How Steep is the Phillips Curve in Developing Economies?

A Sufficient Statistics Approach and Estimates for India

Juan Herreño
UC San Diego

Noémie Pinardon-Touati
Columbia University

Malte Thie
Paris Dauphine

First Version: July 28, 2025

Updated Frequently

Abstract

We estimate the Phillips curve for India to shed light on the output-inflation trade-off in developing economies. We develop a method to estimate the slope of the Phillips curve based on sufficient statistics that apply to a broad class of New Keynesian models. Using portable causal research designs, we estimate the firm-level passthrough of cost shocks into prices at different horizons, and the slope of marginal costs curves at different levels of aggregation. These empirical moments map into the slope of the Phillips curve and yield a decomposition into three terms: price rigidity, micro real rigidities, and macro real rigidities. The slope of the Phillips curve in India is one order of magnitude steeper than in the United States. This difference is explained by weaker macro real rigidities and less rigid prices. Extending the model to allow for input misallocation, we find that the re-allocative effects of monetary policy affect the Phillips curve, but this effect is small.

JEL Classification: E31

We thank conference participants at the Barcelona Summer Forum and the SED Copenhagen meetings, as well as seminar participants at Columbia University, UC San Diego, the Minneapolis Fed, the San Francisco Fed, University of Rochester, Central Bank of Chile, University of Chile, and PUC Chile for valuable comments. We thank Fauzi Estiko for excellent research assistance.

1. Introduction

What role do Keynesian economics play in developing economies? In particular, how do aggregate demand expansions translate into inflationary pressures or increases in real output? At the heart of this inquiry lies the shape of the Phillips curve. Let us consider the New Keynesian version of the Phillips curve (NKPC), now the textbook formulation:

$$\pi_t = \kappa_y \tilde{y}_t + \beta \mathbb{E}_t[\pi_{t+1}] + u_t \tag{1}$$

where π_t is inflation, $\tilde{y}_t$ is the output gap (the difference between real output and its natural level), and u_t is an exogenous cost-push shock. The slope of the Phillips curve κ_y characterizes the sensitivity of inflation to the output gap (i.e., to an increase in demand).

A large body of empirical evidence in advanced economies shows strong evidence for Keynesian mechanisms—the Phillips curve is relatively flat (e.g., Hazell et al. 2022), and monetary shocks have large effects on real output (see Bauer and Swanson 2023 for a review). However, developing economies differ in fundamental ways. The macroeconomic environment is characterized by high and volatile inflation, which may affect the sluggishness of the price level to changes in demand. Production may be less scalable, hindering real output expansions. Finally, developing economies are subject to large allocative distortions, which may interact with the inefficient price dispersion caused by inflation.

In this paper, we quantify the effects of domestic demand expansions on inflation in India, a large developing economy. We develop an estimation approach for the slope of the Phillips curve based on sufficient statistics that applies to the broad class of New Keynesian models that yield the NKPC formulation in (1). Using portable causal research designs, we estimate the firm-level passthrough of cost shocks into prices at different horizons, and the slope of marginal costs curves at different levels of aggregation. We show that these empirical moments suffice to recover the slope of the Phillips curve. We then move beyond this class of models to examine how input misallocation—a defining feature of developing economies—alters the transmission of demand expansions to inflation. In the presence of misallocation, demand shocks affect inflation not only through the output gap but also through potential changes in allocative efficiency. Using a similar sufficient statistics approach, we quantify this additional term in the Phillips curve.

We find that the Phillips curve is steep, with a slope roughly 8 times larger than estimates for the United States. This is due to a combination of less sticky prices and steeper marginal cost curves. Changes in allocative efficiency play only a small role.

Our approach offers three key advantages. First, it circumvents the identification challenges that are pervasive in time series analysis. Second, by focusing on a small set of sufficient statistics—rather than fully parameterizing the demand system or the supply side—it is more robust to model misspecification. In addition, unlike indirect inference methods based on impulse responses of aggregate variables to monetary shocks, our approach requires no assumptions about any part of the model beyond the Phillips curve. Third, it enables a transparent decomposition of the inflation response into distinct and interpretable mechanisms suitable for comparisons of the determinants of inflation across countries and time periods.

We exploit the Indian Annual Survey of Industries (ASI), a large-scale representative survey of formal manufacturing establishments, over the period 1998-2017. This data records firm×product-level output prices and quantities across 1,200 highly disaggregated products. Similarly, on the input side, we obtain input cost and quantity purchased for each disaggregated input. This data allows us to re-construct Indian PPI inflation from the bottom up. Section 2 presents motivating facts in aggregate data and details the data used in our analysis.

We estimate the response of inflation to a change in the output gap, κ_y. Our empirical approach is valid in *any* model of the New Keynesian class that delivers the standard formulation of the Phillips curve in (1). We present the model environment in Section 3. We exploit the fact that in this class of models the slope of the Phillips curve is the product of two sufficient statistics: $\kappa_y = \kappa_{mc} \times \Omega$, where Ω is the sensitivity of real marginal costs to the output gap, and κ_{mc} is the elasticity of inflation to real marginal costs. We estimate these two sufficient statistics in turn.

We first estimate the slope of the marginal cost-based Phillips curve κ_{mc}, which reflects price rigidity and micro-level real rigidities (strategic complementarities in price-setting and/or upward-sloping firm-level marginal costs). Section 4 presents our estimation. We show that this slope can be identified from the pass-through of input cost shocks to product prices. The intuition is that firms pass through idiosyncratic cost shocks at the same rate as aggregate marginal cost shocks. A simple OLS regression of output prices on input costs would yield a biased estimate if unobserved demand shocks cause a firm to increase its price and lead to higher input costs. To circumvent this concern, we instrument the realized change in input costs by a shift-share instrument exploiting variation in firms' pre-determined exposure to intermediate inputs with different price dynamics, similar to Amiti, Itskhoki, and Konings (2019). We find a pass-through equal to 0.22 at the annual frequency, for cost shocks that are well-approximated by an AR(1) with persistence equal to 0.75. Because of our inclusion of product×time fixed effects, these estimates effectively hold competitors' prices constant.

To obtain the slope of the marginal cost-based Phillips curve κ_{mc}, the firm-level pass-through needs to be rescaled by an estimate of the frequency of price change, reflecting the fixed point adjustment from partial to general equilibrium. We estimate the frequency of price change in two ways. First, we measure it directly in the data. Second, we use pass-throughs at longer horizon. Both methods yield a frequency of price changes equal to 0.9 at the annual frequency. This is equivalent to 0.45 at the quarterly frequency, higher than the number reported by Nakamura and Steinsson (2008) for the United States in the recent period, but comparable to estimates for countries and time periods with similar inflation rates as India in our period of study (Gagnon 2009; Nakamura et al. 2018; Alvarez et al. 2019). From this analysis, we obtain a slope $\kappa_{mc} = 0.095$ at the quarterly frequency.

The second key object to estimate is Ω, the elasticity of aggregate real marginal cost to output. The firm-level elasticity of marginal cost to output reflects returns to scale and firm-specific input supply curves, if any. At the aggregate level, Ω additionally accounts for the increase in factor prices that respond to aggregate (but not firm-level) shocks.

We first estimate the firm-level elasticity of marginal costs to changes in quantities—which is distinct from Ω but will prove useful to identify mechanisms. An obvious identification threat is the presence of firm-specific supply shocks that increase marginal costs and reduce quantities. We introduce a demand shifter that exploits product-specific shifts in demand, and heterogeneity in firm exposure to different products. We find that at the firm level, a one percent increase in quantities causes an increase in marginal costs of 0.2 percent, consistent with a returns to scale parameter of 0.86 in the short run.

To estimate Ω, we then use the same design across industries and regions to identify the elasticity of the marginal cost curve at those levels of aggregation. If input markets clear at these levels of aggregation, then we obtain an estimate of the true Ω for the aggregate economy; otherwise this procedure yields a lower bound. Our estimates imply that after a one percent increase in industry or regional output, marginal costs at the corresponding level of aggregation rise by 0.6-0.7 percent.

Putting these estimates together, we find that the slope of the Phillips curve in India is 0.066 at the quarterly frequency, 8 times larger than in developed countries (Hazell et al. 2022; Gagliardone et al. 2023). Our approach allows us to decompose the magnitude of the slope into its components. We find that the slope of the marginal cost-based Phillips curve $\kappa_{mc} = 0.095$ is roughly twice larger than estimates for developed countries (Mavroeidis, Plagborg-Møller, and Stock 2014; Gagliardone et al. 2023). This is fully driven by less rigid prices: if India had the frequency of price changes observed in developed countries, our estimate for κ_{mc} would be similar. Micro-level real rigidities matter: they reduce the slope of the Phillips curve by a factor of 4, but both the degree of strategic complementarities and the slope of firm-level marginal cost curves have similar magnitudes as what has been documented in developed countries. Finally, the aggregate marginal cost curve is significantly steeper in India and accounts for most of the difference: our estimate for Ω is at least three times larger than similar estimates for developed countries (e.g., Boehm and Pandalai-Nayar 2022).

In the broad class of New Keynesian models that yield the NKPC formulation in (1), κ_y fully characterizes the response of inflation to demand shocks. In Section 5, we deviate from this benchmark to investigate how input misallocation —a key feature of developing economies—affects the transmission of demand shocks to inflation. To characterize how input misallocation affects the Phillips curve, we extend our baseline model to allow for steady-state input wedges, and assume that the demand system is Kimball.

Demand expansions have a first-order effect on allocative efficiency, via a reallocation of production to more or less distorted firms. The first of these reallocative effects happens when firms face different demand elasticities, and hence charge different markups and have different pass-throughs. This channel is the same as in Baqaee, Farhi, and Sangani (2024), which we generalize to non-constant returns to scale. The second channel is new. Heterogeneous markups and input wedges generate a distribution of ex-ante combined distortions, and firms with different distortions have different pass-throughs. Demand expansions improve (worsen) allocative efficiency if they reallocate quantities to firms with ex-ante larger (smaller) combined distortions. The allocative effects of demand shocks on inflation enter the Phillips curve as an endogenous cost-push shock: positive (negative) allocative efficiency effects work to flatten (steepen) the Phillips curve.

We fully characterize the response of allocative efficiency to demand shocks as a function of a small number of sufficient statistics that can be identified in our firm-level data. The allocative effects of demand shocks depend on the average and differential effect of input cost shocks on prices and quantities across firms with different ex-ante combined distortions.[1] We find that the allocative effects of changes in aggregate demand are quantitatively small: a 1% increase in the output gap reduces allocative efficiency by 0.01%. The data moments that inform this small elasticity are the compressed price and quantity pass-throughs for firms with different demand elasticities and ex-ante combined distortions. Hence, quantitatively, assuming that the aggregate supply curve is stable in response to a temporary aggregate demand expansion is a good assumption for India, even when the economy suffers from substantial steady-state inefficiencies.

Finally, we shows that inflation in India is well-explained by domestic demand factors. Specifically, in Section 6, we construct a measure of the output gap by obtaining the cyclical component of industrial production using the Hamilton (2018) filter. Using this measure of the output gap and our estimated slope of the Phillips curve, we obtain a time series for predicted inflation. For the vast majority of episodes, our measure of demand-driven inflation tracks realized inflation. The exception are two episodes in which narrative records suggest the Phillips curve shifted due to a combination of supply shocks and shifts in long-run inflation expectations driven by changes in the conduct of monetary policy.

Related literature. First, this work most closely relates to the large body of work on the slope of the Phillips curve. The overwhelming majority of contributions study the United States or other OECD countries. Most papers rely on time series methods, despite ubiquitous identification and statistical issues (Mavroeidis, Plagborg-Møller, and Stock 2014). A relatively recent series of papers exploits cross-sectional variation: Beraja, Hurst, and Ospina (2019); Hazell et al. (2022); Cerrato and Gitti (2022); Gagliardone et al. (2023). Closest to the approach implemented here, Gagliardone et al. (2023) also exploit micro-level data on prices and quantities from Belgium to estimate the slope of the Phillips curve. Compared to their work, we introduce two novel contributions. First, the analysis in Gagliardone et al. (2023) focuses on identifying the slope of the marginal cost-based

[1] Importantly, these effects cannot be, in general, summarized by differential price pass-through estimates across the size distribution, since size is not a sufficient statistic for ex-ante wedges in our setting.

Phillips curve. We develop causal research designs to identify the slope of marginal cost curves in the cross-section of firms, industries, and local labor markets, to recover the slope of the output-based Phillips curve. Second, we characterize and quantify how the slope of the Phillips curve is modified when distortions interact with pricing frictions.

We also complement the few studies that study the slope of the Phillips curve in the context of emerging and developing economies, all using time series methods (Mohanty, Klau et al. 2001; Filardo and Lombardi 2014; Ball, Chari, and Mishra 2016).

Second, we relate to a recent literature studying the effects of monetary shocks on allocative efficiency (Reinelt and Meier 2020; Mongey 2021; Baqaee, Farhi, and Sangani 2024). Our framework builds on Baqaee, Farhi, and Sangani (2024), which we augment to account for potentially non-constant returns to scale and the presence of input wedges. These new ingredients qualitatively affect predictions regarding the effect of monetary shocks for allocative efficiency. We show non-parametric identification of the sufficient statistics that determine the allocative efficiency effect. Finally, we provide a quantitative analysis of this channel in a large developing country—where this type of effects are likely to be strongest.

Third, we complement the literature on pricing decisions in emerging and developing countries (Gagnon 2009; Alvarez et al. 2019; Drenik and Perez 2020). While existing studies focus on hyperinflation episodes, we study a setting with moderately high and persistent inflation, which is representative of "normal times" in these countries. Second, we focus on estimating the effect of domestic demand expansions on inflation.

Finally, our work contributes to the literature on the pass-through of cost shocks into prices (Gopinath and Rigobon 2008; Gopinath and Itskhoki 2011a; Amiti, Itskhoki, and Konings 2019) and the literature on micro- and macro-level cost curves (Shea 1993; Bresnahan and Ramey 1994; Boehm and Pandalai-Nayar 2022). We use these estimates as the building blocks to our estimation of the slope of the Phillips curve.

2. Motivating facts and data

2.1. Inflation in developing countries and the case of India

In this subsection we document a simple fact: average inflation rates decline along the development path. Figure 1 displays a binned scatterplot of the relationship between GDP per capita (constant 2005 dollars, in logarithms) and CPI inflation across countries. For each country, we use the median values of each variable over 1980-2023.[2] We observe a strong negative association between the two variables. Low income countries have, on average, higher inflation rates. The green diamond marks India and shows that India has levels of inflation representative of countries with this level of development.

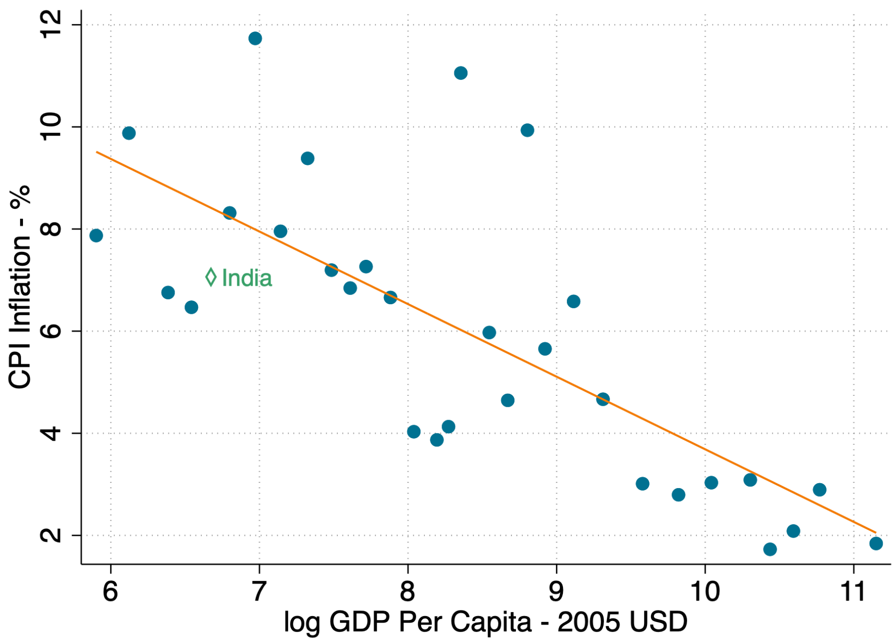


FIGURE 1. Inflation in the cross-section of countries

Note: The Figure shows the association between the median log GDP per capita in a country and its median CPI inflation rate from 1980-2023. Source: World Development Indicators.

2.2. The Phillips correlation in India and identification concerns in aggregate data

Figure 2 shows what Stock and Watson (2020) call the *Phillips correlation* which is an association between inflation and quantities in equilibrium. Note that this differs from the structural interpretation of the Phillips curve which plays the role of the aggregate supply schedule. We use quarterly data from 1996Q4 to 2020Q1. Our measure of inflation is the year-over-year percent change in the manufacturing Wholesale Price Index, and for quantities we use the cyclical component of the manufacturing quantity index that we recover using the Hamilton (2018) filter. Each dot corresponds to a date.

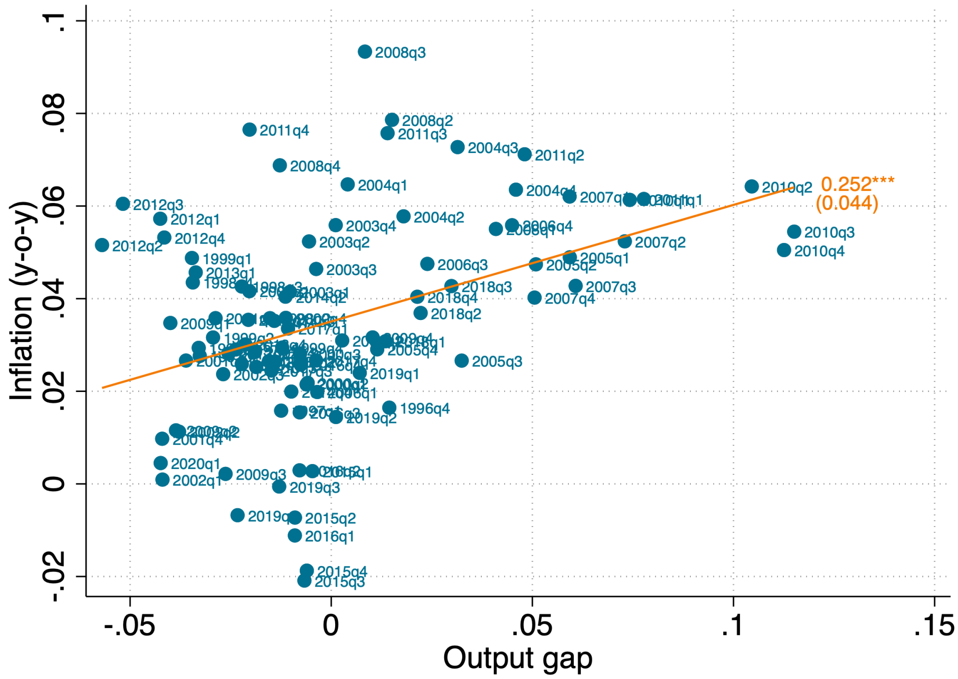
[2]We use the median to filter the effect of influential observations driven by periods of hyperinflations.

Panel (a) shows that on average, there is a positive relationship between inflation and the output gap. It is useful to inspect two episodes of recent economic history in India to understand the severity of the problem with using the time series relationship to uncover the slope of the Phillips curve. Panel (b) shows the same data, but highlights these historical episodes. In orange, we highlight the period after the Great Financial Crisis, from 2010 to 2012, which was characterized by a strong deceleration of the economy together with a high and stable inflation rate of around 7%. Narrative accounts of this episode mention supply headwinds in the agricultural sector, and an anchoring of inflation expectations at high levels. Interestingly, this parallels the discussion of a "missing disinflation" in the United States at the same time. The second period we highlight, in green, is the plummeting of the inflation rate around 2014 after the newly-appointed governor of the Reserve Bank of India Raghuram Rajan announced the adoption of inflation targeting, with potential effects for long-run inflation expectations, as well as the end of adverse supply shocks. This vertical shift in the Phillips correlation reminisces the United States during the early 1980s.

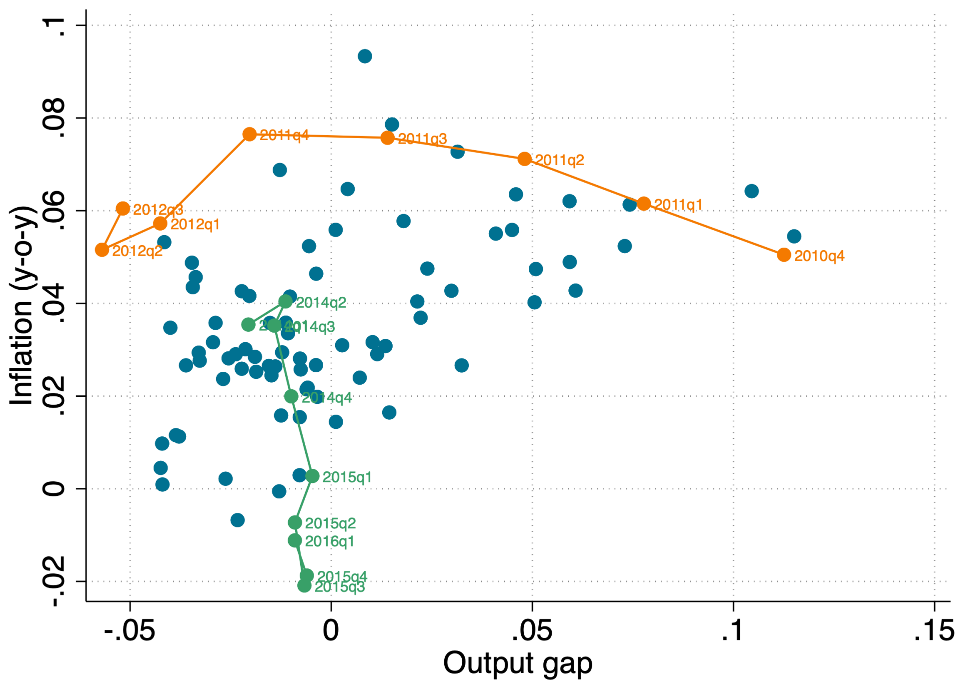
Of course, it is a possibility that the Phillips correlation is reflecting a time-varying slope of the Phillips curve that changed from being almost perfectly flat to almost perfectly vertical in a matter of two years. A more plausible possibility, however, is that the Phillips correlation is reflecting the occurrence of shifters to the Phillips curve, let them be triggered by changes in monetary policy, financial shocks, or changes in relative prices. We develop a method for estimating the structural Phillips curve that is immune to these concerns.

FIGURE 2. Phillips correlation in India

(a) Phillips correlation



(b) Post-GFC stagflation and Rajan disinflation



Note: The x-axis is the transitory component of manufacturing output using the Hamilton (2018) index. The y-axis corresponds for the y-o-y inflation rate for the WPI manufacturing index in India. The solid lines connect the points for two subperiods.

2.3. Micro-data data on prices and quantities: the Annual Survey of Industries

Our main data source is the Indian Annual Survey of Industries (ASI). We use data from 1998 to 2017. The ASI contains information on a representative sample of manufacturing establishments, conditional on them taking part in the organized sector, and either employing more than 20 employees, or employing more than 10 employees and using electricity. The sampling scheme is summarized in Table C.1. In terms of coverage, the value added of establishments in ASI covers 61% of total manufacturing value added in India as reported in the national accounts (the latter includes small establishments that the ASI excludes and the informal sector).

While we do not have firm identifiers and hence cannot aggregate plants under common ownership, less than 7.5% of all plants are part of a multiplant firm with sister plants that file separate survey returns. With that caveat in mind, we call the units of observation in our data "firms."

Variable definitions. The main firm-level variables we use are revenues $\mathcal{R}_{it}$, labor costs $\mathcal{C}^l_{it}$, intermediate input costs $\mathcal{C}^x_{it}$, and capital K_{it}. We construct firm-level revenues as the gross sales value of products sold. Labor costs is the sum of wages, salaries, bonuses and supplemental labor costs. Labor costs $\mathcal{C}^l_{it}$ divided by number of days worked L_{it} yields the daily wage w^l_{it}. The capital stock is the sum of fixed assets.

The official sectoral classification (NIC) changed in 1998, 2004 and 2008. We use official NIC concordances to construct a harmonized classification. We obtain 81 consistently-defined manufacturing industries. Our definition of industries mostly follows 3-digit industries in the NIC 1998 classification but splits some highly populated industries and aggregates others.

Product and input prices and quantities. The main advantage of the ASI is that both for the products that manufacturing plants produce and the inputs they buy, we observe information on sales, quantities, and unit values, at the product level. To exploit this data, we construct a harmonized classification of products and inputs over the whole sample, allowing us to classify all products into around 1,200 product codes (approximately corresponding to the 5-digit level of the Indian NPC classification). Appendix C.1 details the steps of the construction of the harmonized classification. Table C.2 shows an excerpt of the product classification.

We construct a panel of firm-product prices and quantities. We denote by $\Delta \log p_{ijt}$ and $\Delta \log y_{ijt}$ are the change in the log of price p and log of quantity y of product j sold by firm i at time t, respectively. Even working with narrowly-defined product categories, unobserved heterogeneity could prevent a meaningful comparison of prices across firms. We always work with within firm×product changes in the price (or quantity), largely alleviating this concern. We detail additional cleaning steps in Appendix C.1. The set of products for which we observe a valid price (quantity) change, which we denote $\mathcal{J}_i$, account for, on average, 75% of firm-level total sales. We define the firm-level price index as the Törnqvist-weighted change in the observed firm×product-level price changes: $\underline{\Delta} \log p_{it} = \sum_{j \in \mathcal{J}_i} \bar{s}_{ijt} \Delta \log p_{ijt}$. We use the convention of placing a bar on top of the share to denote that these shares are the mid-point of the shares in $t-1$ and t, and the bar under the Δ sign indicates that we take the average price change over the set of *observed* products. We construct the firm-level change in quantities as $\Delta \log y_{it} = \Delta \log \mathcal{R}_{it} - \underline{\Delta} \log p_{it}$.[3]

As a data validation exercise, we compare the inflation series implied by the firm×product-level price changes in ASI to the aggregate producer price index (WPI). The result of this exercise is presented in Figure Figure C.1 . The two series move very closely. We note a discrepancy in 2004 and 2005; for this reason, we present robustness checks of all our results excluding these two years.

Similarly, we observe purchase value, unit price, and quantity purchased for intermediate inputs (materials classified in the same 1,200 products, and energy disaggregated into electricity, oil, and coal). We denote by $\Delta \log w_{ikt}$ and $\Delta \log x_{ikt}$ the log change in prices and quantities of input k used by firm i. We perform the same cleaning steps as described for prices. The inputs for which we observe a valid price (quantity) change, which we denote $\mathcal{K}_i$, account for on average 57% of firm-level total input purchases. We define the firm-level input quantity change as the Törnqvist-weighted change in the observed firm×input-level quantity changes: $\underline{\Delta} \log x_{it} = \sum_{k \in \mathcal{K}_i} \bar{s}_{ikt} \Delta \log x_{ijt}$. We construct the firm-level intermediate input price index as $\Delta \log w^x_{it} = \Delta \log \mathcal{C}^x_{it} - \underline{\Delta} \log x_{it}$.[4]

[3]Because we do not observe the price and quantity change for all total sales, in general $\underline{\Delta} \log p_{it} + \underline{\Delta} \log y_{it} \neq \Delta \log \mathcal{R}_{it}$. We assume that the price change of observed products is on average equal to the price change for all products.

[4]The assumption is that $\underline{\Delta} \log x_{it}$ (the average increase in input quantity for the inputs $\underline{\mathcal{K}}_i$ for which we observe input-level data) is equal to the average change in input quantity for all inputs $\Delta \log x_{it}$. This assumption is the most natural when different material inputs are strong complements (and it is exactly

Sample of analysis. We restrict the sample of analysis to firm$\times$year observations which reported a positive output. We drop firms that report no days worked and no employees throughout their existence. Moreover, we restrict the sample to observations that displayed consistent accounting values, i.e. for which individual input and output components closely summed to their reported aggregate values. Finally, we require that firms report disaggregated sales (purchase) values for at least one product (input). Official sampling weights are used in all of our calculations. Our final sample is an unbalanced panel of 193,352 firms.

3. Model

This section describes the theoretical framework that guides our empirical investigation. The main text provides a succinct description and we leave all details and proofs to Appendix A. The simplified environment we present now is amenable to several extensions and interesting special cases which we relegate to Appendix A as well.

3.1. Environment

The economy is composed of four types of agents. Households consume the final good, save, and supply labor. A final good producer produces the final good using differentiated varieties indexed by $i \in [0,1]$. Producers of each differentiated variety i produce using labor and have sticky prices. A central bank implements monetary policy.

Households. Households choose consumption C and labor L to maximize discounted future utility $\mathbb{E}_0\left[\sum_{t=0}^{+\infty}\beta^t u(C_t, L_t)\right]$ subject to a per-period budget constraint $P_t^Y C_t + Q_t B_t = B_{t-1} + w_t^l L_t + T_t$ where P_t^Y is the price of the consumption bundle, B_t is holdings of one-period risk-free nominal bonds with price Q_t, w_t^l is the wage, and T_t denotes any profits rebated to households as lump-sum as well as any lump-sum taxes paid by the household.

Final good producers. Let Y_t denote aggregate production of the final good. Y_t is used for consumption C_t, so that $Y_t = C_t$.

The final good Y_t is produced by a perfectly competitive firm using a bundle of differentiated intermediate inputs y_{it} for $i \in [0,1]$. We consider an arbitrary, invertible demand

true if production is Leontief).

system that gives rise to a demand curve of the form:

$$y_{it} = \mathcal{D}(p_{it}/\mathcal{P}_t)Y_t. \tag{2}$$

This demand structure nests the popular CES and Kimball aggregators. With the addition of a layer for sectors, it can also accommodate oligopolistic competition a la Atkeson and Burstein (2008). We make this mapping explicit in Appendix A.2.

The price elasticity of demand is given by:

$$\theta_{it} = \theta\left(\frac{y_{it}}{Y_t}\right) = -\frac{\partial \log y_{it}}{\partial \log p_{it}}. \tag{3}$$

Differentiated varieties producers. Each variety i is produced by a single firm. Firms produce with labor as their only input using a production function with potentially decreasing returns to scale

$$y_{it} = e^{z_{it}} l_{it}^{a}. \tag{4}$$

The total nominal cost function is denoted by $\mathcal{C}(y_{it}, w_{it}, z_{it}) = w_{it}\left(\frac{y_{it}}{e^{z_{it}}}\right)^{\frac{1}{a}}$, where w_{it} is the input price, z_{it} is firm-level productivity, and a is the degree of returns to scale. We allow for input prices to depend on firm-level input demand in the spirit of Woodford (2003) and assume an isoelastic specification $w_{it}^{v} = w_{t}^{v}\left(\frac{l_{it}}{L_t}\right)^{a_w}$. The marginal cost function $mc(y_{it}, w_{it}, z_{it}) = \frac{d\mathcal{C}}{dy_{it}}$ is the total derivative of the cost function with respect to firm-level quantities, which allows for the potential dependence of input prices on firm production.

A firm has a probability $1-\alpha$ of being able to reset its price in each period. A firm that can reset its price chooses the price that maximizes:

$$\max_{p_{it|t}} \mathbb{E}_t\left[\sum_{s=0}^{+\infty} \alpha^s \Lambda_{t,t+s}\left[p_{it|t} y_{it+s|t} - \mathcal{C}(y_{it+s|t}, w_{it+s}, z_{it+s})\right]\right]$$

subject to the demand curve $y_{it+s|t} = \mathcal{D}(p_{it|t}/\mathcal{P}_{t+s})Y_{t+s}$ and the cost function. $\Lambda_{t,t+s}$ is the stochastic discount factor of the representative household.

It will be convenient to define the following objects. $\mu_{it}^{f} = \frac{\theta_{it}}{\theta_{it}-1}$ is the desired markup that the firm would choose in a flexible price environment. $\Gamma_{it} = \dfrac{\partial \log \mu_{it}^{f}}{\partial \log \frac{y_{it}}{Y_t}}$ is the elasticity of the flexible price markup with respect to relative size. $\rho_{it} = \frac{1}{1+\Gamma_{it}\theta_{it}}$ is the partial equilibrium pass-through of a marginal cost shock into the firm's price in a flexible price environment.[5] In the CES monopolistic competition case, demand elasticities are primitives, $\Gamma_{it} = 0 \; \forall i$, and $\rho_{it} = 1 \; \forall i$. Away from this particular case, ρ_{it} can be below or above 1 depending on the sign of Γ_{it}. Note that μ^f_{it}, Γ_{it}, and ρ_{it} are only a function of a firm's relative size $\frac{y_{it}}{Y_t}$. Finally, in a sticky price environment, the actual markup of the firm may differ from the desired markup under flexible prices. We denote the actual markup of the firm: $\mu_{it} = \frac{p_{it}}{mc_{it}}$.

Monetary authority. Monetary policy sets the nominal interest rate according to a Taylor rule.

We solve the model by log-linearization around the zero-inflation symmetric (across firms) steady state. We take a first-order expansion for small monetary policy shocks. Quantities without a t subscript refer to the steady-state.

3.2. Characterization

Marginal cost-based Phillips curve. A firm that can reset its price at time t will choose:

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\zeta\rho\hat{mc}_{t+s} + (1-\zeta\rho)\hat{\mathcal{P}}_{t+s}\right)\right] \tag{5}$$

$\hat{mc}_t$ is the change in the aggregate nominal marginal cost. $\zeta = \frac{1}{1+d_{mc,y}\theta\rho}$, where $d_{mc,y} = \frac{1-a+a_w}{a}$ is a constant equal to the elasticity of firm-level marginal costs with respect to firm-level production, and it is equal to one in a textbook model with CRS in production and common labor markets. Note that ρ and ζ are not firm-specific because of our symmetric steady-state assumption, which we relax in Section 5. Aggregating across firms, we obtain the marginal cost-based Phillips curve:

$$\hat{\pi}_t = \varphi\omega(\hat{mc}_t - \hat{\mathcal{P}}_t) + \beta\mathbb{E}_t[\hat{\pi}_{t+1}], \tag{6}$$

[5] This relationship arises from the definition $\log p_{it} = \log \mu_{it} + \log mc_{it}$ and taking partial derivatives with respect to log marginal costs, yielding $\rho_{it} = -\theta_{it}\Gamma_{it}\rho_{it} + 1$.

where $\hat{mc}_t - \hat{\mathcal{P}}_t$ is the log-deviation in the aggregate real marginal cost. $\kappa_{mc} = \varphi\omega$ is the slope of the marginal cost-based Phillips curve. $\varphi = \dfrac{(1-\alpha)(1-\beta\alpha)}{\alpha}$ captures the role of price rigidity. The multiplicative factor $\omega = \rho\zeta$ captures micro real rigidities, due to two distinct economic forces. First, the flexible price partial equilibrium pass-through ρ reflects strategic complementarities in price-setting. Second, ζ captures the fact that when firm-level marginal costs are upward sloping, a cost shock induces an adjustment in size, which dampens the first-round effect on marginal cost. Together, ω is the average flexible price pass-through of an input cost shock into prices. A lower ω, reflecting stronger micro real rigidities, flattens the Phillips curve.

Aggregate marginal costs. From the market clearing conditions, we obtain the solution for aggregate marginal cost:

$$\hat{mc}_t = \underbrace{\left[\frac{1-a+\nu^{-1}}{a} + \upsilon\right]}_{\Omega\,=\,\text{Elasticity of mc wrt output}} \hat{Y}_t + \mathcal{P}_t \tag{7}$$

where $\nu^{-1} = \frac{u_{ll}L}{u_l} - \frac{u_{cl}L}{u_c}$ is the inverse Frisch elasticity of labor supply. υ depends on the chosen assumption on consumption-labor complementarities in the utility function. Under GHH preferences, $\upsilon = 0$. With separable preferences $\upsilon = \sigma^{-1}$ where $\sigma^{-1} = -\frac{u_{cc}C}{u_c}$ is the inverse elasticity of intertemporal substitution.

Output-based Phillips curve. Combining (6) and (7), we obtain the output-based New Keynesian Phillips curve:

$$\hat{\pi}_t = \kappa_y \hat{Y}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}] \tag{8}$$

with $\kappa_y = \varphi\omega\Omega$ the slope of the Phillips curve.

Three-equations New Keynesian model. Additionally solving for the Euler equation (see Appendix equation A.32), we obtain a three-equation version of the New Keynesian model:

$$\text{(NKPC)} \qquad \hat{\pi}_t = \kappa_y \hat{Y}_t + \beta \mathbb{E}_t[\hat{\pi}_{t+1}]$$

$$\text{(Euler equation)} \qquad c\hat{Y}_t = c\mathbb{E}_t\left[\hat{Y}_{t+1}\right] - \sigma\left(\hat{i}_t - \mathbb{E}_t\left[\hat{\pi}_{t+1}\right]\right)$$

$$\text{(MP rule)} \qquad \hat{i}_t = \phi_\pi \hat{\pi}_t + \phi_y \hat{Y}_t + \varepsilon_t^{MP},$$

where $\hat{Y}_t$ the log-deviation of output is also equal to the output gap since we are only considering the presence of monetary policy shocks.

3.3. Generalization

Our objective is to quantify the link between demand expansions and inflation as described by equation (NKPC) by estimating the slope $\kappa_y = \varphi\omega\Omega$. The model introduced above introduces the key concepts and notations used in our empirical analysis. More generally, our estimation approach holds in any model of the New Keynesian class that yields the formulation for the NKPC in (NKPC) (potentially with an exogenous cost-push shock). We present a list of extensions and modifications of our model that still allow for a such representation in Table A.1.

4. Estimation of the slope of the Phillips curve

In this section, we estimate the elasticity of inflation with respect to the output gap in the Phillips curve (NKPC) from a set of sufficient statistics. That is, we estimate $\kappa_y = \varphi\omega\Omega$, which we refer to as the slope of the Phillips curve.

We exploit the fact that in the class of models that yield the formulation of the Phillips curve in (NKPC), we can always write the slope κ_y as the product of two objects. These objects are the slope of the marginal cost-based Phillips curve $\kappa_{mc} = \varphi\omega$, and the elasticity of marginal costs with respect to quantities Ω. Intuitively, a demand-driven quantity expansion raises marginal costs, in proportion to Ω, and the increase in marginal costs then feeds through prices and leads to inflation, captured by κ_{mc}. This decomposition was first introduced by Galı and Gertler (1999).

Our key innovation is in how we take this decomposition to the data. We directly measure each of these sufficient statistics using causal cross-sectional research designs. First, our approach is robust to the presence of unobserved shocks that drive spurious correlation between prices, marginal costs, and quantities. Second, this approach does not require us to take a stance on the specific microfoundations that shape those elasticities, as long as those microfoundations end up producing the Phillips curve in equation (NKPC). For instance, we do not need to take a stance on the structure of preferences that gives rise to a supply curve for inputs, or the specific shape of demand curves that generates strategic complementarities. By contrast, the usual practice in the literature to obtain the slope of the Phillips curve is to impose more rigid functional forms and calibrate all the model parameters using either external evidence (with a risk for model misspecification in the model blocks that shape the Phillips curve) or by matching impulse responses to aggregate shocks (which implicitly leverages the blocks of the model different than the Phillips curve, with a risk for misspecification as well).

This section presents our empirical methodology to estimate the slope of the marginal cost-based Phillips curve κ_{mc} (section 4.1), the elasticity of marginal costs with respect to quantities Ω (section 4.2), and finally computes our estimate for the slope of the Phillips curve in India (section 4.3).

4.1. Slope of the marginal cost-based Phillips curve

The slope of the marginal cost-based Phillips curve is $\kappa_{mc} = \frac{(1-\alpha)(1-\beta\alpha)}{\alpha}\omega$, where $\omega = \rho\zeta$, as before. We show that it can be recovered from two moments: the firm-level partial equilibrium pass-through of cost shocks into prices which identifies $(1-\alpha)(1-\beta\alpha)\omega$, and an estimate of α.

Identification of $(1-\alpha)(1-\beta\alpha)\omega$**.** The firm-level partial equilibrium pass-through of cost shocks into prices identifies $(1-\alpha)(1-\beta\alpha)\omega$. To make this point, in Appendix A.3 we augment our baseline model with firm-specific variable idiosyncratic input cost shocks ϑ_{it}. Variable costs for firm i are given by $\hat{w}^v_{it} = \hat{w}^l_t + \vartheta_{it}$.[6] We first consider the case of a zero-persistence shock. The optimal reset price is now given by:

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\omega\left(\hat{mc}_{t+s} + \vartheta_{it+s}\right) + (1-\omega)\hat{\mathcal{P}}_{t+s}\right)\right]. \tag{9}$$

[6] Note that Appendix A.3 also allows for intermediate inputs in the production function, in line with the data. For ease of exposition, we derive the identification argument in the main text with labor only.

The Calvo assumption implies that the price charged by firm i in period t, is equal to its reset price whenever it gets a chance to reset its price, and equal to its last period's price whenever it does not. Formally, $\hat{p}_{it} = \mathbb{1}^p_{it}\hat{p}_{it|t} + (1 - \mathbb{1}^p_{it})\hat{p}_{it-1}$, where $\mathbb{1}^p_{it}$ is a "Calvo-fairy" dummy that takes the value of 1 if a price adjustment is permissible for firm i in period t. By plugging equation (9) into the Calvo law of motion of individual prices we just stated, we derive an expression for the price change of firm i in period t:

$$\hat{p}_{it} - \hat{p}_{it-1} = \mathbb{1}^p_{it}(1-\beta\alpha)\omega\vartheta_{it} + \mathbb{1}^p_{it}(1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\omega\hat{mc}_{t+s} + (1-\omega)\hat{\mathcal{P}}_{t+s}\right)\right] - \mathbb{1}^p_{it}\hat{p}_{it-1}.$$

Let us assume that we observe $\mathcal{Z}^{\vartheta}_{it}$, a proxy for ϑ_{it} satisfying $\vartheta_{it} = k^{\vartheta}\mathcal{Z}^{\vartheta}_{it}$. Our first identification result of this section, proven in Appendix A.3.3, is that a regression of $\Delta \log p_{it}$ on $\Delta \log w^v_{it}$ instrumented by the exogenous zero-persistence cost shifter $\mathcal{Z}^{\vartheta}_{it}$, yields in population an IV estimate equal to:

$$\beta^{IV}_{p,w} = (1-\alpha)(1-\beta\alpha)\omega. \tag{10}$$

Intuitively, equation (10) states that after an exogenous change in input costs of 1%, firms change their price with an elasticity equal to ρ, which triggers changes in quantities, and due to potentially upward-sloping marginal cost curves, triggers endogenous changes in marginal costs. The fixed point between prices and marginal cost changes is captured by ζ. These two economic mechanisms captured by $\omega = \rho\zeta$ would be the end of the story if firms had flexible prices, but the presence of price rigidity implies that only a fraction $1-\alpha$ of firms is able to reset their prices, and when they do, they take into account that with some probability they will not be able to reset their prices in the future, giving rise to the full term in the equation.

In Appendix A.3.3 we present an extension of our identification argument for the case where the cost disruption follows an AR(1) process with persistence ρ_{ϑ}. In this case, we show that: $\beta^{IV}_{p,w} = \frac{1-\beta\alpha}{1-\beta\alpha\rho_{\vartheta}}(1-\alpha)\omega$.

Empirical strategy. The argument derived in the model lends itself to a research design in a panel of firms. Specifically, we estimate the following regression:

$$\Delta \log p_{ijt} = \Theta_{jt} + \beta_{p,w} \Delta \log w_{it}^{v} + \epsilon_{ijt}, \tag{11}$$

where $\Delta \log p_{ijt}$ is the change in the price charged by firm i for a given product j, $\Delta \log w_{it}^{v}$ is the change in the firm-level input price index, and Θ_{jt} is a set of product×time fixed effects. Product×time fixed effects have two roles. First, they ensure that we only estimate the reaction of $\Delta \log p_{ijt}$ to the firm's own input cost shock and partial out the response to any contemporaneous change in the product-level price index (strategic complementarities). Second, they control for product-level demand shocks, a key threat to identification is this context, as detailed below.

$\Delta \log w_{it}^{v}$ is the change in the price index of variable inputs. We consider that variable inputs are labor and intermediate inputs (materials and energy) and we perform robustness checks where we treat labor as a fixed input in the short-run. We construct the change in the price index of variable inputs as $\Delta \log w_{it}^{v} = \bar{s}_{ixt} \, \Delta \log w_{it}^{x} + \bar{s}_{ilt} \, \Delta \log w_{it}^{l}$. $\bar{s}_{ixt}$ and $\bar{s}_{ilt}$ are the Törnqvist-weighted shares of intermediates and labor in variable inputs, respectively. $\Delta \log w_{it}^{l}$ is the change in the daily wage. $\Delta \log w_{it}^{x}$ is the change in the intermediate inputs price index, which as a reminder is constructed as $\Delta \log w_{it}^{x} = \Delta \log \mathcal{C}_{it}^{x} - \underline{\Delta} \log x_{it}$.[7]

The main concern of estimating equation (11) is if firms experience demand shocks for their products and face upward-sloping supply curves for some or all of their inputs. In this case, a demand shock will induce an increase in the price of the firm's output and input bundle. It will hence generate a positive correlation between prices and input costs, but due to an economic mechanism that is distinct from the pass-through of cost shocks we are aiming to estimate. A second empirical concern is that our estimates of pass-through may be attenuated due to measurement error in input costs.

Instrument. We address this identification concern by using an instrumental variable approach. We consider two instruments. The first instrument $\mathcal{Z}_{it}^{\vartheta}$ (instrument A) follows the methodology in Amiti, Itskhoki, and Konings (2019), but uses all inputs as opposed

[7] The assumption is that $\underline{\Delta} \log x_{it}$ (the average increase in input quantity for the inputs $\underline{\mathcal{K}}_i$ for which we observe input-level data) is equal to the average change in input quantity for all inputs $\Delta \log x_{it}$. In robustness checks, we also present results using $\Delta \log w_{it}^{x} = \underline{\Delta} \log w_{it}^{x}$.

to imported inputs. It is defined as follows. For the set $\underline{\mathcal{K}}_i$ of intermediate inputs of firm i for which we observe a price change $\Delta \log w_{ikt}$, we define:

$$\mathcal{Z}_{it}^{\vartheta} = \sum_{k \in \underline{\mathcal{K}}_i} s_{ikt-1} \Delta \log w_{ikt}, \tag{12}$$

where s_{ikt-1} is the previous period share of input k in all inputs used by firm i. Compared to the construction of $\Delta \log w_{it}^{v}$, the instrument $\mathcal{Z}_{it}^{\vartheta}$: (i) excludes labor, which is the input most likely subject to firm-specific supply curves, (ii) only exploits the inputs for which we observe the price change, (iii) uses lagged weights (as opposed to mid-point weights in the construction of the input price index). This instrument has a strong predictive power. A caveat is that the instrument exploits the change in the price of input k paid by firm i, which may lead to identification concerns if even material inputs other than labor have firm-specific supply curves. To address this concern, we develop a second instrument. The second instrument $\mathcal{Z}_{it}^{\vartheta}$ (instrument B) is defined as follows:

$$\mathcal{Z}_{it}^{\vartheta} = \sum_{k \in \underline{\mathcal{K}}_i} s_{ikt-1} \Delta \log w_{s(i)kt} \tag{13}$$

Instead of using the firm-input specific price change $\Delta \log w_{ikt}$, we use the average price change for this input in the firm's state $\Delta \log w_{s(i)kt}$. We use the state-level price change because markets for many inputs have a local dimension. Given that many firms use the same input, it is unlikely that the price of the input is affected by the demand shock of a single firm.[8] This second instrument, thus, has the advantage of being more immune to the concern related to firm-specific supply curves, though at the cost of lower first stage power.

Regressions are weighted by firm×product-level lagged sales multiplied by the ASI sampling weight to obtain results representative at the aggregate level. Because the distribution of firm size is highly skewed, we winsorize the top 1% of weights to avoid results being overly sensitive to a few large firms. Standard errors are clustered at the firm level.

Results. Table 1 presents our results. Columns (1)-(3) present OLS estimates, columns (4)-(6) present our IV estimates with instrument A, and columns (7)-(9) present our IV

[8] We exclude inputs with less than 10 observations in the state×year cell.

estimates with instrument B.

The IV estimate is around 0.22 and is stable across specifications. This implies that a 10% increase in the input price index leads to a 2.2% price increase in the same year. It is worth noting that our IV estimates are roughly double than our OLS elasticities, suggesting an important role of the instrument in correcting for measurement error or endogeneity. The results are highly similar whether we use instrument A or instrument B.

Table 1. Elasticity of price changes to input cost changes

	$\Delta \log p_{ijt}$								
	OLS			Instrument A			Instrument B		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\Delta \log w_{it}$	0.094***	0.094***	0.092***	0.217***	0.216***	0.209***	0.214***	0.213***	0.186***
	(0.007)	(0.007)	(0.008)	(0.016)	(0.016)	(0.018)	(0.043)	(0.043)	(0.047)
Year × Product FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Firm × Product FE			✓			✓			✓
Controls		✓			✓			✓	
Observations	364,862	364,862	309,493	364,517	364,517	309,186	363,800	363,800	308,488
F-stat				4551.5	4558.8	3399.6	507.7	508.1	373.8
Adj. passthrough				0.200	0.199	0.195	0.197	0.197	0.173

Note: This table reports the results of estimating equation 11. Columns (1)-(3) report OLS results with the independent variable defined as the firm-level change in input costs. Columns (4)-(6) report IV results with the instrument defined in (12). Columns (7)-(9) report IV results with the instrument defined in (13). Regressions are weighted by firm×product-level lagged sales, adjusted for the ASI sampling weight (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

In the limit of perfect flexible price pass-through ($\rho = 1$) and constant returns to scale together with common input markets ($\zeta = 1$), our model predicts that $\beta^{IV}_{p,w} = \frac{1-\beta\alpha}{1-\beta\alpha\rho_\vartheta}(1 - \alpha)$. This term tends to 1 as α tends to zero (fully flexible prices). So our statistic rejects the limit of a standard New Keynesian model with constant returns to scale and full flexible-price pass-throughs, unless prices are very rigid at an annual frequency, a possibility that we will reject in the following paragraphs.

From the pass-through to the slope of the marginal cost-based Phillips curve. The estimates in Table 1 correspond to $\beta^{IV}_{p,w} = \frac{1-\beta\alpha}{1-\beta\alpha\rho_\vartheta}(1 - \alpha)\omega$. Meanwhile, we aim to estimate $\kappa_{mc} = \frac{1}{\alpha} \times (1-\alpha)(1-\beta\alpha)\omega$. Comparing these two expressions shows that we need to make two adjustments. First, we need to adjust for the persistence of the input cost shock and multiply our estimate by $1 - \beta\alpha\rho_\vartheta$ to obtain the zero-persistence firm-level pass-through

$(1-\alpha)(1-\beta\alpha)\omega$. Second, we need to divide this quantity by $\frac{1}{\alpha}$ to convert the partial equilibrium pass-through into its general equilibrium counterpart.

We now show how to estimate ρ_ϑ and α. Note that we need to take a stand on the value of β. We use a yearly frequency $\beta = 0.96$ as a benchmark.

Estimation of ρ_ϑ. We estimate the persistence of the input price disturbance by estimating the autocorrelation of the instrument $\mathcal{Z}_{it}^{\vartheta}$. The estimating equation and results for various versions of the instrument are presented in Figure D.1.

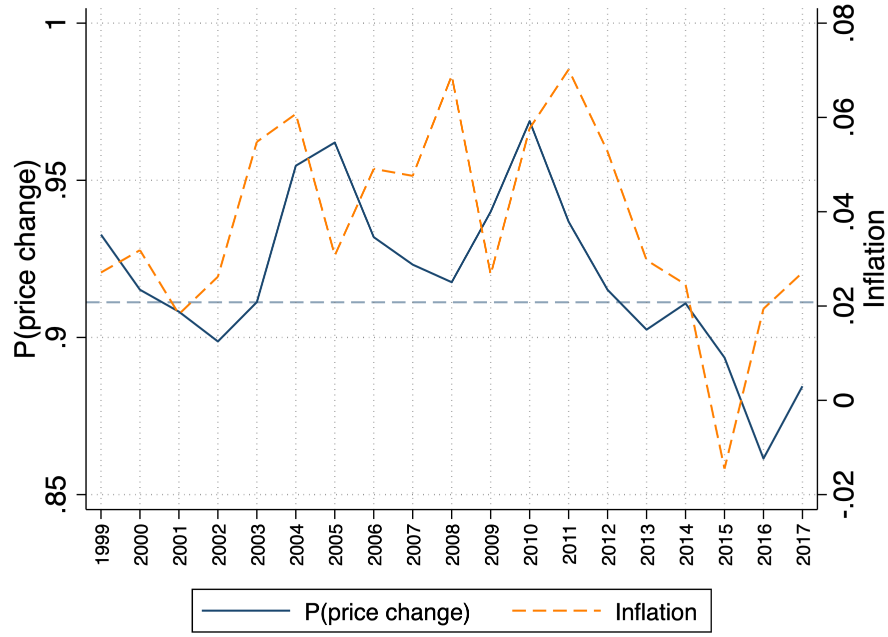
Estimation of α. We estimate α in two ways. First, we estimate the frequency of price changes from the price data directly. We follow standard practice to measure the frequency of price changes and measure an average frequency of price changes at the annual frequency of 0.91, mapping into $\alpha = 0.09$ at the annual frequency. Figure 3 shows the time-series of our measure of price rigidity. On top of marking the average frequency of price changes in a dotted line, the figure validates our measure by showing that the frequency comoves with the official Wholesale Price Index (WPI) inflation rate in India. This is true despite the two series coming from entirely different sources (the WPI is not based on micro-data from the ASI but on an independent data collection). For simplicity our model abstracts from movements in the frequency of price changes, and we use its average value throughout.

A caveat of this measure is that it is subject to mismeasurement in the frequency of price changes. We propose an additional methodology to recover α, leveraging the predictions of the Calvo model for the longer-horizon pass-through coefficients. Let us denote $\beta_{p,w}^{IV,h}$ the IV-local projection coefficient at horizon h, i.e., when the outcome variable is defined as $\log p_{ijt+h} - \log p_{ijt-1}$. It is straightforward to show that:

$$\beta_{p,w}^{IV,1} = \frac{1-\beta\alpha}{1-\beta\alpha\rho_\vartheta}\omega(1-\alpha)\left(\rho_\vartheta + \alpha\right) = (\rho_\vartheta + \alpha)\beta_{p,w}^{IV,0}.$$

This expression is very intuitive. $\frac{1-\beta\alpha}{1-\beta\alpha\rho_\vartheta}\omega$ is the desired pass-through of a firm who can reset its price on impact $h = 0$. If the firm can adjust its price at $h = 1$ (probability $1-\alpha$), then it will adjust its price by $\rho_\vartheta \times \frac{1-\beta\alpha}{1-\beta\alpha\rho_\vartheta}\omega$, reflecting the new value of the cost disturbance at $h = 1$. If the firm cannot adjust its price at $h = 1$ (probability α), then with probability $1-\alpha$ it could reset its price at $h = 0$ with pass-through $\frac{1-\beta\alpha}{1-\beta\alpha\rho_\vartheta}\omega$, and with probability α it could not reset its price at $h = 0$ and the pass-through is 0. This yields the expression above. The last equality shows that the ratio $\beta_{p,w}^{IV,1} / \beta_{p,w}^{IV,0}$ identifies $\rho_\vartheta + \alpha$. We estimate α by

FIGURE 3. Frequency of price changes and the inflation rate in India



Note: This figure illustrate the time series behavior of the frequency of price changes along with the behavior of the wholesale price index (WPI) inflation in India.

combining our estimate of ρ_ϑ with estimates of dynamic pass-throughs estimated in Table D.6. We find that $\alpha = 0.09$ as above is consistent with this alternative estimation method.

Our estimates are consistent with a frequency of price changes of 0.45 per quarter, or 0.18 per month. Interestingly, these are similar to estimates in the literature for Mexico (Gagnon 2009), United States (Nakamura et al. 2018), and Argentina (Alvarez et al. 2019), when those economies faced inflation rates similar to those of India.

Slope of marginal cost-based Phillips curve. First, with our estimates of α and ρ_ϑ in hand, we rescale the firm-level pass-through by $1 - \beta\alpha\rho_\vartheta$ to obtain the pass-through in the case of a zero-persistence shock. This quantity is reported in the last line of Table 1. A zero-persistence shock that increases the cost of inputs by 1% for some firms relative to others leads to a 0.2% increase in the prices of affected firms relative to others at the yearly frequency.[9] This corresponds to a 0.05% increase at the quarterly frequency.

Second, we rescale the firm-level pass-through by α to obtain the slope of the aggre-

[9]Note that this estimate cannot be readily compared with estimates of the pass-through of marginal cost shocks into prices, since the pass-through of input cost shocks to marginal costs is not 1 when returns to scale are decreasing, as we document below.

gate Phillips curve κ_{mc}. Our estimates imply $\kappa_{mc} = 0.095$ when expressed at the quarterly frequency. That is, a 1% aggregate increase in marginal costs leads to a 0.1% increase in inflation in the same quarter. We discuss the mechanisms behind this slope and compare it with existing estimates in section 4.3.

Robustness and extensions. In Appendix D.1, we include a battery of robustness exercises and extensions. In Table D.2, we include additional fixed effects, controls, modify our definition of the input price index, use a different weighting approach. All results are consistent. Table D.3 drops products subject to "reservation" regulations or the year when the RBI conducted a large scale demonetization episode (2016). Table D.1 estimates our specification at the firm-level (as opposed to firm$\times$product-level) and yields similar estimates. Table D.5 investigates non-linearities by excluding observations in increasing bands centered around the point where the cost shock is equal to zero. We do not find any evidence for non-linearities of the pass-through. Table D.4 investigates heterogeneity in the pass-through coefficient between the first subperiod characterized by high inflation (1998-2013) and the following subperiod (2014-2017).

4.2. The elasticity of marginal costs to output changes

The second component of the slope of the Phillips curve is Ω, the elasticity of marginal costs to output changes. Ω reflects two economic forces. At the level of each individual firm, the elasticity of marginal costs to output changes $d_{mc,y} = \frac{1-a}{a} + \frac{a_w}{a}$ reflects the degree of returns to scale a and the response of firm-level input prices when $a_w > 0$. At the aggregate level, a demand shock generates a response of factor prices, so that $\Omega = \frac{1-a}{a} + \frac{\nu^{-1}}{a} + \upsilon$, where ν^{-1} is the inverse price elasticity of the input supply curve and $\upsilon = 0$ if preferences are GHH or $\upsilon = \sigma^{-1}$ if preferences are separable.

We start by estimating the firm-level elasticity of marginal costs $d_{mc,y}$. This parameter carries information for both ζ and Ω and hence will be useful to decompose the slope of the NKPC. Then, we estimate the elasticity of marginal costs at coarser levels of aggregation to account for the input price response and obtain an estimate of Ω.

4.2.1. Firm-level marginal cost curve.

Identification. Firm marginal costs are given (in log-deviations) by:

$$\hat{mc}_{it} = \frac{1-a}{a}\hat{y}_{it} + \frac{a_w}{a}(\hat{y}_{it} - \hat{Y}_t) + \hat{w}^l_t \tag{14}$$

Our second identification result of this section, proven in Appendix A.3.3, is that a regression of $\Delta \log mc_{it}$ on $\Delta \log y_{it}$ instrumented by a demand shifter $\mathcal{Z}^{\xi}_{it}$, yields in population an IV estimate equal to:

$$\beta^{IV}_{mc,y} = \frac{1-a}{a} + \frac{a_w}{a} \tag{15}$$

Empirical strategy. We estimate the regression equation:

$$\Delta \log mc_{it} = \Phi_{s(i),t} + \beta_{mc,y}\Delta \log y_{it} + \epsilon_{i,t}, \tag{16}$$

$\Delta \log y_{it}$ is firm-level output growth. For any variable cost function of the form $\mathcal{C}_{it} \propto y_{it}^{\frac{1}{a}}$, we can write $\Delta \log mc_{it} = \Delta \log \mathcal{C}_{it} - \Delta \log y_{it}$. This implies what we can equivalently estimate (16) with the change in total variable cost as the outcome variable and obtain $\beta_{\mathcal{C},y}$ as estimate of $\frac{1+a_w}{a}$, or estimate (16) using the previous definition of marginal cost and obtain $\beta_{mc,y}$ as estimate of $\frac{1-a+a_w}{a}$. We estimate this regression with $\Phi_{s(i),t}$, a set of industry$\times$time fixed effects.

Instrument. The main identification concern is that firm-level supply shocks will induce a correlation between marginal costs and quantities that are not informative about the slope of the marginal supply curve.

To address this identification concern, we instrument the change in firm-level quantity by a demand shifter. We leverage the fact that firms in our sample are multiproduct firms with heterogeneous exposure to their product portfolio. As a result, changes in the demand for specific products will induce firm-level demand shocks as a function of their pre-existing product shares. We define:

$$\mathcal{Z}^{\xi}_{it} = \sum_j s_{ijt-1}\Delta \log \mathcal{R}_{jt}, \tag{17}$$

where $\Delta \log \mathcal{R}_{jt}$ is the log change in product-level sales, and s_{ijt-1} is the one-lag share of product j in the sales of firm i. Our identifying assumption is that firms that are differentially exposed to more demand-sensitive products are not differentially exposed to firm-level supply shocks.

Results. Table 2 presents our estimates. Columns (1) and (2) present estimates where we consider that variable costs are materials and labor, while columns (3) and (4) exclude labor from our definitions of variable costs. Our average estimate of 0.17 implies that when quantities increase by 10%, marginal costs increase by 1.7%. Assuming no firm-specific input markets, this coefficient maps to the structural object $\frac{1-a}{a}$; then our results imply a short-run firm-level returns to scale parameter $a = 0.86$. Our inference for a is stable regardless of whether we use as a dependent variable a measure of changes in marginal or total costs (column 2 versus column 1). We obtain slightly lower returns to scale when we exclude labor from our cost variables (column 3 versus column 1), consistent with theory. The F-statistic for the first stage is large and stable across specifications.

TABLE 2. Firm-level elasticity of marginal costs to changes in quantities

	Baseline		Excl. labor	
	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$
	(1)	(2)	(3)	(4)
$\Delta \log y_{it}$	0.168**	1.082***	0.256***	1.182***
	(0.076)	(0.075)	(0.082)	(0.081)
Year × Ind. FE	✓	✓	✓	✓
Observations	267,011	267,011	267,010	267,010
F-Stat	171.91	171.91	171.91	171.91
Returns to scale	0.86	0.92	0.80	0.85

Note: This table reports the results of estimating equation (16). It report IV results with the instrument defined in (17). Columns (1) and (3) use the change in marginal costs as outcome variable; columns (2) and (4) use the change in variable costs as outcome variable. In columns (1) and (2), variable costs are materials and labor, while columns (3) and (4) exclude labor from variable costs. Regressions are weighted by firm-level lagged sales (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively. The last row in the Table presents our estimates for the returns to scale parameter a.

4.2.2. Aggregate marginal cost curve.

Our estimate $\hat{\beta}_{mc,y}$ is a partial equilibrium short-run supply elasticity that keeps constant prices and quantities at higher levels of aggregation. In particular, as we derived in Section 3, our model implies that Ω, the short-run elasticity of aggregate marginal costs to aggregate quantity, is given by

$$\Omega = \frac{1 - a + \nu^{-1}}{a} + \upsilon, \tag{18}$$

which makes clear that the reaction of input prices, crucially the real wage, to higher demand for inputs will induce a higher elasticity of marginal costs to quantities in general equilibrium than in partial equilibrium.

Empirical strategy. To obtain an estimate for Ω, we propose to estimate the counterpart of the firm-level marginal cost specification (16) at a higher level of aggregation, namely at the regional (district) or at the industry level. By estimating the response of marginal costs at higher levels of aggregation, these estimates will capture the equilibrium price adjustment occurring at those levels of aggregation.

The key identification concern in the firm-level marginal cost specification also applies when estimating this specification at the district or industry level. We therefore again use an instrumental variable strategy where the district (industry)-level change in quantity is instrumented by a demand-side shifter. For the district-level specification, our demand shifter exploits product-specific shifts in demand and variation in product specialization across districts. For each district d, we define:

$$\mathcal{Z}^{\xi}_{dt} = \sum_j s_{djt-1} \Delta \log \mathcal{R}_{jt}, \tag{19}$$

where $\Delta \log \mathcal{R}_{jt}$ is product-level sales growth, and s_{djt-1} is the one-lag sales share of product j in district d. This is similar to the design in Hazell et al. (2022).

For the industry-level specification, our demand shifter exploits the intuition that increases in sales in downstream industries generates demand shifts in upstream industries.

Formally, for each 3-digit industry k, we define:

$$\mathcal{Z}^{\xi}_{kt} = \sum_j s_{kjt-1} \Delta \log \mathcal{R}_{jt}. \tag{20}$$

s_{kjt-1} is the share of the total sales of industry k that are used as inputs in the production of good j. The set of downstream goods j is composed of the disaggregated product codes for manufacturing goods, and the more aggregated product codes of the IO table for the non-manufacturing goods. The shares sum to less than 1 since a fraction of the good goes to final consumption or capital formation. $\Delta \log \mathcal{R}_{jt}$ is sales growth for good j, obtained from the ASI micro-data for manufacturing goods and from the national accounts otherwise. This is similar to the designs in Shea (1993) and Boehm and Pandalai-Nayar (2022).

Table 3. Aggregate elasticity of marginal costs to changes in quantities

	District		Industry	
	Δ log mc	$\Delta \log \mathcal{C}$	$\Delta \log mc$	$\Delta \log \mathcal{C}$
$\Delta \log y$	0.583***	1.487***	0.703**	1.704***
	(0.144)	(0.140)	(0.310)	(0.311)
Year FE	✓	✓	✓	✓
F Stat	103.08	103.08	26.83	26.83
N	7,707	7,707	1,211	1,211

Note: This table reports the results of estimating the aggregated version of equation 16. Columns (1) and (2) show district-level results with the instrument defined in (19). Columns (3) and (4) show industry-level results with the instrument defined in (20). Columns (1) and (3) use the change in marginal costs as outcome variable; columns (2) and (4) use the change in variable costs as outcome variable. Regressions are weighted by district(industry)-level lagged sales (top 1% winsorized). ***, ** and * indicate significance at the 1%, 5% and 10% levels, respectively.

Results. Estimates in Table 3 imply that after an increase in 1% in quantities, marginal costs increase from 0.6% to 0.7% depending on the source of variation. This directly maps to the Ω parameter.

What if a component of the elasticity of marginal costs to quantities is only operative at the national level, as opposed to the district or the industry level? In this case, our estimate will partial out this fraction of the increase in the marginal cost, and provide a lower bound for Ω.

That we can estimate Ω directly in our data, independent of the specification of other sub-blocks of our model, is one of the main advantages of our research design. It is an improvement over variants of the standard approach of inferring Ω from a moment-matching exercise in which an econometrician infers the value of the structural parameters that pin down Ω by minimizing the distance of the impulse response functions to an aggregate shock, usually a monetary policy shock, in the data and those implied by a model. Notably, that approach is subject to criticisms of mispecification in every block of the structural model that influences the whole shape of an impulse response function. By estimating Ω in the data, our approach is less subject to these concerns.

4.3. The slope of the Phillips curve

With our estimates of κ_{mc} and Ω, we can assemble our estimate for $\kappa_y = \kappa_{mc} \times \Omega$, the elasticity of inflation with respect to the output gap. We find $\kappa_y = 0.066$ at the quarterly frequency. That is, a 10% increase in the output gap raises inflation by 0.66 percentage points.

A key implication of this finding is that the Phillips curve in India is steeper than in developed countries. Our estimate for κ_y is one order of magnitude larger than the most recent estimates in the literature for developed countries ($\kappa_y = .008$ in Hazell et al. 2022; $\kappa_y \in [0.006, 0.021]$ in Gagliardone et al. 2023). An advantage of our methodology is that we can decompose the slope of the Phillips curve into its components, in order to understand what makes the Phillips curve steep.

Slope of the marginal cost-based Phillips curve κ_{mc}. Our estimate $\kappa_{mc} = 0.095$ is almost twice as large as that documented by Gagliardone et al. (2023), who find a quarterly slope of the marginal cost-based Phillips curve equal to 0.05. These estimates are comparable in terms of methodologies. Our estimate is also in the upper range of the estimates from the literature review in Mavroeidis, Plagborg-Møller, and Stock (2014) ($\kappa_{mc} \in [0.005, 0.08]$), although these estimates are less comparable as they all rely on time series methods.

The elasticity of inflation to marginal costs κ_{mc} is determined by two distinct economic forces. The first one is price rigidity (α), and the second is the extent of micro-level real rigidities induced by strategic complementarities and decreasing returns to scale in production ($\omega = \rho\zeta$). A key advantage of our method is that we can separately identify the contribution of these two forces. Indeed, with estimates of α and β, we can separately identify the term capturing micro-level real rigidities ω, and provide counterfactual values for

κ_{mc} if we only change the degree of nominal rigidity α or micro-level real rigidities ω.

The larger slope of the Phillips curve κ_{mc} in India relative to US estimates can be fully rationalized by differences in the frequency of price change: if India had the same frequency of price change as developed economies, the slope of the marginal cost-based Phillips curve would be equal to $\kappa_{mc} = 0.048$, similar to that in Gagliardone et al. (2023) and the midpoint of the range in Mavroeidis, Plagborg-Møller, and Stock (2014).

Separately estimating the term capturing micro-level real rigidities, we find $\omega = 0.25$. We can reject the benchmark value of 1 (constant returns to scale and no strategic complementarities). We can further decompose this parameter to separately investigate the role of decreasing returns to scale and strategic complementarities. Our estimate of the firm-level elasticity of marginal costs to quantities $d_{mc,y}$ is consistent with estimates of returns to scale from developed countries. With our estimate of $d_{mc,y}$ and calibrating the average demand elasticity to match the aggregate markup, we find a markup elasticity equal to 1.56. This is very similar to Gopinath and Itskhoki (2011b) who find a markup elasticity equal to 1.5, and Amiti, Itskhoki, and Konings (2019) who find a markup elasticity equal in the range of $[0.6, 1.2]$. Our estimate is thus consistent with values for developed countries. This markup elasticity maps to an average superelasticity of demand equal to approximately 8. This is consistent with values used in the literature, and consistent with the idea that strategic complementarities strongly amplify monetary non-neutrality. Therefore, both decreasing returns and strategic complementarities matter but seem to operate similarly in India and in developed countries. This is consistent with our previous point that the difference in κ_{mc} between India and developed countries can be fully rationalized by differences in the degree of price rigidity.

Slope of the marginal cost curve Ω. Most of the difference between the slope of the Phillips curve in India and in developed countries comes from the substantially higher output elasticity of marginal costs in India Ω. Our estimates of roughly 0.7 are three to four times larger than the estimates by Shea (1993) and Boehm and Pandalai-Nayar (2022), both using a similar methodology at the industry-level. Our estimate is also three times larger than Gagliardone et al. (2023). From this analysis, it appears that a feature of developing countries is steeper cost curves.

5. Steady-state misallocation and shifts in the Phillips curve

5.1. Model

In this section we extend the model in section (3) to allow for one deviation of the Phillips curve formulation in equation (1) that is salient for developing economies: we consider how input misallocation affects the transmission of demand expansions to inflation. Due to the cross-setional inefficiency of the steady state, demand expansions affect inflation not only through κ_y, but through potential changes in allocative efficiency.

The main text provides a succinct description and we leave all details and proofs to Appendix B. The main difference in the framework is in the problem of intermediate producers. Households, and the final good sector are the same as in the main text although we explicitly model the final good producer as using a Kimball aggregator.

Households. The problem of the household is identical to that in Section 3.

Final good producers. Let Y_t denote aggregate production of the final good. Y_t is used for consumption C_t so that $Y_t = C_t$.

The final good Y_t is produced by a perfectly competitive firm using a bundle of differentiated intermediate inputs y_{it} for $i \in [0, 1]$. Intermediate input varieties are assembled into the final good using the Kimball aggregator:

$$\int_0^1 \Upsilon\left(\frac{y_{it}}{Y_t}\right) di = 1,$$

where the function $\Upsilon(.)$ is strictly increasing, strictly concave, and satisfies $\Upsilon(1) = 1$. The CES aggregator is the special case $\Upsilon(q) = q^{\frac{\theta-1}{\theta}}$ for $\theta > 1$.

Taking the prices p_{it} of the inputs as given and denoting the price of the final good P_t^Y, the final good producer chooses y_{it} to maximize profits. This gives rise to the demand function:

$$\frac{y_{it}}{Y_t} = \Upsilon'^{-1}\left(\frac{p_{it}}{\mathcal{P}_t}\right), \tag{21}$$

where $\frac{p_{it}}{\mathcal{P}_t}$ determines substitution across varieties. The price index $\mathcal{P}_t$ is given by $\mathcal{P}_t = \frac{P_t^Y}{D_t}$.

$P_t^Y = \int_0^1 p_{it}\frac{y_{it}}{Y_t}di$ is the ideal price index. $D_t = \int_0^1 \Upsilon'(\frac{y_{it}}{Y_t})\frac{y_{it}}{Y_t}di$ is a "demand" index. When demand is CES, D_t is a constant equal to $\frac{\theta}{\theta-1}$. Away from the CES case, D_t is not a constant and is increasing in the dispersion of quantity shares. The price elasticity of demand is only a function of firm relative size: $\theta_{it} = \theta\left(\frac{y_{it}}{Y_t}\right)$.

Differentiated varieties producers. Each variety i is produced by a single firm. Firms produce with the production function in equation 4. The unit cost of inputs is w_t^v. We introduce input price distortions as wedges in the tradition of Hsieh and Klenow (2009). Each firm i pays $w_t^v(1+\tau_i)$ per unit of variable input. τ_i is a mean-zero steady-state distortions in firm size.

A firm has a probability $1-\alpha$ of being able to reset its price in each period. A firm that can reset its price chooses the price that maximizes:

$$\max_{p_{it|t}} \mathbb{E}_t \left[\sum_{s=0}^{+\infty} \alpha^s \Lambda_{t,t+s} \left[p_{it|t}y_{it+s|t} - \mathcal{C}(y_{it+s|t}, w_{t+s|t}^v, z_i, \tau_i)\right]\right]$$

subject to the demand curve $y_{it+s|t} = \Upsilon'^{-1}\left(\frac{p_{it|t}}{\mathcal{P}_{t+s}}\right)Y_{t+s}$ and the cost function $\mathcal{C}(y_{it+s|t}, w_{t+s|t}^v, z_i, \tau_i) = (1+\tau_i)w_{t+s}^v e^{-\frac{1}{a}z_i} y_{it+s|t}^{\frac{1}{a}}$. $\Lambda_{t,t+s}$ is the stochastic discount factor of the representative household. $\mu_{it}^f = \frac{\theta_{it}}{\theta_{it}-1}$, $\Gamma_{it} = \frac{\partial \log \mu_{it}^f}{\partial \log \frac{y_{it}}{Y_t}}$, $\rho_{it} = \frac{1}{1+\Gamma_{it}\theta_{it}}$ maintain their meaning of the flexible-price markup, the elasticity of the flexible price markup with respect to relative size, and the partial equilibrium pass-through of a marginal cost shock into prices under flexible prices, respectively.

Monetary authority. The nominal interest rate is set according to a Taylor rule.

Equilibrium. The equilibrium definition and details on the log-linearization are analogous to that in Section 3, with the clarification that steady-state distribution of firm size depends on the joint distribution of (z_i, τ_i). We denote $\lambda_i = \frac{p_i y_i}{P^Y Y}$ sales share in steady-state. Let $\mathbb{E}_\lambda[X_{it}] = \int_0^1 \lambda_i X_{it} di$.

5.2. Characterization

Marginal cost-based Phillips curve. A firm that can reset its price at time t will choose:

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\zeta_i\rho_i\hat{mc}_{t+s} + (1-\zeta_i\rho_i)\hat{\mathcal{P}}_{t+s}\right)\right], \tag{22}$$

where $\hat{mc}_t \equiv \mathbb{E}_\lambda[\hat{mc}_{it}]$ is the change in the aggregate nominal marginal cost and $\zeta_i = \frac{1}{1+\frac{d\hat{mc}}{d\hat{y}}\theta_i\rho_i}$. Aggregating across firms, we obtain the marginal cost-based Phillips curve:

$$\hat{\pi}_t = \varphi\omega(\hat{mc}_t - \hat{P}^Y_t) - \varphi(1-\omega)\hat{D}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}], \tag{23}$$

where $\hat{mc}_t - \hat{P}^Y_t$ is the change in the aggregate real marginal cost, and we define $\omega = \mathbb{E}_\lambda[\zeta_i\rho_i]$ in this extended model, a generalization of the same concept introduced before. φ, ρ_i, ζ_i have analogous definitions to those in Section 3, but they make clear that in this extended model both ρ_i and ζ_i vary across firms as a function of steady state differences in elasticities of substitution driven by differences in firm size.

$\hat{D}_t$ is the change in the demand index. All else equal, a higher $\hat{D}_t$ implies that individual firms compete with a more aggressive price index, which works towards lowering inflation.

Aggregate marginal costs. Let us define aggregate productivity Z_t as satisfying

$$Y_t \equiv Z_t L_t^a. \tag{24}$$

From the market clearing conditions, we obtain the solution for aggregate marginal cost:

$$\hat{mc}_t = \underbrace{\left[\frac{1-a+\nu^{-1}}{a} + \upsilon\right]}_{\Omega = \text{Elasticity of mc wrt output}} \hat{Y}_t - \underbrace{\left[\frac{\nu^{-1}}{a}\right]}_{\Xi = \text{Elasticity of mc wrt TFP}} \hat{Z}_t + \hat{P}^Y_t, \tag{25}$$

where, as before, ν^{-1} is the inverse Frisch elasticity of labor supply, and υ depends on the chosen assumption on consumption-labor complementarities in the utility function.

Output-based Phillips curve. Combining (23) and (25), we obtain the output-based New Keynesian Phillips curve:

$$\hat{\pi}_t = \kappa_y \hat{Y}_t - \varphi\omega\Xi\hat{Z}_t - \varphi(1-\omega)\hat{D}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}]. \tag{26}$$

Allocative efficiency. Let us define the combined allocative distortion as $m_{it} \equiv \mu_{it}(1+\tau_i)$. Let us define the aggregate steady-state distortion as: $\mathcal{M} \equiv a\frac{p^Y Y}{w^v L} = \mathbb{E}_\lambda[m_i^{-1}]^{-1}$. The change in the demand index and aggregate productivity are characterized by:

$$\hat{D}_t = -\text{Cov}_\lambda\left[\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}, \hat{p}_{it}\right], \tag{27}$$

$$\hat{Z}_t = -\text{Cov}_\lambda\left[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}, \hat{y}_{it}\right]. \tag{28}$$

In the simple case where there is no dispersion in input wedges, one can show that $\hat{Z}_t = -\mathcal{M}\hat{D}_t$. The demand index captures misallocation stemming from heterogeneous demand elasticities. When $\text{Cov}_\lambda[\theta_i, \hat{p}_{it}] > 0$, the relative price of firms with initially low markups rises relative to other firms, reallocating resources away from those firms and towards high markup firms. This increases allocative efficiency. When all $\tau_i = 0$, the only cause of allocative inefficiencies is the dispersion in markups due to imperfect competition. In the more general case with non-constant input wedges τ_i, the change in allocative efficiency depends on whether quantities are reallocated towards high or low combined distortion firms.

Using the solution of the firm's pricing problem, we obtain the law of motions for $\hat{D}_t$ and $\hat{Z}_t$ in terms of model parameters and steady-state values:

$$\hat{D}_t = -\frac{\varphi\kappa_D\left(\Omega\,\hat{Y}_t - \Xi\hat{Z}_t\right)}{1+\beta+\varphi(1+\kappa_D)} + \frac{\hat{D}_{t-1}+\beta\hat{D}_{t+1}}{1+\beta+\varphi(1+\kappa_D)} \tag{29}$$

$$\hat{Z}_t = \frac{\varphi\kappa_Z\left(\Omega\,\hat{Y}_t + \hat{D}_t\right)}{1+\beta+\varphi(1+\Xi\kappa_Z)} + \frac{\hat{Z}_{t-1}+\beta\hat{Z}_{t+1}}{1+\beta+\varphi(1+\Xi\kappa_Z)} \tag{30}$$

$$\text{with } \kappa_D = \mathbb{E}_\lambda[\zeta_i\rho_i]\left(\mathbb{E}_\lambda\left[\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}\frac{\zeta_i\rho_i}{\mathbb{E}_\lambda[\zeta_i\rho_i]}\right] - 1\right)$$

$$\kappa_Z = \mathbb{E}_\lambda[\theta_i\zeta_i\rho_i]\left(\mathbb{E}_\lambda\left[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}\frac{\theta_i\zeta_i\rho_i}{\mathbb{E}_\lambda[\theta_i\zeta_i\rho_i]}\right] - \mathbb{E}_\lambda\left[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}\right]\right).$$

The parameters κ_D and κ_Z are the equivalent of the covariances in (27) and (28), noticing that the changes in relative prices $\hat{p}_{it}$ are proportional to the firm-level pass-through coefficients $\zeta_i \rho_i$, and the changes in relative quantities $\hat{y}_{it}$ are proportional to the firm-level coefficients $\theta_i \zeta_i \rho_i$. When ex-ante allocative distortions covary negatively with pass-throughs, output booms $\hat{Y}_t$ driven by monetary shocks are concomitant with improvements in aggregate productivity.

A five-equations New Keynesian model. Solving for the Euler equation (see Appendix equation B.26), we obtain a five-equation version of the New Keynesian model:

$$\text{(NKPC)} \qquad \hat{\pi}_t = \kappa_y \hat{Y}_t - \varphi \omega \Xi \hat{Z}_t - \varphi(1-\omega)\hat{D}_t + \beta \mathbb{E}_t[\hat{\pi}_{t+1}]$$

$$\text{(LOM for D)} \qquad \hat{D}_t = -\frac{\varphi \kappa_D \left(\Omega \hat{Y}_t - \Xi \hat{Z}_t\right)}{1+\beta+\varphi(1+\kappa_D)} + \frac{\hat{D}_{t-1} + \beta \hat{D}_{t+1}}{1+\beta+\varphi(1+\kappa_D)}$$

$$\text{(LOM for Z)} \qquad \hat{Z}_t = \frac{\varphi \kappa_Z \left(\Omega \hat{Y}_t + \hat{D}_t\right)}{1+\beta+\varphi(1+\Xi\kappa_Z)} + \frac{\hat{Z}_{t-1} + \beta \hat{Z}_{t+1}}{1+\beta+\varphi(1+\Xi\kappa_Z)}$$

$$\text{(Euler equation)} \quad c\hat{Y}_t - \tilde{c}\left(\hat{Y}_t - \hat{Z}_t\right) = \mathbb{E}\left[c\hat{Y}_{t+1} - \tilde{c}\left(\hat{Y}_{t+1} - \hat{Z}_{t+1}\right)\right] - \sigma\left(\hat{i}_t - \mathbb{E}_t\left[\hat{\pi}_{t+1}\right]\right)$$

$$\text{(MP rule)} \qquad \hat{i}_t = \phi_\pi \hat{\pi}_t + \phi_y \hat{Y}_t + \varepsilon_t^{MP}$$

5.3. Empirical implementation

We turn to the estimation of the sensitivity of the shifters D_t and Z_t with respects to aggregate demand shocks, which we named κ_D and κ_Z, respectively in equations (29).

Identification of κ_D. Following the results on the identification of the slope of the Phillips curve, $\mathbb{E}_\lambda[\zeta_i \rho_i]$ is obtained from the regression of $\Delta \log p_{it}$ on $\Delta \log w^v_{it}$, instrumented by $\mathcal{Z}^\vartheta_{it}$, and corrected for the factor $(1-\alpha)(1-\beta\alpha)$:

$$\mathbb{E}_\lambda[\zeta_i \rho_i] = \frac{\beta^{IV}_{p,w}}{(1-\alpha)(1-\beta\alpha)}. \tag{31}$$

Looking at $\mathbb{E}_\lambda\left[\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]} \frac{\zeta_i \rho_i}{\mathbb{E}_\lambda[\zeta_i \rho_i]}\right]$, it is the expectation of the product of the relative price pass-through $\frac{\zeta_i \rho_i}{\mathbb{E}_\lambda[\zeta_i \rho_i]}$ and the relative demand elasticity $\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}$. We use the law of iterated expectations and estimate the price pass-through regression by bins of steady-state relative demand elasticity.

To obtain demand elasticities, we estimate markups and use the assumption that on average over the whole sample, markups will be equal to desired flexible price markups, which we invert to obtain demand elasticities. We estimate markups using the production approach, with materials as the flexible input. We rely on the elasticity of marginal costs with respect to quantities estimated above, which identifies the output elasticity of materials, so that our estimation does not suffer from the concern raised by Bond et al. (2021). Our estimation requires that any input wedge on materials is priced. We believe this assumption to be plausible; in particular, Singer (2019) documents that a large fraction of material inputs misallocation in India can be attributed to transportation costs that are reflected in prices recorded in the ASI. We describe the procedure in full details in Appendix B.4. Figure B.1 summarizes the markups and demand elasticities estimated in this way. We find that demand elasticities are decreasing in firm size, consistent with a positive superelasticity of demand.

Identification of κ_Z. $\mathbb{E}_\lambda[\theta_i \zeta_i \rho_i]$ is obtained from the regression of $\Delta \log y_{it}$ on $\Delta \log w^v_{it}$, instrumented by z^ϑ_{it}, and corrected for the factor $(1-\alpha)(1-\beta\alpha)$. An exogenous cost shock creates a relative price adjustment proportional to $\zeta_i \rho_i$, hence the relative quantity change is proportional to $\theta_i \zeta_i \rho_i$, as consumers substitute with demand elasticity θ_i. Therefore, the term $\mathbb{E}_\lambda[\theta_i \zeta_i \rho_i]$ is just the sales-weighted average quantity pass-through of a marginal cost shock (divided by $(1-\alpha)(1-\beta\alpha)$).

The research design is the same as for estimating price pass-throughs, and only requires substituting the outcome variable for a change in quantities. Formally, using the shape of the demand curves and the equation for reset prices, we find an equation for the change in quantities after an idiosyncratic cost shock. Then, a regression of $\Delta \log y_{it}$ on $\Delta \log w^v_{it}$ instrumented by the exogenous cost shifter ϑ_{it}, yields in population an IV estimate equal to:

$$\beta^{IV}_{q,w} = (1-\alpha)(1-\beta\alpha)\mathbb{E}_\lambda\left[\theta_i \rho_i \zeta_i\right]. \tag{32}$$

$\mathbb{E}_\lambda\left[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}\frac{\theta_i \zeta_i \rho_i}{\mathbb{E}_\lambda[\theta_i \zeta_i \rho_i]}\right]$ is the expectation of the product of the relative quantity pass-through $\frac{\theta_i \zeta_i \rho_i}{\mathbb{E}_\lambda[\theta_i \zeta_i \rho_i]}$ and the relative combined distortion $\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}$. We use the law of iterated expectations and estimate the quantity pass-through regression by bins of steady-state relative distortion.

This requires that we identify firm-level combined distortions m_i^{-1}. We show that m_i^{-1} is proportional to revenue productivity (TFPR). This is the same result as in Hsieh and Klenow (2009) where TFPR captures both input and output wedges. TFPR can be readily estimated in the data by combining estimates of the marginal revenue products of different inputs, which with the Cobb-Douglas assumption are just equal to revenues divided by input quantity. We describe the procedure in full details in Appendix B.4. Figure B.2 shows the obtained marginal revenue products for labor, capital, intermediates, and the resulting TFPR, by deciles of firm market shares.

Finally, $\mathbb{E}_\lambda\left[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}\right]$ can be readily estimated from our estimates of distortions and demand elasticities.

Results. We find that $\kappa_D = .00035$. $\kappa_D > 0$ reflects the fact that firms with larger demand elasticities tend to have a larger price pass-through $\rho_i\zeta_i$: firms in the fourth quartile of θ_i (the smallest firms) have a pass-through roughly 1.5 times larger compared to firms in the first quartile of θ_i (the largest firms). Lower pass-through for larger firms is consistent with the existing evidence in Amiti, Itskhoki, and Konings (2019).

Regarding Z_t, we find that $\kappa_Z = -.02791$. The key reason is that the inverse distortion $\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}$ is positively correlated with demand elasticities $\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}$. $\kappa_Z < 0$ imply that demand shocks endogenously reallocate quantities towards less distorted firms, reducing allocative efficiency.

These results imply that demand expansions only have very small effects on allocative efficiency. The elasticity of $\hat{Z}_t$ with respect to $\hat{Y}_t$ is $\frac{\varphi\kappa_Z\Omega}{1+\beta+\varphi(1+\Xi\kappa_Z)}$. We do not estimate Ξ but know $\Xi > 0$, thus we can bound this quantity by $\frac{\varphi\kappa_Z\Omega}{1+\beta+\varphi} = -0.003$. The elasticity of $\hat{D}_t$ with respect to $\hat{Y}_t$ is $\frac{\varphi\kappa_D\Omega}{1+\beta+\varphi(1+\kappa_D)} = 0.00004$.

In addition, $\hat{Z}_t$ enters the Phillips curve with a small coefficient: it is multiplied by $\varphi\omega \approx 0.095$ and $\Xi \leq \Omega < 1$.[10] More broadly, extensive quantitative explorations with different specifications of consumer preferences have always revealed that the effects of misallocation on inflation we estimate are negligible, even after allowing for values of κ_D and κ_Z on the upper end of our confidence intervals. In the aggregate results that follow,

[10]Note that in the generalization of the model with intermediates, $\Xi \leq \Omega$ is always true for GHH preferences only, the case of separable preferences depending on parametric assumptions.

we set $\kappa_D = \kappa_Z = 0$ guided by these results.

6. Monetary non-neutrality in a large developing country

In this section, we illustrate the implications of our estimates for the Phillips curve for the behavior of aggregate inflation.

We use the aggregate measure for the output gap in the manufacturing sector in India that introduced in section 2.2 and that underlies Figure 2, and our estimate for κ_y to check the time series fit of manufacturing inflation as predicted by the Phillips curve. The objective of the exercise is to provide an estimate of the importance of domestic output gap fluctuations in driving inflation, holding constant exogenous supply shifters or shifts in inflation expectations. Formally, we plot the following relationship:

$$\pi_t = \kappa_y \hat{b} \tilde{y}_t + \bar{\pi} \tag{33}$$

$\bar{\pi}$ is average inflation over the sample period. $\hat{b}$ is estimated from $\sum_{j=0}^{5} \beta^j \tilde{y}_{t+j} = a + b\tilde{y}_t + e_t$ and maps the current value of the output gap into the present value of future expected output gaps. This approach is the same as advocated by Hazell et al. (2022). Note that this exercise assumes that the variation in the transitory component of output corresponds to the output gap, and not to transitory changes in the natural rate of output.[11]

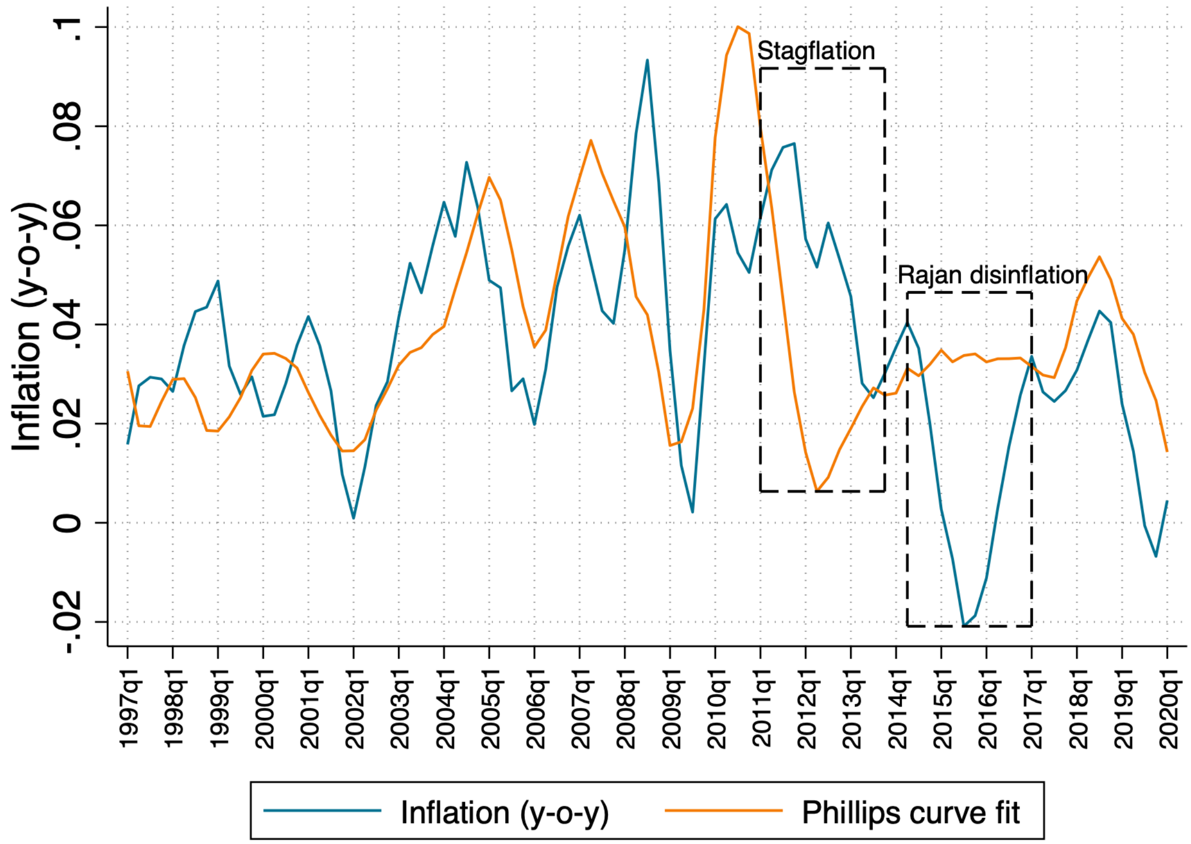
Figure 4 shows that the estimated output gap and our estimated slope of the Phillips curve yield a series for predicted inflation that closely follows actual inflation in India during the period we analyze. This result is not guaranteed by construction: we do not use the official WPI quantity or price indexes in our estimate of κ_y directly.

Conceptually, Figure 4 shows that domestic changes in quantities are an important driver of domestic prices, which is different from suggestions that inflation in developing countries is driven exclusively by cost shocks generated in the rest of the world.

The two periods for which the fit of the Phillips curve is not good are the two periods we highlighted in Figure 2b, the post-GFC stagfaltion and the Rajan disinflation, in which changes in long-run expectations and time-varying cost shocks affected the dynamics of inflation. It is thus natural to expect that other factors, and not fluctuations in demand

[11] In robustness exercises, we residualize our measure of the output gap on known supply shocks like oil price fluctuations and find highly similar results.

FIGURE 4. Fit of the aggregate Phillips Curve



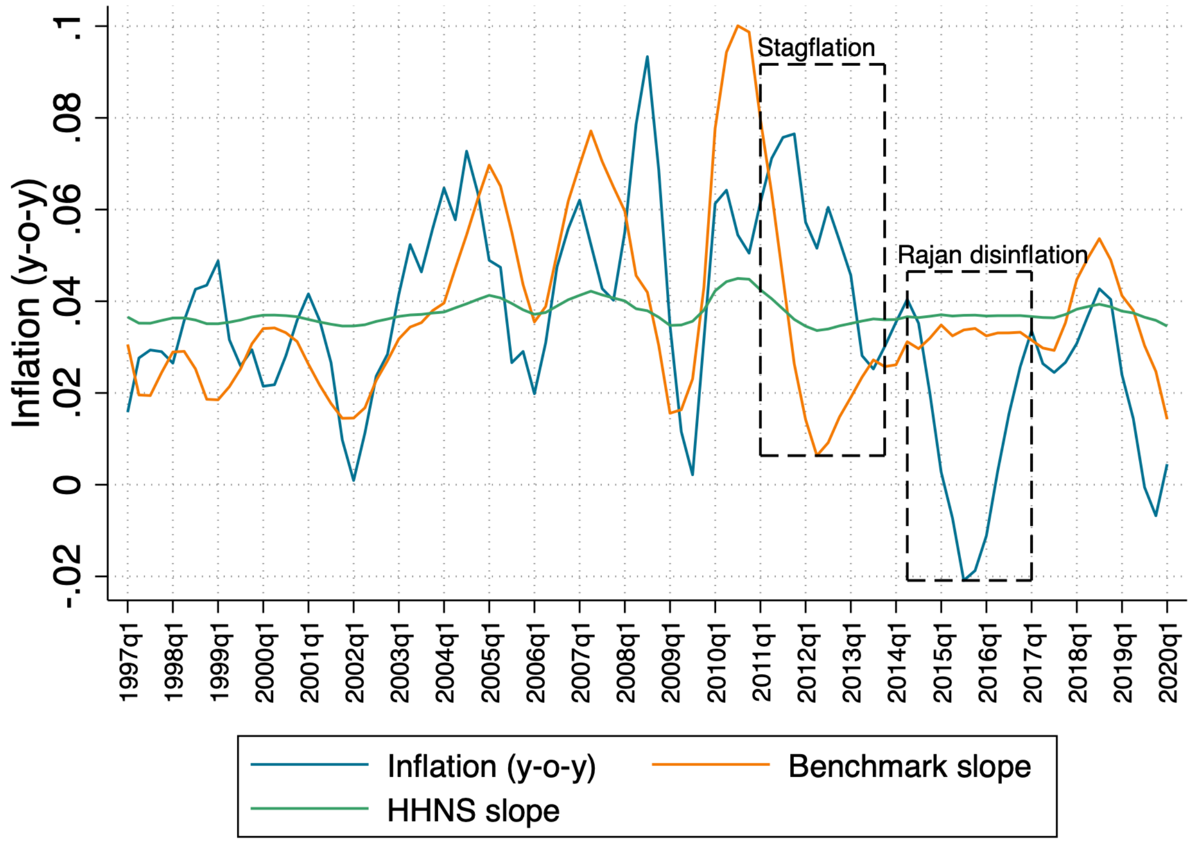
Note: This Figure takes the measure of the output gap we introduced in Section 2.2, and fits the Phillips curve as explained in equation (33) in the main text.

account for the behavior of inflation in these particular episodes.

To highlight the quantitative importance of the difference in the slope we estimate, compared to the estimates produced by the previous literature for the case of developed economies, we repeat the exercise in Figure 4, but adding an additional line, the fit that would come from an estimated κ_y equal to that of Hazell et al. (2022). We keep constant the other inputs in the calculation of equation (33).

Figure 5 shows the results and clarifies that the difference in the slope we estimate is crucial to rationalize the wide movements in inflation in India over the last 20 years. In particular, using a slope equal to that estimated in the United States by Hazell et al. (2022) would yield the conclusion that business cycle demand variation did not contribute meaningfully to the dynamics of inflation. On the contrary, our estimates suggest that the bulk of this variation, with the exception of the two episodes highlighted before, can very well be rationalized by movements in Indian aggregate demand.

FIGURE 5. Fit of the aggregate Phillips Curve



Note: This Figure takes the measure of the output gap we introduced in Section 2.2, and fits the Phillips curve as explained in equation (33) in the main text.

7. Conclusion

We have presented a portable method to estimate the slope of the Phillips curve from firm level data. Our method is free of indirect inference blocks that use information from the shape of the aggregate demand curve or the monetary policy rule, and does not require to take a strong stance on the particular microfoundations that give rise to residual demand curves, the structure of the markets for inputs as long as those microfoundations respect a decomposition of the slope of the Phillips curve into three multiplicative factors: the elasticity of prices to marginal costs, the elasticity of marginal costs to changes in quantities, and the frequency of price changes.

We have also shown the method is amenable to extensions, and we considered the salient effect of input misallocation. In distorted economies, aggregate demand expansions have the potential to reallocate production across firms with different markups and across firms with different input wedges. As a result, the covariance of demand elasticities with firm sizes are not sufficient statistics to characterize the allocative efficiency of demand expansions. Instead, the relevant covariances are those of firm-level pass-throughs with respect to firm TFPRs.

We apply our methodology to India, an economy with vast dispersion in input wedges, and find that the slope of the Phillips curve holding constant the allocation of resources is one order of magnitude larger than in developed countries such as Belgium or the United States. Around 75% of the variation in slopes across these countries is driven by differences in the elasticity of marginal costs with respect to quantities. The remaining 25% is driven by variation in the frequency of price changes across countries. The extent of micro real rigidities is remarkably similar across countries.

References

Alvarez, Fernando, Martin Beraja, Martin Gonzalez-Rozada, and Pablo Andrés Neumeyer. 2019. "From hyperinflation to stable prices: Argentina's evidence on menu cost models." *The Quarterly Journal of Economics* 134 (1): 451–505.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings. 2019. "International shocks, variable markups, and domestic prices." *The Review of Economic Studies* 86 (6): 2356–2402.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review* 98 (5): 1998–2031. https://www.aeaweb.org/articles?id=10.1257/aer.98.5.1998.

Ball, Laurence, Anusha Chari, and Prachi Mishra. 2016. "Understanding inflation in India.", National Bureau of Economic Research.

Baqaee, David R, Emmanuel Farhi, and Kunal Sangani. 2024. "The supply-side effects of monetary policy." *Journal of Political Economy* 132 (4): 1065–1112.

Bauer, Michael D, and Eric T Swanson. 2023. "A reassessment of monetary policy surprises and high-frequency identification." *NBER Macroeconomics Annual* 37 (1): 87–155.

Beraja, Martin, Erik Hurst, and Juan Ospina. 2019. "The aggregate implications of regional business cycles." *Econometrica* 87 (6): 1789–1833.

Boehm, Christoph E, and Nitya Pandalai-Nayar. 2022. "Convex supply curves." *American Economic Review* 112 (12): 3941–3969.

Boehm, Johannes, Swati Dhingra, and John Morrow. 2022. "The Comparative Advantage of Firms." *Journal of Political Economy* 130 (12): 3025–3100. https://doi.org/10.1086/720630.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch. 2021. "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data." *Journal of Monetary Economics* 121: 1–14.

Bresnahan, Timothy F, and Valerie A Ramey. 1994. "Output fluctuations at the plant level." *The Quarterly Journal of Economics* 109 (3): 593–624.

Cerrato, Andrea, and Giulia Gitti. 2022. "Inflation since covid: Demand or supply."

Drenik, Andres, and Diego J Perez. 2020. "Price setting under uncertainty about inflation." *Journal of Monetary Economics* 116: 23–38.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2023. "How costly are markups?" *Journal of Political Economy* 131 (7): 1619–1675.

Filardo, Andrew J, and Marco J Lombardi. 2014. "Has Asian emerging market monetary policy been too procyclical when responding to swings in commodity prices?" *BIS Paper* (77j).

Gagliardone, Luca, Mark Gertler, Simone Lenzu, and Joris Tielens. 2023. "Anatomy of the Phillips Curve: Micro Evidence and Macro Implications."

Gagnon, Etienne. 2009. "Price setting during low and high inflation: Evidence from Mexico." *The Quarterly Journal of Economics* 124 (3): 1221–1263.

Galı, Jordi, and Mark Gertler. 1999. "Inflation dynamics: A structural econometric analysis." *Journal of monetary Economics* 44 (2): 195–222.

Gopinath, Gita, and Oleg Itskhoki. 2011a. "In search of real rigidities." *NBER Macroeconomics Annual* 25 (1): 261–310.

Gopinath, Gita, and Oleg Itskhoki. 2011b. "In Search of Real Rigidities." *NBER Macroeconomics*

Annual 25: 261–310.

Gopinath, Gita, and Roberto Rigobon. 2008. "Sticky borders." *The Quarterly Journal of Economics* 123 (2): 531–575.

Hamilton, James D. 2018. "Why you should never use the Hodrick-Prescott filter." *Review of Economics and Statistics* 100 (5): 831–843.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson. 2022. "The slope of the Phillips Curve: evidence from US states." *The Quarterly Journal of Economics* 137 (3): 1299–1344.

Hsieh, Chang-Tai, and Peter J Klenow. 2009. "Misallocation and manufacturing TFP in China and India." *The Quarterly Journal of Economics* 124 (4): 1403–1448.

Martin, Leslie A., Shanthi Nataraj, and Ann E. Harrison. 2017. "In with the Big, Out with the Small: Removing Small-Scale Reservations in India." *American Economic Review* 107 (2): 354–86. https://www.aeaweb.org/articles?id=10.1257/aer.20141335.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H Stock. 2014. "Empirical evidence on inflation expectations in the New Keynesian Phillips Curve." *American Economic Journal: Journal of Economic Literature* 52 (1): 124–188.

Mohanty, MS, Marc Klau et al. 2001. "What determines inflation in emerging market economies?" *BIS Papers* 8: 1–38.

Mongey, Simon. 2021. "Market structure and monetary non-neutrality.", National Bureau of Economic Research.

Nakamura, Emi, and Jón Steinsson. 2008. "Five facts about prices: A reevaluation of menu cost models." *The Quarterly Journal of Economics* 123 (4): 1415–1464.

Nakamura, Emi, Jón Steinsson, Patrick Sun, and Daniel Villar. 2018. "The elusive costs of inflation: Price dispersion during the US great inflation." *The Quarterly Journal of Economics* 133 (4): 1933–1980.

Reinelt, Timo, and Matthias Meier. 2020. "Monetary policy, markup dispersion, and aggregate TFP."

Rijesh, R. 2022. "Concordance of Manufacturing Industries Using NIC-2008 and NIC-2004: A Disaggregate Level Study.". https://doi.org/10.21203/rs.3.rs-2055633/v1.

Shea, John. 1993. "Do supply curves slope up?" *The Quarterly Journal of Economics* 108 (1): 1–32.

Singer, Gregor. 2019. "Endogenous markups, input misallocation and geographical supplier access."."

Stock, James H, and Mark W Watson. 2020. "Slack and cyclically sensitive inflation." *Journal of Money, Credit and Banking* 52 (S2): 393–428.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*.

Appendix for online publication

Appendix A. Model derivations

A.1. Derivations of baseline model

Households. Households choose consumption C and labor L to maximize discounted future utility $\mathbb{E}_0 \sum_{t=0}^{+\infty} \beta^t u(C_t, L_t)$ subject to a per-period budget constraint $P_t^Y C_t + Q_t B_t = B_{t-1} + w_t L_t + T_t$ where P_t^Y is the ideal price index of the consumption bundle, B_t is holdings of one-period risk-free nominal bonds with price Q_t, w_t is the wage, and T_t denotes any profits rebated to households as lump-sum. From the households' optimization problem we obtain the Euler equation:

$$\frac{1}{1+i_t} = \beta \mathbb{E}_t [\frac{u_c(C_{t+1}, L_{t+1})}{u_c(C_t, L_t)} \frac{P_t^Y}{P_{t+1}^Y}] \tag{A.1}$$

and the labor supply function:

$$\frac{u_l(C_t, L_t)}{u_c(C_t, L_t)} = \frac{w_t^l}{P_t^Y} \tag{A.2}$$

Final good producers. Let Y_t denote aggregate production of the final good. This can be used for consumption C_t, so that $Y_t = C_t$. The final good Y_t is produced by perfectly competitive firms using a bundle of differentiated intermediate inputs y_{it} for $i \in [0, 1]$. Taking the prices p_{it} of the inputs as given and denoting the price of the final good P_t^Y, final good producers choose y_{it} to maximize profits. This gives rise to the demand function:

$$y_{it} = \mathcal{D}(p_{it}/\mathcal{P}_t) Y_t. \tag{A.3}$$

The price elasticity of demand is given by:

$$\theta_{it} = \theta(\frac{y_{it}}{Y_t}) = -\frac{\partial \log y_{it}}{\partial \log p_{it}}, \tag{A.4}$$

where $\mathcal{P}_t$ is the substitution-relevant price index.

Proposition 1. *The price elasticity of demand θ_{it} is a function of relative quantities only*

PROOF. We use the chain rule to express the demand elasticity as follows

$$\theta_{it} = -\frac{\partial \log y_{it}}{\partial \log p_{it}} = -\frac{\partial \log y_{it}}{\partial y_{it}} \frac{Y_t \mathcal{D}'(p_{it}/\mathcal{P})}{P_t} \frac{\partial p_{it}}{\partial \log p_{it}} \tag{A.5}$$

$$= -\frac{1}{y_{it}} \frac{Y_t}{\mathcal{P}_t} \mathcal{D}'(p_{it}/\mathcal{P}) p_{it} \tag{A.6}$$

$$= -\frac{p_{it}}{\mathcal{P}_t} \frac{\mathcal{D}'(p_{it}/\mathcal{P})}{\mathcal{D}(p_{it}/\mathcal{P})}. \tag{A.7}$$

Since $\mathcal{D}$ is an invertible function, then

$$\frac{p_{it}}{\mathcal{P}_t} = \mathcal{D}^{-1}(y_{it/Y_t}) \tag{A.8}$$

Therefore

$$\theta_{it} = \theta\left(\frac{y_{it}}{Y_t}\right) - \frac{\mathcal{D}^{-1}(y_{it/Y_t})\left(\mathcal{D}' \circ \mathcal{D}^{-1}(y_{it/Y_t})\right)}{y_{it}/Y_t} \tag{A.9}$$

concluding θ_{it} is only a function of relative quantities.

□

Differentiated varieties producers. Each variety i is produced by a single firm. Firms produce with the following technology:

$$y_{it} = e^{z_i} l_{it}^a \tag{A.10}$$

The cost function writes:

$$\mathcal{C}(y_{it}, w_{it}^v, z_i) = w_{it}^v \left(\frac{y_{it}}{e^{z_i}}\right)^{\frac{1}{a}} \tag{A.11}$$

w_{it}^v denotes the price index of variable inputs. In the model with only labor, w_{it}^v is equal to the wage paid by firm i w_{it}^l. We allow for the price of inputs to be firm-specific, in the spirit of (Woodford 2003). This may be due to, for example, household preferences over amenities, or any other microfoundation that introduces firm-specific input markets. We adopt the following functional form:

$$w^v_{it} = w^v_t \left(\frac{l_{it}}{L_t}\right)^{a_w} \tag{A.12}$$

The textbook New Keynesian model sets $a = 1, a_w = 0$. Extensions with decreasing returns to scale set $a < 1$, $a_w = 0$, while models with firm-specific input markets allow for $a_w > 0$.

Therefore, the marginal cost function takes the shape

$$mc_{it} = \frac{a_w + 1}{a} w^v_t y_{it}^{\frac{1+a_w-a}{a}} e^{-\frac{1+a_w}{a(1-\gamma)} z_i} \frac{1}{L_t^{a_w}}.$$

A firm has a probability $1 - \alpha$ of being able to reset its price in each period. We denote $x_{it+s|t}$ the $t + s$ value of variable x for a firm that could last reset its price at time t. A firm that can reset its price maximizes choses the price that maximizes:

$$\max_{p_{it|t}} \mathbb{E}_t \left[\sum_{s=0}^{+\infty} \alpha^s \Lambda_{t,t+s} \left[p_{it|t} y_{it+s|t} - \mathcal{C}(y_{it+s|t}, w^v_{it+s|t}, z_i)\right]\right]$$

subject to the demand curve $y_{it+s|t} = \mathcal{D}\left(p_{it|t}/\mathcal{P}_{t+s}\right) Y_{t+s}$ and the cost function $\mathcal{C}(y_{it+s|t}, w^v_{it+s|t}, z_i) = w^v_{it+s|t} e^{-\frac{1}{a} z_i} y_{it+s|t}^{\frac{1}{a}}$. $\Lambda_{t,t+s}$ is the stochastic discount factor.

It will be convenient to define the following quantities. $\mu^f_{it} = \frac{\theta_{it}}{\theta_{it}-1}$ is the desired markup that the firm would choose in a flexible price environment. $\Gamma_{it} = \dfrac{\partial \log \mu^f_{it}}{\partial \log \frac{y_{it}}{Y_t}}$ is the elasticity of the flexible price markup with respect to relative size. ρ_{it} is the partial equilibrium pass-through of a marginal cost shock into the firm's price.

$$d \log p_{it} = d \log \mu^f_{it} + d \log mc_{it} = \Gamma_{it}\, d \log \frac{y_{it}}{Y_t} + d \log mc_{it} = -\Gamma_{it}\theta_{it} \left(d \log p_{it} - d \log \mathcal{P}_t\right) + d \log mc_{it}$$

$$d \log p_{it} = \frac{1}{1 + \Gamma_{it}\theta_{it}} d \log mc_{it} + \frac{\Gamma_{it}\theta_{it}}{1 + \Gamma_{it}\theta_{it}} d \log \mathcal{P}_t$$

Therefore, $\rho_{it} \equiv \dfrac{\partial \log p_{it}}{\partial \log mc_{it}} = \frac{1}{1+\Gamma_{it}\theta_{it}}$. In the CES case, all firms face the same demand elasticities, $\Gamma_{it} = 0\ \forall i$, and $\rho_{it} = 1\ \forall i$. Away from the CES case, ρ_{it} can be below or above 1 depending on the sign of Γ_{it}.

Note that μ^f_{it}, Γ_{it}, and ρ_{it} are only a function of a firm's relative size $\frac{y_{it}}{Y_t}$.

Finally, in a sticky price environment, the actual markup of the firm may differ from the flexible price desired markup. We denote the actual markup of the firm: $\mu_{it} = \frac{p_{it}}{mc_{it}}$.

Monetary authority. Monetary policy sets the nominal interest rate according to a Taylor rule.

Equilibrium. Equilibrium is defined by the following conditions: (i) Consumers choose consumption and labor to maximize utility taking prices and wage as given; (ii) Firms with flexible prices set prices to maximize their value taking the price index and their residual demand curves as given; firms with sticky prices meet demand at fixed prices ; (iii) Monetary policy sets the nominal interest rate; (iv) All resource constraints are satisfied.

We solve the model by log-linearization around a symmetric zero-inflation steady state. We take a first-order expansion for small monetary policy shock. Quantities without a t subscript refer to the steady-state. We assume that the aggregator Y_t is such that $\hat{Y}_t = \int_0^1 \hat{y}_{it} di$. Given this property of the aggregator, and the shape of the demand curves in A.3, yield the result that $\hat{\mathcal{P}}_t = \hat{P}^Y_t$.

A.1.1. Characterization

Marginal cost-based Phillips curve. Taking first order conditions of the objective function and log-linearizing around a zero inflation symmetric steady state we find that a firm that can reset its price at time t will choose:

$$\hat{p}_{it|t} = (1 - \beta\alpha)\mathbb{E}_t \left[\sum_{s=0}^{+\infty} (\beta\alpha)^s (\hat{\mu}^f_{it+s|t} + \hat{mc}_{it+s|t}) \right] \tag{A.13}$$

$\hat{\mu}^f_{it+s|t}$ is the log-deviation of the flexible price markup at $t+s$ of a firm that could last reset its price at t. It is given by:

$$\hat{\mu}^f_{it+s|t} = -\Gamma\theta(\hat{p}_{it|t} - \hat{\mathcal{P}}_{t+s}), \tag{A.14}$$

where Γ and θ are the markup and demand elasticities evaluated at the symmetric steady state. Let $\hat{mc}_t \equiv \mathbb{E}[\hat{mc}_{it}]$ the change in the aggregate nominal marginal cost. We can write:

$$\hat{mc}_{it+s|t} - \hat{mc}_{t+s} = -d_{mc,y}\theta(\hat{p}_{it|t} - \hat{\mathcal{P}}_{t+s}), \tag{A.15}$$

where $d_{mc,y} = \frac{1-a+a_w}{a}$ is the elasticity of firm-level marginal costs with respect to firm-level quantities, where a is the extent of decreasing returns to scale in production, and a_w is the elasticity of unit input costs.

We obtain:

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\zeta\rho\hat{mc}_{t+s} + (1-\zeta\rho)\hat{\mathcal{P}}_{t+s}\right)\right] \tag{A.16}$$

where ρ is the flexible price partial equilibrium pass-through of marginal cost shocks into prices and $\zeta = \frac{1}{1+d_{mc,y}\theta\rho}$, both evaluated at the symmetric steady state. ζ captures the fact that when marginal cost curves slope upward, a cost shock induces an adjustment in size, which dampens the first-round effect on marginal cost. $\omega = \zeta\rho$ combines these two terms and captures the flexible price pass-through of an input cost shock into prices.

We can write this equation recursively as:

$$\hat{p}_{it|t} = (1-\beta\alpha)\left(\omega\hat{mc}_t + (1-\omega)\hat{\mathcal{P}}_t\right) + \beta\alpha\mathbb{E}_t[\hat{p}_{it+1|t+1}] \tag{A.17}$$

Inflation dynamics. The change in the price of the final good is given by:

$$\hat{\mathcal{P}}_t = \mathbb{E}[\hat{p}_{it}] \tag{A.18}$$

Let $\mathbb{1}^p_{it}$ be a dummy equal to 1 if firm i can reset their price at t.

$$\hat{p}_{it} = (1-\mathbb{1}^p_{it})\,\hat{p}_{it-1} + \mathbb{1}^p_{it}\,\hat{p}_{it|t} \tag{A.19}$$

Therefore,

$$\hat{\mathcal{P}}_t = \mathbb{E}[(1-\mathbb{1}^p_{it})\hat{p}_{it-1}] + \mathbb{E}[\mathbb{1}^p_{it}\hat{p}_{it|t}] \tag{A.20}$$

The Calvo fairy is orthogonal to the firm's steady state sales share so that

$$\hat{\mathcal{P}}_t = \alpha\hat{\mathcal{P}}_{t-1} + (1-\alpha)\mathbb{E}[\hat{p}_{it|t}] \tag{A.21}$$

By definition,

$$\hat{\pi}_t = \hat{\mathcal{P}}_t - \hat{\mathcal{P}}_{t-1} \tag{A.22}$$

Aggregating across firms, we obtain the marginal cost-based Phillips curve:

$$\hat{\pi}_t = \varphi\omega(\hat{mc}_t - \hat{\mathcal{P}}_t) + \beta\mathbb{E}_t[\hat{\pi}_{t+1}] \tag{A.23}$$

$\hat{mc}_t - \hat{\mathcal{P}}_t$ is the change in the aggregate real marginal cost. $\varphi = \dfrac{(1-\alpha)(1-\beta\alpha)}{\alpha}$ is the slope of the marginal cost-based Phillips curve in the case of constant returns to scale and CES demand. $\omega = \zeta\rho$ reflects micro-level real rigidities due to decreasing returns to scale ζ and strategic complementarities ρ.

Aggregate marginal costs. Let us define aggregate productivity Z_t as satisfying:

$$Y_t \equiv Z_t L_t^a \tag{A.24}$$

To the first order,

$$\hat{Y}_t = a\hat{L}_t \tag{A.25}$$

where $\hat{Z}_t = 0$ follows from the steady state being efficient (across firms).

The log-linearized labor supply function is derived from the utility maximization problem of the household:

$$\nu^{-1}\hat{L}_t + (\sigma^{-1} - \iota_{lc})\hat{C}_t = \hat{w}_t^{\upsilon} - \hat{\mathcal{P}}_t, \tag{A.26}$$

where $\sigma^{-1} = -\frac{u_{cc}C}{u_c}$ is the inverse elasticity of intertemporal substitution, $\nu^{-1} = \frac{u_{ll}L}{u_l} - \frac{u_{cl}L}{u_c}$ is the inverse Frisch elasticity of labor supply, and $\iota_{lc} \equiv -\frac{u_{lc}C}{u_l}$. We define $\upsilon \equiv (\sigma^{-1} - \iota_{lc})$.

Substituting $\hat{L}_t$ and using $\hat{C}_t = \hat{Y}_t$, we obtain equilibrium prices for labor:

$$\hat{w}_t^v = \frac{\nu^{-1}}{a}\hat{Y}_t + (\sigma^{-1} - \iota_{lc})\hat{Y}_t + \hat{\mathcal{P}}_t. \tag{A.27}$$

Aggregate steady-state marginal costs are equal to

$$\hat{mc}_t = \mathbb{E}\left[\hat{mc}_{it}\right] = \hat{w}_t^v + \frac{1-a}{a}\hat{Y}_t. \tag{A.28}$$

Together with equilibrium prices from equation (A.27), we can then derive the following solution for aggregate marginal cost:

$$\hat{mc}_t = \underbrace{\left[\frac{1-a+\nu^{-1}}{a} + \upsilon\right]}_{\Omega\,=\,\text{Elasticity of mc wrt output}} \hat{Y}_t + \hat{\mathcal{P}}_t. \tag{A.29}$$

Output-based Phillips curve. Combining (A.23) and (A.29), we obtain the output-based New Keynesian Phillips curve:

$$\hat{\pi}_t = \varphi\omega\Omega\hat{Y}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}] \tag{A.30}$$

Euler equation. From the utility maximization problem of households, we obtain the generic (log-linearized) Euler equation as:

$$\hat{C}_t - \sigma\iota_{cl}\hat{L}_t = \mathbb{E}_t\left[\hat{C}_{t+1} - \sigma\iota_{cl}\hat{L}_{t+1}\right] - \sigma\left(\hat{i}_t - \mathbb{E}_t\left[\hat{\pi}_{t+1}\right]\right) \tag{A.31}$$

where $\iota_{cl} = \frac{u_{cl}L}{u_c}$. Knowing the equilibrium input price from equation (A.27), and using $\hat{C}_t = \hat{Y}_t$, we can derive the Euler equation as:

$$\left(1 - \frac{\sigma\iota_{cl}}{a}\right)\hat{Y}_t = \mathbb{E}_t\left[\left(1 - \frac{\sigma\iota_{cl}}{a}\right)\hat{Y}_{t+1}\right] - \sigma\left(\hat{i}_t - \mathbb{E}_t\left[\hat{\pi}_{t+1}\right]\right). \tag{A.32}$$

Three-equations New Keynesian model. We obtain a three-equation version of the New Keynesian model:

$$\hat{\pi}_t = \varphi\omega\Omega\hat{Y}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}] \tag{A.33}$$

$$c\hat{Y}_t = c\mathbb{E}_t\left[\hat{Y}_{t+1}\right] - \sigma\left(\hat{i}_t - \mathbb{E}_t\left[\hat{\pi}_{t+1}\right]\right) \tag{A.34}$$

$$\hat{i}_t = \phi_\pi \hat{\pi}_t + \phi_y \hat{Y}_t + \varepsilon_t^{MP} \tag{A.35}$$

A.2. Special Cases : Kimball and Atkeson-Burstein

In this section, we present two special cases of the environment presented in the main body: monopolistic competition with Kimball demand, and oligopolistic competition as in Atkeson and Burstein (2008). We restrict the exposition to the description of the demand system and the problem of intermediate variety producers, since the rest follows the previous section.

A.2.1. Kimball preferences

Final good producers. The final good Y_t is produced by a perfectly competitive firm using a bundle of differentiated intermediate inputs y_{it} for $i \in [0,1]$. Intermediate input varieties are assembled into the final good using the Kimball aggregator:

$$\int_0^1 \Upsilon\left(\frac{y_{it}}{Y_t}\right) di = 1$$

where the function $\Upsilon(.)$ is strictly increasing, strictly concave, and satisfies $\Upsilon(1) = 1$. The CES aggregator is the special case $\Upsilon(q) = q^{\frac{\theta-1}{\theta}}$ for $\theta > 1$. Taking the prices p_{it} of the inputs as given and denoting the price of the final good P_t^Y, the final good producer chooses y_{it} to maximize profits. This gives rise to the demand function:

$$\frac{y_{it}}{Y_t} = \Upsilon'^{-1}\left(\frac{p_{it}}{\mathcal{P}_t}\right) \tag{A.36}$$

where $\frac{p_{it}}{\mathcal{P}_t}$ determines substitution across varieties. The price index $\mathcal{P}_t$ is given by $\mathcal{P}_t = \frac{P_t^Y}{D_t}$. $P_t^Y = \int_0^1 p_{it} \frac{y_{it}}{Y_t} di$ is the ideal price index. $D_t = \int_0^1 \Upsilon'(\frac{y_{it}}{Y_t}) \frac{y_{it}}{Y_t} di$ is a "demand" index. When demand is CES, D_t is a constant equal to $\frac{\theta}{\theta-1}$. The price elasticity of demand is only a function of firm relative size and is given by:

$$\theta_{it} = \theta\left(\frac{y_{it}}{Y_t}\right) = -\frac{\partial \log y_{it}}{\partial \log p_{it}} = \frac{\Upsilon'\left(\frac{y_{it}}{Y_t}\right)}{-\frac{y_{it}}{Y_t}\Upsilon''\left(\frac{y_{it}}{Y_t}\right)} \tag{A.37}$$

Then, the desired markup $\mu^f_{it} = \frac{\theta_{it}}{\theta_{it}-1}$, the markup elasticity $\Gamma_{it} = \frac{\partial \log \mu^f_{it}}{\partial \log y_{it}}$, and the passthrough $\rho_{it} = \frac{\partial \log p_{it}}{\partial \log mc_{it}}$ only depend on the firm's relative size.

Differentiated varieties producers. As before, a firm that can reset its price chooses the price that maximizes:

$$\max_{p_{it|t}} \mathbb{E}_t \left[\sum_{s=0}^{+\infty} \alpha^s \Lambda_{t,t+s} \left[p_{it|t} y_{it+s|t} - \mathcal{C}(y_{it+s|t}, w^v_{it+s|t}, z_i) \right] \right]$$

subject to the demand curve $y_{it+s} = \Upsilon'^{-1}\left(\frac{p_{it}}{\mathcal{P}_{t+s}}\right) Y_{t+s}$ and the cost function in (A.11).

Characterization of the marginal cost-based Phillips curve. Solving the model by log-linearizing around the symmetric steady state, and following the same derivation steps as above, we obtain the result that a firm that can reset its price at time t will choose:

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t \left[\sum_{s=0}^{+\infty} (\beta\alpha)^s \left(\omega \hat{mc}_{t+s} + (1-\omega)\hat{\mathcal{P}}_{t+s} \right) \right] \tag{A.38}$$

where $\zeta = \frac{1}{1+d_{mc,y}\theta\rho}$. This expression uses the fact that at the symmetric steady state, all firms have the same relative size, and hence θ, ρ and ζ are constant for every firm.

Aggregating across firms, and using the notation $\omega = \zeta\rho$ to summarize micro real rigidities, we obtain the marginal cost-based Phillips curve:

$$\hat{\pi}_t = \varphi\omega(\hat{mc}_t - \hat{P}^Y_t) + \beta \mathbb{E}_t[\hat{\pi}_{t+1}] \tag{A.39}$$

$\hat{mc}_t - \hat{P}^Y_t$ is the change in the aggregate real marginal cost. To derive this last expression, we used the fact that $\hat{P}^Y_t - \hat{\mathcal{P}}_t = \hat{D}_t = 0$ around the symmetric steady state. Indeed, log-linearizing the expression for the demand index yields:

$$\hat{D}_t = -\text{Cov}\left[\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}, \hat{p}_{it} \right] \tag{A.40}$$

In the symmetric steady state, $\theta_i = \theta$ for all i and $\hat{D}_t = 0$.

This expression for the marginal cost-based Phillips curve is the same as (A.23). The rest of the derivations follows and we obtain the same 3-equations New Keynesian model in (A.33)-(A.35).

A.2.2. Atkeson Burstein Oligopolistic Competition

The final goods producer assembles the final good using a CES aggregator across a continuum of sectors, indexed by j. Each sectoral bundle is formed by a CES aggregator of the varieties produced by N distinct firms, indexed by i. N is the same across sectors. The elasticities of substitution across sectors, and across varieties within a sector are denoted by ϕ, and η, respectively.

As in the general model, firms are ex-ante homogeneous, hence produce the same level of output and charge the same relative price in the flexible price steady state.

The assumptions stated before imply that:

$$Y_t = \left(\int_0^1 Y_{jt}^{\frac{\phi-1}{\phi}} ds \right)^{\frac{\phi}{\phi-1}}, \tag{A.41}$$

$$Y_{jt} = \left(\sum_{i=1}^{N} y_{ijt}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}. \tag{A.42}$$

This preference structure gives rise to a firm-level demand curve of the form:

$$y_{ijt} = Y_{jt} \left(\frac{p_{ijt}}{\mathcal{P}_{jt}} \right)^{-\eta}, \tag{A.43}$$

$$Y_{jt} = Y_t \left(\frac{\mathcal{P}_{jt}}{\mathcal{P}_t} \right)^{-\phi}. \tag{A.44}$$

Combining these two layers of demand we find that:

$$y_{ijt} = \left(\frac{p_{ijt}}{\mathcal{P}_{jt}} \right)^{-\eta} \left(\frac{\mathcal{P}_{jt}}{\mathcal{P}_t} \right)^{-\phi} Y_t. \tag{A.45}$$

As opposed to the textbook case with monopolistic competition, individual firm pricing has a non-negligible effect on industry prices and quantities, which is easier to observe through the ideal sectoral price index,

$$\mathcal{P}_{jt} = \left(\sum_{i=1}^{N} p_{ijt}^{1-\eta} \right)^{\frac{1}{1-\eta}}. \tag{A.46}$$

The price elasticity of demand is only a function of firm relative size and is given by:

$$\theta_{ijt} = \theta \left(\frac{y_{ijt}}{Y_{jt}} \right) = -\frac{\partial \log y_{ijt}}{\partial \log p_{ijt}} \tag{A.47}$$

$$\theta_{ijt} = \eta - (\eta - \phi) \left(\frac{p_{ijt}}{\mathcal{P}_{jt}} \right)^{1-\eta} \tag{A.48}$$

$$\theta_{ijt} = \eta - (\eta - \phi) \left(\frac{y_{ijt}}{Y_{jt}} \right)^{\frac{\eta-1}{\eta}}. \tag{A.49}$$

It is worth noting that in a symmetric steady-state equilibrium the elasticity of demand is equal to:

$$\theta = \eta - (\eta - \phi) \frac{1}{N}, \tag{A.50}$$

and that in general, the elasticity of demand incorporates the effect that the firm has, and understand it has, on sectoral aggregates.

Differentiated varieties producers. A firm i in sector j that can reset its price chooses the price that maximizes:

$$\max_{p_{ijt|t}} \mathbb{E}_t \left[\sum_{s=0}^{+\infty} \alpha^s \Lambda_{t,t+s} \left[p_{ijt|t} y_{ijt+s|t} - \mathcal{C}(y_{it+s|t}, w^v_{ijt+s|t}, z_i) \right] \right]$$

subject to the demand curve $y_{ijt+s|t} = p_{ijt|t}^{-\eta} \left(\frac{1}{N} \sum_{k=1}^{N} p_{kjt+s}^{1-\eta} \right)^{\frac{\eta-\phi}{1-\eta}} \mathcal{P}_{t+s}^{-\phi} Y_{t+s}$ and the cost function in equation (A.11). It is implicit in the demand curve that $p_{kjt+s} = p_{ijt|t}$ for $i = k$ where i is an arbitrary firm from sector j that gets to reset its price in period t.

Characterization of the marginal cost-based Phillips curve. We again solve the model by log-linearization around the zero-inflation symmetric steady state. An arbitrary firm i

in sector j that resets its price at time t will choose:

$$\hat{p}_{ijt|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\hat{mc}_{ijt+s|t} + \hat{\mu}^f_{ijt+s|t}\right)\right]. \tag{A.51}$$

A first order approximation of desired markups around a symmetric steady state (within and across sectors) yields $\hat{\mu}^f_{ijt+s|t} = -\Gamma\theta(\hat{p}_{ijt|t} - \hat{\mathcal{P}}_{jt+s})$.

$$\hat{p}_{ijt|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\omega\hat{mc}_{jt+s} + (1-\omega)\hat{\mathcal{P}}_{jt+s}\right)\right] \tag{A.52}$$

where $\hat{mc}_{jt}$ is the change in the aggregate nominal marginal cost in sector j and $\zeta = \frac{1}{1+d_{mc,y}\theta\rho}$. Aggregating across firms within a sector, and using the notation $\omega = \zeta\rho$ to summarize micro real rigidities, we obtain the marginal cost-based Phillips curve:

$$\hat{\pi}_{jt} = \varphi\omega(\hat{mc}_{jt} - \hat{\mathcal{P}}_{jt}) + \beta\mathbb{E}_t[\hat{\pi}_{j,t+1}], \tag{A.53}$$

where $\hat{mc}_{jt} - \hat{\mathcal{P}}_{jt}$ is the log-deviation in the sectoral real marginal cost. Further aggregating over a continuum of equally-sized sectors yields

$$\hat{\pi}_t = \varphi\omega(\hat{mc}_t - \hat{\mathcal{P}}_t) + \beta\mathbb{E}_t[\hat{\pi}_{t+1}], \tag{A.54}$$

where an aggregate variable $\hat{x}_t = \int_0^1 \hat{x}_{jt}dj$.

This expression for the marginal cost-based Phillips curve is the same as (A.23). The rest of the derivations follows and we obtain the same 3-equations New Keynesian model in (A.33)-(A.35).

A.3. Extended model and identification arguments

To consider the mapping between the model and the data, we make three modifications to the baseline model. First, we introduce materials to the production function of intermediate producers with firm-specific shocks to the cost of those materials. Second, we introduce firm-specific demand shocks. Third, we introduce capital as a fixed factor in the production function.

To help the reader move between the main text and this appendix, after the main results, we present simplified expressions for the special case of the model without capital and without intermediates when the cost shock is a cost shock to the only input of production.

A.3.1. Additional assumptions

Firm-specific demand shocks. We use the specification of the demand curve similar in spirit to that of Edmond, Midrigan, and Xu (2023) :

$$\frac{y_{it}}{Y_t} = \mathcal{D}\left(\frac{p_{it}}{\xi_{it} P_t}\right) \tag{A.55}$$

in the particular case of Kimball demand $\mathcal{D} = \Upsilon^{'-1}$.

Intermediate varieties producers.

$$y_{it} = e^{z_i}(k_i^{\gamma} v_{it}^{1-\gamma})^a, \tag{A.56}$$

where

$$v_{it} = l_{it}^{\phi} x_{it}^{1-\phi}. \tag{A.57}$$

We consider firm-specific shocks to the price of inputs, akin to iceberg transportation cost shocks, denoted $e^{\vartheta_{it}^l}$ and $e^{\vartheta_{it}^x}$. ϑ_{it}^l and ϑ_{it}^x are mean-zero idiosyncratic shocks. The model in the main text is the special case of this model where $\gamma = 0$, and $\phi = 1$.

Materials producers. Materials are produced by a perfectly competitive representative firm with the following production function: $X_t = (Y_t^X)^{\frac{\eta}{1+\eta}}$ where Y_t^X are units of the final good used for intermediate inputs production and $\eta \geq 0$. This implies that the price of the material good is given by: $w_t^x = P_t^Y(1+\eta^{-1})X_t^{\eta^{-1}}$. Y_t can be used for consumption C_t or as an input to produce materials Y_t^X so that $Y_t = C_t + Y_t^X$.

A.3.2. Characterization

Final good producers. We again denote the price elasticity of demand by θ_{it}, a function of relative size only.

$$\hat{y}_{it} - \hat{Y}_t = -\theta_i(\hat{p}_{it} - \hat{P}_t) + \frac{\theta_i}{Y}\hat{\xi}_{it} \tag{A.58}$$

Intermediate varieties producers. We first consider the firm's cost-minimization problem, taking output and input prices as given. We can first solve for the optimal choice of labor and materials, taking variable inputs and prices as given:

$$\min_{l_{it},x_{it}} [e^{\vartheta^x_{it}} w^x_{it} x_{it} + e^{\vartheta^l_{it}} w^l_{it} l_{it}] \quad \text{subject to} \quad (l^{\phi}_{it} x^{1-\phi}_{it}) \geq \bar{v}, \tag{A.59}$$

where

$$w^x_{it} = w^x_t \left(\frac{x_{it}}{X_t}\right)^{a_w} \tag{A.60}$$

$$w^l_{it} = w^l_t \left(\frac{l_{it}}{L_t}\right)^{a_w}. \tag{A.61}$$

We obtain that for any v_{it}, the choice of labor and material will be given by:

$$l_{it} = \left(\frac{w^x_{it} e^{\vartheta^x_{it}}}{w^l_{it} e^{\vartheta^l_{it}}} \frac{\phi}{1-\phi}\right)^{1-\phi} v_{it} \tag{A.62}$$

$$x_{it} = \left(\frac{w^x_{it} e^{\vartheta^x_{it}}}{w^l_{it} e^{\vartheta^l_{it}}} \frac{\phi}{1-\phi}\right)^{-\phi} v_{it} \tag{A.63}$$

This defines a price index for variable inputs, inclusive of the iceberg cost shock:

$$w^v_{it} = \left(\frac{w^l_{it} e^{\vartheta^l_{it}}}{\phi}\right)^{\phi} \left(\frac{w^x_{it} e^{\vartheta^x_{it}}}{1-\phi}\right)^{1-\phi} \tag{A.64}$$

$$= w^v_t e^{(\phi \vartheta^l_{it} + (1-\phi)\vartheta^x_{it})} \left(\frac{v_{it}}{V_t}\right)^{a_w} \tag{A.65}$$

$$= w^v_t e^{\vartheta_{it}} \left(\frac{v_{it}}{V_t}\right)^{a_w}, \tag{A.66}$$

where $\vartheta_{it} \equiv \phi\vartheta^l_{it} + (1-\phi)\vartheta^x_{it}$ is the effective iceberg cost faced by firm i, and $V_t = L_t^{\phi} X_t^{1-\phi}$. The special case in the main text sets $\phi = 1$.

We then solve for the optimal choice of variable inputs, for a given level of output:

$$\min_{v_{it}} (w^v_{it} v_{it} + w^k_t k_i) \quad \text{subject to} \quad e^{z_i}(k_i^{\gamma} v_{it}^{1-\gamma})^a \geq \bar{y} \tag{A.67}$$

Using A.64 and the production function, we can write the total cost function as

$$\mathcal{C}(y_{it}, w^v_{it}, w^k_{it}, z_i) = w^v_t e^{\vartheta_{it}} y_{it}^{\frac{1+a_w}{a(1-\gamma)}} \left(e^{z_i} k_i^{a\gamma}\right)^{-\frac{1+a_w}{a(1-\gamma)}} \frac{1}{V_t^{a_w}} + w^k_{it} k_i. \tag{A.68}$$

The marginal cost function is then given by

$$mc_{it} = \frac{a_w + 1}{a(1-\gamma)} w^v_t e^{\vartheta_{it}} y_{it}^{\frac{1+a_w-(1-\gamma)a}{a(1-\gamma)}} \left(e^{z_i} k_i^{a\gamma}\right)^{-\frac{1+a_w}{a(1-\gamma)}} \frac{1}{V_t^{a_w}}.$$

Marginal cost-based Phillips curve. We log-linearize around a symmetric steady state with zero inflation. A firm that can reset its price at time t will choose:

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s(\hat{\mu}^f_{it+s|t} + \hat{mc}_{it+s|t})\right] \tag{A.69}$$

$\hat{\mu}^f_{it+s|t}$ is the optimal markup at $t+s$ of a firm that could last reset its price at t. It is given by:

$$\hat{\mu}^f_{it+s|t} = -\Gamma\theta(\hat{p}_{it|t} - \hat{\mathcal{P}}_{t+s}) + \Gamma\frac{\theta}{Y}\hat{\xi}_{it+s|t}, \tag{A.70}$$

where Γ and θ are common across firms due to our assumption of a symmetric steady state.

Let $\hat{mc}_t \equiv \mathbb{E}[\hat{mc}_{it}]$ the change in the aggregate nominal marginal cost, and $d_{mc,y} = \frac{1+a_w-(1-\gamma)a}{a(1-\gamma)}$ the elasticity of marginal costs with respect to firm scale. We can write:

$$\hat{mc}_{it+s|t} - \hat{mc}_{t+s} = \vartheta_{it+s} + d_{mc,y}(\hat{y}_{it+s|t} - \hat{Y}_{t+s}) \tag{A.71}$$

$$= \vartheta_{it+s} - d_{mc,y}\theta(\hat{p}_{it|t} - \hat{\mathcal{P}}_{t+s}) + d_{mc,y}\frac{\theta}{Y}\hat{\xi}_{it+s|t} \tag{A.72}$$

Then,

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\rho\zeta\hat{mc}_{t+s} + \rho\zeta\vartheta_{it+s} + (1-\rho\zeta)\hat{\mathcal{P}}_{t+s} + (1-\rho\zeta)\frac{\hat{\xi}_{it+s|t}}{Y}\right)\right] \tag{A.73}$$

where ρ is the flexible price partial equilibrium pass-through of marginal cost shocks into prices and $\zeta = \frac{1}{1+d_{mc,y}\theta\rho}$. We will use the notation $\omega = \zeta\rho$ to summarize micro real rigidities. Re-writing this equation recursively and aggregating across firms, we obtain the marginal cost-based Phillips curve:

$$\hat{\pi}_t = \varphi\omega(\hat{mc}_t - \hat{\mathcal{P}}_t) + \beta\mathbb{E}_t[\hat{\pi}_{t+1}] \tag{A.74}$$

Aggregate marginal costs. Let us define aggregate productivity Z_t as satisfying

$$Y_t \equiv Z_t\left(K^\gamma(L_t^\phi X_t^{1-\phi})^{1-\gamma}\right)^a \tag{A.75}$$

$$\hat{Y}_t = a(1-\gamma)\hat{V}_t = a(1-\gamma)(\phi\hat{L}_t + (1-\phi)\hat{X}_t) \tag{A.76}$$

From the cost-minimization problem of the producers of differentiated varieties, we obtain the log-linearized input demands:

$$\hat{l}_{it} = (1-\phi)(\hat{w}_t^x - \hat{w}_t^l) + (1+a_w)\hat{v}_{it} - a_w\hat{V}_t + (1-\phi)(\vartheta_{it}^x - \vartheta_{it}^l) \tag{A.77}$$

$$\hat{x}_{it} = -\phi(\hat{w}_t^x - \hat{w}_t^l) + (1+a_w)\hat{v}_{it} - a_w\hat{V}_t - \phi(\vartheta_{it}^x - \vartheta_{it}^l) \tag{A.78}$$

Using the first-order conditions of the cost-minimization, and aggregating across firms, we can derive aggregate input demand functions as:

$$\hat{L}_t = (1-\phi)(\hat{w}_t^x - \hat{w}_t^l) + \hat{V}_t = (1-\phi)\left(\hat{w}_t^x - \hat{w}_t^l\right) + \frac{1}{a(1-\gamma)}\hat{Y}_t, \tag{A.79}$$

$$\hat{X}_t = -\phi(\hat{w}_t^x - \hat{w}_t^l) + \hat{V}_t = -\phi\left(\hat{w}_t^x - \hat{w}_t^l\right) + \frac{1}{a(1-\gamma)}\hat{Y}_t. \tag{A.80}$$

The (log-linearized) labor supply function is derived from the utility maximization problem of the household:

$$\nu^{-1}\hat{L}_t + (\sigma^{-1} - \iota_{lc})\hat{C}_t = \hat{w}_t^l - \hat{\mathcal{P}}_t, \tag{A.81}$$

From the output maximization of the materials producer, we obtain the materials supply curve, again in log-linear form:

$$\eta^{-1}\hat{X}_t = \hat{w}_t^x - \hat{P}_t^Y \tag{A.82}$$

From the market clearing conditions, we obtain equilibrium prices for labor and materials as:

$$\hat{w}_t^l = \frac{\nu^{-1}(1+\eta^{-1})}{1+\eta^{-1}\phi+\nu^{-1}(1-\phi)}\frac{1}{a(1-\gamma)}\hat{Y}_t + \frac{(\sigma^{-1}-\iota_{lc})(1+\eta^{-1}\phi)}{1+\eta^{-1}\phi+\nu^{-1}(1-\phi)}\hat{C}_t + \hat{\mathcal{P}}_t, \tag{A.83}$$

$$\hat{w}_t^x = \frac{\eta^{-1}(1+\nu^{-1})}{1+\eta^{-1}\phi+\nu^{-1}(1-\phi)}\frac{1}{a(1-\gamma)}\hat{Y}_t + \frac{\eta^{-1}\phi(\sigma^{-1}-\iota_{lc})}{1+\eta^{-1}\phi+\nu^{-1}(1-\phi)}\hat{C}_t + \hat{\mathcal{P}}_t. \tag{A.84}$$

Aggregate marginal costs are equal to

$$\hat{mc}_t = \mathbb{E}\left[\hat{mc}_{it}\right] = \phi\hat{w}_t^l + (1-\phi)\hat{w}_t^x + \frac{1-a(1-\gamma)}{a(1-\gamma)}\hat{Y}_t. \tag{A.85}$$

From the goods market clearing condition we can derive

$$\hat{C}_t = \upsilon_0\left[1-(1-\lambda_c)\frac{1+\psi}{a(1-\gamma)}\right]\hat{Y}_t, \tag{A.86}$$

where λ_c denotes the share of the final good that is used for consumption in steady-state and where we defined $\upsilon_0 \equiv \frac{1+\eta^{-1}\phi+\nu^{-1}(1-\phi)}{\lambda_c(1+\eta^{-1}\phi+\nu^{-1}(1-\phi))+(1-\lambda_c)(1+\eta^{-1})\phi(\sigma^{-1}-\iota_{lc})}$. Together with equilibrium prices from equations (A.83) and (A.84), we can then derive the following solution for aggregate marginal cost:

$$\hat{mc}_t = \underbrace{\left[\frac{1-a(1-\gamma)+\psi}{a(1-\gamma)} + \upsilon\left(1-(1-\lambda_c)\frac{(1+\psi)}{a(1-\gamma)}\right)\right]}_{\Omega\,=\,\text{Elasticity of mc wrt output}}\hat{Y}_t + \hat{\mathcal{P}}_t. \tag{A.87}$$

$\psi = \frac{\phi\nu^{-1}+(1-\phi)\eta^{-1}+\nu^{-1}\eta^{-1}}{1+\nu^{-1}(1-\phi)+\phi\eta^{-1}}$ is the slope of the variable input supply curve, reflecting the slope of the labor supply curve ν^{-1} and the slope of the materials supply curve η^{-1}. We define $\upsilon \equiv \frac{(\sigma^{-1}-\iota_{lc})(1+\eta^{-1})\phi}{\lambda_c(1+\eta^{-1}\phi+\nu^{-1}(1-\phi))+(1-\lambda_c)(1+\eta^{-1})\phi(\sigma^{-1}-\iota_{lc})}$. We then obtain the output-based Phillips curve:

$$\hat{\pi}_t = \varphi\omega\Omega\hat{Y}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}], \tag{A.88}$$

A.3.3. Identification results: slope

Pass-through of cost shock into prices. Take the case of a shock with zero persistence. First, write $\hat{p}_{it} = \mathbb{1}^p_{it}\hat{p}_{it|t} + (1-\mathbb{1}^p_{it})\hat{p}_{it-1}$, where $\mathbb{1}^p_{it}$ is a "Calvo-fairy" dummy that takes the value of 1 if a price adjustment is permissible for firm i in period t. Using the formula for the optimal reset price, and using $\omega = \rho\zeta$,

$$\begin{aligned}\hat{p}_{it} - \hat{p}_{it-1} &= \mathbb{1}^p_{it}(1-\beta\alpha)\omega\vartheta_{it} + \mathbb{1}^p_{it}(1-\beta\alpha)(1-\omega)\frac{1}{Y}\hat{\xi}_{it} \\ &\quad + \mathbb{1}^p_{it}(1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s\left(\omega\hat{mc}_{t+s} + (1-\omega)\hat{\mathcal{P}}_{t+s}\right)\right] - \mathbb{1}^p_{it}\hat{p}_{it-1}\end{aligned} \tag{A.89}$$

Proposition 2. *Let us assume that $\mathbb{1}^p_{it} \perp \vartheta_{it}$, and $\vartheta_{it} \perp \hat{\xi}_{it}$, $\vartheta_{it} \perp \hat{p}_{it-1}$. Finally, assume that we observe $\mathcal{Z}^\vartheta_{it}$ a proxy for ϑ_{it} satisfying $\vartheta_{it} = k^\vartheta\mathcal{Z}^\vartheta_{it}$. Then, the coefficient of the regression*

$$\Delta\log p_{it} = \alpha_t + \beta_0^{RF}\mathcal{Z}^\vartheta_{it} + \varepsilon_{it} \tag{A.90}$$

identifies $\beta_0^{RF} = k^\vartheta(1-\beta\alpha)(1-\alpha)\omega$. The coefficient of the regression:

$$\Delta\log w_{it} = \alpha_t + \beta_0^{FS}\mathcal{Z}^\vartheta_{it} + \varepsilon_{it} \tag{A.91}$$

identifies $\beta_0^{FS} = k^\vartheta$. Consequently, the IV coefficient identifies $\beta_0^{IV} = (1-\beta\alpha)(1-\alpha)\omega$.

Proof. The key behind this result is the orthogonality of price adjustment in Calvo with respect to the cost shock. Specifically, the reduced form results come from

$$\beta_0^{RF} = \frac{\text{Cov}\left(\Delta \log p_{it}, \mathcal{Z}_{it}^{\vartheta}\right)}{\text{Var}\mathcal{Z}_{it}^{\vartheta}} = \omega k^{\vartheta}(1-\beta\alpha)\frac{\text{Cov}\left(\mathbb{1}_{it}^{p}\mathcal{Z}_{it}^{\vartheta}, \mathcal{Z}_{it}^{\vartheta}\right)}{\text{Var}\mathcal{Z}_{it}^{\vartheta}}$$
$$= \omega k^{\vartheta}(1-\beta\alpha)\mathbb{E}\left(\mathbb{1}_{it}^{p}\right)\frac{\text{Cov}\left(\mathcal{Z}_{it}^{\vartheta}, \mathcal{Z}_{it}^{\vartheta}\right)}{\text{Var}\mathcal{Z}_{it}^{\vartheta}}$$
$$= \omega k^{\vartheta}(1-\beta\alpha)(1-\alpha),$$

while the first stage comes from

$$\beta_0^{FS} = \frac{\text{Cov}\left(\Delta \log w_{it}, \mathcal{Z}_{it}^{\vartheta}\right)}{\text{Var}\mathcal{Z}_{it}^{\vartheta}} = k^{\vartheta}\frac{\text{Cov}\left(\mathcal{Z}_{it}^{\vartheta}, \mathcal{Z}_{it}^{\vartheta}\right)}{\text{Var}\mathcal{Z}_{it}^{\vartheta}}$$
$$= k^{\vartheta},$$

where the first equality uses A.64.

In population $\beta_0^{IV} = \frac{\beta_0^{RF}}{\beta_0^{FS}}$, obtaining the result. □

We now allow for persistence of ϑ_{it} by assuming it follows an AR(1) with persistence ρ_{ϑ}.

PROPOSITION 3. *Let us denote β_h the local projection coefficient at horizon h. Under the same notations and assumptions, $\beta_0^{RF} = k^{\vartheta}\frac{1-\beta\alpha}{1-\beta\alpha\rho_{\vartheta}}(1-\alpha)\rho\zeta$, $\beta_1^{RF} = k^{\vartheta}\frac{1-\beta\alpha}{1-\beta\alpha\rho_{\vartheta}}\rho\zeta(1-\alpha)\left(\rho_{\vartheta}+\alpha\right)$, $\beta_2^{RF} = k^{\vartheta}\frac{1-\beta\alpha}{1-\beta\alpha\rho_{\vartheta}}\mathbb{E}_{\lambda}\rho\zeta(1-\alpha)(\rho_{\vartheta}^2+\alpha\rho_{\vartheta}+\alpha^2)$.*

Slope of marginal cost curve. Define $d_{y,\xi} = \theta_i \frac{1}{Y}$, such that:

$$\hat{y}_{it} - \hat{Y}_t = -\theta(\hat{p}_{it} - \hat{\mathcal{P}}_t) + d_{y,\xi}\hat{\xi}_{it} \tag{A.92}$$

We repeat (A.72) for convenience:

$$\hat{mc}_{it+s|t} - \hat{mc}_{t+s} = \vartheta_{it} - d_{mc,y}\theta(\hat{p}_{it|t} - \hat{\mathcal{P}}_{t+s}) + d_{mc,y}d_{y,\xi}\hat{\xi}_{it+s|t}. \tag{A.93}$$

Then,

$$\hat{p}_{it|t} = (1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s(\omega(\hat{mc}_{t+s} + \vartheta_{it} + \left[d_{mc,y}+\Gamma\right]d_{y,\xi}\hat{\xi}_{it+s|t} - \hat{\mathcal{P}}_{t+s}) + \hat{\mathcal{P}}_{t+s})\right]$$

Note that because of the approximation around a zero-inflation symmetric steady state, firms that cannot reset their price keep the same relative price and the same relative quantity as a consequence. Using the shape of the demand curve,

$$\begin{aligned}\hat{y}_{it|t} &= \hat{Y}_t + d_{y,\xi}\hat{\xi}_{it} + \theta\hat{\mathcal{P}}_t \\ &\quad - \theta(1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s(\omega(\hat{mc}_{t+s} + \vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it+s|t} - \hat{\mathcal{P}}_{t+s}) + \hat{\mathcal{P}}_{t+s})\right]\end{aligned}$$

and the marginal cost equation

$$\begin{aligned}\hat{mc}_{it|t} &= \hat{mc}_t + \vartheta_{it} + d_{mc,y}d_{y,\xi}\hat{\xi}_{it} + d_{mc,y}\theta\hat{\mathcal{P}}_t \\ &\quad - d_{mc,y}\theta(1-\beta\alpha)\mathbb{E}_t\left[\sum_{s=0}^{+\infty}(\beta\alpha)^s(\omega(\hat{mc}_{t+s} + \vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it+s|t} - \hat{\mathcal{P}}_{t+s}) + \hat{\mathcal{P}}_{t+s})\right]\end{aligned}$$

For any of these variables, we now store all the time-specific variables in δ_t and consider the case of one-time ϑ_{it} and $\hat{\xi}_{it+s|t}$ shocks without persistence,

$$\begin{aligned}\hat{p}_{it|t} &= \delta_t^p + (1-\beta\alpha)\omega\left(\vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it}\right) \\ \hat{y}_{it|t} &= \delta_t^y + d_{y,\xi}\hat{\xi}_{it} - \theta(1-\beta\alpha)\omega\left(\vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it}\right) \\ \hat{mc}_{it|t} &= \delta_t^{mc} + \vartheta_{it} + d_{mc,y}\left(d_{y,\xi}\hat{\xi}_{it} - \theta(1-\beta\alpha)\omega(\vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it})\right)\end{aligned}$$

Let us denote $\mathbb{1}_{it}^p$ the dummy variable for whether a firm can reset its price. Omitting the δ_t terms, which will be absorbed by time fixed effects:

$$\begin{aligned}\hat{p}_{it} &= \mathbb{1}_{it}^p\left((1-\beta\alpha)\omega(\vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it})\right) + (1-\mathbb{1}_{it}^p)\hat{p}_{it-1} \\ \hat{y}_{it} &= \mathbb{1}_{it}^p\left(d_{y,\xi}\hat{\xi}_{it} - \theta(1-\beta\alpha)\omega(\vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it})\right) + (1-\mathbb{1}_{it}^p)\hat{y}_{it-1} \\ \hat{mc}_{it} &= \mathbb{1}_{it}^p\left(\vartheta_{it} + d_{mc,y}\left(d_{y,\xi}\hat{\xi}_{it} - \theta(1-\beta\alpha)\omega(\vartheta_{it} + \left[d_{mc,y} + \Gamma\right] d_{y,\xi}\hat{\xi}_{it})\right)\right) + (1-\mathbb{1}_{it}^p)\hat{mc}_{it-1}\end{aligned}$$

Let us assume that we have a proxy for the demand shifter $\mathcal{Z}_{it}^\xi$ satisfying $\kappa^\xi\mathcal{Z}_{it}^\xi = \hat{\xi}_{it}$. In addition, let us assume $\hat{\xi}_{it} \perp \vartheta_{it}$, $\hat{\xi}_{it} \perp \mathbb{1}_{it}^p$.

Let us now denote by $\beta_{y,\mathcal{Z}^\xi}$ the first stage of a regression of quantities on the demand shifter z^{ξ}_{it}. In population,

$$\beta_{y,z^{\xi}} = \kappa^{\xi}(1-\alpha)\mathbb{E}\left[d_{y,\xi} - \theta(1-\beta\alpha)\omega(\left[d_{mc,y}+\Gamma\right] d_{y,\xi})\right]. \tag{A.94}$$

Similarly let us denote by $\beta_{mc,z^{\xi}}$ a regression of marginal costs on the demand shifter z^{ξ}_{it}. In population,

$$\beta_{mc,z^{\xi}} = \kappa^{\xi}(1-\alpha)d_{mc,y}\mathbb{E}\left[d_{y,\xi} - \theta(1-\beta\alpha)\omega(\left[d_{mc,y}+\Gamma\right] d_{y,\xi})\right]. \tag{A.95}$$

Therefore, the IV where z^{ξ}_{it} is used as an instrument for $\hat{y}_{it}$ yields:

$$\beta^{IV} = d_{mc,y} \equiv \frac{1+a_w-(1-\gamma)a}{a(1-\gamma)}, \tag{A.96}$$

which is the curvature of the marginal cost function accounting for the fixed input.

In a data generating process where firms do not hold fixed capital, the IV coefficient would be given by the simplified expression

$$\beta^{IV} = \frac{1+a_w-a}{a}, \tag{A.97}$$

and in a data generating process where firms do not hold fixed capital and input markets are common, the IV coefficient would be given by the simplified expression

$$\beta^{IV} = \frac{1-a}{a}. \tag{A.98}$$

A.4. Additional extensions

In this section we consider extensions in which the following two log-linear relations hold

- $\hat{mc}_{it} = \hat{mc}_t + \Phi(\hat{y}_{it} - \hat{Y}_t)$
- $\hat{mc}_t - \hat{P}^Y_t = \Omega\hat{Y}_t$

Under these two conditions plus our marginal-cost based Phillips curve, we can represent the output-based Phillips curve as

$$\hat{\pi}_t = \varphi\omega\Omega\hat{Y}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}]. \tag{A.99}$$

A.4.1. Regional markets

There are N local labor markets in the national economy indexed by j. There is no worker mobility in the short run, and there is no home bias, so that the consumption basket of households across regions is the same. This assumption is not important, and can be relaxed to allow for home bias in tradeables or the existence of a non-tradeable sector. We keep the assumption that labor markets are common within regions. It is easy to extend them to the assumption of firm-specific input markets.

We assume a local marginal-cost Phillips curve can be written as:

$$\pi_{jt} = \beta\mathbb{E}_t\pi_{j,t+1} + \varphi\omega(\hat{mc}_{jt} - \hat{\mathcal{P}}_{jt}) \tag{A.100}$$

where the important assumption is that φ, β, ω are invariant across regions.

CPI inflation is the population-weighted average of local PPI inflation

$$\pi_t = \frac{1}{N}\sum_j \pi_{jt} \tag{A.101}$$

The aggregated Phillips curve is given by:

$$\pi_t = \beta\mathbb{E}_t\pi_{t+1} + \varphi\omega\frac{1}{N}\sum_j(\hat{mc}_{jt} + \hat{\mathcal{P}}_{jt}) \tag{A.102}$$

Region-level nominal marginal costs (denominated in local goods) are given by:

$$\hat{mc}_{jt} = \hat{w}_{jt} + \frac{1-a}{a}\hat{Y}_{jt} \tag{A.103}$$

which we can expressed as

$$\hat{mc}_{jt} = (\hat{w}_{jt} - \hat{\mathcal{P}}_t) + \hat{\mathcal{P}}_t + \frac{1-a}{a}\hat{Y}_{jt} \tag{A.104}$$

Households are in their labor supply curve:

$$\hat{w}_{jt} - \hat{\mathcal{P}}_t = \nu^{-1}\hat{L}_{jt} + (\sigma^{-1} - \iota_{lc})\hat{C}_{jt} \tag{A.105}$$

so population-weighted sums are given by

$$\frac{1}{N}\sum_j \hat{mc}_{jt} = \frac{1}{N}\sum_i (\nu^{-1}\hat{L}_{jt} + (\sigma^{-1} - \iota_{lc})\hat{C}_{jt}) + \hat{\mathcal{P}}_t + \frac{1-a}{a}\hat{Y}_{jt} \tag{A.106}$$

$$\frac{1}{N}\sum_j \hat{mc}_{jt} = \frac{\nu^{-1}}{a}\hat{Y}_t + (\sigma^{-1} - \iota_{lc})\hat{C}_t + \frac{1-a}{a}\hat{Y}_t + \hat{\mathcal{P}}_t \tag{A.107}$$

$$\hat{mc}_t - \hat{\mathcal{P}}_t = \left(\frac{\nu^{-1}}{a} + \frac{1-a}{a} + \upsilon\right)\hat{Y}_t \tag{A.108}$$

Slope of the marginal cost curve across regions. It is useful to stop here and look at the determination of regional marginal costs as they relate to regional output. In particular,

$$\hat{mc}_{jt} = \nu^{-1}\hat{L}_{jt} + (\sigma^{-1} - \iota_{lc})\hat{C}_{jt} + \frac{1-a}{a}\hat{Y}_{jt} + \hat{\mathcal{P}}_t \tag{A.109}$$

Using the production function for regional output we obtain:

$$\hat{mc}_{jt} = \frac{\nu^{-1} + 1 - a}{a}\hat{Y}_{jt} + (\sigma^{-1} - \iota_{lc})\hat{C}_{jt} + \hat{\mathcal{P}}_t \tag{A.110}$$

A regression of deviations of regional marginal costs on real output would yield:

$$\beta_{mc,y} = \frac{Cov\left(\hat{mc}_{jt} - \hat{\mathcal{P}}_t, \hat{Y}_{jt}\right)}{Var\hat{Y}_{jt}} \tag{A.111}$$

$$= \frac{\nu^{-1} + 1 - a}{a} + \upsilon\frac{Cov\left(\hat{C}_{jt}, \hat{Y}_{jt}\right)}{Var\hat{Y}_{jt}} \tag{A.112}$$

$$= \frac{\nu^{-1} + 1 - a}{a} + \upsilon\beta_{y,c} \tag{A.113}$$

This coefficient is, in general, not equal to $\Omega = \frac{\nu^{-1}+1-a}{a} + \upsilon$. However, we can bound $\beta_{mc,y}$ to be equal to, or a lower bound for Ω for a series of interesting cases.

Formally, $\beta_{mc,y}$ is weakly less or equal to Ω as long as

$$\upsilon(\beta_{mc,y} - 1) \leq 0. \tag{A.114}$$

The first case where this is true, is when $\upsilon = 0$, which occurs for the popular GHH preferences. When there are no wealth effects in labor supply, consumption does not play a role in labor supply, so there is no bias in our estimation of Ω.

The second case is when $\beta_{y,c} \leq 1$. This case holds for a specification with complete markets, where a set of contingent assets would make sure that consumption does not covary across regions after a shock that moves local output. In that case $\beta_{y,c} = 0$, and a regional regression of marginal costs on regional output estimates a lower bound for Ω. In the case of financial autarky, the polar opposite case to complete markets, local regions cannot hold any financial asset in positive supply, and therefore there are not any trade deficits. As a consequence $\hat{C}_{jt} = \hat{Y}_{jt}$, so that $\beta_{c,y} = 1$, yielding zero bias.

The set of data generating process where our estimates are not a lower bound for Ω are such in which in reaction to a transitory production boom, a region runs a trade deficit so that $\beta_{y,c} > 1$. This is not the standard case, where instead regions would save in response to a temporary bonanza in production.

A.4.2. Model with imported intermediates

In the paper we have considered the case of intermediate inputs in a roundabout production function. Here instead we consider the case where intermediates comes from the ROW. Firms still use local intermediates, but we will assume for simplicity that these are the same final good.

Production: Firms produce with a DRS production function on variable inputs $y_j = v_j^a$. Intermediate inputs are a CRS bundle of labor, domestic intermediates, and foreign intermediates $v = l^{\phi_l} x^{\phi_x} m^{\phi_m}$, where $\phi_l + \phi_x + \phi_m = 1$.

We first state the problem of minimizing the total cost of variable inputs subject to a target value for variable input demand, which yields an expression for the unit cost of variable inputs $w^v = \left(\frac{w^l}{\phi_l}\right)^{\phi_l} \left(\frac{P^Y}{\phi_x}\right)^{\phi_x} \left(\frac{P_m}{\phi_m}\right)^{\phi_m}$. And the marginal cost of production

$$mc_{jt} = \frac{1}{a} w_t^v y_{jt}^{(1-a)/a}, \tag{A.115}$$

or in log-linear terms

$$\hat{mc}_{jt} = \hat{w}_t^v + \frac{1-a}{a}\hat{y}_{jt} \tag{A.116}$$

aggregate marginal costs are given by

$$\hat{mc}_t = \hat{w}_t^v + \frac{1-a}{a}\hat{Y}_t \tag{A.117}$$

I will now use the reset equation in the paper which applies equally to this case so we can derive a marginal-based Phillips curve as in the paper.

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \varphi\omega(\hat{mc}_t - \hat{P}_t^Y) \tag{A.118}$$

We can rewrite the real marginal cost equation

$$\hat{mc}_t - \hat{P}_t^Y = \phi_l(\hat{w}_t^l - \hat{P}_t^Y) + \phi_m(\hat{P}_t^m - \hat{P}_t^Y) + \frac{1-a}{a}\hat{Y}_t \tag{A.119}$$

To fully solve the model we need to specify a supply curve for imported intermediates. We will assume that the supply curve for imported intermediates takes the form of

$$M_t = \xi_t^m \left(\frac{P_t^m}{P_t^Y}\right)^{\varepsilon_m} \tag{A.120}$$

Where ξ_t are cost-push shocks in the supply of foreign intermediates, and we allow the possibility that the ROW is elastic in supplying more goods when the relative price of imported intermediates rises. This setting of course nests a constant supply of intermediates, a case where the supply is exogenous. The important assumption is that the supply curve depends on the relative price of intermediates with respect to the domestic price index, as opposed to for example, the relative price of intermediates with respect to the CPI of the ROW.

Firm-level demand for foreign intermediates is given by

$$\hat{m}_{jt} = \hat{w}^v_t + \frac{1}{a}\hat{y}_{jt} - \hat{P}^m_t \tag{A.121}$$

which after integrating over firms implies that total demand is given by

$$\hat{M}_t = \hat{w}^v_t + \frac{1}{a}\hat{Y}_t - \hat{P}^m_t. \tag{A.122}$$

Market clearing in foreign intermediates then implies that:

$$\hat{w}^v_t + \frac{1}{a}\hat{Y}_t - \hat{P}^m_t = \hat{\xi}^m_t + \varepsilon^m(\hat{P}^m_t - \hat{P}^Y_t) \tag{A.123}$$

using the assumption of the price for intermediates that comes from the Cobb Douglass assumption on variable inputs we can rewrite this expression in terms of the relative price of imported intermediates the real wage, output and the import supply shock.

$$(\hat{P}^m_t - \hat{P}^Y_t) = \frac{\phi_l}{\varepsilon^m + 1 - \phi_m}(\hat{w}^l_t - \hat{P}^Y_t) + \frac{1}{a(\varepsilon^m + 1 - \phi_m)}\hat{Y}_t - \frac{1}{\varepsilon + 1 - \phi_m}\hat{\xi}^m_t \tag{A.124}$$

We can then replace this expression in the determination of marginal costs in the aggregate, finding

(A.125)

$$\hat{mc}_t - \hat{P}^Y_t = \phi_l \frac{\varepsilon^m + 1}{\varepsilon^m + 1 - \phi_m}(\hat{w}^l_t - \hat{P}^Y_t) + \frac{1}{a}\left(\frac{(\varepsilon^m + 1 - \phi_m)(1-a) + \phi_m}{\varepsilon^m + 1 - \phi_m}\right)\hat{Y}_t - \frac{\phi_m}{\varepsilon^m + 1 - \phi_m}\hat{\xi}^m_t$$

We just need to check whether we can write the real wage as a function of output only. We will use GHH preferences and use a system of two equations and two unknowns. Two supply curves, and two demand curves from our system

$$\hat{L}_t = \nu(\hat{w}^l_t - \hat{P}^Y_t) \tag{A.126}$$

$$\hat{M}_t = \varepsilon^m(\hat{P}^m_t - \hat{P}^Y_t) + \xi^m_t \tag{A.127}$$

$$\hat{L}_t = \hat{w}^v_t + \frac{1}{a}\hat{Y}_t - \hat{w}_t \tag{A.128}$$

$$\hat{M}_t = \hat{w}^v_t + \frac{1}{a}\hat{Y}_t - \hat{P}^m_t \tag{A.129}$$

We equalize supply and demand for each input and subtract the two resulting equations to find an expression of the relative price for intermediates as a function of the relative price of labor and the presence of supply shocks.

$$(\hat{P}^m_t - \hat{P}^Y_t) = \frac{\nu+1}{\varepsilon^m+1}(\hat{w}^l_t - \hat{P}^Y_t) - \frac{\varepsilon^m}{\varepsilon^m+1} \tag{A.130}$$

and we can plug result into the labor market clearing condition, finding

(A.131)

$$(\hat{w}^l_t - \hat{P}^Y_t) = \frac{1}{a}\left(\nu + 1 - \phi_l - \phi_m\frac{\nu+1}{\varepsilon^m+1}\right)^{-1}\hat{Y}_t + \frac{\phi_m\varepsilon^m}{\varepsilon+1}\left(\nu + 1 - \phi_l - \phi_m\frac{\nu+1}{\varepsilon^m+1}\right)^{-1}\hat{\xi}_t$$

together with the marginal cost equation A.125, these two equations make clear that marginal costs are a function of only one endogenous variable $\hat{Y}_t$ and an exogenous cost shifter ξ^m_t.

A.4.3. Summary of extensions

Table A.1 summarizes several dimensions in which the Phillips curve in 1 holds as it is or with minor modifications, such as the timing of the output gap, or the presence of exogenous cost-push shocks.

Table A.1. Summary of extensions

Category	Baseline model	Extensions
Pricing	Calvo with freq. $1-\alpha$	Rotemberg, Taylor
Prod. function	DRS $a \leq 1$ in labor	Roundabout intermediates, rented or fixed capital, imported intermediates
Input markets	National, flexible input prices	Firm-specific, regional, partial adjustment
Competition	Monopolistic competition	Oligopolistic (Atkeson-Burstein)
Demand	Kimball: demand elasticity θ_i, markup elasticity w.r.t. relative price Γ_i	CES, HDIA, HSA textbook: CES $\theta_i = \theta$, $\Gamma_i = 0$
Household	Discount factor β, labor supply curve	textbook: SEP $\hat{w}_t^r = (\nu^{-1} + \sigma^{-1})\hat{Y}_t$

Note: This Table summarizes a set of admissible extensions or modifications of our benchmark model in which the Phillips curve we use in the main text holds.

Appendix B. Extended model with steady-state misallocation

B.1. Environment

The economy is composed of four sectors. Households consume the final good, save, and supply labor. A final good producer produces the final good using differentiated varieties indexed by $i \in [0,1]$. Producers of each differentiated variety i produce using labor and have sticky prices. A monetary authority sets the nominal interest rate.

Households. The household block is the same as in the baseline model. From the households' optimization problem we obtain the Euler equation:

$$\frac{1}{1+i_t} = \beta \mathbb{E}_t [\frac{u_c(C_{t+1}, L_{t+1})}{u_c(C_t, L_t)} \frac{P_t^Y}{P_{t+1}^Y}] \tag{B.1}$$

and the labor supply function:

$$\frac{u_l(C_t, L_t)}{u_c(C_t, L_t)} = \frac{w_t^l}{P_t^Y} \tag{B.2}$$

Final good producers. Let Y_t denote aggregate production of the final good. The final good Y_t is produced by perfectly competitive firms using a bundle of differentiated intermediate inputs y_{it} for $i \in [0,1]$. We use the Kimball aggregator introduced in Appendix A.2.1:

$$\int_0^1 \Upsilon\left(\frac{y_{it}}{Y_t}\right) di = 1$$

where the function $\Upsilon(.)$ is strictly increasing, strictly concave, and satisfies $\Upsilon(1) = 1$. This gives rise to the demand function:

$$\frac{y_{it}}{Y_t} = \Upsilon'^{-1}\left(\frac{p_{it}}{\mathcal{P}_t}\right) \tag{B.3}$$

where $\frac{p_{it}}{\mathcal{P}_t}$ determines substitution across varieties. The price index $\mathcal{P}_t$ is given by $\mathcal{P}_t = \frac{P_t^Y}{D_t}$. $P_t^Y = \int_0^1 p_{it}\frac{y_{it}}{Y_t} di$ is the ideal price index. $D_t = \int_0^1 \Upsilon'(\frac{y_{it}}{Y_t})\frac{y_{it}}{Y_t} di$ is a "demand" index. When demand is CES, D_t is a constant equal to $\frac{\theta}{\theta-1}$. Away from the CES case, D_t is not a constant and is increasing in the dispersion of quantity shares.

The price elasticity of demand is only a function of firm relative size and is given by:

$$\theta_{it} = \theta(\frac{y_{it}}{Y_t}) = -\frac{\partial \log y_{it}}{\partial \log p_{it}} = \frac{\Upsilon'\left(\frac{y_{it}}{Y_t}\right)}{-\frac{y_{it}}{Y_t}\Upsilon''\left(\frac{y_{it}}{Y_t}\right)} \tag{B.4}$$

Differentiated varieties producers. Each variety i is produced by a single firm. Firms produce with the following technology:

$$y_{it} = e^{z_i} l_{it}^a \tag{B.5}$$

The cost function writes:

$$\mathcal{C}(y_{it}, w_{it}^v, z_{it}, \tau_i) = (1+\tau_i) w_{it}^v \left(\frac{y_{it}}{e^{z_{it}}}\right)^{\frac{1}{a}} \tag{B.6}$$

The marginal cost for firm i writes:

$$mc_{it} = (1+\tau_i) w_{it}^v \frac{1}{a} e^{-\frac{1}{a} z_{it}} y_{it}^{\frac{1-a}{a}} \tag{B.7}$$

A key difference with the baseline model is that we allow for generic input wedges, as opposed to imposing that firms are symmetric in the steady state.

A firm has a probability $1-\alpha$ of being able to reset its price in each period. A firm that can reset its price maximizes choses the price that maximizes:

$$\max_{p_{it|t}} \mathbb{E}_t \left[\sum_{s=0}^{+\infty} \alpha^s \Lambda_{t,t+s} \left[p_{it|t} y_{it+s|t} - \mathcal{C}(y_{it+s|t}, w^v_{i,t+s|t}, z_i, \tau_i) \right] \right]$$

subject to the demand curve $y_{it+s} = \Upsilon'^{-1}\left(\frac{p_{it}}{\mathcal{P}_{t+s}}\right) Y_{t+s}$ and the cost function $\mathcal{C}(y_{it+s}, w^v_{t+s}, z_i, \tau_i) = (1+\tau_i) w^v_{t+s} e^{-\frac{1}{a} z_i} y_{it+s}^{\frac{1}{a}}$. $\Lambda_{t,t+s}$ is the stochastic discount factor.

As above, it will be convenient to define the following quantities. $\mu^f_{it} = \frac{\theta_{it}}{\theta_{it}-1}$ is the desired markup that the firm would choose in a flexible price environment. $\Gamma_{it} = \dfrac{\partial \log \mu^f_{it}}{\partial \log \frac{y_{it}}{Y_t}}$ is the elasticity of the flexible price markup with respect to relative size. ρ_{it} is the partial equilibrium pass-through of a marginal cost shock into the firm's price: $\rho_{it} \equiv \dfrac{\partial \log p_{it}}{\partial \log mc_{it}} = \frac{1}{1+\Gamma_{it}\theta_{it}}$. Note that μ^f_{it}, Γ_{it}, and ρ_{it} are only a function of a firm's relative size $\frac{y_{it}}{Y_t}$. Finally, in a sticky price environment, the actual markup of the firm may differ from the flexible price desired markup. We denote the actual markup of the firm: $\mu_{it} = \frac{p_{it}}{mc_{it}}$.

Monetary authority. Monetary policy sets the nominal interest rate according to a Taylor rule.

Equilibrium. Equilibrium is defined by the following conditions: (i) Consumers choose consumption and labor to maximize utility taking prices and wage as given; (ii) Firms with flexible prices set prices to maximize their value taking the price index and their residual demand curves as given; firms with sticky prices meet demand at fixed prices ; (iii) Monetary policy sets the nominal interest rate; (iv) All resource constraints are satisfied.

We solve the model by log-linearization around the zero-inflation steady state. The steady-state distribution of firm size depends on joint distribution of (z_i, τ_i). We take a first-order expansion for small monetary policy shock. Quantities without a t subscript refer to the steady state.

We denote $\lambda_i = \frac{p_i y_i}{P^Y Y}$ sales share in steady state. Let $\mathbb{E}_\lambda[X_{it}] = \int_0^1 \lambda_i X_{it} di$.

B.2. Characterization

It is useful to the following expressions. First, linearizing the definition of the Kimball aggregator:

$$0 = \mathbb{E}_\lambda \left[\widehat{\frac{y_{it}}{Y_t}} \right] \quad \Leftrightarrow \quad \hat{Y}_t = \mathbb{E}_\lambda[\hat{y}_{it}] \tag{B.8}$$

Second, linearizing the price index:

$$\hat{P}_t^Y = \mathbb{E}_\lambda \left[\hat{p}_{it} \right] \tag{B.9}$$

Third, linearizing $\mathcal{P}_t$:

$$\hat{\mathcal{P}}_t = \hat{P}_t^Y - \hat{D}_t \tag{B.10}$$

Fourth, linearizing the demand index:

$$\hat{D}_t = -\mathbb{E}_\lambda[(\theta_i - 1)(\hat{p}_{it} - \hat{\mathcal{P}}_t)] = -\frac{\text{Cov}_\lambda[\theta_i, \hat{p}_{it}]}{\mathbb{E}_\lambda[\theta_i]} \tag{B.11}$$

Finally, we can write:

$$\hat{\mathcal{P}}_t = \frac{\mathbb{E}_\lambda[\theta_i \hat{p}_{it}]}{\mathbb{E}_\lambda[\theta_i]} \tag{B.12}$$

Marginal cost-based Phillips curve. Following the same steps as above, a firm that can reset its price at time t will choose:

$$\hat{p}_{it|t} = (1 - \beta\alpha)\mathbb{E}_t \left[\sum_{s=0}^{+\infty} (\beta\alpha)^s \left(\zeta_i \rho_i \hat{mc}_{t+s} + (1 - \zeta_i \rho_i) \hat{\mathcal{P}}_{t+s} \right) \right] \tag{B.13}$$

where ρ_i is the flexible price partial equilibrium pass-through of marginal cost shocks into prices and $\zeta_i = \frac{1}{1+d_{mc,y}\theta_i\rho_i}$. ζ_i captures the fact that when returns to scale are below 1, a cost shock induces an adjustment in size, which dampens the first-round effect on marginal cost. $\zeta_i \rho_i$ combines these two terms and captures the flexible price pass-through of an input cost shock into prices. The difference with the baseline model is that both parameters are indexed by i. $d_{mc,y} = \frac{1-a}{a}$ is, as before, the elasticity of marginal costs with respect to firm scale.

We can write this equation recursively as:

$$\hat{p}_{it|t} = (1-\beta\alpha)\left(\zeta_i\rho_i\hat{mc}_t + (1-\zeta_i\rho_i)\hat{\mathcal{P}}_t\right) + \beta\alpha\mathbb{E}_t[\hat{p}_{it+1|t+1}] \tag{B.14}$$

Inflation dynamics. Aggregating across firms similarly to the baseline model, we obtain the marginal cost-based Phillips curve:

$$\hat{\pi}_t = \varphi\omega(\hat{mc}_t - \hat{P}_t^Y) - \varphi(1-\omega)\hat{D}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}] \tag{B.15}$$

$\hat{mc}_t - \hat{P}_t^Y$ is the change in the aggregate real marginal cost. $\hat{D}_t$ is the change in the demand index. $\varphi = \dfrac{(1-\alpha)(1-\beta\alpha)}{\alpha}$ is the slope of the marginal cost-based Phillips curve in the case of constant returns to scale and CES demand. $\omega = \mathbb{E}_\lambda[\zeta_i\rho_i]$ reflects micro-level real rigidities due to decreasing returns to scale ζ_i and strategic complementarities ρ_i.

Aggregate marginal costs. Let us define aggregate productivity Z_t as satisfying

$$Y_t \equiv Z_t L_t^a \tag{B.16}$$

Solving for aggregate marginal costs $\hat{mc}_t = \mathbb{E}_\lambda\left[\hat{mc}_{it}\right]$ as in the baseline model, we obtain:

$$\hat{mc}_t = \underbrace{\left[\frac{1-a+\nu^{-1}}{a} + \upsilon\right]}_{\Omega\text{ = Elasticity of mc wrt output}} \hat{Y}_t - \underbrace{\left[\frac{\nu^{-1}}{a}\right]}_{\Xi\text{ = Elasticity of mc wrt TFP}} \hat{Z}_t + \hat{P}_t^Y. \tag{B.17}$$

$\sigma^{-1} = -\frac{u_{cc}C}{u_c}$ is the inverse elasticity of intertemporal substitution, $\nu^{-1} = \frac{u_{ll}L}{u_l} - \frac{u_{cl}L}{u_c}$ is the inverse Frisch elasticity of labor supply, and $\iota_{cl} = \frac{u_{cl}L}{u_c}$. We define $\upsilon \equiv (\sigma^{-1} - \iota_{lc})$.

Output-based Phillips curve. Combining (B.15) and (B.17), we obtain the output-based New Keynesian Phillips curve:

$$\hat{\pi}_t = \varphi\omega\left(\Omega\hat{Y}_t - \Xi\hat{Z}_t\right) - \varphi(1-\omega)\hat{D}_t + \beta\mathbb{E}_t[\hat{\pi}_{t+1}] \tag{B.18}$$

Allocative efficiency. Let us defined the combined allocative distortion as $m_{it} \equiv \mu_{it}(1+\tau_i)$. Using the definition of the markup and of the marginal cost,

$$m_{it} = a\frac{p_{it}y_{it}}{w_t^v l_{it}} \tag{B.19}$$

We define the aggregate distortion as solving:

$$\mathcal{M}_t \equiv a\frac{P_t^Y Y_t}{w_t^v L_t} \tag{B.20}$$

Step 1: First, we show that:

$$\mathcal{M}_t = \mathbb{E}_\lambda\left[m_{it}^{-1}\right]^{-1} \tag{B.21}$$

Step 2:

$$\hat{Z}_t = a\left[\hat{\mathcal{M}}_t - \mathbb{E}_\lambda[\hat{m}_{it}]\right] \tag{B.22}$$

By definition of the aggregate distortion,

$$\hat{\mathcal{M}}_t = \widehat{P_t^Y Y_t} - \widehat{w_t^v L_t}$$

By definition of the firm level distortion

$$\hat{m}_{it} = \hat{\mu}_{it} = \hat{p}_{it} - \hat{mc}_{it} = \hat{p}_{it} - \hat{w}_t^v - \frac{1-a}{a}\hat{y}_{it}$$
$$\mathbb{E}_\lambda[\hat{m}_{it}] = \hat{P}_t^Y - \hat{w}_t^v - \frac{1-a}{a}\hat{Y}_t$$

Therefore,

$$a(\hat{\mathcal{M}}_t - \mathbb{E}_\lambda[\hat{m}_{it}]) = \hat{Y}_t - a\hat{L}_t = \hat{Z}_t$$

Step 3:

$$\hat{Z}_t = -\mathcal{M}\mathrm{Cov}_\lambda[m_i^{-1}, \hat{y}_{it}] = \mathcal{M}\mathrm{Cov}_\lambda[m_i^{-1}, \theta_i(\hat{p}_{it} - \hat{\mathcal{P}}_t)] \tag{B.23}$$

Log-linearizing the expression for $\mathcal{M}_t$ in (B.21) yields:

$$\hat{\mathcal{M}}_t - \mathbb{E}_\lambda[\hat{m}_{it}] = -\mathcal{M}\mathbb{E}_\lambda[(m_i)^{-1}(\hat{\lambda}_{it} - \hat{m}_{it})] - \mathbb{E}_\lambda[\hat{m}_{it}] = -\mathcal{M}\text{Cov}_\lambda[m_i^{-1}, \hat{\lambda}_{it} - \hat{m}_{it}]$$

In addition,

$$\hat{\lambda}_{it} - \hat{m}_{it} = (\hat{p}_{it} + \hat{y}_{it} - (\hat{P}_t^Y + \hat{Y}_t)) - (\hat{p}_{it} - \hat{mc}_{it}) = -(\hat{P}_t^Y + \hat{Y}_t) + \hat{w}_t^v + \frac{1}{a}\hat{y}_{it}$$

Therefore, $\hat{Z}_t = -\mathcal{M}\text{Cov}_\lambda[m_i^{-1}, \hat{y}_{it}]$.

Step 4: We have two equations that characterize $\hat{D}_t$ and $\hat{Z}_t$ as a function of the change in relative prices:

$$\hat{D}_t = -\frac{\text{Cov}_\lambda[\theta_i, \hat{p}_{it}]}{\mathbb{E}_\lambda[\theta_i]}$$
$$\hat{Z}_t = \mathcal{M}\text{Cov}_\lambda[m_i^{-1}, \theta_i(\hat{p}_{it} - \hat{\mathcal{P}}_t)]$$

where the first equation is (B.11) and the second equation is (B.23).

We know show how to express $\hat{D}_t$ and $\hat{Z}_t$ as a function of model parameters and steady-state objects.

$$\hat{p}_{it} = \mathbb{1}_{it}^p \hat{p}_{it|t} + (1 - \mathbb{1}_{it}^p)\hat{p}_{it-1}$$

Taking the expectation over the realization of the Calvo fairy, we obtain:

$$\text{(B.24)} \qquad (\hat{p}_{it} - \hat{p}_{it-1}) - \beta(\hat{p}_{it+1} - \hat{p}_{it}) = \varphi\left(\zeta_i\rho_i\left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) + \hat{\mathcal{P}}_t\right) - \varphi\hat{p}_{it}$$

Let us first derive the equation for $\hat{D}_t$. Multiplying by θ_i and applying the $\mathbb{E}_\lambda$ operator yields:

$$(\mathbb{E}_\lambda[\theta_i\hat{p}_{it}] - \mathbb{E}_\lambda[\theta_i\hat{p}_{it-1}]) - \beta(\mathbb{E}_\lambda[\theta_i\hat{p}_{it+1}] - \mathbb{E}_\lambda[\theta_i\hat{p}_{it}]) = \varphi\mathbb{E}_\lambda\left[\theta_i\left(\zeta_i\rho_i\left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) + \hat{\mathcal{P}}_t\right)\right] - \varphi\mathbb{E}_\lambda[\theta_i\hat{p}_{it}]$$
$$(\hat{\mathcal{P}}_t - \hat{\mathcal{P}}_{t-1}) - \beta(\hat{\mathcal{P}}_{t+1} - \hat{\mathcal{P}}_t) = \varphi\frac{\mathbb{E}_\lambda\left[\theta_i\zeta_i\rho_i\right]}{\mathbb{E}_\lambda[\theta_i]}\left(\hat{mc}_t - \hat{\mathcal{P}}_t\right)$$

where we repeatedly use equation (B.12). Substracting the NKPC equation,

$$[\hat{D}_t - \hat{D}_{t-1} - \beta(\hat{D}_{t+1} - \hat{D}_t)] = -\varphi \frac{\text{Cov}_\lambda\left[\theta_i, \zeta_i \rho_i\right]}{\mathbb{E}_\lambda[\theta_i]} \left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) - \varphi \hat{D}_t$$

In addition,

$$\hat{mc}_t - \hat{\mathcal{P}}_t = \Omega \hat{Y}_t - \Xi \hat{Z}_t + \hat{D}_t$$

We can rewrite this as equation (29).

$$\begin{aligned}
[\hat{D}_t - \hat{D}_{t-1} - \beta(\hat{D}_{t+1} - \hat{D}_t)] &= -\varphi \kappa_D \left(\Omega \hat{Y}_t - \Xi \hat{Z}_t + \hat{D}_t\right) - \varphi \hat{D}_t \\
\hat{D}_t(1 + \beta + \varphi(1 + \kappa_D)) &= -\varphi \kappa_D \left(\Omega \hat{Y}_t - \Xi \hat{Z}_t\right) + \hat{D}_{t-1} + \beta \hat{D}_{t+1}
\end{aligned}$$

We use the same logic to derive the expression for $\hat{Z}_t$.

$$\begin{aligned}
&(\hat{p}_{it} - \hat{p}_{it-1}) - \beta(\hat{p}_{it+1} - \hat{p}_{it}) - (\hat{\mathcal{P}}_t - \hat{\mathcal{P}}_{t-1}) - \beta(\hat{\mathcal{P}}_{t+1} - \hat{\mathcal{P}}_t) = \varphi \left(\zeta_i \rho_i \left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) + \hat{\mathcal{P}}_t\right) - \varphi \hat{p}_{it} - (\hat{\mathcal{P}}_t - \hat{\mathcal{P}}_{t-1}) - \beta(\hat{\mathcal{P}}_{t+1} - \hat{\mathcal{P}}_t) \\
&- [(\hat{y}_{it} - \hat{y}_{it-1}) - \beta(\hat{y}_{it+1} - \hat{y}_{it}) - (\hat{Y}_t - \hat{Y}_{t-1}) - \beta(\hat{Y}_{t+1} - \hat{Y}_t)] = \varphi \left(\theta_i \zeta_i \rho_i \left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) + \theta_i \hat{\mathcal{P}}_t\right) - \varphi \theta_i \hat{p}_{it} - \theta_i(\hat{\mathcal{P}}_t - \hat{\mathcal{P}}_{t-1}) - \beta \theta_i(\hat{\mathcal{P}}_{t+1} - \hat{\mathcal{P}}_t) \\
&(\hat{Z}_t - \hat{Z}_{t-1}) - \beta(\hat{Z}_{t+1} - \hat{Z}_t) = -\mathcal{M}\text{Cov}_\lambda[m_i^{-1}, \varphi \theta_i \zeta_i \rho_i \left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) - \theta_i(\hat{\mathcal{P}}_t - \hat{\mathcal{P}}_{t-1}) - \beta \theta_i(\hat{\mathcal{P}}_{t+1} - \hat{\mathcal{P}}_t)] - \varphi \hat{Z}_t \\
&(\hat{Z}_t - \hat{Z}_{t-1}) - \beta(\hat{Z}_{t+1} - \hat{Z}_t) = -\varphi \mathcal{M}\text{Cov}_\lambda[m_i^{-1}, \theta_i \zeta_i \rho_i] \left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) + \mathcal{M}\text{Cov}_\lambda[m_i^{-1}, \theta_i]((\hat{\mathcal{P}}_t - \hat{\mathcal{P}}_{t-1}) + \beta(\hat{\mathcal{P}}_{t+1} - \hat{\mathcal{P}}_t)) - \varphi \hat{Z}_t \\
&(\hat{Z}_t - \hat{Z}_{t-1}) - \beta(\hat{Z}_{t+1} - \hat{Z}_t) = -\varphi \mathcal{M} \left(\text{Cov}_\lambda[m_i^{-1}, \theta_i \zeta_i \rho_i] - \text{Cov}_\lambda[m_i^{-1}, \theta_i] \frac{\mathbb{E}_\lambda\left[\theta_i \zeta_i \rho_i\right]}{\mathbb{E}_\lambda[\theta_i]}\right) \left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) - \varphi \hat{Z}_t \\
&(\hat{Z}_t - \hat{Z}_{t-1}) - \beta(\hat{Z}_{t+1} - \hat{Z}_t) = -\varphi \mathbb{E}_\lambda[\theta_i \zeta_i \rho_i] \left(\mathbb{E}_\lambda[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]} \frac{\theta_i \zeta_i \rho_i}{\mathbb{E}_\lambda[\theta_i \zeta_i \rho_i]}] - \mathbb{E}_\lambda[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]} \frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}]\right) \left(\hat{mc}_t - \hat{\mathcal{P}}_t\right) - \varphi \hat{Z}_t
\end{aligned}$$

Then,

$$(\hat{Z}_t - \hat{Z}_{t-1}) - \beta(\hat{Z}_{t+1} - \hat{Z}_t) = -\varphi \kappa_Z \left(\Omega \hat{Y}_t - \Xi \hat{Z}_t + \hat{D}_t\right) - \varphi \hat{Z}_t$$

from which we can obtain the expression in (29).

Euler equation. From the utility maximization problem of households, we obtain the generic (log-linearized) Euler equation as:

$$\hat{C}_t - \sigma \iota_{cl} \hat{L}_t = \mathbb{E}_t \left[\hat{C}_{t+1} - \sigma \iota_{cl} \hat{L}_{t+1}\right] - \sigma \left(\hat{i}_t - \mathbb{E}_t \left[\hat{\pi}_{t+1}\right]\right) \tag{B.25}$$

Knowing the equilibrium input price and using the expression for $\hat{C}_t$ in (B.25) we can derive the Euler equation as:

$$\hat{Y}_t - \frac{\sigma \iota_{cl}}{a}\left(\hat{Y}_t - \hat{Z}_t\right) = \mathbb{E}\left[\hat{Y}_{t+1} - \frac{\sigma \iota_{cl}}{a}\left(\hat{Y}_{t+1} - \hat{Z}_{t+1}\right)\right] - \sigma\left(\hat{i}_t - \mathbb{E}_t\left[\hat{\pi}_{t+1}\right]\right). \tag{B.26}$$

B.3. Identification results in the extended model: slope

The slope of the Phillips curve is $\kappa_y = \varphi \mathbb{E}_\lambda[\zeta_i \rho_i]\Omega$, compared to $\kappa_y = \varphi\omega\Omega$ in the baseline model. It is straightforward to extend our identification proof for the firm-level passthrough to the case of steady-state heterogeneity. Therefore, the slope of the marginal cost-based Phillips curve can be identified using the same steps as in the baseline model. The extended model yields similar predictions for Ω, so that our identification strategy for this term is unchanged.

We now develop the identification proof for the firm-level passthrough. We consider the model with intermediates, capital, and supply and demand shocks, for comparability with Appendix A.3.

Proposition 4. *Let us assume that $\mathbb{1}^p_{it} \perp \vartheta_{it}$, $\mathbb{1}^p_{it} \perp \lambda_i$, and $\vartheta_{it} \perp \hat{\xi}_{it}$ (where $\hat{\xi}_{it}$ is any firm-level demand shock), $\vartheta_{it} \perp \hat{p}_{it-1}$, $\vartheta_{it} \perp \lambda_i$. Finally, assume that we observe $\mathcal{Z}^\vartheta_{it}$ a proxy for ϑ_{it} satisfying $\vartheta_{it} = k^\vartheta \mathcal{Z}^\vartheta_{it}$. Then, the coefficient of the regression*

$$\Delta \log p_{it} = \alpha_t + \beta_0^{RF} \mathcal{Z}^\vartheta_{it} + \varepsilon_{it} \tag{B.27}$$

identifies $\beta_0^{RF} = k^\vartheta(1-\beta\alpha)(1-\alpha)\omega$ for $\omega = \mathbb{E}_\lambda[\rho_i\zeta_i]$. The coefficient of the regression:

$$\Delta \log w_{it} = \alpha_t + \beta_0^{FS} \mathcal{Z}^\vartheta_{it} + \varepsilon_{it} \tag{B.28}$$

identifies $\beta_0^{FS} = k^\vartheta$. Consequently, the IV coefficient identifies $\beta_0^{IV} = (1-\beta\alpha)(1-\alpha)\omega$.

Proof. The key behind this result is the orthogonality of price adjustment in Calvo with respect to the cost shock. Specifically, the reduced form results come from

$$\beta_0^{RF} = \frac{\text{Cov}_\lambda\left(\Delta \log p_{it}, \mathbb{Z}^\vartheta_{it}\right)}{Var_\lambda \mathbb{Z}^\vartheta_{it}} = \omega k^\vartheta (1-\beta\alpha)\frac{\text{Cov}_\lambda\left(\mathbb{1}^p_{it}\mathcal{Z}^\vartheta_{it}, \mathcal{Z}^\vartheta_{it}\right)}{Var_\lambda \mathcal{Z}^\vartheta_{it}}$$

$$
\begin{aligned}
&= \omega k^{\vartheta}(1-\beta\alpha)\mathbb{E}_{\lambda}(^{p}_{it})\frac{\text{Cov}_{\lambda}\left(z^{\vartheta}_{it}, z^{\vartheta}_{it}\right)}{Var_{\lambda}z^{\vartheta}_{it}} \\
&= \omega k^{\vartheta}(1-\beta\alpha)(1-\alpha),
\end{aligned}
$$

while the first stage comes from

$$
\beta^{FS}_{0} = \frac{\text{Cov}_{\lambda}\left(\Delta \log w_{it}, z^{\vartheta}_{it}\right)}{Var_{\lambda}z^{\vartheta}_{it}} = k^{\vartheta}\frac{\text{Cov}_{\lambda}\left(z^{\vartheta}_{it}, z^{\vartheta}_{it}\right)}{Var_{\lambda}z^{\vartheta}_{it}} = k^{\vartheta},
$$

where the first equality uses A.64. In population $\beta^{IV}_{0} = \frac{\beta^{RF}_{0}}{\beta^{FS}_{0}}$, obtaining the result. □

B.4. Identification results in the extended model: allocative efficiency

B.4.1. Identification strategy

The goal is to identify:

$$
\begin{aligned}
\kappa_{D} &= \mathbb{E}_{\lambda}[\zeta_i\rho_i]\left(\mathbb{E}_{\lambda}\left[\frac{\theta_i}{\mathbb{E}_{\lambda}[\theta_i]}\frac{\zeta_i\rho_i}{\mathbb{E}_{\lambda}[\zeta_i\rho_i]}\right]-1\right) \\
\kappa_{Z} &= \mathbb{E}_{\lambda}[\theta_i\zeta_i\rho_i]\left(\mathbb{E}_{\lambda}\left[\frac{m_i^{-1}}{\mathbb{E}_{\lambda}[m_i^{-1}]}\frac{\theta_i\zeta_i\rho_i}{\mathbb{E}_{\lambda}[\theta_i\zeta_i\rho_i]}\right]-\mathbb{E}_{\lambda}\left[\frac{m_i^{-1}}{\mathbb{E}_{\lambda}[m_i^{-1}]}\frac{\theta_i}{\mathbb{E}_{\lambda}[\theta_i]}\right]\right)
\end{aligned}
$$

We first detail how to estimate each objects, assuming m_i^{-1} and θ_i are known. We then detail the measurement of m_i^{-1} and θ_i.

Identification of $\mathbb{E}_{\lambda}[\theta_i\zeta_i\rho_i]$ and $\mathbb{E}_{\lambda}[\zeta_i\rho_i]$. $\mathbb{E}_{\lambda}[\zeta_i\rho_i]$ is obtained from the regression of $\Delta \log p_{it}$ on $\Delta \log w^{v}_{it}$, instrumented by ϑ_{it}, combined with our estimate of $(1-\alpha\beta)(1-\alpha)$. Similarly, $\mathbb{E}_{\lambda}[\theta_i\zeta_i\rho_i]$ is obtained from the regression of $\Delta \log y_{it}$ on $\Delta \log w^{v}_{it}$, instrumented by ϑ_{it}.

Identification of $\mathbb{E}_{\lambda}\left[\frac{\theta_i}{\mathbb{E}_{\lambda}[\theta_i]}\frac{\zeta_i\rho_i}{\mathbb{E}_{\lambda}[\zeta_i\rho_i]}\right]$ and $\mathbb{E}_{\lambda}\left[\frac{m_i^{-1}}{\mathbb{E}_{\lambda}[m_i^{-1}]}\frac{\theta_i\zeta_i\rho_i}{\mathbb{E}_{\lambda}[\theta_i\zeta_i\rho_i]}\right]$. We obtain these quantities by estimating the price and quantity regressions by bins of $\frac{\theta_i}{\mathbb{E}_{\lambda}[\theta_i]}$ and $\frac{m_i^{-1}}{\mathbb{E}_{\lambda}[m_i^{-1}]}$, respectively, and using the law of iterated expectations.

Identification of $\mathbb{E}_{\lambda}\left[\frac{m_i^{-1}}{\mathbb{E}_{\lambda}[m_i^{-1}]}\frac{\theta_i}{\mathbb{E}_{\lambda}[\theta_i]}\right]$. $\mathbb{E}_{\lambda}\left[\frac{m_i^{-1}}{\mathbb{E}_{\lambda}[m_i^{-1}]}\frac{\theta_i}{\mathbb{E}_{\lambda}[\theta_i]}\right]$ is estimated directly from the data.

B.4.2. Measurement of m_i^{-1} and θ_i

Estimation of the demand elasticities θ_i. We estimate demand elasticities by inverting the formula for desired markups: $\theta_i = \frac{\mu_i^f}{\mu_i^f - 1}$. This strategy is infeasible for individual firms for two reasons: (i) individual markup measures are very noisy, with many values very close to or below 1, making the estimate of demand elasticities extremely sensitive to measurement error ; (ii) with sticky prices, individual firm markups will diverge from ideal markups. To circumvent this problem, we take seriously the prediction of the model that within an industry, variation in the ideal markup only comes from variation in firm-level market shares. Define Q bins of market shares. We assume that $\forall i \in \mathcal{Q}_q,\ \mu_i^f = \mu_q^f = \mathbb{E}_\lambda[\mu_i^f | \mathcal{Q}_q]$. With the assumption that the markup is constant within $\mathcal{Q}_q$, we then obtain $\mathbb{E}_\lambda[\theta_i | \mathcal{Q}_q] = \frac{\mu_q^f}{\mu_q^f - 1} = \theta_q$. Using the law of iterated expectations, we obtain $\mathbb{E}_\lambda[\theta_i]$ and $\mathbb{E}_\lambda\left[\frac{m_i^{-1}}{\mathbb{E}_\lambda[m_i^{-1}]}\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}\right]$.

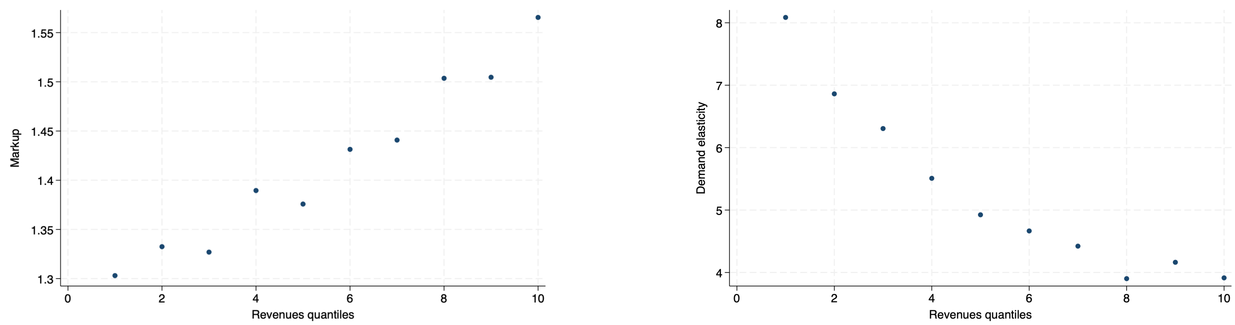
We estimate markups and use the assumption that on average over the whole sample, markups will be equal to desired flexible price markups, which we can invert to obtain demand elasticities. We estimate markups using the production approach, with materials as the flexible input. We rely on the elasticity of marginal costs with respect to quantities estimated in Table 2, which identifies the output elasticity of materials, so that our estimation does not suffer from the concern raised by Bond et al. (2021). Our estimation requires that any input wedge on materials is priced. We believe this assumption to be plausible; in particular, Singer (2019) documents that a large fraction of material inputs misallocation in India can be attributed to transportation costs that are reflected in prices recorded in the ASI.

Figure B.1 illustrates the markups and demand elasticities estimated in this way. We find that markups are increasing in firm's market shares, consistent with existing evidence. As a result, demand elasticities decline in firm size.

Note that because all the terms depend on $\frac{\theta_i}{\mathbb{E}_\lambda[\theta_i]}$, our procedure is robust to mis-measurement in the level of demand elasticities.

FIGURE B.1. Markups and demand elasticities by size

(a) Markups (b) Demand elasticities



Note: Panel (a) plots average markups by deciles of firms' market shares. Panel (b) plots demand elasticities by deciles of firms' market shares.

Estimation of the allocative distortion m_i^{-1}. As a reminder, we defined $m_i \equiv (1+\tau_i^v)\frac{\theta_i}{\theta_i-1}$. It is then straightforward to show that:

$$\text{(B.29)} \qquad m_i \propto \frac{p_i y_i}{v_i} = \left(\frac{p_i y_i}{l_i}\right)^{\phi} \left(\frac{p_i y_i}{m_i}\right)^{1-\phi} = \text{MRPL}^{\phi}\text{MRPM}^{1-\phi}$$

This formula shows that in the presence of both input-side distortions and markups, TFPR measures the combined distortion, as in Hsieh and Klenow (2009).

When taking this to the data, we consider two points. First, we integrate capital to our definition of wedges. While our baseline model omits capital, consistent with the focus on the response of marginal costs to monetary shocks, we do account for the fact that there are persistent distortions in marginal revenue products of capital across firms. Hence, we measure:

$$\text{(B.30)} \qquad m_i \propto \left(\frac{p_i y_i}{l_i}\right)^{\phi_l} \left(\frac{p_i y_i}{m_i}\right)^{\phi_m} \left(\frac{p_i y_i}{k_i}\right)^{\phi_k}$$

where ϕ_l, ϕ_m, ϕ_k are obtained from estimating production functions.

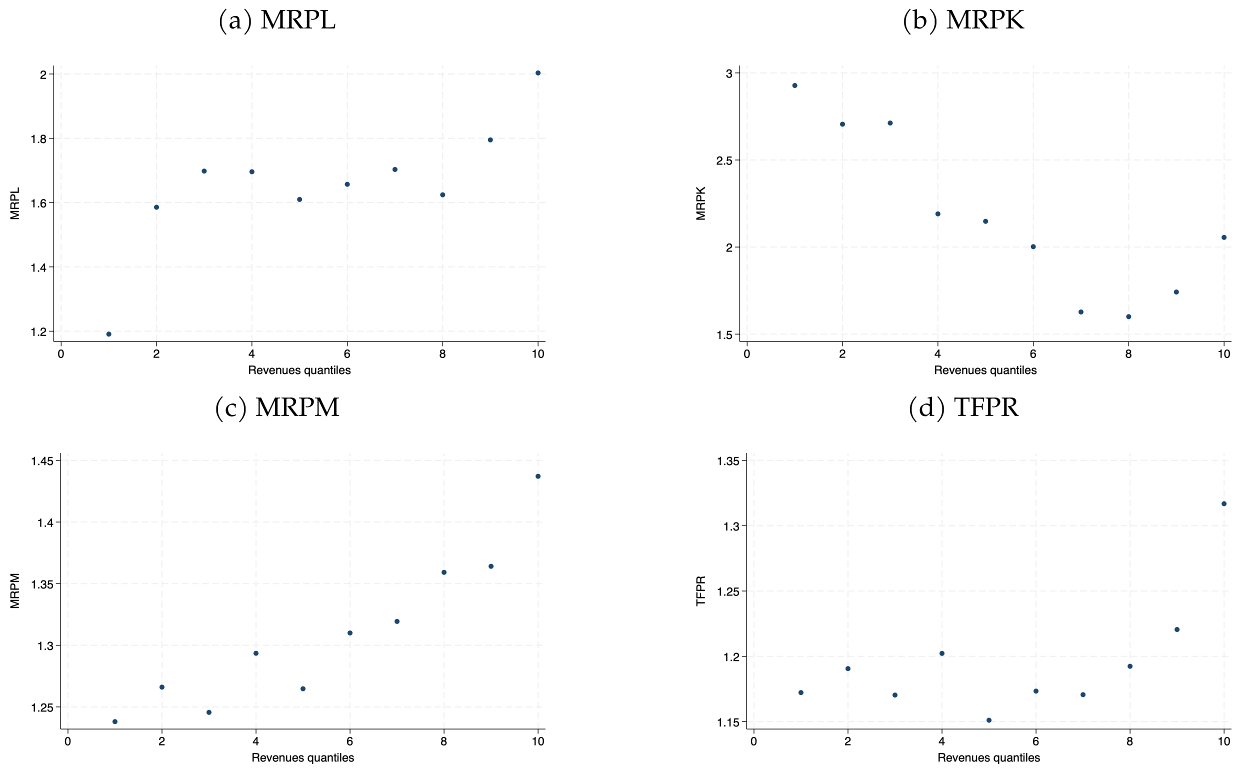
Second, constructing the material wedge $\frac{p_{it} y_{it}}{m_{it}}$ requires to divide sales by material quantity. Typical income statement items typically only report the purchase value of materials. This is a major issue: if material wedges are priced, the ratio $\frac{p_{it} y_{it}}{w_{it}^m m_{it}}$ would incorporate the wedge in the denominator of this ratio.

We overcome this issue by exploiting the decomposition of input purchases into quantities and unit values. This exercise requires that we have comparable quantities and unit values for each material inputs *across* firms, which is a more demanding requirement than our baseline tests exploiting within firm×product comparability. We therefore restrict our attention to inputs for which quantities and unit values are reported based on physical quantities: length, areas, volume, mass, and energy units. In our sample, 91% of the total purchase value of material inputs is recorded in physical units.

In practice, we proceed as follows. Let $\mathcal{K}_{PU} \subseteq \mathcal{K}$ be the subset of inputs denominated in a physical unit. Here we consider the set $\mathcal{K}$ to be the original input classification: 5-digit ASICC codes until 2010, and 7-digit NPC codes afterwards. We do so because units are defined at this level, and vary within the more harmonized codes we use in the main empirical exercises (because here we only compare inputs used by different firms in the same year, we do not need harmonized classifications). For each input $k \in \mathcal{K}_{PU}$, we deflate the input purchase value $m_{kit}w^m_{kit}$ by the firm-specific component of the input price w^m_{kit}/w^m_{kt}, where the denominator is computed as the weighted median of the price of input k across all firms in year t. We can then compute the ratio $\frac{p_{it}y_{it}}{m_{kit}w^m_{kt}}$. The denominator includes a constant across all firms, which is innocuous. We proceed in this fashion for all inputs $k \in \mathcal{K}_{PU}$ and aggregate at the firm level using input shares.

Figure B.2 shows the obtain marginal revenue products for labor, capital, intermediates, and the resulting TFPR, by deciles of firm market shares.

FIGURE B.2. MRPXs by size



Note: This figure plots averages of our estimates of marginal revenue products by bins of firm size. Within each industry, we define 10 bins of equal sales density.

Appendix C. Data

C.1. Indian Annual Survey of Industries (ASI)

The ASI is a dataset put together by India's Ministry of Statistics and Programme Implementation (MOSPI). The reference period for each survey is the accounting year, which in India begins on the 1st of April and ends on the 31st of March the following year. Throughout the paper we reference the surveys by the earlier of the two years covered.

Coverage and sampling methodology. The ASI contains information on a representative sample of manufacturing establishments, conditional on them taking part of the organized sector, and either employing more than 20 employees, or employing more than 10 employees and using electricity. We call the subpopulation of firms satisfying this criteria the ASI population. Within the ASI population, ASI defines a Census sector which is sampled exhaustively and a Sample sector for which the micro-data contains only a representative sample. Details of how the sampling methodology for the ASI changes over time are shown in Table C.1. ASI provides sampling weights, which we use to weight all data moments.

Table C.1. Sampling Methodology for Indian ASI

Period	Census Sector	Sample Sector
1998	Complete enumeration states, plants with >200 workers, all joint returns	Stratified within state $\times$ 4-digit industry (NIC-98), minimum of 8 plants per stratum
1999-2003	Complete enumeration states, plants with ≥ 100 workers, all joint returns	Stratified within state $\times$ 4-digit industry (NIC-98), 12% sampling fraction (20% in 2002), minimum of 8 plants per stratum
2004-2006	6 less industrially developed states, 100 or more workers, all joint returns, all plants within state $\times$ 4-digit industry with <4 units	Stratified within state $\times$ 4-digit industry, 20% sampling, minimum of 4 plants
2007	5 less industrially developed states, 100 or more workers, all joint returns, all plants within state $\times$ 4-digit industry with <6 units	Stratified within state $\times$ 4-digit industry, minimum 6 plants, 12% sampling fraction: exceptions
2008-2013	6 less industrially developed states, 100 or more workers, all joint returns, all plants within state $\times$ 4-digit industry with <4 units	Stratified within district $\times$ 4-digit industry, minimum 4 plants, 20% sampling fraction

Note: Baseline sampling fractions are shown, not accounting for state-specific exceptions.

We compare total value added by establishments in the ASI population to total manufacturing value added in India. The latter includes value added by the ASI population, value added by the organized sector establishments below the size threshold, and value added in the informal manufacturing sector. We find that ASI covers 61% of total manufacturing value added on average across years.

Sample selection. We start with 1,068,114 plant $\times$ year observations. We subsequently employ multiple sample selection rules. First, we restrict the sample to factory $\times$ year observations with either positive reported gross sales, or positive reported sales at the factory gate. This drops one third of all observations (360,145). Next, we disregard all observations that exactly copied their sales from the previous year, suspecting these plants to be actually closed. This drops 1,179 additional observations. Third, we drop all plant × years that reported either no days worked, or no persons employed, dropping a supplementary 206 observations. These cleaning steps follow Martin, Nataraj, and Harrison (2017).

Moreover, we restrict the sample to observations with correct accounting: we drop all observations where the difference between aggregate items and the absolute value of the sum of corresponding sub-items exceeded 10% of the aggregate items. We regard four aggregate items: (i) the purchase value of basic items (including imports), (ii) the purchase value of non-basic items, (iii) the purchase value of total inputs, and (iv) the gross sales value of output. This accounting rule drops an extra 4,098 observations. Finally, disregarding all factory × years without any reported positive output or input (including energy) values at the product level, excludes an additional 49,166 observations. The final sample thus includes 193,352 unique plants for a total of 653,320 individual plants × year observations.

Industry classification. Our data relies on three distinct industry classification systems: NIC-98 (1998–2003), NIC-04 (2004–2007), and NIC-08 (2008 and beyond). We first address issues with the 5-digit industry codes in NIC-98, where codes are sometimes masked with zeroes or absent from the official documentation, by replacing them with the most frequent 5-digit code within each 4-digit grouping following the approach of Martin, Nataraj, and Harrison (2017). Next, we apply concordances from NIC-08 to NIC-04 using the mapping provided by Rijesh (2022), and from NIC-04 to NIC-98 using the mapping provided by Martin, Nataraj, and Harrison (2017). We manually supplement the mappings for industries not covered by these concordances. In case of 1:m mappings, we select the appropriate industry based on transition matrices in the micro-data.

Product classification. Our analysis standardizes product classifications across four distinct classifications used in our sample: NPCMS 2015 (2016-2017), NPCMS 2011 (2010-2015), ASICC 2009 (2008-2009), and ASICC 2008 (pre-2008). We harmonize all product codes to NPCMS 2011, as it provides a well-defined five-digit structure that balances granularity and coverage. Given the absence of an official concordance between NPCMS 2015 and NPCMS 2011, we constructed a mapping using fuzzy matching (based on product codes, descriptions, and units) and semantic embeddings (OpenAI's AA2 model). For ASICC 2009, we utilize the official concordance to NPCMS 2011 but address its limitations (such as missing mappings and invalid classification) by leveraging ASI data and semantic embeddings. The harmonization of ASICC 2008 to ASICC 2009 follows the concordance from Boehm, Dhingra, and Morrow (2022). Table C.2 shows an excerpt of the product classification.

Table C.2. Example of NPC-MS 2011 5-digit classification

Code	Description
35	Other chemical products; man-made fibres
351	Paints and varnishes and related products; artists' colours; ink
35110	Paints and varnishes and related products
35120	Artists', students' or signboard painters' colours, modifying tints, amusement colours and the like
35130	Printing ink
35140	Writing or drawing ink and other inks
352	Pharmaceutical products
353	Soap, cleaning preparations, perfumes and toilet preparations
354	Chemical products n.e.c.
355	Man-made fibres
36	Rubber and plastics products
361	Rubber tyres and tubes
36111	New pneumatic tyres, of rubber, of a kind used on motor cars
36112	New pneumatic tyres, of rubber, of a kind used on motorcycles or bicycles
36113	Other new pneumatic tyres, of rubber
36114	Inner tubes, solid or cushion tyres, interchangeable tyre treads and tyre flaps, of rubber
36115	Camel back strips for retreading rubber tyres
36120	Retreaded pneumatic tyres, of rubber
362	Other rubber products
36210	Reclaimed rubber
36220	Unvulcanized compounded rubber, in primary forms or in plates, sheets or strip; unvulcanized rubber in forms other than primary forms or plates, sheets or strip
36230	Tubes, pipes and hoses of vulcanized rubber other than hard rubber
36240	Conveyor or transmission belts or belting, of vulcanized rubber
36250	Rubberized textile fabrics, except tyre cord fabric
36260	Articles of apparel and clothing accessories (including gloves) of vulcanized rubber other than hard rubber
36270	Articles of vulcanized rubber n.e.c.; hard rubber; articles of hard rubber
363	Semi-manufactures of plastics
364	Packaging products of plastics
369	Other plastics products

Note: For codes other than 351, 361 & 362, the 5-digit classifications are not shown.

District identifier. There are two versions of the ASI data: panel data and cross-sectional data. Throughout our analysis, we use the panel version because it allows us to track establishments over time. However, the main drawback of the panel data is the absence of district identifiers. To address this, we follow the methodology proposed by Martin, Nataraj, and Harrison (2017), with some minor modifications, to obtain the district identifiers based on 1998 district borders. The 1998 borders represent the most aggregated administrative division during our sample period. Using this approach, we are able to identify 497 unique districts in our final sample.

The first step is to merge the ASI panel data (1998–2009) with the ASI cross-sectional data to obtain district codes by year. Since the firm IDs differ between the two versions, the merge is performed based on a series of specific factory characteristics, similar to what is used in Martin, Nataraj, and Harrison (2017). To obtain a consistent set of district identifiers based on 1998 boundaries, we use the concordance provided by Martin, Nataraj, and Harrison (2017). This gives us district identifiers for 98% of firm-year observations before 2010 in the final sample. For 2010-2017 the ASI cross-sectional files mask the district code, so we assign districts only where the establishment appears at least once before 2010. This yields valid district identifiers for about 63% of observations starting in 2010 in the final sample.

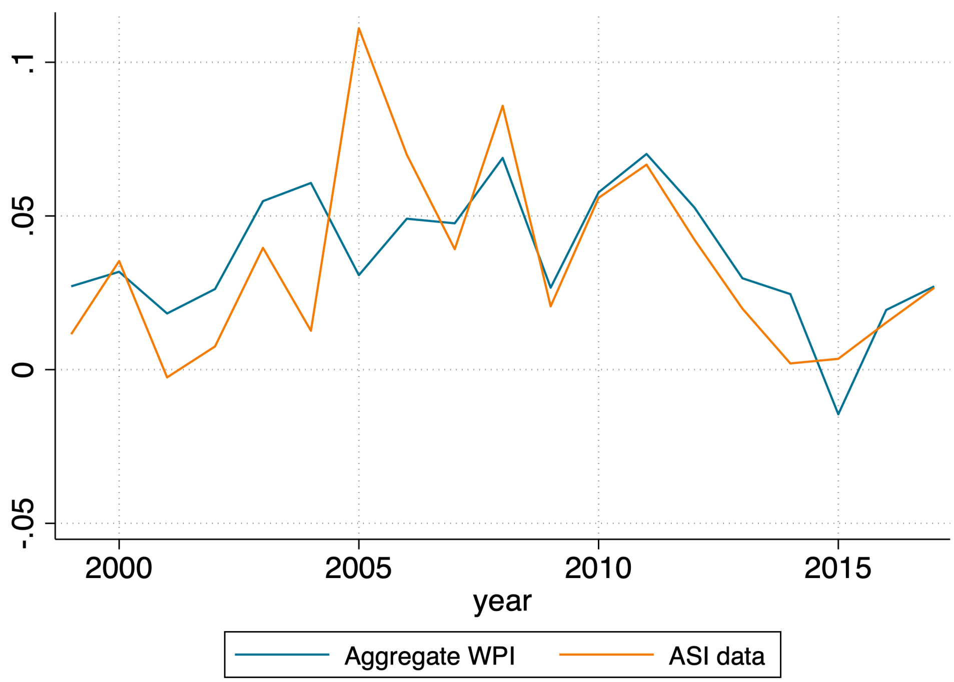
Building firm$\times$product- and firm$\times$input-level price changes. The construction of firm $\times$ product(input)-level price change involves harmonizing product codes and adjusting prices and quantities.

Harmonization of product codes. First, we notice that firms often report distinct product codes within a narrow category (e.g., a 4-digit grouping) in consecutive years, even when they produce a single product within this category in each of the years. Investigating these cases, we conclude that these cases often correspond to misreported codes. We alleviate this issue by harmonizing product codes within firms by assigning new product codes based on the most common existing codes in cases where a single consistent product code exists per 4- or 3-digit classification. This affects 7% observations in the products data and 6.6% observations in the inputs data.

Cleaning price changes. Next, we address discrepancies between reported and imputed price and quantity variables. Imputed price (quantity) values are constructed by dividing the output value by the quantity (price). Since it is unclear which value is accurate, we implement several adjustments: replacing zero reported prices or quantities with imputed values, using imputed prices when the reported prices appear to be calculated using the wrong formula, and substituting reported prices with imputed ones when manufactured and sold quantities are very similar but reported and imputed prices differ significantly. Lastly, we use the imputed price (or quantity) if the difference between the reported price and its within-firm mean is greater than that of the quantity (and vice versa), provided both are within the same order of magnitude. For the input dataset, we apply fewer techniques due to the limited number of variables available for verification (i.e., there is no manufactured quantity). Finally, as noted by (Boehm, Dhingra, and Morrow 2022), the data contains unit mistakes due to misplaced commas. To address this, we rescale values up or down when the price was multiplied by 10^n and the quantity was multiplied by 10^{-n} with respect to the previous year, for $n \in [-9..9]$.

Data quality check. Figure C.1 matches the yearly price index for the manufacturing sector from WPI, constructed as the average of quarterly price indices, to the price index generated from ASI data. The latter is constructed as the sales-weighted median of log-price changes, with weights being constructed as Törnqvist weights. While the ASI data appears to have a little more volatility, especially between 2005 and 2008, overall the two lines show a very similar pattern.

FIGURE C.1. WPI inflation and inflation from ASI micro-data



Note: This figure plots two inflation series. The blue line is WPI inflation (manufacturing). The pink line is price growth obtained from the micro-data. It is constructed as the sales-weighted median of log-price changes, with weights being constructed as Törnqvist weights.

Table C.3. Summary statistics

Panel A: Firm level

	P10	P50	P90	Mean	SD
Price change $\Delta \log p_{it}$	-0.388	0.025	0.445	0.028	0.430
Output change $\Delta \log y_{it}$	-0.627	0.019	0.580	-0.005	0.672
Variable cost change $\Delta \log \mathcal{C}_{it}$	-0.312	0.069	0.389	0.037	0.439
Marginal cost change $\Delta \log mc_{it}$	-0.507	0.034	0.609	0.047	0.647
Sales change	-0.326	0.060	0.368	0.017	0.468
Raw mats. purchase value change	-0.361	0.066	0.422	0.029	0.529
Energy purchase value change	-0.387	0.061	0.471	0.048	0.640
Observations	653,320				

Panel B: Firm $\times$ product level

	P10	P50	P90	Mean	SD
Price change $\Delta \log p_{ijt}$	-0.451	0.022	0.510	0.027	0.603
Output change $\Delta \log y_{it}$	-0.764	0.015	0.724	-0.003	0.997
Observations	982,083				

Note: This table presents summary statistics of firm and firm $\times$ product level variables.

Appendix D. Additional results

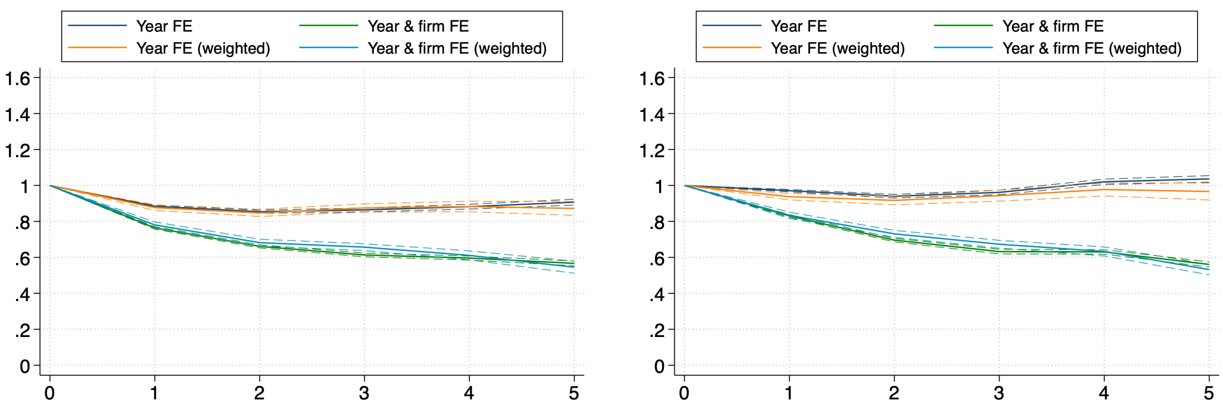
D.1. Firm-level pass-through of cost shocks into prices

Autocorrelation of the input cost shock. In the model, ϑ_{it}^{m} is the exogenous %-change in the price paid for inputs between the steady-state (denoted t_0) and time t. The empirical counterpart is the sum of instruments from $t_0 + 1$ to t (in each period, the instrument shifts the $t-1$ to t growth rate). To estimate the autocorrelation of the cost disturbance, we therefore estimate the following regressions:

$$\sum_{s=0}^{h} \mathcal{Z}_{it+s}^{\vartheta} = \beta_h \mathcal{Z}_{it}^{\vartheta} + \varepsilon_{it} \tag{D.1}$$

Figure D.1. Autocorrelation of input cost shock

(a) Instrument A (b) Instrument B



Robustness checks. The following tables present robustness checks of the firm-level pass-through of cost shocks into prices.

TABLE D.1. Firm level elasticity of price changes to input cost changes

	$\Delta \log p_{it}$								
	OLS			Instrument A			Instrument B		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\Delta \log w_{it}$	0.106***	0.088***	0.087***	0.281***	0.215***	0.215***	0.468***	0.272***	0.271***
	(0.007)	(0.007)	(0.007)	(0.017)	(0.017)	(0.017)	(0.040)	(0.048)	(0.048)
Year FE	✓			✓			✓		
Year × Ind. FE		✓	✓		✓	✓		✓	✓
Controls			✓			✓			✓
Observations	270,356	267,001	267,001	270,046	266,719	266,719	269,498	266,218	266,218
F-stat				5153.4	4758.7	4771.9	777.5	533.8	533.9
Adj. passthrough				0.258	0.198	0.198	0.429	0.249	0.249

Note: This table presents robustness checks of the main results shown in 1. All regressions are estimated at the firm level. Columns (4)–(6) use the instrument defined in (12), while columns (7)–(9) use the instrument defined in (13). Regressions are weighted by firm-level lagged sales and adjusted using ASI sampling weights (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table D.2. Elasticity of price changes to input cost changes: Additional controls and fixed effects

Panel A: Instrument A

	$\Delta \log p_{ijt}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log w_{it}$	0.208***	0.218***	0.219***	0.222***	0.214***	0.217***	0.200***	0.190***
	(0.018)	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)	(0.015)	(0.017)
Year × Product FE	✓	✓	✓	✓	✓		✓	✓
Year × State FE			✓					
Year × State × Product FE				✓				
Year × Ind. FE					✓			
Year × Alt. Product FE						✓		
Firm × Product FE								✓
Markup Controls	✓							
Demand Control		✓						
Price Index	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Alt.	Alt.
Observations	301,052	363,951	364,517	316,320	360,576	365,090	364,517	309,186
F-stat	3780.9	4534.9	4657.7	4688.3	4603.3	4578.4	141561.5	117916.7
Adj. passthrough	0.192	0.201	0.202	0.205	0.197	0.200	0.184	0.178

Panel B: Instrument B

	$\Delta \log p_{ijt}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log w_{it}$	0.210***	0.214***	0.217***	0.206***	0.218***	0.213***	0.183***	0.159***
	(0.048)	(0.043)	(0.043)	(0.059)	(0.044)	(0.043)	(0.035)	(0.039)
Year × Product FE	✓	✓	✓	✓	✓		✓	✓
Year × State FE			✓					
Year × State × Product FE				✓				
Year × Ind. FE					✓			
Year × Alt. Product FE						✓		
Firm × Product FE								✓
Markup Controls	✓							
Demand Control		✓						
Price Index	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Alt.	Alt.
Observations	300,558	363,234	363,800	315,800	359,908	364,373	363,800	308,488
F-stat	397.9	511.0	474.8	262.1	501.9	510.3	2202.7	1958.0
Adj. passthrough	0.194	0.197	0.200	0.190	0.201	0.196	0.168	0.148

Note: This table presents robustness checks of the main results shown in 1. All regressions are estimated at the firm×product level. Different columns include alternative combinations of fixed effects and additional controls. Alt. Product FE refers to the bunched similar NPCMS 2011 product codes. Demand control refers to the instrument defined in (17). The Baseline Price Index is computed as the difference between the change in total variable cost (materials, energy, and labor) and the change in quantity, while the Alt. Price Index is a weighted average of input price changes. Panel A reports IV results using the instrument defined in (12), while Panel B uses the instrument defined in (13). Regressions are weighted by firm-by-product-level lagged sales and adjusted using ASI sampling weights (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

TABLE D.3. Elasticity of price changes to input cost changes: Sample cuts

Panel A: Instrument A

	$\Delta \log p_{ijt}$					
	Dereservation Policy		Demonetization Eps. Drop 2016		Omitting 2004 & 2005	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	0.220***	0.214***	0.221***	0.221***	0.221***	0.212***
	(0.016)	(0.018)	(0.017)	(0.021)	(0.016)	(0.019)
Year × Product FE	✓	✓	✓	✓	✓	✓
Firm × Product FE		✓		✓		✓
Observations	351,655	297,062	330,269	271,873	334,611	281,454
F-stat	4963.9	3785.1	3928.2	2702.9	4640.6	3423.9
Adj. passthrough	0.203	0.200	0.204	0.207	0.205	0.198

Panel B: Instrument B

	$\Delta \log p_{ijt}$					
	Dereservation Policy		Demonetization Eps. Drop 2016		Omitting 2004 & 2005	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	0.220***	0.204***	0.215***	0.187***	0.219***	0.193***
	(0.042)	(0.047)	(0.045)	(0.051)	(0.044)	(0.048)
Year × Product FE	✓	✓	✓	✓	✓	✓
Firm × Product FE		✓		✓		✓
Observations	350,952	296,376	329,635	271,289	333,938	280,797
F-stat	562.3	424.3	451.7	314.6	529.9	392.5
Adj. passthrough	0.203	0.188	0.198	0.174	0.202	0.181

Note: This table presents robustness checks of the main results presented in 1. Different columns represent different sample cuts and fixed effects. "Dereservation Policy" drops product×year cells when a product loses its legal restriction to be produced only by small-scale firms. "Demonetization Eps." drops product×year cells in 2016, the year India invalidated 500 and 1000 rupee notes. "Omitting 2004 & 2005" drops product×year cells in 2004 and 2005, the years where, as shown in Figure C.1, inflation computed from ASI micro-data is far from the actual WPI inflation (manufacturing). Panel A reports IV results with the instrument defined in (12). Panel B reports IV results with the instrument defined in (13). Regressions are weighted by firm×product-level lagged sales, adjusted for the ASI sampling weight (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table D.4. Elasticity of price changes to input cost changes: Disinflation episode

Panel A: Instrument A

	$\Delta \log p_{ijt}$			
	Disinflation Eps.			
	1998–2013		2014–2017	
	(1)	(2)	(3)	(4)
$\Delta \log w_{it}$	0.256***	0.253***	0.173***	0.148***
	(0.025)	(0.030)	(0.021)	(0.026)
Year × Product FE	✓	✓	✓	✓
Firm × Product FE		✓		✓
Observations	223,972	177,672	140,545	112,533
F-stat	1609.6	1084.0	4883.0	3945.5
Adj. passthrough	0.235	0.236	0.161	0.141

Panel B: Instrument B

	$\Delta \log p_{ijt}$			
	Disinflation Eps.			
	1998–2013		2014–2017	
	(1)	(2)	(3)	(4)
$\Delta \log w_{it}$	0.247***	0.177**	0.168***	0.167**
	(0.065)	(0.074)	(0.052)	(0.069)
Year × Product FE	✓	✓	✓	✓
Firm × Product FE		✓		✓
Observations	223,594	177,315	140,206	112,180
F-stat	189.2	131.2	709.0	498.3
Adj. passthrough	0.228	0.165	0.155	0.158

Note: This table presents robustness checks of the main results presented in 1. Different columns represent different sample cuts and fixed effects. "Disinflation Eps." separates pre- and post-periods of India's disinflation episode under Governor Rajan. Panel A reports IV results with the instrument defined in (12). Panel B reports IV results with the instrument defined in (13). Regressions are weighted by firm×product-level lagged sales, adjusted for the ASI sampling weight (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table D.5. Elasticity of price changes to input cost changes: Non-linearity

Panel A: Instrument A

	$\Delta \log p_{ijt}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	0.217***	0.213***	0.212***	0.213***	0.210***	0.215***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
Year × Product FE	✓	✓	✓	✓	✓	✓
Observations	364,517	351,301	341,720	314,344	273,493	206,476
F-stat	4,551.5	4,522.0	4,426.3	4,345.1	4,118.2	3,769.6
Excl. band	None	[-0.005,0.005]	[-0.01,0.01]	[-0.025,0.025]	[-0.05,0.05]	[-0.10,0.10]
Adj. passthrough	0.200	0.197	0.196	0.197	0.193	0.198

Panel B: Instrument B

	$\Delta \log p_{ijt}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	0.214***	0.206***	0.211***	0.204***	0.213***	0.222***
	(0.043)	(0.042)	(0.043)	(0.043)	(0.042)	(0.044)
Year × Product FE	✓	✓	✓	✓	✓	✓
Observations	363,800	350,610	341,041	313,719	272,948	206,022
F-stat	507.7	505.2	497.1	475.9	448.7	368.1
Excl. band	None	[-0.005,0.005]	[-0.01,0.01]	[-0.025,0.025]	[-0.05,0.05]	[-0.10,0.10]
Adj. passthrough	0.197	0.190	0.194	0.188	0.196	0.204

Note: This table presents robustness checks of the main results presented in 1. Each column drops observations where $\Delta \log w_{it}$ is within certain values, as indicated in the row "Excl. band". Panel A reports IV results with the instrument defined in (12). Panel B reports IV results with the instrument defined in (13). Regressions are weighted by firm×product-level lagged sales, adjusted for the ASI sampling weight (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

TABLE D.6. Elasticity of price changes to input cost changes: Dynamic effects

Panel A: Instrument A

	$\Delta \log p_{ijt}$					
	h=0		h=1		h=2	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	0.217***	0.209***	0.236***	0.196***	0.219***	0.168***
	(0.016)	(0.018)	(0.025)	(0.027)	(0.034)	(0.032)
Year × Product FE	✓	✓	✓	✓	✓	✓
Firm × Product FE		✓		✓		✓
Observations	364,517	309,186	210,999	187,614	136,034	122,790
F-stat	4551.5	3399.6	2337.5	1874.0	1416.7	1097.4
β^i/β^0			1.088	0.939	1.009	0.802

Panel B: Instrument B

	$\Delta \log p_{ijt}$					
	h=0		h=1		h=2	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	0.214***	0.186***	0.127*	0.131*	0.161*	0.168*
	(0.043)	(0.047)	(0.069)	(0.073)	(0.089)	(0.091)
Year × Product FE	✓	✓	✓	✓	✓	✓
Firm × Product FE		✓		✓		✓
Observations	363,800	308,488	210,605	187,237	135,783	122,541
F-stat	507.7	373.8	244.5	191.4	151.8	113.3
β^i/β^0			0.593	0.705	0.755	0.903

Note: This table presents robustness checks of the main results presented in 1. Different columns represent dynamic effects at different horizons ($h = 0, 1, 2$). Panel A reports IV results with the instrument defined in (12). Panel B reports IV results with the instrument defined in (13). Regressions are weighted by firm×product-level lagged sales, adjusted for the ASI sampling weight (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

TABLE D.7. Elasticity of quantity changes to input cost changes: Dynamic effects

Panel A: Instrument A

	$\Delta \log y_{ijt}$					
	h=0		h=1		h=2	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	-0.038*	-0.076***	-0.074**	-0.075**	-0.138***	-0.159***
	(0.022)	(0.024)	(0.034)	(0.035)	(0.048)	(0.047)
Year × Product FE	✓	✓	✓	✓	✓	✓
Firm × Product FE		✓		✓		✓
Observations	364,517	309,186	210,999	187,614	136,034	122,790
F-stat	4551.5	3399.6	2337.5	1874.0	1416.7	1097.4

Panel B: Instrument B

	$\Delta \log y_{ijt}$					
	h=0		h=1		h=2	
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \log w_{it}$	-0.061	-0.079	-0.041	-0.053	-0.238*	-0.278**
	(0.061)	(0.065)	(0.102)	(0.099)	(0.133)	(0.129)
Year × Product FE	✓	✓	✓	✓	✓	✓
Firm × Product FE		✓		✓		✓
Observations	363,800	308,488	210,605	187,237	135,783	122,541
F-stat	507.7	373.8	244.5	191.4	151.8	113.3

Note: This table presents robustness checks of the main results presented in 1. Different columns represent dynamic effects at different horizons ($h = 0, 1, 2$). Panel A reports IV results with the instrument defined in (12). Panel B reports IV results with the instrument defined in (13). Regressions are weighted by firm×product-level lagged sales, adjusted for the ASI sampling weight (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

D.2. Elasticity of marginal costs with respect to quantities

TABLE D.8. The partial equilibrium elasticity of marginal costs to changes in quantity, additional controls, and fixed effects

	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log y_{it}$	0.151**	1.069***	0.211**	1.129***	0.351***	1.216***	0.151**	1.065***
	(0.074)	(0.073)	(0.100)	(0.099)	(0.107)	(0.102)	(0.075)	(0.074)
Year × Ind. FE	✓	✓			✓	✓	✓	✓
Year × State FE	✓	✓						
Year × Ind. × State FE			✓	✓				
Markup controls					✓	✓		
Cost shock control							✓	✓
Observations	267,011	267,011	260,894	260,894	224,470	224,470	266,719	266,719
F-Stat	180	180	112	112	112	112	172	172
Returns to scale	0.87	0.94	0.83	0.89	0.74	0.82	0.87	0.94

Note: This table presents robustness checks of the main results presented in 2. Different columns include alternative combinations of fixed effects and additional controls. Cost shock control refers to instruments defined in (12). Regressions are weighted by firm-level lagged sales (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively

Table D.9. The partial equilibrium elasticity of marginal costs to changes in quantity, sample cuts

	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$	$\Delta \log mc_{it}$	$\Delta \log \mathcal{C}_{it}$
	Dereservation Policy		Demonetization Eps. Drop 2016		Disinflation Eps.			
					1998–2013		2014–2017	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\Delta \log y_{it}$	0.170**	1.082***	0.203**	1.128***	0.225**	1.156***	0.068	0.951***
	(0.077)	(0.075)	(0.087)	(0.086)	(0.109)	(0.108)	(0.097)	(0.089)
Year × Ind. FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	256,292	267,011	242,974	242,974	169,418	169,418	97,593	97,593
F-stat	166.9	171.9	142.7	142.7	97.2	97.2	88.2	88.2
Returns to scale	0.85	0.92	0.83	0.89	0.82	0.87	0.94	1.05

Note: This table presents robustness checks of the main results presented in 2. Different columns represent different sample cuts. "Dereservation Policy" drops product×year cells when a product loses its legal restriction to be produced only by small-scale firms. "Demonetization Eps." drops product×year cells in 2016, the year India invalidated 500 and 1000 rupee notes. "Disinflation Eps." separates pre- and post-periods of India's disinflation episode under Governor Rajan. Regressions are weighted by firm-level lagged sales (top and bottom 1% winsorized). Standard errors are clustered at the firm level. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively