The Demand Drivers of (Unequal) Growth:

Evidence from 25 years of structural change in India

Jonas Gathen* Gautam Narayanan†

August 19, 2026

Abstract

How important is domestic demand in driving and directing growth over the course of development? And given that demand for different products varies strongly over the income distribution, how biased is demand-driven growth towards the rich? We study these questions quantitatively in the context of India's economic development since the 1980s. Leveraging household consumption data, we show that cross-sectional Engel curves predict future changes in product demand. Using this 'Engel demand shifter' as instrument and linking consumption to production data, we quantify how demand drives productivity growth and structural change. Linking the empirical results to a tractable multi-industry endogenous growth model, we show that domestic demand drove around 30% of India's growth since the 1980s. However, growth is strongly biased towards relative luxury products, leading the rich to experience 25% higher real consumption growth. Redistribution can offset part of the pro-rich growth bias by directing technical change towards products the poor consume more.

*Tilburg University, Email: j.gathen@tilburguniversity.edu.

†Toulouse School of Economics (TSE), Email: gautam.narayanan@tse-fr.eu. We are grateful for helpful discussions and feedback from Matteo Bobba, Martí Mestieri, Ezra Oberfield, Michael Peters, Tommaso Porzio, Sjak Smulders, Robert Ulbricht, and Has van Vlokhoven. We also thank seminar and workshop participants at Cemfi, Cologne, LSE, TSE, Tilburg and the STEG workshop. We thank Hala Tobji for outstanding research assistance and a STEG PhD grant for generous support. All remaining errors are our own.

1 Introduction

How important is domestic demand in driving and directing economic growth in developing countries? And if it is important, how biased is this growth towards the preferences of the rich given that demand for different products varies strongly over the income distribution? Modern growth theories assign demand a central role in driving growth; firms enter and invest in R&D to gain consumer demand and entire industries see productivity growth from Learning-by-Doing in response to increases in demand. Yet whether these forces are important in developing countries is not well understood, let alone their distributional consequences as growth may be biased towards products the rich consume more.

We study these questions quantitatively in the context of India's economic development since the 1980s, bringing together a structural model and novel empirical evidence on how domestic demand drives economy-wide growth. The key empirical challenge for identifying the causal effect of demand is that demand and industry responses – such as productivity growth or entry and exit – are jointly determined. In the empirical part of the paper, we address this identification challenge by exploiting predictable changes in consumption patterns implied by Engel's law: as households become richer, their relative demand systematically shifts across goods. These Engel curves can be directly observed from detailed cross-sectional household consumption data and we show that Engel curves in the 1980s predict changes in product-level demand decades later. Linking household-level consumption to production data, we then quantify how changes in domestic demand drive productivity growth, reallocation and input choices at the firm- and industry level. We find evidence that changes in demand drive strong reallocation across industries and (measured) productivity growth. However, we only find muted responses to demand changes for entry and exit, the overall number of firms, and variety growth.

We then link these empirical results to a multi-industry endogenous growth model to decompose the Indian growth experience and study the role of policy. The model tractably combines the two main modeling blocs that mimic our empirical setup: (1) heterogeneous households that differ in their income and whose demand changes across industries as they become richer, and (2) a production side where firms in different industries endogenously enter and improve their productivity given changes in demand. We then calibrate our model using our empirical results, showing how the causal estimates pin down the key growth elasticities of the model. With the model in hand, we show that domestic demand drove around 30% of India's growth since the 1980s. However, growth is strongly biased towards relative luxury products, leading the rich to experience 25% higher real consumption growth. Finally, we show that redistribution can offset part of the pro-rich growth bias by directing technical change towards products the poor consume more.

India is the ideal setting for our study due to a combination of data availability and context. High quality household-level consumption expenditure data from the National Sample Survey (NSS) allows us to measure changes in consumer demand across roughly 300 broad products over a period of 25 years from 1987 to 2011. We then link the demand side to production data at the industry-level drawing on manufacturing micro data from the Annual Survey of Industries (ASI).[1] Besides the availability of representative and comparable micro data, India is also a fascinating context to study how demand drives growth and whether this growth was biased in favor of the rich. Between 1987 and 2011, demand changed dramatically across sectors and industries in the economy; aggregate income per capita increased almost 4-fold, the population increased by more than 50% and the agricultural expenditure share decreased by almost 20 percentage points from 62 to 44%. At the same time, India is highly unequal and income inequality even further soared; with the income share of the top 10% increasing from around 35% in 1987 to 55% by 2011 (Bharti et al., 2026).

Empirically, we extend the methodology in Acemoglu and Linn (2004) and especially Beerli et al. (2020) by exploiting cross-sectionally revealed income effects to predict future changes in demand. We start estimating product-level Engel curves in 1987 for roughly 100 products that we can consistently link over time. Engel curves trace out how expenditure shares vary over the expenditure distribution in the cross-section. We estimate Engel curves non-parametrically and separately by Indian state, to allow for well-documented differences in preferences across different regions in India (Atkin, 2013). To construct our instrument, we fix the estimated Engel curves and shift every household's previous income by a common factor to match aggregate per capita household expenditure increases. The resulting 'Engel demand shifters' arise from the interaction of (1) income effects as households move along their nonlinear, product-specific Engel curves, and (2) shifts in the expenditure distribution, which determine how different portions of the Engel curve are weighted for aggregate product demand. Importantly, we construct our instrument by only drawing on baseline, cross-sectional variation and a proportional shift in the *level* of household expenditures. By fixing Engel curves in the 1980s, we predict demand in the 2010s ignoring any future changes in relative prices, as these would be contaminated by any future productivity changes that are unrelated to demand.

[1] Our main production data is only for manufacturing, but includes food processing and other agriculture-related products. We have additional data and analyses on agriculture and services that are not included in the current paper version, including the representative Cost of Cultivation surveys (CC) for agriculture and the CMIE Prowess database for services by large, listed firms. India, unfortunately, has otherwise poor coverage of firm-level data for the services sector. Within manufacturing, we also have results for informal establishments drawing on the NSS unorganized manufacturing data.

The key assumption is that our 'Engel demand shifters' only affect production through their effect on (expected) demand. The main threat to this identification assumption are other changes in the economy that happen between the 1980s and 2010s, that are correlated with our instrument and also affect production. To gauge the potential importance of such threats, we run a battery of empirical tests. First, we show that our instrument is largely uncorrelated with baseline characteristics of the Indian economy; predicting our instrument using a large set of potential baseline covariates produces an R^2 of close to 5%. Second, we look at specific changes in the Indian economy that have been highlighted in the literature and may be correlated with our instrument. We show that our instrument is uncorrelated with important other "shocks" to the Indian economy, such as various tariff changes.

On the model side, we tractably combine (1) heterogeneous households with non-homothetic preferences (Comin et al., 2021) who differ in their income, and (2) a multi-industry endogenous growth model in which firms produce heterogeneous varieties in different industries and industry-level entry and productivity grows in response to industry-level demand changes. While the model captures different growth mechanisms, the main driver of endogenous industry-level productivity growth in response to demand changes is industry-wide Learning-by-Doing (LBD) in the tradition of Arrow (1962). We believe Learning-by-Doing is the simplest explanation for the empirical results we find and realistic for an empirical context in which purposeful R&D spending is very low. Besides LBD, the model also captures growth from: (1) endogenous firm-level innovation, (2) exogenous sources that are orthogonal to the mechanism we study (e.g. foreign technologies, the opening up of the economy and institutional improvements), and (3) growth in human capital and the overall population, which induces new demand growth. The main technical difficulty is that combining multi-industry endogenous growth with non-homothetic demand means that the economy is not on a balanced growth path since demand is changing systematically across industries over time. We deal with this issue by building an analytically tractable model that we can study outside the steady state, featuring size-independent productivity growth as in Atkeson and Burstein (2010), free entry, and a Pareto distribution, which generates traveling-wave productivity dynamics as in Luttmer (2007) and Sampson (2016). We then calibrate the model on the observed (non-balanced) growth path in the data, targeting key regression coefficients from our reduced-form results to identify the main model parameters.

Using the model, we establish three sets of results. First, we decompose the Indian growth experience and find that demand forces from population growth and Learning-by-Doing together explain about 27.5% of real per capita consumption growth. Other sources of growth that are exogenous to domestic demand explain about 40%, increases in human capital around 30% and the small remainder is explained by endogenous innovation, in line with the small role for R&D in India. Second, we find that growth in India was strongly biased towards the preferences of the rich, with the rich experiencing about 25% higher real consumption gains than the poor. However, we find that this bias is mainly driven by the high baseline level of inequality rather than by changes in inequality over time. At last, we find that redistribution policy can be effective in letting the poor benefit more from growth. Interestingly, we find that redistributing income towards the poor even has positive growth effects over most of the income distribution as it would redirect firm entry and productivity growth towards sectors that most households care about. However, we also find that the dynamic growth effects from redistribution are small, with the static welfare gains from redistribution trumping any dynamic growth effects by orders of magnitude.

Related Literature We contribute to three main strands of literature. First, our paper sheds light on and helps quantify the drivers of economy-wide growth in a development context. Empirically, we show novel causal evidence on the importance of domestic demand in driving growth across a wide set of industries, demonstrating the power of directed technical change for the aggregate economy. This extends evidence for specific industries in developed countries to a broader set of industries and a longer time period (e.g. Acemoglu and Linn, 2004; Ilzetzki, 2024). More importantly, however, our empirical methodology allows us to provide evidence on how demand drives productivity growth and reallocation at the industry- and firm-level. We show empirically that demand drives changes in measured TFP, but not in varieties nor entry. As we show, these empirical results are crucial to discipline quantitative growth models. For example, our results on entry dynamics are in contrast to standard workhorse models of endogenous growth that would falsely predict that changes in industry-level demand are "eaten up" by entry. We show how to extend a tractable quantitative growth model to account for our empirical facts and then use the model to decompose growth drivers and inform policy.

Second, our paper links to a rich, mostly theoretical, literature that studies the link between inequality and growth by combining non-homothetic preferences and endogenous growth (Foellmi and Zweimüller, 2006). Our contribution with respect to the literature is quantitative. We build a tractable multi-industry endogenous growth model that features non-homothetic preferences and household heterogeneity. Technically, the combination of non-homothetic preferences, household heterogeneity and endogenous growth is difficult because demand changes across industries as the economy grows. Apart from special tractable cases (e.g. Bohr et al., 2021; Oberfield, 2023), this requires to study the economy off a balanced growth path. Our model provides a tractable way to study the economy outside the steady state and – importantly – without having to specify the final steady state, opening up further possibilities for quantitative applications.

Finally, we also contribute to the literature at the intersection of growth and structural change. Empirically, we extend the empirical methodology by Beerli et al. (2020) that allows to disentangle demand and productivity growth, drawing on the importance of income effects as has been highlighted in the structural change literature (Boppart, 2014; Comin et al., 2021) to construct our demand instrument. The idea of using predictable changes in demand to study productivity and innovation responses has a longer tradition. For example, Acemoglu and Linn (2004) studied innovation in the US pharmaceutical industry in response to changes in demand, which they instrument using changes in the age composition of the population. Beerli et al. (2020) instrument demand growth in China by exploiting changes in product-level demand from shifts in China's income distribution. We extend and adapt their methodology in two important dimensions. On the data side, they use only a crude measure of income based on four different income bins and a measure of household ownership over a limited set of 16 durable goods. We draw on a continuum measure of household expenditure and a large set of consumption goods to construct much richer product-level Engel curves. Second, Beerli et al. (2020) instrument demand using changes in the level and the entire distribution of income. We only use changes in the level of income since we believe that changes in the distribution are likely to be endogenous and can thus invalidate the empirical approach.

The rest of the paper is structured as follows. The next section gives an overview of the data, while section 3 shows the main empirical results. Section 4 develops the structural model and discusses our model calibration. In Section 5, we quantify the demand drivers of growth and study the role of inequality and policy. The last section concludes.

2 Data

Our analysis relies on household and production data.

Household demand data.— On the household side, we draw on the consumption modules in the National Sample Survey, and for the main analysis focus on data from the years 1987 & 2011 (the "NSS data"). The NSS is a high quality, representative survey collected by the Ministry of Statistics and Programme Implementation going back to 1950, making it one of the oldest continuing household sample surveys in the world (Deaton, 2005). We start with the 1987 NSS wave because this is the first year in which rich households are oversampled, ensuring a much better coverage at the right tail, a feature that is key to construct our demand shifters. We end in 2011, the last year in which publicly-available

NSS microdata was published.[2]

Production data.— For our main analysis, we draw on plant-level manufacturing data from the Annual Survey of Industries for the years 1989 & 2011 (the "ASI data"). ASI is the primary data source on India's manufacturing sector and is representative of registered manufacturing plants with 10+ workers. As robustness, we also link data from the NSS unorganized manufacturing schedule ("UM") which covers smaller and informal manufacturing establishments, but which is only available in selected years (e.g. 1989 & 2005, but not 2011). We use the ASI data to measure our main production-side variables at the firm and industry level (where the latter are aggregated across firms using ASI-provided survey weights). Most of the variables we use are standard, including measures of total sales, inputs and input costs. To test the importance of variety effects, we draw on the number of unique products that firms report producing, similar to Goldberg et al. (2010).

We also measure firm- and industry-level total factor productivity using a cost-shares approach assuming Cobb-Douglas production functions and perfect competition. Static optimization allows us to obtain elasticities from nominal expenditure shares of inputs (capital, labor, materials and fuel) in total output. To obtain real variables, we deflate nominal variables at the industry-level following a similar approach as Allcott et al. (2016) using the micro data from the Indian wholesale price index (WPI), which are effectively producer prices. To account for labor quality, we deflate plant level wage-bills using state-year labor prices, which are derived using Mincer regressions run on data from the NSS employment surveys for the year 1989 and 2011.

Linking demand and production.— We link the consumption data to the manufacturing micro data by linking NSS consumer products to 5-digit NIC industries. We do so using a many-to-many linkage that we construct by hand (details are in the Appendix). Given that the demand side is for final consumer products, we drop unmatched industries on the production side. These are industries that produce purely intermediate goods or goods that are fully exported (e.g. Mica). In the end, we are left with 112 product-industry pairs, examples of which are "bread & biscuits", "beer", "knitted garments", "medicine" and "motorcycles".

3 Empirical Evidence

This section establishes our two main empirical results. The first result is that we can predict future changes in product-level demand on the consumer side. We do so using our

[2]The NSS 2022 wave also has consumption expenditure, but was not available at the start of this project.

novel 'Engel demand shifters' that exploit cross-sectional Engel curves and the presence of important income effects in households' consumption behavior. In the second step, we link our demand shifters to production data and establish that Indian production responds to increased demand by increasing production, increasing productivity and reallocating factors of production. However, we find no evidence for increases in varieties, nor for net firm entry.

Result 1: Income effects drive changes in consumer demand

We start by formalizing aggregate household demand D_t^j for product j at time t as:

$$D_t^j = \int_{i \in H_t} d_i^j \left(\overrightarrow{p_t}, Y_{it}, X_{it}, \varepsilon_{it}^j\right) \quad \text{with:} \quad d_i^j \left(\overrightarrow{p_t}, Y_{it}, X_{it}, \varepsilon_{it}^j\right) \equiv \omega_i^j \left(\overrightarrow{p_t}, Y_{it}, X_{it}, \varepsilon_{it}^j\right) \cdot Y_{it} \tag{1}$$

where households are denoted by i and the household's general Marshallian demand function is denoted by d_i^j, which can also be written in its budget share form ω_i^j. Demand depends on the entire vector of goods prices $\overrightarrow{p_t}$, household expenditure (or "income") Y_i, a vector of observable household characteristics X_i (e.g. household size) and ε_i^j, household i's idiosyncratic taste for product j. H_t denotes the household distribution over $\{Y, X, \varepsilon^j\}$ at time t. The setup nests the special case of *homothetic* household demand in which case dependence on income Y_i drops out. In the general setup we laid out, aggregate final demand for product j can change due to changes in (relative) prices, due to changes in the distribution of tastes, household characteristics or income.

We start out by providing evidence for the importance of *non-homothetic preferences*, that is, the importance of income in explaining variation in household demand. We do so in two substeps. First, we show that in the cross-section, relative demand for different products varies systematically with income; a well-known fact known as Engel's law. In the second substep – and more importantly for the goal of this paper – we show that income effects can be used to predict future demand changes using initially estimated Engel curves.

Engel curves

Income effects are formally given by changes in demand as income varies: $\frac{\partial \omega_i^j}{\partial Y_i}$. Proposition 1 explicitly spells out the assumptions needed such that cross-sectional Engel curves – that trace out changes in demand over the household income distribution – reflect income effects.

Proposition 1 (Engel curves & non-parametric identification of demand)**.** *Assuming that:*

1. *(**Tastes & characteristics are additive**):*

$$\omega_i^j\left(\vec{p}_t, Y_{it}, X_{it}, \varepsilon_{it}^j\right) = \widetilde{\omega}_i^j\left(\vec{p}_t, Y_{it}\right) + \beta X_{it} + \varepsilon_{it}^j$$

2. *(**Tastes are conditionally independent**):* $(\omega^j \mid X) \perp\!\!\!\perp \varepsilon^j$

Then, $\widetilde{\omega}_i^j$ and income effects $\frac{\partial \omega_i^j}{\partial Y_i}$ are non-parametrically identified over the support of Y_i from cross-sectional data that features household expenditures ω_i^j, observables X_i and variation in household income Y_i.

We further define the (residualized) ***Engel curve*** *for product j at time τ as:*

$$\mathcal{E}_\tau^j(y) \equiv \widetilde{\omega}_i^j\left(\vec{p}_t, Y_{it} = y\right)|_{t=\tau}$$

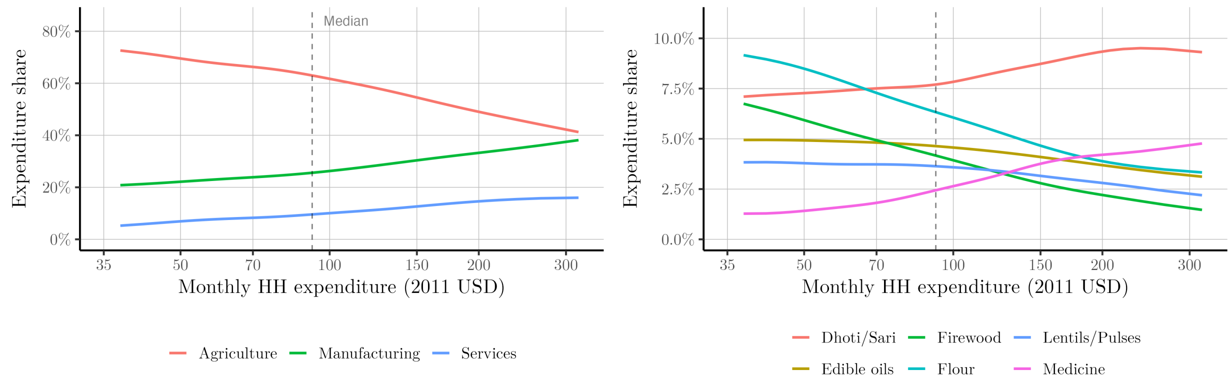
Proof. The proof is in the Appendix. □

That is, Engel curves identify income effects in the case where idiosyncratic household taste shifters do not systematically vary with income conditional on observables X; for example, suppose that households with special tastes for milk are more likely to become rich, then the milk Engel curve would partly pick up selection rather than income effects. When interpreting estimated Engel curves and also for the subsequent model, we abstract from such (conditional) selection effects. However, we note that these assumptions are not strictly needed for our main results given that selection would simply weaken the power of Engel curves to predict future changes in demand.

We start by plotting estimated Engel curves for the year 1987 for two levels of aggregation to show that demand varies systematically with income. Throughout, we estimate all Engel curves with a generalized additive model (GAM), a semi-parametric regression that controls linearly for observables X, while allowing expenditure shares to vary non-parametrically in income using splines. As controls, we always use fixed effects for household size, religion, social group, and whether the household is urban or rural, which are meant to capture differences in preferences as well as differences in relative prices that households may face. Whenever we pool across multiple Indian states, we also control for state fixed effects.

Figure 1 Panel A shows Engel curves for expenditure shares on sectoral aggregates: agriculture, manufacturing & services.[3] Agricultural expenditure shares decline strongly in income from around 75% for the poorest households to 40% for the richest Indian households in 1987. Similarly, demand for manufacturing and services increases strongly with income.

[3]The Appendix provides details on how we classify these sectors following the final expenditure approach according to Herrendorf et al. (2013).



(A) Engel curves for 3 broad sectors

(B) Engel curves for Top 6 products

Figure 1: Engel curves for sector and for top 6 consumer products

Notes: Data from 1987 NSS, sample size: $N = 124,767$. Panel A: Estimated sectoral Engel curves for Agriculture, Manufacturing and Services. Sectors are classified according to the final expenditure approach following Herrendorf et al. (2013), details in Appendix. Sectoral data includes all expenditures, not only those that can be linked to manufacturing. Panel B: Estimated Engel curves for six top consumer products based on their average expenditure shares. Products are homogenized across NSS waves and restricted to products that have some manufacturing content such that they can be matched to manufacturing production data (details in the Appendix). The dotted line in each plot reports the median expenditure level. All Engel curves are estimated based on a generalized additive model (GAM), linearly controlling for fixed effects for state, household size, religion, social group and whether the household is urban or rural.

Panel B shows Engel curves at the product-level. After restricting to products that we can match to our production data (which excludes most services and primary agricultural goods such as rice and wheat), we plot estimated Engel curves for the six most important products based on their average household expenditure share. Engel curves vary considerably across products. For example, flour and firewood are strong necessity goods whose demand is high for the poorest but declines by a factor of three for the richest households. Edible oils and lentils/pulses are also necessity goods but their relative demand declines much less strongly with income. And finally, medicine and clothing (dhotis & saris) are more luxury goods whose relative demand rises with income. This is particularly strong for medicine, whose relative demand more than triples.

Our main empirical specification draws on even more granular Engel curves at the product-state-level, which allow for different demand patterns by Indian state that capture differences in tastes across regions in India (e.g. the prevalence of rice in West Bengal versus wheat in Rajastan as highlighted in Atkin (2013)) and differences in relative prices due to imperfect trade integration. Since these are a total of about 1,000 different curves that are difficult to visualize in bulk, we instead decompose the cross-sectional variation in expenditure shares. We start by noting that our product-state-level Engel curves can explain 44% of the total variation in observed household-product-level expenditure shares.

Throughout this paper, we abstract from the remaining 56% of idiosyncratic household-level variation in expenditure shares, which may come from sizable measurement error, lumpy durable expenditures that we smooth out, and idiosyncratic household preferences that we do not aim to explain. Of the 44% of the total variation that we do explain, about 40% is driven by our observables X, on which we residualize and which we also do not aim to explain. The remaining variation is our true state-product-level Engel curve variation that we subsequently rely on for predicting changes in aggregate state-product-level demand. We further ask how much of our state-product-level variation in Engel curves is coming from product-level variation, which we answer by looking at how much variation in our predicted (residualized) expenditure shares is explained by estimated product-level Engel curves that pool across states. The answer is: 94%. That is, by far most of our explained cross-sectional variation stems from variation across products.

Engel demand shifters

In the second substep, we ask whether cross-sectional Engel curves are predictive of aggregate changes in future demand. For this, we construct our *Engel demand shifters* as follows:

$$\widehat{D}^j_{t'} = \int_{i \in H_0} \left[\widehat{\omega}^j_{i,0}\left(\overrightarrow{p_0}, Y_i \cdot \bar{y}\right) \cdot Y_i \cdot \bar{y} + \hat{\beta}\overline{X}_{i0} \right] di \tag{2}$$

where $\widehat{D}^j_{t'}$ gives aggregate predicted demand at time t for product j using the Engel curve estimates at initial time 0 (which we fix to 1987) and the initial distribution H_0. However, we now shift the initial income distribution proportionally for everyone by a common factor $\bar{y}$, where $\bar{y}$ is chosen such that we exactly match average household expenditure in $t = 2011$. Importantly, the demand shifter varies from the interaction of (1) nonlinear, product-specific Engel curves along which households move, and (2) the shape of the initial state-specific expenditure distribution which we shift proportionally and which determines how different portions of the Engel curve are weighted for aggregate product demand. $\bar{y}$ is then important to match the 'level' increase in incomes to trace out the correct level of income effects. The approach is meant to predict relative changes in demand across different products, but does not necessarily predict the correct level change in aggregate demand. For example, we explicitly fix the initial household distribution H_0 and thus do not feed in the massive demographic changes that India has seen since 1987, which would shift up aggregate demand across all products.

To show that the *Engel demand shifters* are predictive of actual changes in aggregate demand, we again start by showing evidence at the sectoral level and then move to the

Table 1: Predicting aggregate demand at sectoral level

	Agriculture		Manufacturing		Services	
Year	Data	Predicted	Data	Predicted	Data	Predicted
1987	61.3%	61.3%	28.9%	28.9%	9.8%	9.9%
2011	43.9%	46.3%	37.5%	39.3%	18.6%	14.5%

more granular product-state variation. Table 1 shows that our demand shifter predicts large changes in sectoral expenditure shares over a period of close to 25 years. In the data, the aggregate agricultural expenditure share declined by close to 20 percentage points between 1987 & 2011; our demand shifter can account for 85% of that change. This is astonishing, because our approach only uses income effects as revealed by cross-sectional variation in 1987, we do not incorporate any of the massive changes in prices and technologies that likely contributed to structural change out of agriculture. While astonishing, the dominance of income effects in driving structural change is a common finding in the structural change literature (Boppart, 2014; Comin et al., 2021), lending credence to our demand instrument. At the same time, our demand shifter also misses important changes: we slightly overestimate the increase in manufacturing and strongly underestimate the increase in services. This is perfectly in line with mediocre manufacturing performance and strong service-led growth over this time period (e.g. Fan et al., 2023), which pushed towards much more services as would have been expected purely based on relative demand in 1987.

At last, Table 2 provides regression evidence that allows us to condition predicted variation in state-product-level Engel demand shifters on sets of fixed effects. Column 1 reports the unconditional correlation, giving a strong positive correlation with an R^2 of 75%. However, this greatly overestimates the predictability of the demand instrument for two reasons. First, it also predicts values for 1987, which the demand instrument will fit almost by construction as long as Engel curves are specified correctly. To address this point, column 2 only predicts out-of-sample for the year 2011. The fit is less good, but the instrument still correctly predicts around 65% of the total variation in demand. Second, columns 1 and 2 still pick up on a level effect: aggregate demand for rice is generally much higher than aggregate demand for beer, both in 1987 and in 2011, and our demand shifter would get this right even without predicting any changes in demand. Thus, columns 3 and 4 both look at changes in demand within products. The two columns only differ in whether they additionally control for year or state-year fixed effects. Reassuringly, estimated coefficients stay almost unchanged and both specifications have large explanatory power for residual variation in household demand. Columns 5 and 6 – our most preferred specifications – draw

Table 2: First stage demand results

Dependent Variable:	log(demand actual)					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
log(demand predict)	1.0***	0.77***	0.63***	0.89***	0.78***	0.96***
	(0.009)	(0.02)	(0.05)	(0.07)	(0.07)	(0.07)
Fixed-effects						
Product			Yes	Yes		
Year			Yes		Yes	
State-year				Yes		Yes
State-product					Yes	Yes
Fit statistics						
Observations	3,894	1,947	3,894	3,894	3,894	3,894
R^2	0.478	0.689	0.958	0.959	0.964	0.964
Within R^2			0.062	0.083	0.078	0.093
F-test	1,789.8	4,314.3	476.5	344.5	25,802.9	1,254.5
F-test (projected)			252.2	338.5	164.4	195.7

Clustered (State-product) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

only on within state-product variation, exploiting how households move along their product-state Engel curves in combination with shifts in the state-specific income distribution. While product-state fixed effects increase the number of fixed effects by a factor of 10 and decrease the residual variation that can be explained, the F-stat in these specifications is still very strong, with values well above 100. Still, the difference to columns 3 & 4 shows that a sizable share of the variation in our Engel demand shifters comes from within products across states, rather than within product-states. Since our identification argument is mostly about within product-state variation, we use product-state fixed effects throughout the main subsequent results.

Result 2: Demand drives reallocation and growth

We now move to linking changes in consumer demand to changes in production. Throughout, we denote establishments by i, states by s and industries – since they are linked to products – by j. Throughout this section, we consider two types of regressions that differ in their level of aggregation:

$$y_{j,s,t} = \mathrm{FE}_{j,s,t} + \beta log(\mathrm{demand})_{j,s,t} + \epsilon_{j,s,t} \quad \text{(Industry-level regressions)} \tag{3}$$

$$y_{i,j,s,t} = \mathrm{FE}_{i,j,s,t} + \beta log(\mathrm{demand})_{j,s,t} + \epsilon_{i,j,s,t} \quad \text{(Establishment-level regressions)} \tag{4}$$

where y captures any outcome variable and FE refers to different sets of fixed effects that we make explicit in each regression. Importantly, demand always varies at the industry-state-level where we use either *predicted demand* (reduced form) or *instrumented demand* (IV). The two sets of regressions differ in their interpretations: while the industry-level regressions give the average effect for industry-state combinations, the establishment-level regressions report results for the average establishment for industry-state combinations. Since there is entry and exit, these two sets of results generally differ. Our first step is to show that our demand instrument actually affects production. In this step, we also further validate our instrument, providing balance tests and further evidence consistent with the instrument's exclusion restriction. In the second step, we show the mechanisms through which demand affects output, looking at inputs, changes in the composition of firms as well as productivity and variety growth.

Step 1: Validating the instrument

We start out by showing that our predicted demand indeed leads to changes in production. In a closed economy without intermediate goods, consumer demand equals production. However, Indian states – our unit of observation – are not fully closed economies and intermediate goods production could also play an important role in mediating the effect of final demand on production, potentially weakening the strength of our first stage on output. Table 3 reports regression results for the effect of predicted demand on firm output. Predicted demand changes drive sizable changes in output. The strong positive first stage shows that Indian states are far from perfectly open economies in which changes in local demand would not transmit differentially to local production.

Next, we provide further evidence for the instrument's exclusion restriction, which states that the 'Engel demand shifters' only affect outcomes Y through their effect on (expected) consumer demand. The economic argument in favor of the exclusion restriction is that the instrument by construction excludes changes in relative prices, tastes, the household distribution over observables X and changes in the expenditure distribution, all of which may plausibly be endogenous to other changes that have happened in India over the period 1987 to 2011 and which independently drive changes in Y. The key economic assumption is that demand in 1987 is a static choice in which households do not strategically consume

Table 3: Main output results

Dependent Variable:	log(sales)	
Model:	(1)	(2)
Variables		
log(demand predict)	0.86***	0.91***
	(0.15)	(0.17)
Fixed-effects		
State-product	Yes	Yes
Year	Yes	
State-year		Yes
Fit statistics		
Observations	36,731	36,731
Size of the 'effective' sample	1,123	1,123
F-test (projected)	83.5	67.5

Notes: Data is at the industry-state-year level. Sales are total revenue (inclusive of other receipts) from ASI data (1987 and 2011). We predict demand based on the state-product-year-level Engel demand shifter. Clustered (State-product) standard-errors in parentheses.
Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

based on expectations over future (quality) changes in products. Given that our products are relatively broad (e.g. rice and motorscooters) and most are non-durables, we believe this to be an uncontroversial assumption. For example, a violation of our exclusion restriction would be if the rich households in 1987 were 'trendsetters' in the sense that they systematically consume more goods that – independently of demand – see more quality or productivity improvements in the future.[4]

The plausibly much bigger threat to the exclusion restriction is that other (supply-side) changes in the economy that happen between the 1980s and 2010s and affect outcomes Y are potentially correlated with our instrument. To gauge the importance of such threats, we run a battery of empirical tests. First, we evaluate whether our instrument is correlated with any baseline characteristics of the Indian economy in 1987. Table 4 reports the results of an omnibus test, predicting our demand instrument using a larger set of baseline characteristics. Reassuringly, Column 1 shows that – even unconditionally – our instrument is largely uncorrelated with baseline characteristics, with an overall R^2 that is less than 9%. However, a number of baseline characteristics are significantly correlated with our instrument, such

[4]While there may be some evidence that households strategically consume durables in specific contexts (e.g. (De Groote and Verboven, 2019)) we believe this to be second order here. Also, these strategic considerations would have to be differentially correlated with household income.

Table 4: Balance test for instrument on baseline variation

Dependent Variable:	Predicted Demand Growth			
Model:	(1)	(2)	(3)	(4)
Variables				
log(nb of firms)	-0.03	-0.14	0.15***	0.004
	(0.09)	(0.09)	(0.04)	(0.01)
log(nb of workers)	0.09**	0.05	0.04*	0.0003
	(0.04)	(0.04)	(0.02)	(0.006)
log(total VA output + 1)	-0.0005	0.002	-0.005	0.001
	(0.007)	(0.007)	(0.004)	(0.001)
log(nb of products + 1)	-0.05	0.07	-0.16***	-0.02
	(0.08)	(0.08)	(0.04)	(0.01)
Top 10 output share	0.21**	0.18*	0.07	0.01
	(0.09)	(0.09)	(0.04)	(0.02)
Top 10 product share	0.27*	0.20	0.10	0.007
	(0.15)	(0.14)	(0.06)	(0.02)
Top 10 empl share	-0.24	-0.11	-0.09	0.01
	(0.17)	(0.18)	(0.07)	(0.02)
Labor elasticity	1.9**	1.8**	0.48	-0.07
	(0.84)	(0.84)	(0.32)	(0.08)
Capital elasticity	2.9***	2.2**	1.4***	0.09
	(0.82)	(0.87)	(0.39)	(0.15)
Input elasticity	1.4**	1.2**	0.47*	-0.06
	(0.56)	(0.58)	(0.28)	(0.07)
Share of firms import	0.05	0.10	-0.18**	-0.06*
	(0.17)	(0.18)	(0.07)	(0.03)
Avg plant age	-0.01***	-0.02***	-0.001	-0.0001
	(0.005)	(0.005)	(0.002)	(0.0005)
Productivity (TFPR)	-0.03	-0.03	0.0009	-0.004
	(0.02)	(0.02)	(0.008)	(0.003)
Fixed-effects				
State		Yes		Yes
Product			Yes	Yes
Fit statistics				
Observations	918	918	918	918
R^2	0.087	0.248	0.819	0.979
Within R^2		0.100	0.074	0.031
F-test (projected), p-value		3.32×10^{-14}	8.13×10^{-9}	0.021

Clustered (Product) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

as employment (total number of workers and the employment share in the top 10 firms), capital intensity (as measured by the capital elasticity)and average plant age. Crucially, our exclusion restriction holds conditional on a set of fixed effects. Columns 2 and 3 report the results of omnibus tests conditional on state and product fixed effects separately. While we can reject the null hypothesis of joint significance in these cases, Column 4 shows that once we account for both product and state fixed effects, baseline characteristics explain only 3.1% of the within product-state variation – the F-test fails to reject joint insignificance at the 1% level.

Second, we look at specific changes in the Indian economy that have been highlighted in the literature and may be correlated with our instrument. We first look at tariff changes across our products over the period 1989-2003. The concern here is that the liberalization episode in the 1990s was biased towards certain products, and these changes are in the same direction as the predicted growth in demand for our products. We use tariff datasets made publicly available by De Loecker et al. (2016) and calculate a host of measures to reflect the changes in tariffs over the period of our study. These measures include average changes, minimum changes and maximum changes in both input and output tariffs. While maximum and minimum changes will cover both extreme increases and decreases in tariffs, long-run changes can be better captured using averages.

Specifically, we test whether changes in predicted demand growth can be predicted using our varied measures of tariff changes. Table 5 provides the estimates from these regressions. We find that most tariff measures are not predictive of changes in demand for products. Average changes in input tariffs are correlated with predicted demand growth (See col. 5). However, our instrument relies on product and state variation while tariffs are uniform across states. To account for this, we also construct interactions of tariff measures and state level shares of imports. Results for this are presented in the appendix in Table 13, where we find qualitatively the same results. Finally, we can control for tariff measures in our main regression as a robustness check.

To account for product-state variation in our instrument, we construct tariff measures that use the share of states in the total imports of the product.In Table 13, we present a balance test wherein we regress predicted demand growth at the product-state on the earlier mentioned tariff measures, now interacted with the state shares. We find that, while predicted demand growth cannot be predicted using our host of tariff measures (within R-squared of 0.005), our measure of average output tariff changes has a significant correlation with the instrument. Intuitively, this can be expected as the reduction in output tariffs leads to more foreign products in the Indian market (pro-competitive effect). A common practice in this case is to control for variables that violate the balance test, in this case average output

Table 5: Regression of Predicted Demand Growth on Tariff Measures

Dependent Variable:	Predicted Demand Growth					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Min Δ Output Tariff	-0.0092					
	(0.3728)					
Avg Δ Output Tariff		-1.974				
		(2.424)				
Max Δ Output Tariff			-0.6469			
			(0.5643)			
Min Δ Input Tariff				-0.3231		
				(2.198)		
Avg Δ Input Tariff					-21.64**	
					(9.751)	
Max Δ Input Tariff						-1.673
						(1.890)
Fit statistics						
Observations	84	84	84	84	84	84
R^2	7.46×10^{-6}	0.00802	0.01577	0.00026	0.05666	0.00946

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

tariffs.

Step 2: The mechanisms through which demand drives production

So far, we validated the instrument showing the strength of the first stage (does actual demand respond to predicted demand), providing evidence in favor of the exclusion restriction, and providing evidence for the reduced form (does predicted demand explain production). We now move to the *causal evidence* of how demand drives production. For this, we focus on IV regressions where we instrument actual demand with predicted demand and look at outcomes on the production side. We focus on results for the industry-level regressions here and leave the establishment-level regressions for the appendix.

Table 6 reports the effect of demand on value added output and total sales. The effect is strong and significant, with an elasticity of close to 1, which is in line with the previous reduced-form evidence. This means that at the local (state) level, industrial output reacts almost one-to-one to changes in demand over longer periods of time.

Next, we show that industries increase their output in part simply by increasing their input usage. Table 7 reports the corresponding IV regression results. Increases in industry-

Table 6: IV Regressions for Output (industry level, 1989–2011)

Dependent Variables:	VA	VA (Incl.)	Sales	Sales (Incl.)
Model:	(1)	(2)	(3)	(4)
Variables				
Demand	1.167**	1.129**	0.9430**	0.8727**
	(0.4950)	(0.4684)	(0.3923)	(0.3831)
Fixed-effects				
State × Product	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
Fit statistics				
Observations	1,626	1,716	1,706	1,758
F-stat (1st stage)	37.841	36.125	35.408	36.909

Data is at the state-industry-year level. Results are based on IV regressions where state-industry-year-level HH demand is instrumented using predicted demand. "Inclusive" means including income from non-primary activities. Clustered (State × Product) standard-errors in parentheses.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

level demand lead industries to hire more labor with the number of workers (column 1) and total labor costs (column 2) increasing, use more capital (column 4) and more intermediates (column 6). While the effect for capital is not significant given that the point estimate is slightly smaller and capital is more noisily observed, we cannot reject that input response elasticities are identical. Establishment-level regression results confirm a significant increase in capital usage, especially for bigger firms. Importantly, elasticities for the input response are around 0.5, roughly half of the corresponding output elasticity, indicating productivity improvements to which we turn next. Finally, we find that the increase in labor mostly comes from non-managerial labor, as the employment of managers does not seem to respond to demand increases.

The last set of results digs deeper into the growth mechanisms through which industries and firms respond to demand changes. As shown in columns 2 and 3 in Table 8, we find an increase in productivity in response to demand changes using measures of TFPR and labor productivity (total output per worker).[5] The latter is not statistically significant, but is similar in magnitude to the TFPR estimate and not significantly different. In terms of magnitude, productivity gains are sizable; a doubling of demand (100% increase) implies a roughly 60% increase in productivity growth at the state-industry level. While only imperfect measures of productivity growth, these results are consistent with models of vertical

[5]As discussed in the data section, we construct TFPR based on static cost shares, consistent with a Cobb-Douglas production function and static input choices under perfect or monopolistic competition. Details are in Appendix A.1.

Table 7: IV Regressions for Inputs (industry level, 1989-2011)

Dependent Variables:	Employment	Wages	Managers	Capital	K/L	Int. Inputs
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Demand	0.4763**	0.5416**	-0.0570	0.3994	-0.0769	0.7207**
	(0.1954)	(0.2232)	(0.1838)	(0.2578)	(0.1627)	(0.2886)
Fixed-effects						
State × product	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,758	1,758	1,634	1,758	1,758	1,754
F-test (1st stage)	36.909	36.909	36.094	36.909	36.909	36.424

Notes: Data is at the state-industry-year level (1989–2011). Results are based on IV regressions where state-industry-year-level demand is instrumented using predicted demand. Clustered (State × Product) standard-errors are in parentheses.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

innovation in response to demand growth.

Next, we look at the response for measures of horizontal innovation, which could come from a combination of an increase in product lines and firms. Surprisingly, we find no evidence for horizontal innovation. We neither find an increase in the number of product lines (column 1), nor in the number of firms (column 4). These results are in line with firms simply expanding existing production in response to demand changes and are also consistent with not seeing an expansion of managerial labor (which would likely be needed if new product lines were introduced). Finally, we can also look at a measure of overall firm turnover to test whether demand changes lead to more competition and thus more entry and exit. We also do not find evidence for that; average plant age remains roughly constant over time and – if anything – increases, not decreases. Overall, our results are thus most consistent with firms expanding production and their (revenue) productivity without increasing product lines or strongly affecting entry and exit.

Discussion

The empirical section showed strong causal empirical evidence in favor of the basic growth mechanism that changes in consumer demand (due to changes in consumer income) drives differential industry growth which in turn spurs differential industry productivity growth. We now move to a quantitative model that serves the dual purpose of (1) decomposing the Indian growth experience, helping us to quantify the overall importance of demand-driven

Table 8: Regressions for Mechanisms (industry level, 1989-2011)

Dependent Variables:	Nb. Products	TFPR	Lab. Productivity	Nb. firms	Avg. Plant age
Model:	(1)	(2)	(3)	(4)	(5)
	Poisson	OLS	OLS	OLS	OLS
Variables					
Predicted Demand	0.0676				
	(0.1511)				
Demand		0.5808**	0.3964	-0.1303	0.0723
		(0.2749)	(0.2785)	(0.1342)	(0.1010)
Fixed-effects					
State×product	Yes	Yes	Yes	Yes	Yes
year	Yes	Yes	Yes	Yes	Yes
Fit statistics					
Observations	1,760	1,758	1,758	1,758	1,638
F-test (1st stage), Demand		36.909	36.909	36.909	35.285

Data is at the state-industry-year level. Results of IV regressions where aggregate HH demand at state-industry-year level is instrumented using predicted demand . Clustered (State × Product) standard-errors in parentheses.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

growth which the reduced-form results do not allow to do, and (2) study the role of policy.

4 Structural model

The model combines two main modeling blocs that mimic our empirical setup. On the demand side, we have heterogeneous households that differ in their income and have non-homothetic preferences over different final consumption goods (Comin et al., 2021), which means their demand shifts across products as their income grows. On the production side, we consider an endogenous growth framework in which firms produce heterogeneous varieties in different industries and industry-level productivity grows in response to industry-level demand changes. While we allow for different drivers of growth in our model, the main driver of endogenous industry-level productivity growth in response to demand changes is industry-wide Learning-by-Doing (LBD) in the tradition of Arrow (1962). We believe Learning-by-Doing is the simplest explanation for the empirical results we find and realistic for an empirical context in which purposeful R&D spending is very low.[6] Besides LBD, we capture three main other sources of growth: (1) endogenous firm-level innovation, (2) exogenous growth that is orthogonal to the mechanism we study such as growth due to improvements

[6] As we show in the calibration section, the aggregate R&D spending share among Indian ASI firms is tiny, at below 0.5%. Comparable numbers for manufacturing in OECD countries is above 8%, more than 20 times larger (see OECD STAN database, 2019).

in foreign technologies, or institutional improvements and the opening up of the Indian economy, and finally (3) growth in human capital due to educational improvements and growth in the overall population, which induces new demand growth.

The main technical difficulty in combining non-homothetic demand with endogenous growth is that this economy is not on a balanced growth path since demand is changing across industries as the economy grows. We deal with this issue by building an analytically tractable model of heterogeneous-firm dynamics and endogenous productivity growth. In particular, we follow Atkeson and Burstein (2010) in modeling size-independent productivity growth and combine this with endogenous entry and a Pareto distribution, which generate traveling-wave productivity dynamics as in Luttmer (2007) and Sampson (2016).

The model economy is set in discrete time and a model period is one year. We model India as a small open economy with a fixed international interest rate R. We start describing the household side and then move to the firm side.

4.1 Household & demand side

There is a continuum of households i whose exogenous mass at time t is denoted by L_t, the size of the population. We allow for exogenous changes in L_t to capture the large population changes that India underwent over the time period of our study and which also drives growth in demand over time. Households differ in their human capital h_{it} which they supply inelastically to the market at the economy-wide wage rate w_t. We denote the human capital distribution by $\mathcal{H}_t$ which we assume to be log-normal with parameters (μ_H, σ_H). With a slight abuse of notation, we write the measure of the distribution as: $|\mathcal{H}_t| = L_t$. Total labor supply is given by H_t and average skills are given by $\overline{H}_t$. Households have non-homothetic CES preferences for broad products $j \in \{1, 2, .., J\}$ such as "motorcycles" and "beer", which we also denote as "industries". Within-industries, households have standard CES preferences for varieties (i.e. different varieties of motorcycles and beer). Specifically, we assume that household i's preferences are given (implicitly) by:

$$\sum_{j=1}^{J} \Omega_j^{1/\bar{\sigma}} \left(\frac{C_{ijt}}{C_{it}^{\varepsilon_j}} \right)^{\frac{\bar{\sigma}-1}{\bar{\sigma}}} = 1 \quad \text{(Across industries: Non-homothetic CES)} \tag{5}$$

$$C_{ijt} = \left(\int_{\mathcal{N}_{jt}} c_{ikjt}^{\frac{\sigma-1}{\sigma}} dk \right)^{\frac{\sigma}{\sigma-1}} \quad \text{(Within industries: standard homothetic CES)} \tag{6}$$

where C_{ijt} and c_{ikjt} denote respectively the demand for industry product j and for industry-j-specific variety k. $\Omega_j > 0$ are the taste shifters that scale consumption choices at the industry level, ε_j are industry-specific non-homotheticity parameters that control *relative* income elasticities (Comin et al., 2021) and $\bar{\sigma}$ denotes the elasticity of substitution across industries. $\mathcal{N}_{jt} \subset [0, N_{jt}]$ is the (endogenous) set of active product varieties within industry j at time t with total measure N_{jt}, and σ is the elasticity of substitution across varieties within an industry. Households optimally choose their industry-level and variety-level consumption by minimizing costs:

$$\min_{\{C_{ijt}\}} \sum_{j=1}^{J} P_{jt} C_{ijt} \quad s.t. \quad \sum_{j=1}^{J} \Omega_j^{1/\bar{\sigma}} \left(\frac{C_{ijt}}{C_{it}^{\varepsilon_j}} \right)^{\frac{\bar{\sigma}-1}{\bar{\sigma}}} = 1 \tag{7}$$

$$\forall j: \quad \min_{\{c_{ikjt}\}} \int_{\mathcal{N}_{jt}} p_{kjt} c_{ikjt} \quad s.t. \quad C_{ijt} = \left(\int_{\mathcal{N}_{jt}} c_{ikjt}^{\frac{\sigma-1}{\sigma}} dk \right)^{\frac{\sigma}{\sigma-1}} \tag{8}$$

taking as given industry-level prices P_{jt}, variety-level prices p_{kjt}, and total household income I_{it} given by: $I_{it} = h_{it} \times \left(w_t + \Pi_t / (\overline{H}_t \cdot L_t) \right)$, where we have assumed that profits also scale proportionally with human capital h to capture large inequality in ownership of businesses in India. Note that we abstract from household savings and capital in our model. The following proposition characterizes optimal household consumption choices.

Proposition 2 (Household demand, optimal consumption choices and aggregation)**.** *The household consumption problem can be separated into a within-industry and an across-industry problem.*

1. ***Across industries:*** *households take industry-specific prices P_{jt} and their income I_{it} as given. Optimal industry-specific consumption choices are then given by:*

$$C_{ijt} = \Omega_j \left(\frac{P_{jt}}{I_{it}} \right)^{-\bar{\sigma}} C_{it}^{\varepsilon_j (1-\bar{\sigma})}$$

with the individual consumption aggregator C_{it} implicitly defined by:

$$I_{it} = \left[\sum_{j=1}^{J} \Omega_j \left(C_{it}^{\varepsilon_j} P_{jt} \right)^{1-\bar{\sigma}} \right]^{\frac{1}{1-\bar{\sigma}}}$$

Aggregate expenditures on industry j are then a function of prices $\{P_j\}_j$ and the distribution over household-level income I_{it}:

$$E_{jt} = \Omega_j P_{jt}^{1-\bar{\sigma}} \int_i I_{it}^{\bar{\sigma}} \cdot \underbrace{C_{it}^{\varepsilon_j (1-\bar{\sigma})}}_{f(I_{it}, \{P_j\}_j)} di$$

2. ***Within industries:*** *households take both industry-specific prices P_{jt}, variety-specific prices p_{kjt} and industry-level expenditures E_{ijt} as given. Optimal household-level and aggregate variety-specific demand are then given by:*

$$c_{ikjt} = p_{kjt}^{-\sigma} P_{jt}^{\sigma-1} E_{ijt} \quad \Rightarrow \quad c_{kjt} \equiv \int_i c_{ikjt} di = p_{kjt}^{-\sigma} P_{jt}^{\sigma-1} E_{jt} \tag{9}$$

Proof. Detailed derivations are given in the Appendix. □

4.2 Firm production & endogeneous growth

On the production side, we consider heterogeneous firms within an industry who each produce a variety of the industry-level good.[7] The only source of heterogeneity across firms is their variety-specific productivity or quality q, which evolves endogenously due to firm innovation, industry-level Learning-by-Doing and exogenous growth. At production, firms solve a static within-period problem, taking their productivity as given and choosing their variety-specific price and production quantity to maximize per-period profits. Outside the production stage, firms invest in their productivity to gain market shares, make entry choices and exit.

For tractability, we follow the modern endogenous growth literature by assuming all costs scale with firms' productivity (Atkeson and Burstein, 2010). As we will describe in more detail, this shuts down some important margins of within-industry firm dynamics, but maintains that industry-level demand affects industry-level growth dynamics. Our model will be able to replicate muted effects of changes in demand on entry and exit, but as we will make clearer below, modeling entry in particular is important for solving the model tractably. We introduce the within-period production problem first and then move to the dynamic entry, investment and exit problems.

4.2.1 Within-period choices: profits, prices and quantities

At any given point in time, firms in industry j produce output of variety k using a linear technology in labor h:

$$y_{kj} = q_{kj} h_{kj} \quad \text{with:} \quad q_{kj} = A \cdot z_{kj} \tag{10}$$

where A denotes aggregate technology, which grows exogenously over time and captures the exogenous growth part of the model, and z_{kj} is firm-specific productivity. Given the standard monopolistic competition setup, firms take demand from Equation 9 as given, but choose their variety-specific price and quantity to maximize profits according to:

[7] We abstract from firms producing multiple varieties given that we do not find evidence for variety effects.

Proposition 3 (Within-period firm problem & optimal choices)**.** *Define industry-specific demand as* $\tilde{D}_j \equiv P_j^{(\sigma-1)/\sigma} E_j^{1/\sigma}$*. Then firms solve:*

$$\max_{\{h\}} (q_{kj} \cdot h)^{(\sigma-1)/\sigma} \tilde{D}_j - w \cdot h$$

Leading to the following optimal choices:

$$\begin{aligned}
h_{kj}^* &= \Big(((\sigma-1)/\sigma) \cdot w^{-1} q_{kj}^{(\sigma-1)/\sigma} \cdot \tilde{D}_j\Big)^\sigma \\
p_{kj}^* &= w \cdot (\sigma/(\sigma-1)) \cdot q_{kj}^{-1} \\
y_{kj}^* &= \Big(((\sigma-1)/\sigma) \cdot w^{-1} \cdot q_{kj} \cdot \tilde{D}_j\Big)^\sigma \\
\pi_{kj}^* &= \frac{1}{\sigma} \cdot \tilde{q}_{kj} \cdot \big(((\sigma-1)/\sigma) \cdot w^{-1}\big)^{\sigma-1} \cdot \tilde{D}_j^\sigma = \frac{1}{\sigma} \cdot p_{kj}^* \cdot y_{kj}^* \quad \textit{with:} \quad \tilde{q}_{kj} \equiv q_{kj}^{\sigma-1} = \overbrace{A^{\sigma-1}}^{\equiv \tilde{A}} \overbrace{z_{kj}^{\sigma-1}}^{\equiv \tilde{z}_{kj}}
\end{aligned}$$

Proof. Results follow from standard algebra. □

Proposition 3 thus clarifies that increases in firms' productivity and industry-level demand (ceteris paribus) lead firms to increase their inputs, their output and within-period profits.

4.2.2 Dynamic choices: Endogenous technology, entry & exit

The timing of the firm problem is as follows: At the beginning of a period, new potential entrants decide whether to enter or not. After that, incumbents and new entrants decide to invest in their productivity to gain future market shares taking into account changes in future demand. Firms then produce with their improved productivity, also benefiting from Learning-by-Doing (LBD) and exogenous growth. At the end of the period, firms exit exogenously. We discuss each in turn.

Entry In each industry, there is a large pool of potential entrants who can enter by paying an entry cost $c_{j,t}^E$ which is denoted in units of labor. By paying the entry cost, firms get to draw their initial productivity $\tilde{z}^-$ from the productivity distribution of incumbent producing firms. We denote firms' productivity before their productivity investments and production by $\tilde{z}^-$. The entry problem then writes:

$$\max \Big\{\mathbb{E}_{\tilde{z}}\big[V(\tilde{z}^-, j, t)\big] - w_t \cdot c_{j,t}^E, 0\Big\} \qquad \text{with:} \quad c_{j,t}^E = c^E \cdot \mathbb{E}[\tilde{z}^-]_{j,t} \cdot \tilde{A}_t^{-\xi} \tag{11}$$

Entry costs scale with the industry-level technology $\mathbb{E}[\tilde{z}^-]_{j,t}$, because it is harder to come up with a new idea for a firm in a more advanced industry and which ensures that entry costs scale in productivity. At the same time, costs also decline with $\tilde{A}_t$, which captures aggregate institutional reforms or better access to foreign technologies that make entry easier, and whose strength is parameterized by ξ. c^E denotes the level of entry costs.

Endogenous productivity growth Incumbents and new entrants then decide on their investments i in productivity $\tilde{z}$, taking as given expectations over future profits and any industry-level productivity growth from Learning-by-Doing $LBD_{j,t}$. At the same time, aggregate growth in $\tilde{A}$ improves overall productivity $\tilde{q}$ according to: $\tilde{A}_t/\tilde{A}_{t-1} = 1 + u$, which we assume is constant over time and across industries to capture any other sources of growth that are empirically important but plausibly orthogonal to the demand mechanism in our model. For firm k's productivity $\tilde{z}$ we assume that the law of motion is multiplicative:

$$\tilde{z}^+_{k,j,t} = \tilde{z}^-_{k,j,t} \cdot (1+i) \cdot (1 + \mathrm{LBD}_{j,t}) \tag{12}$$

with Learning-by-Doing specified in more detail below. Without loss of generality, we measure i in units of productivity and denote the costs of investing in i by $c_z(\tilde{z}^-, i, j, t)$, again denoted in units of labor. The firm innovation investment problem can then be written as:

$$V(\tilde{z}^-, j, t) = \max_i \left\{ \mathbb{E}_{u,j}\left[V^P(\tilde{z}^- \cdot (1+i) \cdot (1 + \mathrm{LBD}_{j,t}), j, t) \right] - w_t \cdot \tilde{z}^- \cdot \tilde{A}_t^{-\xi} \cdot \bar{c} \cdot \frac{i^2}{2} \right\} \tag{13}$$

where $V^P(\tilde{z}^+, j, t)$ denotes the value of starting production period t with productivity $\tilde{z}^+$ in industry j. Following the tractable setup in Acemoglu et al. (2018), we assume costs are quadratic in investments and scale with idiosyncratic productivity. This ensures bounded, closed-form productivity growth rates that are independent of firm size, consistent with Gibrat's law. We again assume that overall institutional improvements $\tilde{A}_t$ also reduce the costs of innovation, for example by making it easier to access foreign technologies. $\bar{c}$ denotes the level of innovation costs.

Finally, we specify the Learning-by-Doing technology. We assume that LBD growth is proportional to growth in industry-level physical production according to:

$$\mathrm{LBD}_{j,t} = \chi \cdot \max\left\{ \left((y_{j,t} - y_{j,t-1})/y_{j,t-1} \right), 0 \right\} \quad \text{with:} \quad y_{j,t} = \frac{Y_{j,t}}{N^{\nu}_{j,t}} \quad \& \quad Y_{j,t} = \frac{E_{j,t}}{A_t P_{j,t}} \tag{14}$$

The idea is that as industries scale up production and firms add new production lines, new opportunities for learning arise, similar to learning from new investments as in Arrow (1962). Importantly, χ captures the key strength of the Learning-by-Doing mechanism and will be central to capturing the demand-induced productivity growth in our model.

Two key implicit assumptions of the functional form are important to highlight. First, to reflect a large literature that emphasizes Learning-by-Doing growth from *within firm* growth in output (Levitt et al., 2013; Ilzetzki, 2024; Barwick et al., 2025), we allow LBD growth to only imperfectly scale in the number of firms as captured by the parameter $\nu \in [0, 1]$. For $\nu = 0$, all that matters for Learning-by-Doing is aggregate industry output $Y_{j,t}$, no matter if this growth comes solely from having more firms. On the other extreme, $\nu = 1$ signifies the case where all that matters is within-firm growth in physical output $Y_{j,t}/N_{j,t}$. The second assumption is on how we specify physical output. Specifically, by dividing nominal output also by A_t, we assume that Learning-by-Doing only comes from industry-specific growth drivers and not aggregate improvements in institutions. We believe this gives a clearer separation between the different drivers of growth and – if anything – gives a conservative role to Learning-by-Doing.

Exit At the end of the period, an (exogenous) share of firms λ survives and the rest exits.

4.2.3 Firm value functions & aggregation

Firm value functions Firm value functions in our setup are tractable and can be summarized in the following proposition:

Proposition 4 (Firm value function). *The value function of a firm with productivity $\tilde{z}^-$ in industry j at the beginning of period t can be written as:*

$$V(\tilde{z}^-, j, t) = \max_i \left\{ V^P(\tilde{z}^- \cdot (1+i) \cdot (1+LBD), j, t) - w_t \cdot c_z(\tilde{z}^-, i, j, t) \right\} \tag{15}$$

$$V^P(\tilde{z}^+, j, t) = \tilde{A}_t \cdot \tilde{z}^+ \cdot \tilde{\pi}_{j,t} + \frac{\lambda}{R} V(\tilde{z}^+, j, t+1) \tag{16}$$

Then V and V^P are homothetic in $\tilde{z}$: $V(\tilde{z}, j, t) \equiv \tilde{z} \cdot \tilde{V}(j, t)$ and $V^P(\tilde{z}, j, t) \equiv \tilde{z} \cdot \tilde{V}^P(j, t)$ such that:

$$\tilde{V}(j, t) = \max_i \left\{ (1+i) \cdot (1+LBD) \tilde{V}^P(j, t) - w_t \cdot \tilde{A}_t^{-\xi} \cdot \bar{c} \cdot \frac{i^2}{2} \right\} \tag{17}$$

$$\tilde{V}^P(j, t) = \tilde{A}_t \cdot \tilde{\pi}_{j,t} + \frac{\lambda}{R} \tilde{V}(j, t+1) \tag{18}$$

where optimal productivity investments i^* and $\tilde{V}(j, t)$ are given by:

$$i^* = \frac{(1+LBD)\tilde{V}^P(j, t)}{w_t \cdot \tilde{A}_t^{-\xi} \cdot \bar{c}} \tag{19}$$

$$\tilde{V}(j, t) = (1+LBD)\tilde{V}^P(j, t) \cdot \left[1 + \frac{(1+LBD)\tilde{V}^P(j, t)}{2 \cdot w_t \cdot \tilde{A}_t^{-\xi} \cdot \bar{c}} \right] \tag{20}$$

As long as there is a positive mass of entrants in equilibrium, the free entry condition implies that:

$$\tilde{V}(j,t) = w_t \cdot c^E \cdot \tilde{A}_t^{-\xi} \tag{21}$$

which is increasing in entry costs c^E and declining in the quality of institutions $\tilde{A}$. Taken together, this implies that the productivity investment rate is given by:

$$i^* = \frac{(1+LBD)}{w_t \cdot \tilde{A}_t^{-\xi} \cdot \bar{c}} \left[\tilde{\pi}_{j,t} + \frac{\lambda}{R} w_{t+1} \cdot c^E \cdot \tilde{A}_{t+1}^{-\xi} \right] \tag{22}$$

which is increasing in contemporaneous profits $\tilde{\pi}_{jt}$ (and thus demand), entry costs c^E, and decreasing in innovation costs $\bar{c}$.

Proof. Results follow from simple algebra. □

How does this model account for changes in demand driving productivity growth? As households become richer and shift their expenditures more towards industries that sell relative luxury goods, demand for this industry's varieties increases, which raises incumbents' and potential entrants' expected profits. Hence, more firms start to enter, however, the additional gains are only partially competed away by entrants because more production also induces stronger learning-by-doing gains, which reduces prices, profits and hence entry. Importantly, the growth induced by learning-by-doing is not internalized by firms nor by households when making their consumption choices. At the same time, the model features strong path dependence in productivity: past changes in demand still affect productivity today as the industry's level of technology slowly builds up over time, as new productivity growth builds on the shoulder of previous progress. This means that in this framework, industries that saw large demand in the past will still be very productive today (e.g. think agriculture). It also implies that consumption choices of households today have dynamic growth effects as they affect future productivity. The additional nuance in this economy is that the households who induce industry-level productivity growth in the first place may not be the same households who consume the good in the future since households' consumption moves towards relative luxury products over time.

The model also captures interesting industry dynamics. Firm-level endogenous choices aggregate up to industry dynamics and aggregate dynamics. While there is direct interaction between firms within industries due to price competition, there is also indirect interaction across industries and sectors via prices: If some industries develop quickly, this drives up aggregate wages, reallocates labor towards well-developed industries and may hinder the development of other industries who now face high wages. This tractable study of industry dynamics over the course of development comes at the cost of abstracting from some important firm dynamics. Most importantly, homothetic value functions come at the cost of identical within-industry productivity investment rates (Gibrat's law). Empirically, Gibrat's law holds well except for small firms.

Firm aggregation To keep track of the distribution of firms, we follow the endogenous growth literature (e.g. Sampson, 2016) by assuming that the initial industry-specific productivity ($\tilde{q}$) distribution of firms at time $t = 0$ is Pareto distributed with (industry-specific) support $\tilde{q} \in [\tilde{q}_{j,0}^{min}, \infty]$ and shape parameter α, such that the probability density function is given by: $f_{j,0}(\tilde{q}) = \frac{\alpha\left(\tilde{q}_{j,0}^{min}\right)^{\alpha}}{\tilde{q}^{\alpha+1}}$. With this choice, the productivity distribution of firms within each industry follows a traveling wave over time, with both the speed at which it changes and the mass of firms potentially changing over time and across industries. Specifically, we have that:

$$\mathbb{E}_{\tilde{q}}[\tilde{q}]_t = \begin{cases} \frac{\alpha \cdot \tilde{q}_{j,0}^{min}}{\alpha-1}, & \text{if: } t = 0 \\ \frac{\alpha \cdot \tilde{q}_{j,t-1}^{min} \cdot (1+i_{j,t}^*) \cdot (1+\text{LBD}_{j,t}) \cdot (1+u)}{\alpha-1}, & \text{if: } t > 0 \end{cases} \tag{23}$$

for $i_{j,t}^*$ giving optimal productivity investments as given by Proposition 4. The reason is that the industry-specific productivity distribution stays Pareto with the same shape parameter because exit is independent of productivity and entrants are a random sample from the incumbent productivity distribution.

4.3 Equilibrium

We now characterize the Markov perfect recursive competitive equilibrium path of the economy over time. We define the industry-level (productivity) distribution of firms at time t by $\mathcal{Q}_{jt}^{-}$ with $|\mathcal{Q}_{jt}^{-}| = N_{j,t}$ denoting its mass.

Equilibrium Starting from an initial state of the economy given by $\{L_0, H_0, \mathcal{H}_0, \{\mathcal{Q}_0^j\}_j\}$, an equilibrium is given by: (1) an exogenous path of household labor supply $\{L_t, H_t, \mathcal{H}_t\}_{t=0}^{\infty}$, (2) a path of endogenous prices $\{w_t^*, \{P_{jt}^*\}_j\}_{t=0}^{\infty}$, and (3) a path of optimal household expenditure functions $\left\{\{C_{ijt}^*, C_{ikjt}^*, E_{jt}^*\}_{i,j,k}\right\}_{t=0}^{\infty}$ such that at every time t:

- Household's optimal expenditures $\{C_{ijt}, C_{ikjt}, E_{jt}\}_{i,j,k}$ are characterized by Proposition 2 and equal to: $\{C_{ijt}^*, C_{ikjt}^*, E_{jt}^*\}_{i,j,k}$.

- Producing firms make (within-period) choices based on Proposition 3 and optimally invest in their productivity based on Equation (22) and Proposition 4. In particular, optimal firm-variety-specific prices and corresponding industry aggregate price indices $\{P_{jt}\}_j$ are consistent with the equilibrium prices: $\{P_{jt}^*\}_j$.

- The labor market clears. Namely:

$$H_t = \sum_{j=1}^{J} \left(\underbrace{N_{j,t}^E c_{j,t}^E}_{\text{entry costs}} + N_{j,t} \left(\underbrace{\mathbb{E}_{j,t}\left[\tilde{z}^-\right] \tilde{A}_t^{-\xi} \frac{\bar{c}_{j,t}}{2} (i_{j,t}^*)^2}_{\text{innovation costs}} + \underbrace{\mathbb{E}_{j,t}\left[\tilde{q}^+\right] \tilde{h}^*(w_t, P_{j,t}^*, E_{j,t}^*)}_{\text{production}} \right) \right) \tag{24}$$

- Firms enter based on Proposition 4 and exit exogenously at rate $1 - \lambda$.
- The distribution of firms updates based on optimal firm entry and exit decisions, optimal firm productivity investments, and the idiosyncratic shock process.
- Aggregate firm profits, which are rebated back to households, are given by:

$$\Pi_t = \sum_{j=1}^{J} \left(N_{j,t} \left(\underbrace{\mathbb{E}_{j,t}\left[\tilde{q}^+\right] \tilde{\pi}^*(w_t, P_{j,t}^*, E_{j,t}^*)}_{\text{production}} - \underbrace{w_t \mathbb{E}_{j,t}\left[\tilde{z}^-\right] \tilde{A}_t^{-\xi} \frac{\bar{c}_{j,t}}{2} (i_{j,t}^*)^2}_{\text{innovation costs}} \right) - \underbrace{w_t N_{j,t}^E c_{j,t}^E}_{\text{entry costs}} \right) \tag{25}$$

4.4 Solving the model computationally

To solve for the equilibrium, we use an iterative fixed point algorithm that jointly determines both investments in productivity and firm dynamics across industries and time. Our model is computationally tractable, especially along the time dimension, allowing us to solve for arbitrary aggregate transition dynamics. Namely, even in a single-industry model where all firms' productivity grows at the same rate, a difficulty is that if future demand enters the value function, then a general solution requires guessing over the entire path of future demand (which varies off a balanced growth path). In a multi-industry model, one would then have to guess over the entire path of future demand for all industries. Our model sidesteps this issue in that the value function depends on current period's demand and entry, and past production. This implies that one can start from an arbitrary initial equilibrium and simply solve iteratively, one period at a time. Analogously, it implies one can also end at an arbitrary point in time, neither having to know nor needing to make an assumption on the final steady state that the economy will reach (apart from regularity conditions that ensure the value function is bounded). In the industry dimension, the traveling-wave result implies that one only needs to keep track of the average productivity level to describe the entire distribution of firms within an industry. Still, the algorithm requires to solve a fixed point over J objects per time period, which means computational costs scale with the number of industries J and time periods. Appendix E provides a detailed description of the algorithm.

Table 9: Overview of directly calibrated parameters

Parameter	Description	Identification idea	Value
$\{L_t\}_t$	Population growth	Fully observed	$\{1.0, ..., 1.56\}$
$\{\sigma_{H,'87}, \sigma_{H,'11}\}$	Variance in HH skills	Variance in HH income	$\{0.83, 1.22\}$
$\{\mu_{H,'87}, \mu_{H,'11}\}_t$	Level of HH skills	Mean skills (Mincer regression)	$\{4.42, 4.27\}$
σ	CES w/in industry	Average profit share	5.13
α	Pareto shape parameter	Revenue distribution (MLE)	1.55
λ	Survival probability	Observed & standard value	8%
R	Gross interest rate	Observed (RBI)	1.08

Details:

4.5 Calibration

To calibrate the model, we find parameters such that the distance between a set of empirical moments and their model counterparts are as close as possible. As targets to pin down the key mechanisms of the model, we use the main regression results from Section 2 as well as moments from household-, industry- and firm-level data. Throughout, we follow the empirical results by solving the model for 25 years, starting from 1987 and ending in 2011, constructing model-based moments based on the exact same time period as in the data. To keep the calibration and model computation tractable, we aggregate industries into ten broader industry groups based on their income elasticities of demand; i.e. some industry groups produce more luxury and others more necessity goods. For the calibration, we note that one set of model parameters can be calibrated directly without solving the full model, while the remaining parameters are estimated jointly by solving the model.

4.5.1 Direct calibration

Population growth. For changes in L_t, we simply enforce observed changes in the Indian working population taken from the World Development Indicators, which increased by more than 50% from around 800 million to 1.26 billion over the period. We normalize $L_1 = 1$.

Household skills and changes in inequality. Inequality across households is defined by the time-varying distribution of skills $\mathcal{H}_t$ defined by its parameters $(\mu_{H,t}, \sigma_{H,t})$. We allow this distribution to vary over time to capture two distinct features: (1) changes in average skills that raise the total supply of efficiency units due to large educational gains in India over the time period of study (capturing changes in $\mu_{H,t}$), and (2) large changes in inequality as reported among others in Bharti et al. (2026) (captured by changes in $\sigma_{H,t}$).

For this, we choose the variance $\sigma_{H,t}$ to match the dramatic changes in the relative income share of the bottom 50%, the 50-90%, and the top 10% as reported in Bharti et al. (2026).[8] We separately estimate $\sigma_{H,t}$ for 1987 and 2011 and then linearly interpolate $\sigma_{H,t}$ in between to capture changes over time. We find that $\sigma_{H,t}$ increased by roughly 50% between 1987 and 2011 from 0.83 to 1.22, reflecting dramatic increases in inequality. For example, the income share of the top 10% alone increased from 35% to 55%.

For changes in the mean $\mu_{H,t}$, we enforce that average skills increased by 27% between 1987 and 2011 as measured by a standard Mincer regression, the same Mincer regression we also use to deflate labor costs in Section 2. We then pick the initial level of $\mu_{H,t}$ such that average household income matches India's (real) income in 1987, denoted in 2011 USD. This normalization ensures that all subsequent parameters have the same interpretable level. Given the log-normal distribution, average skills are given by: $\mathbb{E}_t(h) = exp(\mu_{H,t} + \frac{1}{2}\sigma_{H,t})$. We thus enforce our estimates of $\sigma_{H,t}$ and changes in $\mathbb{E}_t(h)$ to back out the resulting values for $\mu_{H,t}$. In line with large increases in the variance $\sigma_{H,t}$ and modest increases in average skills $\mathbb{E}_t(h)$, we find that the corresponding $\mu_{H,t}$ declined slightly over time (see Table 9).

Elasticity of substitution σ. Through the lens of the model and as is standard for models of monopolistic competition, the elasticity of substitution σ maps directly to the profit share of firms. Namely: $\mathbb{E}\left[\frac{\text{Input costs}}{\text{Revenue}}\right] = \frac{\sigma-1}{\sigma}$. We estimate the average share by industry using our firm data for the first (baseline) year and then average across industries. We find a value of $\sigma = 5.13$, well within the range of standard estimates. In principle, we could also directly feed in heterogeneous σ_j by industry, but we abstract from additional heterogeneity for our baseline results. Using the profit share ensures that our estimates are robust to other input costs that firms incur in the data, but our model abstracts from for simplicity.

Pareto shape α. Through the lens of the model, the revenue distribution has the same Pareto shape parameter as physical productivity with: $\tilde{f}_{j,0}(p^*_{k,j} y^*_{k,j}) = \frac{\alpha(\tilde{\kappa}_{j,0}\tilde{q}^{min}_{j,0})^{\alpha}}{\left(wh^*_{k,j}\right)^{\alpha+1}}$. We estimate α using standard maximum likelihood estimation using the cross-sectional revenue distribution in 1987. Again, we estimate α for each industry separately and then average across industries. From this we find that $\alpha = 1.55$, indicating a sizably right-tailed productivity distribution.

[8] Mathematically, we only need one ratio to identify $\sigma_{H,t}$ at any point in time. In practice we compute two shares: $\sigma^{(50)}_{H,t} = -\Phi^{-1}(\text{share}_{\text{bottom 50}})$ and $\sigma^{(90)}_{H,t} = \Phi^{-1}(0.9) - \Phi^{-1}(1 - \text{share}_{\text{top 10}})$. We then take the average of the two for our estimate of $\sigma_{H,t}$. Discrepancy of the two measures points to the log-normal distribution being a bad fit and observed income shares indeed point to a heavier right-tailed income distribution. However, since our model mechanism is about the consumption distribution and consumption tends to be less right-tailed than income, we believe the log-normal distribution provides a good approximation. For example, the consumption distribution as measured in the NSS (which would miss the right tail 10%) closely follows a log-normal distribution.

Table 10: Indirect inference: Model vs. Data

Param.	Description	Value	Target moment	Data	Model
$\bar{\sigma}$	CES across industries	0.25	First stage coefficient	0.96	1.01
u	Exogenous growth	0.078	Growth in income pc	0.88	0.89
χ	LBD growth source	1.0	IV TFPR regression	0.58	0.56
ν	LBD scale in firms	0.50	IV N firms regression	-0.13	-0.11
$\bar{c}/c^E$	Relative innovation costs	150	Aggr R&D cost share	0.0036	0.0036
ξ	Elasticity of cost declines	0.50	Overall increase in N	2.12	2.12
By industry: HH preferences & initial conditions					
$\{\epsilon_j\}_j$	Rel. income elasticity	Appendix	Engel curves		
$\{\Omega_j\}_j$	Preference weight	Appendix	Expenditure shares t=1		
$\{N_{j,0}\}_j$	Initial nb. of firms	Appendix	Model inversion		
$\{\tilde{q}^{min}_{j,0}\}_j$	Initial size of firms	Appendix	Normalized = 1		

Details: For the target moments we use the first stage coefficient reported in Table 2, column 6. Growth in (real) income per capita (in log differences) is as reported in National Accounts and our model replicates the way national accounts measures it. The IV TFPR regression coefficient is based on Table 8, column 2. The IV N regression coefficient is based on Table 8 column 4.

Survival probability. We choose $\lambda = 0.92$, corresponding to an annual exit rate of 8%, a standard value in the literature.[9]

Gross interest rate R. At last, we choose an annual interest rate of $r = 8\%$, which captures the average real bank lending rate in India over the time period, with $R = 1 + r$.

4.5.2 Indirect calibration: "Outer loop"

For the indirect calibration step, we solve the full model to find remaining model parameters such that the model matches key empirical target moments. In practice, we do so using an "outer SMM loop" where we estimate the main parameters related to the growth dynamics of the model, and an "inner loop" where we show how one can pin down remaining parameters conditional on the outer loop parameters. Results are summarized in Table 10. We start with the five outer loop parameters: $\Theta^{\text{outer}} = \{\bar{\sigma}, u, \chi, \nu, \bar{c}/c^E, \xi\}$ and assume that given Θ^{outer}, the remaining "inner loop" parameters of the model Θ^{inner} are given. Technically, we

[9]However, we note that this is larger than what carefully validated cohort-level analyses find in the more recent ASI data (e.g. see Chatterjee et al., 2025). Currently, setting a lower exit rate is problematic for our model because lower exit also implies lower entry rates, which means the model can more easily hit zero entry for specific industry-year combinations, which currently breaks our algorithm.

solve for Θ^{outer} solving:

$$\min_{\Theta^{\text{outer}}} \sum_{m=1}^{S} \frac{|M_m(\Theta) - M_m^E|}{|M_m^E|}.$$

where M^E denotes the empirical target moments and $M(\Theta)$ their counterparts in our simulated model economy. While parameters are jointly identified, we note that some moments still speak more closely to specific parameters.

Elasticity of substitution across industries $\bar{\sigma}$. On the demand side, the key elasticity of interest is $\bar{\sigma}$, which governs the strength of substitution effects compared to income effects in our model. That is, how strong does consumption respond to changes in relative prices as industries develop differentially? To identify $\bar{\sigma}$, we directly target the coefficient of the first stage of our empirical results (Table 2, column 6), which captures the elasticity with which actual demand (net of income and substitution effects) responds to changes in demand predicted only based on income effects. We replicate the same regression in our model, estimating model-implied industry-specific Engel curves in 1987 using the same spline-based estimator and shifting income as in the data to construct our model-implied demand shifter. Our model fits the first stage only if industry goods are strong complements ($\bar{\sigma} = 0.25$). The reason is that for higher levels of substitution, demand would be higher for industries that see stronger productivity growth, implying a much larger first stage. Comin et al. (2021) find sectoral estimates for $\bar{\sigma}$ that range from 0.25-0.62 for different country groups, putting as at the low end of their estimates.

Exogenous growth. On the supply side, the model features three main sources of growth: endogenous growth from innovation i, Learning-by-Doing and exogenous growth u. Below we discuss how the two endogenous sources of growth are pinned down and note that the importance of exogenous growth u is then identified as the residual that matches the overall growth rate of the economy. We choose as target moment the growth in real income per capita that is usually measured in National Accounts and make sure that we measure the same growth object in our model, with a further discussion of measurement in Section 5.1. In the data, real income per capita increased by 241%, or by 0.88 log points. To rationalize this, our model requires a growth rate of $\tilde{A}$ ($u = 0.078$), which corresponds to an annual growth rate in A of 1.84%. Mechanically, u therefore explains roughly 40% of the growth of real income per capita.

Strength of Learning-by-Doing χ. On the supply side, endogenous growth responds to demand changes via the strength of Learning-by-Doing as parameterized by χ. We pin down χ by targeting the main IV TFPR coefficient (Table 8, column 2) that summarizes the net productivity response to changes in demand. For this, we carefully replicate the same regression in our model. We construct instrumented demand as discussed above, but instead of taking a direct measure of firm- or industry-level productivity $\mathbb{E}[\tilde{q}^{+}_{j,t}]$ as a measure of TFPR, we follow closely how we construct TFPR in the data. Specifically, we use:

$$\widehat{TFPR}_{jt} = \mathbb{E}\frac{py/\hat{P}}{h^{(\sigma-1)/\sigma}}$$

where py is observed firm-level revenue and h is labor as the sole model-implied input (which in our empirical strategy includes also capital and intermediate inputs). Importantly, $\hat{P}$ is the industry-level output price deflator that is constructed by statistical agencies and weighs average price increases without measuring gains from variety, a point we discuss further in Section 5.1. Given this implementation, we show that the TFPR IV regression moment is still informative about changes in productivity in response to changes in demand. In fact, our model requires a very high Learning-by-Doing strength ($\chi = 1.0$) to produce a strong TFPR-moment response.

Within-firm Learning-by-Doing weight ν. To pin down how much LBD depends on within-firm changes in physical output versus overall industry-level output that includes the effect of more firms, we target the main IV regression moment on the number of firms. In the data, this moment is close to zero, which is at odds with standard growth models where demand changes would be "eaten up" by entry. In combination with a large Learning-by-Doing strength χ, our model is able to correctly capture the close-to-zero response of entry with $\nu = 0.5$, implying that half of Learning-by-Doing stems from within-firm output growth and the other half from industry-wide output growth that includes entry.

Relative innovation costs $\bar{c}/c^E$. Innovation costs relative to entry costs directly govern the level of endogenous innovation i^* and the R&D cost share in our model. As empirical target, we thus use the aggregate R&D cost share (over value added) in the ASI data for the year 2015, the first year where R&D expenditures are reported in the data.[10] This share is small, at 0.36%, in line with a low level of R&D activity in India. Our model is able to exactly fit this small share, which also implies a small corresponding endogenous innovation rate i^*. We found that what matters quantitatively is the ratio of the costs and not their level; we currently normalize the level at $c^E = 1$.

Elasticity of cost declines ξ. At last, we pin down ξ by fitting the overall increase in the number of firms over time. Specifically, in the data we measure the number of factories,

[10] Unfortunately, this also means we cannot study how R&D causally responds to changes in demand.

which more than doubled over the time period from 1987 to 2011, and which our model is able to fit well. We find a corresponding elasticity of entry cost declines of $\xi = 0.5$, which governs the economy-wide response of entry to growth.

4.5.3 Indirect calibration: "Inner loop"

Conditional on the "outer loop" parameters, we now show how to back out the remaining "inner loop" parameters.

Demand-side parameters. The remaining demand-side parameters of the economy are the industry-specific relative income elasticities $\{\epsilon_j\}_j$ and industry-specific expenditure weights $\{\Omega_j\}_j$. To estimate the elasticities $\{\epsilon_j\}$, we draw on the implicit Marshallian demand equations of our model-implied expenditure system in the first model period:

$$log\left(\frac{\omega_{i1}^j}{\omega_{i1}^b}\right) = \alpha_j + \underbrace{(\epsilon_j - 1)}_{\equiv\beta_j}\underbrace{\Big((1-\bar{\sigma})log\left(E_{i1}\right) + log(\omega_{i1}^b)\Big)}_{\equiv X_{i,j,t}} \tag{26}$$

where α_j captures an industry-specific constant that depends on relative prices and the expenditure weights Ω_j,[11] and where b is the base industry so that all elasticities ϵ_j and industry-weights Ω_j are relative to industry b. We take toilet soap as the base good ($\epsilon_b = 1$), given that it is widely consumed everywhere, but note that the choice of the base good only affects interpretation of Ω_j and ϵ_j and does not affect the model dynamics. We estimate Equation (26) as a linear regression industry by industry (except b). Income elasticities are then identified from the Engel curve; i.e. how expenditure shares vary with the level of total expenditures. Conditional on $\bar{\sigma} < 1$ (the case where industries are complements), higher ϵ_j means a good's (relative) expenditure share is increasing in total expenditures. For computational tractability, we then aggregate industries into ten broader industry groups. We do so by ranking industries by $\epsilon_j(\bar{\sigma})$ and creating ten bins with a roughly equal number of industries. Table 11 shows our income elasticity estimate for each of the industry groups, with full industry-level estimates in the Appendix. The elasticities broadly align with economic intuition: vehicles, more expensive household appliances, jewellery, watches and furniture are all relatively luxury products. On the other end of the spectrum, the goods with the lowest (relative) income elasticities are food, firewood, cooking oil and basic clothing products.

Finally, we back out industry-specific expenditure weights Ω_j model-consistently by exactly matching each industry group's initial expenditure share in period 1 of the model,

[11] That is: $\alpha_j \equiv log(\Omega_j) + (1-\bar{\sigma})log\left(\frac{P_{j1}}{P_{b1}}\right) + (1-\bar{\sigma})(1-\epsilon_j)log\left(P_{b1}\right)$. The Marshallian demand equation corresponds to Equation (11) in Comin et al. (2021).

Table 11: Overview of industry-specific demand parameters

Group	ϵ_j	$E_{j,1}/E_1$ (%)	Ω_j
Staple foods and firewood	0.018	17.40	10.87
Fish, pork, spices and tobacco	0.401	16.93	10.82
Cooking oils, beef and toiletries	0.981	23.64	15.08
Rubber footwear, sarees, and personal care products	1.481	21.61	12.02
Assorted nuts, household textiles and stationery products	2.151	4.25	1.39
Confectionery, medicines and glassware	2.581	8.38	2.57
Knitted garments, lighting and stainless steel products	2.949	3.71	0.82
Leather footwear, spectacles, furniture and watches	3.512	1.54	0.22
Musical instruments, jewellery, and electrical houseware	5.386	1.20	0.05
Alcohol, cooling appliances and vehicles	10.008	1.34	0.001

Details: The table reports main demand-side parameters for the 10 broad industry groups of the model. The ϵ_j are (relative) income elasticities of demand. Since $\bar{\sigma} = 0.4$ in the baseline calibration, higher ϵ_j implies the good is relatively a more luxury good. Column 2 reports the initial expenditure shares for each of the sectors $E_{j,1}/E_1$, which we target to obtain the demand weights Ω_j.

enforcing model-consistent prices. This implies high weights for the industries with the highest expenditure shares such as food, firewood, tobacco, toiletries and basic clothing. Sectors with lower weights are more luxury products whose initial expenditure shares are small such as vehicles and electrical appliances.

Initial conditions. Conditional on the remaining model parameters, we back out initial conditions of the economy in 1987. We do so by initializing the economy five years earlier to ensure that the starting point in 1987 is consistent with subsequent model dynamics. The state of the economy at any point in time is given by the stock of knowledge $\{\tilde{q}_{j,t}^{min}\}_j$ and the stock of firms $\{N_{j,t}\}_j$. We normalize the initial stock of knowledge $\{\tilde{q}_{j,1}^{min}\}_j$ in the first pre-period such that $\mathbb{E}[\tilde{z}^+]_{j,pre} = 1$ and $\tilde{A}_{pre} = 1$. This is without loss of generality because the "level" of quality/productivity $\tilde{q}$ across different industries is undefined; what matters for *changes* in prices and demand are *changes in* $\tilde{q}$ over time. Any differences in the initial level of industry demand is then simply eaten up by differences in preferences Ω_j.

For the initial number of firms by industry, we note that as long as there is positive entry, the equilibrium number of firms in industry j at any point t is given by:

$$N_{j,t} = \frac{E_{j,t}}{\sigma \mathbb{E}[\tilde{z}^+]_{j,t}} \left[\frac{\bar{c}}{1 + LBD_t} \left(\sqrt{1 + 2\frac{\bar{c}}{c^E}} - 1 \right) - \frac{\lambda}{R} \frac{c^E}{1+u} \right]^{-1} \tilde{A}_t^{\xi}$$

The main reason we initialize the economy a few years before 1987 is that two objects are generally unknown when initializing the economy in period t_0 and have to be chosen in an ad-hoc way that can affect first year dynamics: (1) the initial growth rate $LBD_t(G^y)$, which depends on previous production via growth G^y, and (2) the previous number of firms N_{j,t_0-1} which determines entry. In practice, we initialize the economy assuming that growth G^y in the first pre-period is equal across industries and equals the initial growth in household expenditure H, which is 2.3% in the data. For the previous number of firms N_{j,t_0-1}, we assume that $N_{j,t_0-1} = N_{j,t_0}$, so that entry and exit rates are exactly equal in the first pre-year. We believe this gives a good starting point as it ensures that initial profits (which depend on entry costs) are also reasonable. Both assumptions imply that initial growth in pre-period 1 is equal across industries, but industry growth rates immediately vary endogenously afterwards. This gives a parsimonious and robust way to calibrate initial conditions that imply model-consistent growth dynamics by 1987.

5 Quantitative Results

We now use the estimated model to quantitatively answer two sets of questions. First, we investigate the drivers of growth in India over the period from 1987-2011, decomposing aggregate growth into the contributions from population growth, endogenous innovation, learning-by-doing, and growth in human capital. In the second step, we turn our eyes towards inequality: Who benefited most from growth over time and how biased was growth towards the rich? As part of this analysis, we also look at the role of policy: can redistributive policies make growth more equal? And what are the effects on welfare: should policy redistribute income to direct technological change towards goods that the poor consume more? Or is a bias of growth towards goods that the rich consume beneficial for the poor as they become richer and eventually consume these same goods?

5.1 A Note on Measuring Growth

To answer these questions, we consider two main measures of growth throughout. First, we construct "measured growth" as the growth in real expenditure per capita that is usually measured in National Accounts. This measure is constructed by dividing per capita household expenditures by an aggregate price index P_t. To be comparable to how Indian National Accounts measure consumption growth, we use a chain-weighted Laspeyres price index as the measure for P_t, which weights price changes across industries based on aggregate expenditure weights. Importantly, our model-based prices $P_{j,t}$ include gains from variety that National Accounts do not generally incorporate; we thus construct two different versions: "measured growth" which uses prices without variety gains, and "measured growth with N"

Table 12: Decomposing the main drivers of Indian growth, 1987-2011

Scenario	Measured growth	Measured growth with N	Welfare growth	Growth contribution
Baseline	0.89	1.01	0.56	100%
No exogenous growth	0.51	0.49	0.22	42.7%
No growth in human capital	0.62	0.74	0.38	30.3%
No learning-by-doing	0.69	0.96	0.52	22.5%
No population growth	0.84	0.91	0.50	5.6%
No endogenous innovation	0.84	1.00	0.55	5.6%

Notes: "Measured growth" is (log) growth in real expenditure per capita constructed using a chain-weighted Laspeyres price index as deflator P_t, to follow Indian National Accounts. "with N" includes gains from variety and "without N" excludes gains from variety, the latter is the closest equivalent to National Accounts. Entries report log differences: i.e. "Measured growth" $= log(E_T/P_T) - log(E_{t_0}/P_{t_0})$ with $t_0 = 1987$ and $T = 2011$. Household welfare is measured by the average consumption index C_{ht} implied by the NH-CES consumption aggregator. The last column computes the overall growth contribution of the respective factor using: $1 - g_B/g_{CF}$ where g_B gives the baseline measured growth, and g_{CF} denotes counterfactual measured growth.

that incorporates variety gains.

Our second measure of growth is based more directly on welfare and takes seriously that the growth rate of the economy depends on whose preferences we take as reference. Given that households consume systematically different goods, aggregate growth not only differs for households due to different changes in their incomes, but also because different goods see different price changes over time. As the main measure of welfare growth, we use growth in the average consumption index C_{it} implied by the NH-CES demand system of our model. Alternatively, we look directly at the cross-sectional distribution of C_{it}.

5.2 Decomposing the Indian growth experience

We now decompose the Indian growth experience using our calibrated model. Table 12 reports our main growth decomposition results. Focusing on measured real output growth as the main result of interest, we find that exogenous growth explains the lion's share of growth (42.7%), followed by increases in human capital (30.3%), Learning-by-Doing (22.5%), and population growth and endogenous innovation each explaining roughly 5% of growth. To derive each of these contributions, we solve a separate counterfactual growth path of the Indian economy from 1987 to 2011 shutting down the specific mechanism of interest.[12] Note

[12] To shut down exogenous growth we set $u = 0$. For endogenous innovation, we make innovation prohibitively expensive ($\bar{c} \to \infty$). For learning-by-doing, we set $\chi = 0$. For "no population growth", we solve the counterfactual growth path without changes in the population L_t. Finally, for "no growth in human capital" we assume that average human capital stayed constant and re-solve for the corresponding $\{\mu_t, \sigma_t\}$.

that these contributions do not sum to 100% because different growth drivers interact.

Learning-by-Doing – the main driver of demand-driven growth – is thus about 4x as important as R&D-based innovation in the Indian context, in line with the small quantitative role of R&D investments, but only about half as important as other sources of growth (u). Importantly, the contributions of growth differ quite a bit depending on the measure of growth taken. Including variety gains as in "measured growth with N" and in "welfare growth" reduces the contribution of Learning-by-Doing to 5-7%. The main reason is that within-firm expansion due to strong Learning-by-Doing in the baseline economy crowds out further entry in industries whose demand grows fast. In the counterfactual growth path of the economy without LBD, more entry and the ensuing variety gains compensate for a sizable part of the physical productivity losses.

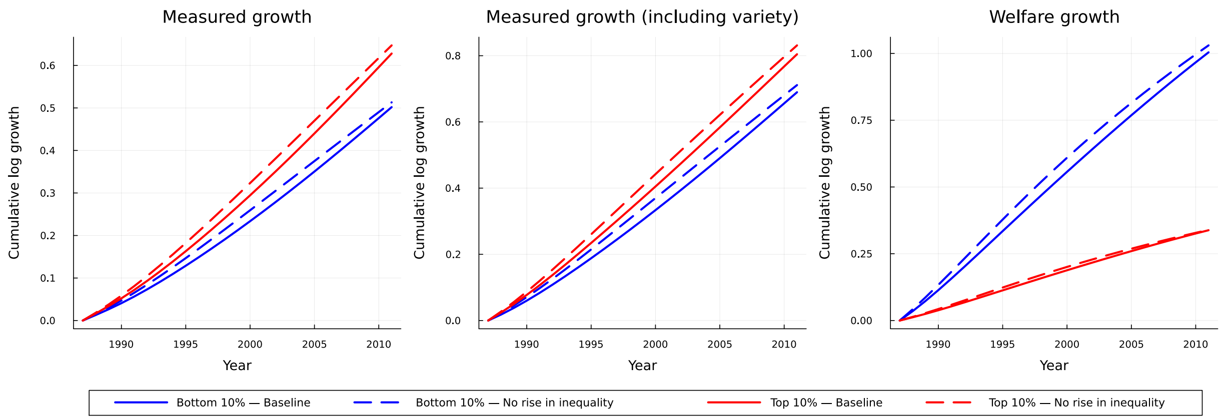
5.3 How biased is growth towards the rich?

We now study how growth played out across the income distribution, how inequality itself affected the growth process by shifting productivity growth and entry towards the preferences of the rich, and whether policy should redirect demand for growth.

How biased was the growth process towards the preferences of the rich? We find that the rich in 1987 experience about 25% higher subsequent real consumption growth than the poor. To derive this number, we consider the following intuitive exercise: we start from the distribution of households in 1987 and look at the bottom and top 10% of income. We then fix these households and their corresponding skills h and ask how both groups fared differently over India's growth experience. Figure 2 plots their consumption growth under our three main measures: "measured growth", "measured growth with variety gains" and "welfare growth". Solid lines describe growth for the baseline economy. In line with growth being biased towards luxury goods which the rich demand more, we find a 25% growth bias for the rich by 2011 based on measured growth (0.63/0.50). Including variety gains leads to a slightly lower relative growth bias (16.5%), because the level of growth is higher as everyone benefits from more varieties over time and variety growth is slightly less biased towards the preferences of the rich. The big difference is welfare growth, which seems to be pro-poor as shown in Panel (C). However, this is simply because the consumption aggregator C is concave, which implies that the same real consumption expenditure gains mechanically lead to stronger gains for poor households.

How did increases in inequality exacerbate the growth bias? Perhaps surprisingly, we find that the dramatic rise in inequality in India over the past 30 years did not system-

Figure 2: India's growth bias in favor of the rich: Bottom 10% vs. Top 10%



atically exacerbate the bias of growth towards the rich, apart from the very top. To show this, we study a counterfactual growth path assuming that inequality had instead remained constant at its 1987 level.[13] Since this counterfactual also changes the distribution of skills in the economy, we cleanly separate the effect by focusing on the same households (with their fixed h) as before. As shown by the dashed lines in Figure 2, growth without further increases in inequality would have been less biased in the early years but eventually ended up with a similar bias by 2011. The reason is due to differential industry growth whose incidence plays out differently over time. In the counterfactual economy without growing inequality, demand for more necessity industries would have been higher, raising entry and productivity growth, while only the top two (relative) luxury industries would have grown less due to fewer rich people demanding vehicles and electrical appliances. In the early years, stronger relative price declines in more necessity industries benefits the poor more, reducing the growth bias. However, as everyone becomes richer, these initial price declines benefit the poor less as they increasingly demand more luxury products which would have already been slightly cheaper if inequality had been higher earlier on. The interesting question is why even the rich would have benefited if inequality had not increased, as evidenced by the slight positive gains for welfare growth in Figure 2. The reason is two-fold. First, even the rich do not have large expenditure shares for luxury industries. For example, the vehicle industry, which alone saw 20% higher relative price declines due to growing inequality, only has a 6% expenditure share by 2011 by the top 10%. So the rich also strongly benefit from higher growth in relative necessity industries. The second reason is more subtle: in the counterfactual without growing inequality, demand for basic food industries would have been higher, leading to positive Learning-by-Doing growth which would have otherwise not

[13]Specifically, we keep $\sigma_{H,t}$ fixed at its 1987 value and solve for the corresponding $\mu_{H,t}$ series such that average skills increased as for the baseline economy.

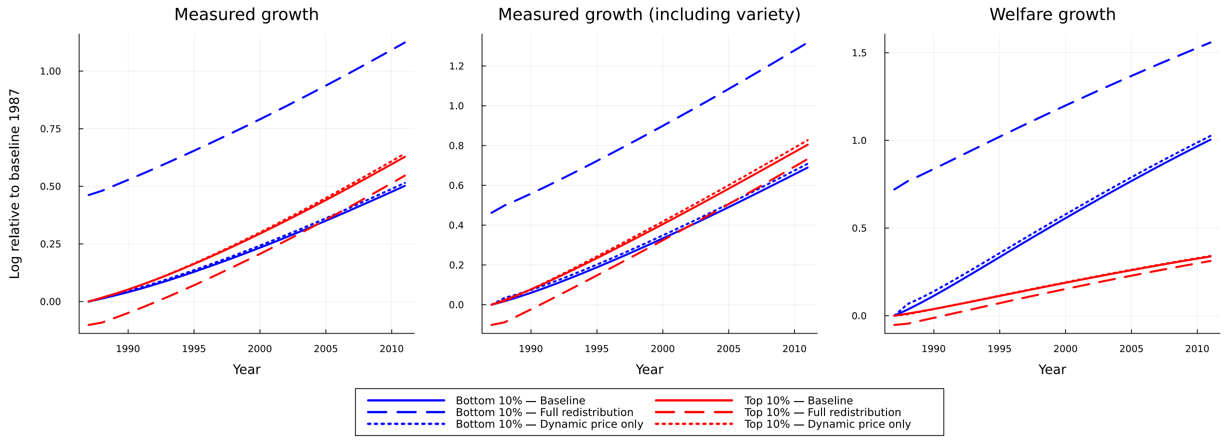


Figure 3: Static and Dynamic benefits of redistribution: Bottom 10% vs. Top 10%

seen any growth. This benefit of managing demand is an externality that households do not take into account when making consumption choices.

Should policy redirect demand for growth? Since the distribution of income affects which industries grow more over time, a natural question is whether policy – by redistributing income – should actively try to counter the pro-rich bias of growth? And whether redistribution policy is a powerful tool to do so. Quantitatively, we find that by far the main benefit of redistribution are its static gains: marginal utility of consumption for the poor is simply much larger than for the rich, which provides a strong rationale for redistribution. However, we also find a small positive dynamic growth benefit from redistribution, as entry and Learning-by-Doing align closer with the preferences of the poor.

To show this, we consider the following exercise. We introduce a proportional labor income tax τ whose proceeds are redistributed lump-sum to households via transfers T_t, so that household income is:

$$I_{it} = (1-\tau) \times h_{it} \times \left(w_t + \tilde{\Pi}_t\right) + T_t$$

This introduces redistribution by shifting income across the distribution to the bottom, with higher τ implying more redistribution and $\tau = 1$ denoting the case of full equality across households. Figure 3 shows results for the counterfactual growth path of the economy under a 20% tax rate, again distinguishing effects for the top and bottom 10% and showing effects for our three measures of growth. The solid line shows the baseline economy, while the dashed line shows counterfactual growth under the 20% redistribution case. The main effect from redistribution is a permanent static gain for poor households whose consumption gains jump up with the beginning of the policy. To dinstinguish the pure dynamic effect on growth, the dotted line shows growth using counterfactual prices that are affected from redirecting entry and Learning-by-Doing but fixing household income to its baseline path without redistribution. The difference between the dotted and solid line indicates small positive growth effects that accumulate over time. However, static gains clearly trump these dynamic effects by orders of magnitude. What is important to highlight is that these positive dynamic growth effects from redistribution stand in stark contrast to the idea that growth driven by the preferences of the rich eventually benefits the poor as they get richer (e.g. Oberfield, 2023). Instead, we find that redistribution affects entry and productivity growth immediately and this weighs stronger than any long-term benefits from cheaper luxury products.

6 Conclusion

This paper quantified the importance of domestic demand in driving and directing growth in a development context, and investigated how biased this growth is towards the rich. We provided novel empirical evidence for the role of domestic demand in driving structural change and productivity growth at the industry level, but with little effects on net entry and product varieties. Using a structural model of endogenous growth calibrated on the empirical results, the paper then showed that domestic demand via population growth and Learning-by-Doing drove about 30% of growth in India over the period from 1987 to 2011. However, this growth was strongly biased towards the preferences of the rich, with the rich experiencing about 25% higher real consumption gains than the poor. At last, we find that redistribution policy can be effective in letting the poor benefit more from growth, with redistribution even having positive growth effects by redirecting firm entry and productivity growth towards more broadly consumed industries.

At the end, at least two important caveats are in order. First, the main results are based on formal manufacturing firms only, leaving important parts of the economy outside the analysis, including a large informal sector and services for which productivity growth responses may operate differently. Second, this paper has focused on final, domestic consumption goods only, leaving intermediate good production and trade outside the scope of this paper. Better understanding how the production network or trade interact with the demand drivers of growth highlighted in this paper are fascinating questions that we leave for future work.

References

Acemoglu, Daron and Joshua Linn, "Market size in innovation: theory and evidence from the pharmaceutical industry," *The Quarterly Journal of Economics*, 2004, *119* (3), 1049–1090.

_ **, Ufuk Akcigit, Harun Alp, Nicholas Bloom, and William Kerr**, "Innovation, Reallocation, and Growth," *American Economic Review*, 2018, *108* (11), 3450–3491.

Allcott, Hunt, Allan Collard-Wexler, and Stephen D O'Connell, "How do electricity shortages affect industry? Evidence from India," *American Economic Review*, 2016, *106* (3), 587–624.

Arrow, Kenneth J, "The economic implications of learning by doing," *The review of economic studies*, 1962, *29* (3), 155–173.

Atkeson, Andrew and Ariel Tomas Burstein, "Innovation, firm dynamics, and international trade," *Journal of Political Economy*, 2010, *118* (3), 433–484.

Atkin, David, "Trade, Tastes, and Nutrition in India," *American Economic Review*, 2013, *103* (5), 1629–1663.

Barwick, Panle Jia, Hyuk-Soo Kwon, Shanjun Li, and Nahim B Zahur, "Drive down the cost: Learning by doing and government policies in the global EV battery industry," Technical Report, National Bureau of Economic Research 2025.

Beerli, Andreas, Franziska J Weiss, Fabrizio Zilibotti, and Josef Zweimüller, "Demand forces of technical change evidence from the Chinese manufacturing industry," *China Economic Review*, 2020, *60*, 101157.

Bharti, Nitin Kumar, Lucas Chancel, Thomas Piketty, and Anmol Somanchi, "Income and wealth inequality in India, 1922–2023: The rise of the Billionaire Raj," *The World Bank Economic Review*, 2026, p. lhag013.

Bohr, Clement E, Martí Mestieri, and Emre Enes Yavuz, "Engel's Treadmill: The Perpetual Pursuit of Cornucopia," Technical Report, Working Paper 2021.

Boppart, Timo, "Structural Change and the Kaldor Facts in a Growth Model with Relative Price Effects and Non-Gorman Preferences," *Econometrica*, 2014, *82* (6), 2167–2196.

Chatterjee, Shoumitro, Kala Krishna, Kalyani Padmakumar, and Yingyan Zhao, "No country for dying firms: Evidence from india," Technical Report, National Bureau of Economic Research 2025.

Comin, Diego, Danial Lashkari, and Marti Mestieri, "Structural change with long-run income and price effects," *Econometrica*, 2021, *89* (1), 311–374.

Deaton, Angus, "Measuring poverty in a growing world (or measuring growth in a poor world)," *Review of Economics and statistics*, 2005, *87* (1), 1–19.

Fan, Tianyu, Michael Peters, and Fabrizio Zilibotti, "Growing like India—the unequal effects of service-led growth," *Econometrica*, 2023, *91* (4), 1457–1494.

Foellmi, Reto and Josef Zweimüller, "Income distribution and demand-induced innovations," *The Review of Economic Studies*, 2006, *73* (4), 941–960.

Goldberg, Pinelopi K, Amit K Khandelwal, Nina Pavcnik, and Petia Topalova, "Multiproduct firms and product turnover in the developing world: Evidence from India," *The Review of Economics and Statistics*, 2010, *92* (4), 1042–1049.

Groote, Olivier De and Frank Verboven, "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems," *American Economic Review*, 2019, *109* (6), 2137–2172.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi, "Two perspectives on preferences and structural transformation," *American Economic Review*, 2013, *103* (7), 2752–2789.

Ilzetzki, Ethan, "Learning by Necessity: Government demand, capacity constraints, and productivity growth," *American Economic Review*, 2024, *114* (8), 2436–2471.

Levitt, Steven D, John A List, and Chad Syverson, "Toward an understanding of learning by doing: Evidence from an automobile assembly plant," *Journal of political Economy*, 2013, *121* (4), 643–681.

Loecker, Jan De, Pinelopi K Goldberg, Amit K Khandelwal, and Nina Pavcnik, "Prices, Markups, and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Luttmer, Erzo GJ, "Selection, growth, and the size distribution of firms," *The Quarterly Journal of Economics*, 2007, *122* (3), 1103–1144.

Oberfield, Ezra, "Inequality and measured growth," Technical Report, National Bureau of Economic Research 2023.

Sampson, Thomas, "Dynamic Selection: an Idea Flows Theory of Entry, Trade, and Growth," *The Quarterly Journal of Economics*, 2016, *131* (1), 315–380.

A Data and Empirical Evidence

A.1 Estimation of TFPR

Estimation of Productivity (TFPR): We first outline the method used to estimate TFPR (Total Factor Productivity Revenue), which is our preferred productivity measure. Assume that product j is produced using a Cobb-Douglas production function as follows:

$$Y_{jst} = A_{jst} K_{jst}^{\alpha_K} L_{jst}^{\alpha_L} M_{jst}^{\alpha_M} F_{jst}^{\alpha_F} \tag{27}$$

where Y_{jst} is the output of product j, in state s at time t, A_{jst} is the Hicks neutral productivity, K_{jst}, L_{jst}, M_{jst} F_{jst} are the capital, labor, material inputs and fuels respectively. Log-linearising yields the usual TFP residual:

$$\log(A_{jst}) = \log(Y_{jst}) - \alpha_K \log(K_{jst}) - \alpha_L \log(L_{jst}) - \alpha_M \log(M_{jst}) - \alpha_F \log(F_{jst}) \tag{28}$$

Note that we allow for state-product level input elasticities $\alpha_{s,j}^X$ where $X \in \{K, L, M, F\}$, and we have suppressed the subscripts s and j on input elasticities. From standard static optimization, we can derive the following mapping between the elasticities of different inputs α_X and the input cost shares for input X_{jt} where $x \in \{K, L, M, F\}$:

$$\alpha^X = \frac{p_{jt}^X X_{jt}}{p_{jt}^Y Y_{jt}} \tag{29}$$

We calculate input elasticities at the state-product level by first calculating weighted medians of the nominal expenditure shares within a state-product-year cell in the ASI establishment level data[14], and then averaging across years. Once we have identified the input elasticities, we deflate all nominal values using the price deflators which have been normalized to 1 in the base year (1989). We deflate the labor input as well to obtain a quality adjusted stock of labor at the establishment level. To do this, we first estimate a Mincer style regression using data from the NSS employment survey (year 1989). This gives us a state-year level quantity of labor which we then multiply to total employment from the ASI data to obtain state-year level measures of quality-adjusted labor quantity and implied prices of labor. Dividing the establishment level wage bills with the state-year prices gives us the labor quantity at the plant level as well. Note that here we are making the assumption of uniformity in price of labor across industries within the state and year.

[14] weights are sample weights provided by the ASI

The final step involves plugging in the calculated input elasticities and the deflated input values to compute the residual Log(A_{jst}) as :

$$\log(A_{jst}) = \log(Y_{jst}) - \hat{\alpha}_K \log(K_{jst}) - \hat{\alpha}_L \log(L_{jst}) - \hat{\alpha}_M \log(M_{jst}) - \hat{\alpha}_F \log(F_{jst}) \tag{30}$$

A.2 Input-Output tables

The Input-Output transaction tables (IOTT) of the Indian economy are officially constructed by the Central Statistical Organisation (CSO) as part of the National Accounts statistics and are updated in intervals of five years. They are created following the basic principles enunciated in the United Nations System of National Accounts. Over time, the IOTTs became more detailed by including more sectors and commodities of the economy. The 1989 & 1993 IOTTs had 115 sectors, while the IOTTs for (give years) were extended to 130 sectors.

Method of construction: (add some additional details here)

Data sources: To construct the IOTTs, the CSO draws on a very similar combination of data sources as we do to obtain a representative picture of the entire Indian economy.

For agricultural crop production, the CSO also draws on the Cost of Cultivation Studies (CCS). For the cost of various plantation crops, the CSO additionally draws on direct information provided by crop-specific plantation boards. The crop-specific input usage also relies on the main input categories that the CCS provides. Some of the reported inputs in the CCS are then further broken down using additional sources. For example, in the case of government-provided irrigation services, these services are further broken down into sub-inputs using government cost reports. The CSO also integrates all by-products that share the same input costs and also integrates downstream industries in case these do not show up as separate industries in the IOTTs. Specifically, the respective milling of rice, flour, dal, and other grains are integrated with their crops.

Besides crop production, major agricultural sectors include animal husbandry, forestry and fishing sectors. For animal husbandry, the major data sources are the Livestock Censuses to obtain estimates of the animal populations, studies carried out by the Indian Agricultural Statistics Research Institute (IASRI) for animal-specific input uses, and the input costs of eight downstream sectors to obtain estimates of output after deducting market charges. Output and input costs of the forestry sector are obtained via the Chief Conservator of the Forests, through the DES in the states and through State Forest Departments. For fishing, the CSO builds on a combination of data from State Fisheries Departments and the Livestock Censuses.

For manufacturing, which captures more than 60 sectors in the IOTT, sector-specific entries are constructed separately for registered activities using the Annual Survey of Industries (ASI) and for unregistered activities using (less regular) surveys on unregistered manufacturing by the NSSO. This mapping is arguably the cleanest given that both surveys ask detailed questions on outputs and inputs and given correspondences between the IOTT sectors and NIC industry classifications.

For the remaining sectors, mostly service sectors but also mining & quarrying, the IOTT uses a combination of data sources. A key source of information is the Enterprise Survey, which covers many services. For other sectors such as construction and transport, detailed sector-specific information is available from a number of different sources. For mining and quarrying as for utility sectors, data is taken directly from annual reports by the major private or government-owned providers.

Imports & Exports: Exports are measured on f.o.b. basis and for merchandise are directly taken from the DGCI&S publication 'Monthly Statistics of Foreign Trade of India, Vol. I, Exports'. For services, exports are directly from the RBI. Similarly for imports, item-wise details of imports of merchandise at c.i.f. values are available in the DGCI&S publication 'Monthly Statistics of Foreign Trade of India, Vol. II, Imports', while imports of services are obtained from the RBI.[15]

A.3 Prices

We draw on wholesale prices collected for the India-wide Wholesale Price Index (WPI) by the Ministry of Commerce & Industry (MCI).[16] The WPI is a measure of the average change of prices of a fixed set of goods at the first point of bulk sale in a commercial transaction in the domestic Indian economy over a given period of time. In practice the WPI is close to a producer price index.[17] The reported micro-level price data are monthly prices for roughly 500 products, which correspond to NIC 4-digit level (check this!), going as far back as the

[15] Imports are reported at c.i.f. values and are exclusive of import duties and domestic taxes. The commodity-wise custom duties (both on imports and exports) are available from the DGCI&S. Both have been aggregated to import duties at the sector level as in the IOTT. Adjustments have been made for refunds & withdrawals to arrive at net import duties.

[16] In principle, a number of Indian states also collect prices for state-level Wholesale Price Indices, with a focus on agricultural products. These are: Assam, Bihar, Haryana, Karnataka, Punjab, Uttar Pradesh and West Bengal. We currently do not use the state-level price information.

[17] For agricultural products, the prices capture a combination of farm harvest prices and prices at the village market procured by Agricultural Marketing Produce Committees (AMPC), which safeguard Indian farmers from excessively-high retail price spreads. For manufactured products, the wholesale prices also capture prices for bulk transactions at the ex-factory gate, ex-mill or ex-mine level (that is, excluding transport costs). They exclude rebates, taxes and levies apart from the excise duty. Services are currently not included in the WPI.

1950s.

Selection of items, varieties, grades and markets: While the reported micro data is monthly for roughly 500 products, the monthly price of each product is actually aggregated from weekly item-level price quotes, with each product having multiple item specifications for which prices are administered. For example, the 1981-82 series included 447 distinct products supported by 2371 individual price quotes, roughly five price quotes per product. For each price-quoted item, the MCO agrees on a precise specification of a product that includes make, model, features along with the unit of sale, type of packaging etc. for which prices are collected every Friday. In case of changes in quality and specifications, the MCO states that "due adjustments are made as per the standard procedures" (unclear what exactly is done). To aggregate prices across items/price quotes at the product-level, the WPI procedure uses simple unweighted averages across relative within-item changes in prices.[18]

More specifically, the MCO agrees on an item specification in consultation with source agencies. Actual price quotes are obtained on a voluntary basis and are partly collected through administrative agencies, the Chambers of Commerce, Trade Associations, Business Houses and leading manufacturing plants. For example, for agricultural products, product specifications are agreed upon with each commodity-specific production board and the Directorate of E&S (M/O Agriculture), while prices are obtained through administrative agencies directly. For manufacturing, separate sets of item-level prices are collected for the organized and unorganized part of manufacturing. The former is based on the ASI and price quotes for selected specifications are obtained from the largest manufacturers . For the unorganized sector more price quotes are used due to the higher price variability and potentially larger measurement error in collecting prices.

Weights: While price aggregation into an index is not required whenever we draw on product-level prices, to construct the WPI, the MCO aggregates individual prices using a Laspeyres index and using value weights for a given base year. In contrast to the CPI where expenditure surveys can be used to construct weights, the WPI constructs weights from Macro to Micro: starting from weights at the broad sectoral and industry level drawing on national accounts data and then using the various individual data sources to construct item-level weights.

[18] The unweighted average is taken because of the difficulty of obtaining accurate weights at that level of aggregation.

A.4 Linking prices and NSS products

To link prices reported in the WPI micro data to NSS products, IO products and broad sectoral goods, we match product-by-product by hand at the most disaggregated product level first and then aggregate prices. For IO products that were not initially matched to NSS products we complete the matching by hand. With more than 500 products, the WPI micro data is in general slightly more disaggregated as the NSS product-level. This masks heterogeneity across sectors: For agricultural products such as vegetables and pulses, the WPI is less comprehensive than the NSS, while the WPI is much more comprehensive within manufacturing.

Many-to-one matching: In the case where an NSS product cannot be directly matched to a corresponding price in the WPI price series, we follow the following approach: We first check for close products that the WPI price series includes. For example, the NSS includes split gram, whole gram and gram products, while the WPI only includes gram. In these cases, we enforce the same price trend across all gram-related products, assuming that while their price level may be different, the products experienced the same price changes. Second, if no close products are available, but an over-category exists, we enforce the price index for the over-category. For example, the WPI price series does not include a number of vegetables that are NSS products, such as carrot and radish. However, the WPI price series builds a price index for the over-category "vegetables" using the specific vegetables and their weights that are included. We enforce the same price series for all other vegetables in the NSS that we cannot match directly, assuming that the other vegetables follow the same price movements as a "representative vegetable" in the WPI price series. Using the category-specific WPI price index enforces the WPI-specific weights that are a function of its base year. Lastly, if neither a close product nor an over-category exists, we do not match the price. We rarely encounter this case in the agriculture and manufacturing sectors, as only four products are without a match, nevertheless, WPI does not include any prices for services, thus we can not match any of the services.

Linking over time as base prices change: The WPI price series aggregates prices using Laspeyres price indices with weights given by base year values (not quantities), which are updated at least every 10 years. This introduces the issue of how to link WPI price series across vintages with different base years. For individual products, this linking is simple since it does not depend on any product-specific weights: we simply ensure that the last overlapping period has the same price, normalizing the price of the next vintage. (what if this is not available?) In the case of no overlap we assume no change in prices between the last period and the first period of the following vintage. For NSS products for which we draw on price indices, we only ensure that price indices align in the overlapping period, even if the underlying weights change, potentially introducing a time aggregation bias common to Laspeyres price indices. At last, in the case where new vintages allow a better product-level matching, we always move to the new matching, but normalize prices to ensure that prices align in the overlapping period.

Aggregating monthly prices: At last, the WPI price series are reported at the monthly level, while we only require them at the annual level. To aggregate, we take simple arithmetic averages between January and December as annual prices. This potentially introduces an aggregation bias as different products are consumed seasonally and prices in some months are more representative of annual prices than others. We ignore this issue, as this would require looking at monthly variation in consumption across NSS products or monthly variation in production across a variety of data sources from which the WPI price quotes are obtained. We only note that while the approach may introduce a bias in the level of the price, if this bias stays the same over time, then changes in prices are still observed unbiased.

A.5 NSS employment

The NSS employment data is based on the Employment and Unemployment module from the National Sample Survey (NSS). We use two waves of employment data, the first is conducted in 1987 and the second in 2011. In addition to the wage and type of activity, we select individual characteristics including education. To make the two waves comparable, we aggregate the education categories in both waves, for example: we aggregate all categories above secondary schooling as the cut-offs are not the same in both waves. Some of the respondents' IDs in the 1987 wave were not unique, we select among those who are repeated the ones that are workers and with the most information available. We transform the data from activity level to individual level, if an individual report multiple activities we aggregate their wages such that we have the total weekly wage for each respondent. Finally we only select wage workers excluding unpaid family labor and employers.

A.6 Linking NSS employment data IO products

The 1987 wave of the employment data links the principal activity of the respondents to the National Industrial Classification (NIC) codes from 1970. The 2011 wave includes NIC codes from 2008. Thus we match each wave separately – using the corresponding codes in the NIC – to IO products by hand. The NIC 1970 has a three digits level of classification apposed to the NIC 2008 that has 5 digits level of classification and is more exhaustive.

Many-to-one and one-to-many matching: In the case where multiple IO products link to one NIC code, we aggregate the IO products such that they represent one category. For example, we aggregate inorganic and organic chemicals on the IO side as they are grouped in the NIC 1970 under the same code. In the case of many-to-one matches, meaning multiple NIC codes matching one IO product, we do not consider this as a problem, multiple industries could be used to produce one product. For example, the IO product ”construction” is very broad and not specific to one job, thus it can match multiple industries in the NIC.

A.7 Linking NIC codes between different years

In this section of the Appendix we detail how we link NIC codes at the most disaggregated level across different revisions over time and how we link NIC codes to NSS products. We start out by describing a general and novel linkage algorithm that exploits Large Language Models (LLM) such as OpenAI's ChatGPT. The linkage algorithm allows to link across revisions of the same classification and potentially across classifications using one-to-many, many-to-one or many-to-many linking. The second part of the section describes in more detail how we apply the algorithm for linking NIC revisions over time in the Indian context.

A.7.1 OpenAI algorithm

Classification systems – for example the UN's ISIC for industries, the HS for products or the ISCO for occupations – are a key input in empirical work in Economics. A similarly common requirement in empirical work is to link classification systems, either across different classification systems (e.g. linking industries to products) or within a classification system across revisions. Unfortunately, standard classification systems undergo regular revisions over time to adapt to changes in the economy, and both the classification system and rounds of revision often differ across countries, leading to the coexistence of many different classification systems at any point in time. Thus, correctly linking different classifications is an important task in a lot of empirical work.

Existing empirical work usually uses a combination of official correspondences and the hard work of (many) research assistants to link classification systems, making the linking a costly and time-consuming exercise. We propose a new method that is fast and cheap exploiting the recent advances of Large Language Models (LLM) such as OpenAI's ChatGPT. LLMs such as ChatGPT are by now powerful enough to understand technical descriptions, summarise their meaning and even pick up on small nuances that distinguish among groups, classes, sub-classes and individual items within classification systems. The algorithm we propose allows for one-to-many, many-to-one and importantly also many-to-many linking.

We start out by outlining the key reasons for disparities in entries of classification systems that prevent a simple one-to-one linking in practice. We distinguish between disparities in revisions over time and disparities across different classification systems. To fix ideas, we take the example of linking industry classification revisions over time and the example for linking industries to products. Throughout, the setup is one where we have two datasets (classification 1 & classification 2), each featuring at least a textual description of the items on which one links. The first disparity – call this "true one-to-one" – is the case where an item shows up uniquely in both classifications, but the descriptions vary in wording. For example, "salt for common use" & "production of salt" for changes in descriptions after a revision. Next, we have "one-to-many" links, where an item in classification 1 maps to multiple items in classification 2. This can happen frequently when the classification splits up more items in further revisions, such as further disaggregating a growing IT industry or IT products. For example, "computers and computing machines for office use" may further get split up into two products: "computers for office use" and "computing machines for office use".[19] Similarly, we can have "many-to-one" links, which happens for example when revisions aggregate previous items, or when the linkage is done at different levels of aggregation. A further complication arises when items cannot be linked, either because there is no possible link available or because of the appearance and disappearance of items across revisions as would be common in detailed product classifications. At last, a final complication that may arise frequently are "many-to-many" links, which require to aggregate up items for both classifications.

To address all of these linkage issues, we suggest the following algorithm:

Step 1: We loop through the categories in the first classification. We create a new dataset that includes the first classification and empty columns to store the linking results. The aim of the loop is to select category i from the first classification, and then ask the LLM to find the j category that fits the most in the second classification. We do so by writing a prompt that indicates the task and the input to the OpenAI and we detail the output that we want. The wording for the prompt matters, an issue discussed in the literature as "prompt engineering", and we discuss our specific choices in the subsection below.

Step 2: Dealing with one to many and many to one links can be complex, for AI and humans. To solve this issue we run the loop described in step 1 twice. We first loop through the first classification and then we do the exact symmetric task by looping through the second classification. By running the loop twice we first allow for many to one links and then for one to many.

[19] Interestingly, in the Indian case, IT industry classifications became more aggregated over time, rather than more disaggregated.

Step 3: We merge the results from the two loops to get many to many links. We aggregate categories to form new groups where each category belongs to one and only one group.

Step 4: We create unique descriptions for the new formed groups by using OpenAI. We achieve this by running a loop that filters the descriptions of the categories that are in group i, we provide this information to OpenAI and ask to return a concise common description.

A.7.2 Application of OpenAI algorithm

Our analyses require a common industrial classification across time in India. Fortunately, the Indian *National Industrial Classification* (hereafter: NIC) is used consistently across different India-wide surveys and censuses at any point in time, including the surveys we are interested in: the Annual Survey of Industries (ASI) and Unorganized Manufacturing surveys for 1989 and 2005, respectively. However, the NIC codes are not consistent over time, since the industry codes and descriptions changed between revisions of NIC in 1987 and 2004. For our empirical analyses, we are interested in a common, harmonized classification at the most disaggregated level, which is at the 4-digit level of NIC revision 1987 and the 5-digit level of NIC revision 2004. That is, we are looking for a "many-to-many" linkage. To the best of our knowledge, this linkage does not exist to date. For example, the Ministry of Statistics and Programme Implementation (MoSPI), only provides official one-to-many concordances between successive years between the 3- and 4-digit levels.

This linkage is difficult in our context. One reason is that some products were further split up in 2004 compared to 1987. For example, in 1987 we have the following industry description "Manufacture of starch" that was split into "Manufacture of starch" and "Manufacture of other starch products" in 2004. We also encounter the symmetric case, where an industry was aggregated in 2004. For example, "Processing and blending of tea" and "Manufacture of instant tea" are aggregated into "Processing and blending of tea including manufacture of instant tea" in the NIC 2004 revision.

For the many-to-many linkage, we use the algorithm described above using ChatGPT. In the following, we describe in more detail the specific steps we took to implement the algorithm. Throughout, we directly work with the API of ChatGPT, which we access through R using an API key that we generate from our OpenAI account.[20]

[20] Details on setting up ChatGPT's API in R can be found [here](https://www.r-bloggers.com/2023/03/call-chatgpt-or-really-any-other-api-from-r/) (Link accessed on 16th August, 2024). Specifically, we create a function that calls OpenAI and takes as arguments a prompt and an authentication key. Multiple parameters can be added to this function to control the creativity (randomnesses) of the model. We recommend to use a low temperature (close to zero) which makes the model more deterministic and a seed for replication. We note that ChatGPT, as other LLMs, does not guarantee replicability even when using a seed.

In the first step, we needed to ensure that we had the right baseline datasets enumerating the industry codes and descriptions. While the ASI surveys provided official pdfs of the NIC codes and descriptions, we encountered duplicate codes with different descriptions and duplicate descriptions. We thus digitized the codes and descriptions and additionally drew on a version of the NIC codes provided by MoSPI to correct the mistakes by hand.

In the next step, we sought to facilitate the linkage exercise for ChatGPT by providing a pre-linked correspondence at a higher level of aggregation. The idea is that letting ChatGPT try to link a given description to all possible descriptions in the dataset requires too much input, raising usage costs and increasing the probability of an error. We thus constructed our own NIC concordance for 1987 and 2004 by hand at a higher level of aggregation (3 digit to 4 digit level). In practice, this means that for a given 4-digit NIC 1987 description i that ChatGPT has to link to any 5-digit NIC 2004 description, we pre-select only the 5-digit NIC 2004 descriptions that are in the 4-digit heading that maps to the corresponding 3-digit NIC 1987 industry of i. For our own higher level NIC concordance, we build on the concordance table made by Alcott et al. (2016) and extend it by hand to many-to-many links (this step could in principle also be done by ChatGPT). We restrict the concordance to only manufacturing products including repairs and the energy sector. Finally, we exclude textile industries from the concordance, as for these industries even a link at the 3 to 4 digit level is difficult by hand, letting ChatGPT handle textiles separately.

The next step is then the "many-to-one" linkage: For each industry description at the 4 digit level in 1987, we select the related 3-digit 1987 classification, we use the concordance table to find the corresponding 4-digit heading for NIC 2004. We then ask ChatGPT to find the best link among all the 5-digit NIC 2004 industries that are under the corresponding 4-digit heading. The exact prompt that we use is: "Task: Find the best match for a specific industry description from 1987 within a list of industry descriptions from 2004" and "Output: Provide the number of the row of the best matching description from 2004 and no more information". If the result returned from this prompt is NA. We ask the model to try again after making it sleep for 5 seconds. We provide the following prompt: "Please try again, I know you can do it. Task: Find the best match for a specific industry description from 1987 within a list of industry descriptions from 2004". This repeated questioning turned out to be a successful tweak, allowing for much better results. In the case where we initially did not receive a reply from ChatGPT, we flagged these cases as they could have meant that there does not exist a good match. Asking ChatGPT to only return the index of the best response rather than the full description greatly minimizes the length of output that ChatGPT provides, reducing the costs of the task.

Next, we use the same algorithm to provide a "one-to-many" linkage, now exchanging the input datasets such that we loop through each 5-digit description in the NIC 2004 classification to obtain the uniquely best match among the 4-digit NIC 1987 descriptions (after pre-selecting on the headings). Finally, we need to aggregate both linkages to a common many-to-many linkage. For this, we first collect all linkages in a single, long data frame. We then find aggregated groups that are unique for each 4-digit 1987 code and 5-digit 2004 code. To obtain our final unified classification, we use OpenAI one last time to produce homogeneous descriptions for the new groups. We use the exact following prompt: "Aggregate the following manufacturing industry descriptions into a concise summary (no more than 6 words)". By creating new codes and new descriptions we obtain our final homogenized concordance table.

B Empirical Results and Balance Tests

B.1 Balance test: State-Industry level Tariff measures

C Model

D Calibration

D.1 Estimated demand and supply side parameters

Table 14: Estimates of demand and supply parameters

Industry	ϵ_j	Ω_j	σ_j	α_j
Accurate digital clinical thermometer			4.145	1.258
Allopathic medicine	3.177	14.796	6.438	0.735
Assorted nuts: groundnut, cashew,	2.367	0.321	12.333	1.377
Atta and maida flour products	0.171	1.051	12.766	1.369
Beer, 1 litre bottle	14.016	3881.710	4.446	2.674
Books, journals, newspapers, periodicals	7.380	210.429	5.441	0.722
Bread and biscuits	3.167	1.810	6.094	0.776
Cheroot: small, cylindrical cigar	-0.984	0.002	8.867	2.451
Chocolate, sweets, cake, and drinks	3.441	2.243	6.971	0.953
Coffee cups for daily use	2.989	0.292	3.324	3.280
Coir and rope products	2.517	0.021	6.883	1.061
Coke, coal, coal gas, charcoal	1.310	0.175	7.693	0.522

Table 14 – continued from previous page

Industry	ϵ_j	Ω_j	σ_j	α_j
Cold bottled or canned beverage	10.775	217.200	4.937	1.045
Cooling devices: air conditioner, cooler	13.795	1122.811	3.939	2.290
Crystalized granulated sugar			5.613	3.209
Curry powder and assorted spices	0.427	0.226	3.329	
Dairy products: baby food, milk, butter	2.911	1.390	6.063	1.222
Dhoti, sari, various clothing fabrics	1.733	8.687	4.796	0.628
Dried fish products	0.170	0.023	8.764	1.841
Durable and versatile plastic products	3.940	0.090	5.935	0.984
Durable, corrosion-resistant metal alloy			6.070	0.649
Durable, glossy, protective coating			7.816	0.633
Earthen ware products	0.494	0.005	2.068	1.355
Efficient, high-capacity washing machine	11.693	20.692	7.366	0.728
Efficient, quiet, adjustable cooling fan	9.467	210.399	8.029	0.752
Efficient, spacious, modern cooling appliance	16.209	35602.467	5.645	1.320
Electric iron, hair drier, shaver	12.986	746.376	7.627	2.012
Electricity services and supply	3.085	3.222	4.252	0.639
Fire-starting wooden sticks in box	0.091	0.043	7.928	1.640
Firewood and wood chips	-0.341	0.366	4.830	1.418
Footwear: rubber, PVC, and other	1.338	0.336	6.057	1.124
Foreign or refined liquor (litre)	13.092	9839.458	3.326	1.295
Frozen water for cooling drinks			1.607	1.208
Fruit juice, pickles, sauce, jam	3.475	0.273	5.939	1.142
Fuel and lighting oils, candles	1.055	0.709	5.010	0.744
Ghee, Vanaspati, and margarine	3.712	10.031	11.414	1.567
Glassware and crockery products	2.960	0.056	2.051	0.827
Gold, silver, jewels, pearls, ornaments	9.481	530.014	4.883	0.846
Hand-rolled traditional Indian cigarettes	0.069	0.184	7.974	0.729
Khandsari: Traditional unrefined cane sugar			5.499	2.479
Knitted garments and knitting wool	3.678	1.093	9.737	0.988
L.P.G. (kg) fuel container	12.082	17220.633	3.074	1.229
Leaf-shaped non-stick frying pan	1.073	0.085	5.588	0.973
Leather boots, sandals, and footwear	3.753	5.533	6.110	0.945
Lighter and comb set	1.126	0.003	5.631	1.179
Lighting products: bulbs, tubelights, lantern	3.591	0.323	7.895	0.672
Lightweight, durable, corrosion-resistant metal			5.307	0.484

Table 14 – continued from previous page

Industry	ϵ_j	Ω_j	σ_j	α_j
Lock for secure fastening	3.559	0.044	5.567	0.799
Medical aids and orthopedic supports	6.502	1.359	4.458	1.162
Methylated spirit, 1 litre bottle			4.110	2.463
Mosquito net for insect protection	3.098	0.043	5.090	0.672
Motor car, jeep	13.545	735.827	1.332	32.619
Motorcycle, scooter options available	19.798	14456933.456	4.320	0.630
Musical instruments: piano, harmonium, more	9.604	3.259	8.221	0.537
Natural cane sugar (Gur)			8.020	1.756
Natural remedy using homeopathic principles			4.012	0.744
Pan finished lime katha	2.829	0.766	4.594	1.493
Pan ingredients and tobacco products	0.908	0.113	8.403	1.284
Personal care and shaving products	2.018	0.460	6.830	0.673
Precision shaving blades for smooth results			4.789	0.634
Radio and radiogram devices			1.853	0.788
Sea salt and other salts	-0.245	0.019	3.244	1.335
Sewai, noodles, salted refreshment,	2.418	1.506	8.697	0.866
Snuff: Finely ground smokeless tobacco	-0.229	0.002	4.840	2.075
Sports goods and toys	6.060	2.886	6.346	1.164
Stainless steel, bell metal, iron	3.726	1.344	5.789	0.753
Stationery articles, including fountain pen	2.759	1.005	5.487	0.948
Tea served in cups	1.431	0.897	3.655	2.646
Television with video capabilities	13.851	63826.239	4.585	1.004
Tender, flavorful, premium quality mutton	2.807	0.533	15.157	1.197
Textiles: shawls, bed sheets, rugs	2.726	1.547	5.771	0.804
Therapeutic appliances and equipment			5.633	0.961
Timekeeping devices: clock, watch	5.724	3.115	4.574	0.872
Tobacco-filled smoking sticks	9.840	1807.314	3.422	1.425
Toddy, traditional country liquor	1.587	0.618	4.931	1.353
Toilet soap, washing soap, soda, requis			6.795	0.638
Torch light and electric batteries	3.605	0.698	5.941	0.653
Traditional herbal medicinal treatments			5.471	0.986
Transport equipment including perambulator	6.371	1.342	5.612	0.703
Two-wheeled pedal-powered transportation device	3.488	2.233	8.270	0.621
Tyres and tubes for vehicles	4.081	0.580	5.092	0.556
Umbrella, rain coat, walking stick	6.200	3.212	6.565	1.127

Table 14 – continued from previous page

Industry	ϵ_j	Ω_j	σ_j	α_j
Variety of rice-based food products	-0.446	0.009	8.810	1.854
Various edible oils and oil seeds	0.967	1.986	11.421	0.897
Various essential household utensils			5.522	1.058
Various furniture items and fixtures	4.746	3.371	7.763	1.033
Various grain and pulse products	-1.227	0.081	6.182	1.085
Various types of lentils and pulses	0.968	1.540	14.948	2.275
Versatile, conductive, reddish-brown metal			6.652	0.756
Versatile, efficient, user-friendly sewing machine	6.301	2.134	8.679	1.007
Vintage audio and video players	11.554	516.688	4.099	0.679
Vintage mechanical writing device	1.141	0.000	231.584	1.392
Zarda, kimam, surt (gm)	0.708	0.008	7.958	3.229
Chicken, turkey and other birds	2.524	0.600		

Notes: Estimates using 1987 ASI and NSS data. ϵ_j: income elasticity; Ω_j: taste parameter; σ_j: Elasticity of substitution within industry j; α_j: pareto shape parameter (Estimated using top 25% of sample).

D.2 Details on replicating main regressions inside the model

In the following, we describe in detail how we replicate the main reduced-form regressions inside the model. Model-based regressions are produced for a given set of parameters Θ, which include also the initial conditions of the economy and which we assume as given here. Across the different regression specifications, the following steps are always the same. First, we solve the equilibrium path of the economy from the initial conditions in period 0 until the end period $T = 24$. We assume that period 0 is the year 1987 in the data, which means T maps to the year 2011. Second, we generate industry-specific data from $t = 0$ and the last year $t = T$. This is simulated firm-level data in the case of firm-level regressions and exact industry-level aggregates in the case of industry-level regressions. Finally, we replicate our regression estimates using model-generated data and the model counterpart to the variables of interest. We start discussing the implementation of the IV approach and then move to discussing data-consistent measurement for each relevant variable.

We start with the main regression on the demand side: the first stage of predicting actual demand with predicted demand (Table 2). For this, we build our instrument analogous to how we do so in the data: We sample a large set of households from the initial distribution in $t = 0$ and estimate their Engel curves (non-parametrically) as: $\omega_{ij0} = f_j(E_{i0}) + \epsilon_{ij0}$. We then

Table 13: Regression of Predicted Demand Growth on Tariff measures

Dependent Variable:	Predicted Demand Growth
Model:	(1)
Variables	
Avg Δ Input Tariff	7.337
	(4.858)
Avg Δ Output Tariff	-2.101**
	(1.013)
Min Δ Input Tariff	-0.5289
	(0.5752)
Max Δ Input Tariff	-0.2704
	(0.6922)
Min Δ Output Tariff	0.2705*
	(0.1372)
Max Δ Output Tariff	0.3243
	(0.3079)
Fixed-effects	
state_code_final	Yes
product	Yes
Fit statistics	
Observations	875
R^2	0.97677
Within R^2	0.00511

Clustered (product) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

use the estimated $\widehat{f}_j(\cdot)$ and the average change in E_{i0} (denoted by $\bar{y}^M$) to construct predicted aggregate industry-level demand for $t = T$. Since we abstract from state-level variation in the model, this gives $2xJ$ observations for which we regress log(predicted demand) on log(actual demand). In the case of many industries, we additionally add year and product fixed effects, as in the reduced form regressions. For all subsequent IV regressions, we save actual demand and predicted demand instruments.

We then proceed with the main industry-level IV regressions on the supply side. These are the IV effects of demand on value-added (Table 6), on employment (Table 7), on productivity and the number of plants (Table 8). For this, we also follow closely the approach we took in the data. For each of the two years $t \in \{0, T\}$, we draw a large number of firms N^S across industries using N_{jt} as industry-specific weights. We then construct industry-level value-added by summing up revenue across firms in a given industry and year (discuss whether we want to deflate by P_{jt}, sum or take average?). For employment, we focus on the wage bill effect reported in Table 7 (as this corrects for selection by worker quality) and use the total model-implied labor $H_{jt} = \sum_i h^*_{ijt}$. For productivity, we draw on the reported results for TFPR. To be in line with the data, we measure TFPR in our model using total revenue Rev_{jt} and total inputs (which is simply the employment H_{jt}) such that: $TFPR_{jt} = \text{Rev}_{jt} / \left(H_{jt}^{(\sigma_j - 1)/\sigma_j} \right)$. (Discuss deflating by P_{jt}). For the number of plants we simply take N_{jt}. Analogous to the demand side, this gives $2xJ$ observations on which we run our regression. In the case of many industries, we additionally add year and product fixed effects, as in the reduced form regressions.

(Add actual results)

(Add firm-level regressions)

E Model computation

In this part of the appendix, we provide further details on the computational algorithm to solve for an equilibrium growth path of our economy. We use an iterative fixed point algorithm that jointly determines both investments in productivity and firm dynamics across industries and time. Our model is computationally tractable, especially along the time dimension, allowing us to solve for arbitrary aggregate transition dynamics. Namely, even in a single-industry model where all firms' productivity grows at the same rate, a difficulty is that if future demand enters the value function, then a general solution requires guessing over the entire path of future demand (which varies off a balanced growth path). Our model sidesteps this issue in that the value function only depends on the current period's demand and (past) entry. This implies that we can start from an arbitrary initial equilibrium and simply solve iteratively, one period at a time. In the industry dimension, the traveling-wave result implies that one only needs to keep track of the average productivity level to describe the entire distribution of firms within an industry. Still, the algorithm requires to solve a fixed point over J objects per time period, which means computational costs do scale with the number of industries J and time periods.

The algorithm features two main loops: (1) an outer (time) loop that iterates over time t starting from t_0 until T, and (2) an inner (industry) loop that solves simultaneously for the set of J expenditures $\{E_{jt}\}_j^J$ at a given point in time t that satisfy the optimality conditions of the model for each industry.

We start by expressing all prices P_j relative to the wage w, which we denote as numéraire so that $w = 1$ for all periods. We also denote by $N_{j,t}^S$ the number of *surviving* firms in industry j at the beginning of period t, by $N_{j,t}^E$ the number of *entering* firms in industry j at the beginning of period t and by $N_{j,t}$ the number of producing firms in industry j in period t such that: $N_{j,t} = N_{j,t}^S + N_{j,t}^E$. In the algorithm below, we focus on the case of exogenous exit where the survival probability λ is fixed exogenously. The algorithm can be readily extended for endogenous exit.

Starting from $t = 1$ with initial conditions $\{\tilde{q}_{j,0}^{min}, N_{j,0}, N_{j,0}^E\}_j$:

Step 1: Solve for $\tilde{V}_{j,t}^P$ & $i_{j,t}^*$.

- Since $\tilde{V}_{j,t} = c^E \left(N_{j,t-1}^E\right)^{\phi_E}$ is given, solve for $\tilde{V}^P$ using:

$$\tilde{V}_{j,t} = (1+u)\tilde{V}_{j,t}^P + \frac{(1+u)^2 \left(\tilde{V}_{j,t}^P\right)^2}{2\bar{c}}$$

- Since $\tilde{V}^P > 0$, we can focus on the positive root, to find:

$$\tilde{V}^P = \left(\frac{\bar{c}}{1+u}\right)\left[\sqrt{1 + \frac{2\tilde{V}_{j,t}}{\bar{c}}} - 1\right]$$

- Which implies the following innovation rate:

$$i_{j,t}^* = \left[\sqrt{1 + \frac{2\tilde{V}_{j,t}}{\bar{c}}} - 1\right]$$

- And leads to an update of:

$$\tilde{q}_{j,t}^{min} = \tilde{q}_{j,t-1}^{min}(1+u)(1+i_{j,t}^*) \quad \Rightarrow \quad \mathbb{E}[\tilde{q}_{j,t}^+]$$

Step 2: Starting from a guess of $\{E_{j,t}\}_j$ and known $\mathbb{E}[\tilde{q}^+_{j,t}]$ solve for entry $N^E_{j,t}$:

- For each j: Solve for the (stable) root in $N^E_{j,t}$ of the nonlinear Bellman equation:

$$\tilde{V}^P_{j,t} = \underbrace{\frac{E_{j,t}}{\sigma\mathbb{E}[\tilde{q}^+_{j,t}]\left(N^S_{j,t} + N^E_{j,t}\right)}}_{\text{Profits: } \pi_{j,t}(N^E_{j,t})} + \frac{\lambda}{R} c^E \left(N^E_{j,t}\right)^{\phi_E} = \left(\frac{\bar{c}}{1+u}\right)\left[\sqrt{1 + \frac{2\tilde{V}_{j,t}}{\bar{c}}} - 1\right]$$

- Note: Without further restrictions on parameters, this equation can in principle have multiple positive roots in $N^E_{j,t}$ or none. For the problem to be well-behaved, we focus on the case of $\phi_E \geq 1$, which ensures that there are generally two positive roots, one stable root in the downward sloping part of $\tilde{V}^P_{j,t}$ and one in the exploding upward sloping part. For the downward sloping part ($\partial\tilde{V}^P_{j,t}/\partial N^E_{j,t} < 0$), the rent dissipation mechanism of entry exceeds the dynamic gain of higher entry for future profits. For the problem to be well-defined, this is the relevant economic case; ruling out the case where the dynamic gain takes over the static rent dissipation effect is the transversality condition of the problem.
- Back out: $N_{j,t} = N^S_{j,t} + N^E_{j,t}$.
- Compute industry-level prices (relative to wages):

$$P_{j,t} = \frac{\sigma}{\sigma - 1}\left(\mathbb{E}_{\tilde{q}}[\tilde{q}^+_{j,t}] \cdot N_{j,t}\right)^{\frac{1}{1-\sigma}}$$

Step 3: Update expenditure guess. Given $P_{j,t}$, $E_{j,t}$, and $\mathbb{E}[\tilde{q}_{j,t}]$, use the profit equation by industry to obtain industry-profits $\Pi_{j,t}$, which sum to aggregate profits Π_t. Use this to construct household income and solve for (updated) implied household expenditure.

- Specifically, to calculate updated industry profits, we need to calculate entry & exit, as well as production and total innovation costs. Profits are the sum of:
  - Production profits: $\left(N_{j,t}\mathbb{E}_{\tilde{q}}\left[\tilde{q}^+_{j,t}\right]\tilde{\pi}^*(P^*_{j,t}, E^*_{j,t})\right)$ given $(N_{j,t}, P^*_{j,t}, E^*_{j,t})$.
  - Exit costs: $-\left(N_{j,t}\mathbb{E}_{\tilde{q}}\left[\tilde{q}^+_{j,t}\right]\mathbb{E}_{j,t}[c^F_j|\text{stay}]\right)$ given $N_{j,t-1}$ from the previous period and $(\lambda(j,t), \mathbb{E}_{j,t}[c^F_j|\text{stay}])$ given $\tilde{V}(j, t+1)$. That is, the survival probability $\lambda(j,t)$ is given by the Gumbel cdf.

$$\lambda(j,t) = \exp(-\exp\left(-\frac{\theta_{jt} - \mu_x}{\sigma_x}\right)) \quad \text{where} \quad \theta_{jt} = \tilde{V}_{jt+1}$$

  This part is absent when exit is exogenous.

- Innovation costs: $-\left(N_{j,t}\mathbb{E}_{\tilde{q}}\left[\tilde{q}^{+}_{j,t}\right]\frac{\bar{c}_{j,t}}{2}\left(i^{*}_{j,t}\right)^{2}\right)$
- Entry costs: $-\left(N^{E}_{j,t}c^{E}_{j,t}\right)$ given $N^{E}_{j,t}$.

- Compute industry-level expenditures $E_{j,t}$ as a function of prices $\{P_{j,t}\}_j$ and the household distribution over income $I_{i,t}$ with income given by: $I_{i,t} = h_{it} + \Pi_t/L_t$. Note that the sum of guessed profits across industries equals aggregate profits. Specifically, we start with consumption of product j by household i: $C_{i,j,t} = \Omega_j \left(\frac{P_{j,t}}{I_{i,t}}\right)^{-\bar{\sigma}} C_{i,t}^{\epsilon_j(1-\bar{\sigma})}$ where $I_i = \sum_j P_j C_{ij}$. Let us redefine $\tilde{C}(I_i) \equiv C_i^{1-\bar{\sigma}}$ such that (omitting subscript t)

$$\sum_j \Omega_j P_j^{1-\bar{\sigma}} \tilde{C}(I_i)^{\epsilon}_j = I_i^{1-\bar{\sigma}}$$
$$\sum_j \Omega_j \tilde{P}_j \tilde{C}^{\epsilon_j} = I_i^{1-\bar{\sigma}}$$

This allows to solve for $\tilde{C}$ using a nonlinear solver (e.g. bijection or Newtonian solver). Once one obtains $\tilde{C}(I)$, the industry j expenditure writes as

$$E_{i,j,t} \equiv P_{j,t}C_{i,j,t} = \Omega_j \tilde{P}_{j,t} I_{i,t}^{\bar{\sigma}} \tilde{C}_{i,j,t}(I)^{\epsilon_j}$$

Aggregate expenditure $E_{j,t}$ is then given by

$$E_{j,t} = \int e_{i,j,t}\, d\mathrm{H}_t = \int \tilde{P}_{j,t} I_t^{\bar{\sigma}} \tilde{C}(I)^{\epsilon}_j \partial \mathrm{H}_t$$

Assuming a known distribution for human capital, one can approximate this integral using standard quadrature/discretization methods.

Step 4. **Check for convergence** Check that $\{E^{\text{update}}_{j,t}\}_j = \{E_{j,t}\}_j$ (up to convergence criteria). If not: Update guesses and return to Step 0. Otherwise, continue.

Step 5: If $t = T$ stop. Otherwise, set $t = t + 1$. Form new guess of $\{E_{j,t}\}_j$ based on previous equilibrium values and return to Step 0.

F Results