(0.047)	(0.045)	(0.032)	(0.036)
$\ln \text{Wages per GS}_{i,pre}$	0.052	0.083	0.421	-0.176	0.343	0.398	0.347
	(0.121)	(0.197)	(0.262)	(0.439)	(0.560)	(0.302)	(0.370)
$\ln \text{Intermediates over inputs}_{i,pre}$	0.151**	-0.634***	-0.096	-0.891***	-0.611*	-0.637***	-0.253
	(0.074)	(0.167)	(0.216)	(0.311)	(0.357)	(0.209)	(0.251)
$\ln \text{Exports}_{i,pre}$	0.000	-0.000	0.000	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\Delta\tau_i^{O,ASEAN}$	0.226	-1.153	-1.877	-2.228	-1.938	-0.089	-1.770
	(0.545)	(1.490)	(1.432)	(2.531)	(2.440)	(1.935)	(2.264)
$\Delta\tau_i^{I,ASEAN}$	0.173	-0.823	-1.352	-1.716	-2.484	0.474	-0.039
	(0.527)	(1.664)	(1.768)	(2.783)	(2.693)	(2.349)	(2.502)
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State × 2d-product FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,197	3,129	3,098	1,213	1,202	1,726	1,710
R-squared	0.158	0.185	0.185	0.295	0.304	0.263	0.266

Notes: OLS estimation. Only manufacturing single-product plants. $\mathbb{1}\left[\Delta Y_i^{IMP}\right]$ is a dummy variable taking value 1 if the plant imports less in the post-reform period, and zero otherwise. ΔY_i^{IMP} is the log change of average imports before and after the period. ΔY_i^{IND} is the log change of average domestic input sourcing before and after the period. $\text{Exposure}_{ij}^{GST,HIGH}$, $\text{Exposure}_{ij}^{GST,MID}$ and $\text{Exposure}_{ij}^{GST,LOW}$ are three-category discrete versions of the continous exposure variable defined in equation 3 (see appendix B). X_{it} is a vector of time-varying plant characteristics containing log capital per labor, wages per sales, log gross sale value and whether the plant is an exporter. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$