Firm productivity and derived factor demand: when market power leads to a decoupling

Filippo Biondi*

Job Market Paper

Preliminary and still evolving draft. Click here for the latest version.

August 16, 2022

Abstract

The common presumption in the literature is that after a positive productivity shock firms expand production and demand more input factors. While this is true in perfectly competitive markets, I argue that this is not always the case under imperfect competition. Whenever firms that produce more manage to set higher markups, their derived factor demand becomes gradually less responsive to productivity shocks and may even turn negative, leading to a decoupling of factor demand from productivity growth. As this ultimately depends on the price elasticity of demand faced by each firm, I characterize the theoretical conditions for this result to emerge both in terms of the features of output demand and market structure. I show that many widely-used demand functions lead to a non-monotonic relationship between input and productivity and this occurs in both monopolistic and oligopolistic settings, especially if firms face low competitive pressures. To assess the empirical relevance of this mechanism, I develop a new approach to detect whether larger firms scale back in terms of input use after a productivity increase. Applying it to Chinese manufacturing firm-level data, I find evidence of it in more than 20% narrowly-defined industries, markedly where larger firms set higher markups.

Keywords: *productivity, derived factor demand, firm-level markups, incomplete pass-through.*

*PhD Candidate. KU Leuven and Research Foundation Flanders. E-mail: *filippo.biondi@kuleuven.be*

I am extremely grateful to Jo Van Biesebroeck and Jan De Loecker for their continuous mentoring and support, to Frank Verboven, Mathieu Parenti, Fabio Bertolotti, John Morrow, Nicola Fontana and participants at the KU Leuven IO seminars, the CEPR Applied IO conference for their fruitful comments and suggestions. Financial support from the Research Foundation Flanders (FWO) is gratefully acknowledged. Any remaining errors are solely my responsibility.

1 Introduction

Firm productivity growth is crucial in the analysis of many economic phenomena. This is notably the case in the study of industry dynamics, where firm-level productivity shocks are thought to be key determinants of firms' behavior, influencing both their output and input decisions. In recent years, however, new evidence has emerged on a decline in the responsiveness of firms' factor demand to productivity shocks (Ilut, Kehrig, and Schneider, 2018; Decker, Haltiwanger, Jarmin, and Miranda, 2020).

Identifying the mechanisms underlying this result has become the focus a rapidly growing area of research at the intersection of industrial organization and macroeconomics (Kehrig and Vincent, 2021; De Loecker, Eeckhout, and Mongey, 2021). On one hand, because this decline in responsiveness raises a fundamental distributional concern: if productivity growth is not completely passed on through higher output, lower price and higher input factors demand, either consumers or suppliers of factors are likely to be affected. On the other, because the (un)responsiveness of factor demand to productivity is puzzling from a theoretical point of view. The common presumption in the literature on firm heterogeneity (Syverson, 2011; De Loecker and Syverson, 2021), indeed, is that a input factor demand is always monotonic increasing in firm productivity. While this is true in workhorse models of perfect competition (Levinsohn & Petrin, 2003), in this paper I prove that this is not always the case under imperfect competition.

When firms can increase their markups by producing more, I prove that their factor demand becomes gradually less responsive to Hicks-neutral productivity shocks and may even turn negative, leading to a decoupling of input from productivity growth. The contribution of my paper is to characterize the theoretical conditions that lead to this result, propose a new approach to detect it with standard firm-level data, and show its empirical relevance.

To provide the intuition for this overlooked result, consider a firm that becomes more productive. How would this firm adjust its derived factor demand, i.e. the amount of input used to maximize its profits? This is the crucial comparative static I focus on throughout the paper. A positive productivity shock generates two opposing effects. On one hand, a higher level of productivity means that the firm can produce the same level of output with fewer inputs. On the other, it means that also its marginal costs are lower, which creates the incentive to expand output to maximize its profits. To meet these higher production targets, the firm needs more input factors. In general, the net effect on factor demand depends on how much the firm decides to expand its output, which is an equilibrium outcome influenced by the features of the market in which it operates.

If a firm exerts some market power over its customers and faces a less elastic demand whenever it produces more, I show that the incentive to expand its output after a productivity shock gradually declines. This is because at a lower price elasticity of demand, the marginal revenue generated by an additional unit of output gets lower and lower. At high levels of output, demand becomes nearly-satiated and to convince customers to buy 1% more output, the price of the firm's product must fall so much that marginal revenue starts decreasing by more. In such a situation, the firm takes its "foot off the gas" and expands output by less than 1%. It does so precisely to prevent its marginal revenue from declining too rapidly and to keep maximizing its profits (by raising its markups). Since the productivity improvement is more than enough to produce the extra output, no additional input is needed. As a result, input demand halts and starts decreasing.

This incomplete pass-through of productivity to output is the key mechanism through which market power can mute the responsiveness of factor demand to productivity shock. As it is ultimately driven by the price elasticity of demand faced by a firm, I characterize the conditions for this result both in terms of the features of market demand and the nature of competition.

Starting from a monopoly, I identify for each functional form of demand the level of price elasticity of demand below which a firm expands its output less than proportionally to a productivity shock. The value of the curvature of demand determines this level since it governs the rate at which the elasticity declines with output. To understand which demand systems lead to this result, I bring this insight into the demand manifold framework developed by Mrázová and Neary (2017). This framework allows comparing demand functions based only on their implied relationship between the values of the elasticity and curvature, the so-called demand manifold. Building on it, I show that a non-monotonic relationship between derived factor demand and productivity can arise in many demand specifications that are commonly used in the literature. Among others, it arises in linear demand, the Bulow and Pfleiderer (1983) demand, the Stone-Geary/Linear expenditure system, the CARA system used by Behrens, Mion, Murata, and Suedekum (2020), the Logistic demand (Cowan, 2016), the Klenow and Willis (2016) specification of Kimball demand, and many other demand functions considered by Mayer, Melitz, and Ottaviano (2021) that follow the 2$^{\text{nd}}$ Marshall's law of demand.

Characterizing this result at such a level of generality allows me to connect this result with many other comparative statics predictions that are driven by the values of elasticity and curvature. In particular, I show that the derived factor demand halts and starts declining after a productivity shock in correspondence to high levels of markups and low (incomplete) pass-through rates of cost to price. For example, with linear demand this result occurs at a level of elasticity of 3, which implies a level of markups equal to 1.5. For concave demands, the level of markups is even lower. To my knowledge, this connection between the responsiveness of factor demand and pricing behavior has been overlooked in the literature so far.

Beyond monopoly, I argue that a similar mechanism is at play also in standard models of monopolistic competition with free entry and of oligopolistic competition in quantity.[1] In these settings, however, the relevant elasticity for firms' behavior is not the price elasticity of market demand but the elasticity of the *residual* demand faced by each firm. For this reason, in addition to the features of market demand, I show that any factor that reduces the competitive pressures perceived by each firm will reduce the responsiveness of its factor demand to productivity shock. The main implication is that, within a market, firms react very differently to the same productivity shock, with the largest firms (in terms of output) having the lowest and potentially negative responsiveness. In a monopolistic competitive setting, this prediction also holds in levels implying that more productive firms are not necessarily the largest firms in terms of input. Second, I show that for a given demand function, a decoupling of factor demand from productivity is more likely to emerge in less competitive markets. This is particularly clear under oligopoly, where I show that in markets with a fewer competitors and/or a more collusive conduct the responsiveness is lower for all firms. Moreover, in oligopoly, I find that even CES demand can lead to a decoupling of factor demand to productivity growth at lower elasticity. This is because strategic interactions lead to variable elasticity of residual demand even if the elasticity of market demand is constant.

Throughout the paper, I consider a simple setting with a single input factor, a technology with constant returns to scale, and price-taking behavior in the input markets. However, I prove that the same insights hold in more general environments with more than one variable input factor, with non-constant returns to scale and with market power in the input market. In doing so, I find that decreasing returns to scale and monopsony power in the factor market further reduces the responsiveness of derived factor demand to productivity shocks.

Having established the generality of this result from a theoretical point of view, I then assess its empirical relevance. To detect whether a firm operates in the range of elasticities where this non-monotonicity has kicked-in (or not), the ideal approach would be to observe a firm decreasing its input use while increasing its output after a well-identified positive Hicks-neutral productivity shock. However, the presence of other con-

[1] After all in these other market structures, firms behave like a monopolist on their residual demands, even in the presence of horizontal product differentiation or strategic interactions among them.

temporaneous demand and cost shocks makes it difficult (if not impossible) to do that since these other shocks always lead a firm to increase both its input and output. An additional challenge is empirical and posed by the fact that in most firm-level datasets, output is reported in terms of revenue rather than physical quantities. In an imperfect competitive setting with variable price elasticity of demand, this prevents researchers from estimating productivity and even more identifying productivity shocks.

Given these challenges, I develop two approaches that can be applied by researchers even if firm-level productivity cannot be estimated and without taking a stance on the specific functional form of demand faced by the firms. Under certain assumptions, indeed, I show that it is possible to infer directly from data on revenue and input (in levels and in changes) whether firms in a given market react to a productivity shock by decreasing their input use.

Finally, I apply these two detection tests to the Chinese manufacturing firm-level data. Building on the work by Brandt, Van Biesebroeck, Wang, and Zhang (2017), I use the data from the census of "above-scale" manufacturing establishments conducted by the National Bureau of Statistics (NBS) over the period 1998-2007.[2] This is a period of intense productivity growth and structural transformation in which China emerged as "the world's factory", which is ideal for studying the impact of productivity shocks. The dataset contains information for hundreds of narrowly-defined industries. Moreover, it provides text descriptions for (up to) three main products of each firm. This allows me to restrict the analysis to single-product firms that did not introduce new products over a period, reducing the sources of potential biases in the application of the tests. In preliminary analyses, I find evidence of larger firms decreasing their input use after a productivity shock in more than 20% of the industries. Consistently with my theoretical prediction, significantly more in industries where larger firms set higher markups.

Related literature. This paper is related and contributes to a growing literature, at the intersection of industrial organization and macroeconomics, on the origins and consequences of the declining responsiveness of factor demand, particularly labor demand, to productivity shocks. This paper contributes to the debate by highlighting that the features of output demand and market structure heavily impact the connection between product and factor markets. In particular, it is mostly related to the recent work by De Loecker et al. (2021). In their model, the authors show that in an oligopolistic setting with nested-CES demand, productivity shocks lead to smaller changes in output by firms with higher markups, which in turn reduce the rate of adjustment of their employment. While the focus of their analysis is on the role of strategic interactions among firms, I prove that this is a much more general result and the mechanism can be so strong that the firms' derived factor demand halts and even decreases after a productivity shock.

More generally, the results of my paper challenge the common presumption, ubiquitous in the literature on firm productivity, that a firm always expands its factor demand when it becomes more efficient. The fact that this has been proven in perfect competition by Levinsohn and Petrin (2003) and monopolistic competition with CES demand by De Loecker (2011) has led to a deep-rooted belief that productivity and input factor use always are monotonically related. Both in levels (i.e. productive firms $\approx$ large in terms of employment) and in changes (i.e. more productive $\approx$ demanding more inputs). Among others, illustrative and authoritative examples of this can be found in the most-cited review article on productivity by Syverson (2011), in the chapter on firm productivity by De Loecker and Syverson (2021) within the latest IO Handbook or in the introduction of the model of superstar firms by Autor, Dorn, Katz, Patterson, and Van Reenen (2020). In the presence of imperfect competition, in this paper I show that is not necessarily the case. The fact that derived factor demand has always been presumed to be monotonically related to productivity shocks originates from this oversight.

[2] The NBS implements a census of all state-owned manufacturing enterprises and all non-state manufacturing firms with sales exceeding RMB 5 million, or about $600,000 over that period.

From a methodological point of view, this monotonic assumption is particularly important for estimating production functions, as the validity of the control function approach critically depends on it. In my paper, I show why the presence of variable markups can pose a fundamental challenge to this approach.

To my knowledge, there have been glimpses of this result in few working papers by Bakhtiari (2009), Edmond, Midrigan, and Xu (2018), and Matsuyama and Ushchev (2022). However, their analyses focused only on monopolistic competition, on a specific functional form of demand and do not contain any reference to primitives of output demand. The contributions of my paper is to identify the declining price elasticity of output demand as the key mechanism behind this result, to prove its generality in terms of the features of both output demand and market structure and to assess its empirical relevance. The only empirical study to provide empirical evidence of a non-monotonic relationship between productivity and labor is Bakhtiari (2012) who focus on the ready-mix concrete industry in the US.

The theoretical contributions of my paper build on the new insights on the role of demand elasticity and curvature recently formalized by Mrázová and Neary (2017, 2019, 2020). I further develop their analysis by deriving the comparative static predictions about derived factor demand into the manifold framework and by extending it to oligopolistic settings.

In so doing, I also contribute to the literature on the (incomplete) pass-through as an economic tool (Weyl & Fabinger, 2013). While most of the literature on the theoretical determinants of pass-through focused on marginal costs (taxes, input price shocks, exchange rates, tariff, etc), in my paper I highlight the peculiarity of productivity shocks concerning their factor market implications. Moreover, I show that that level of markups and pass-through rates can be informative about the corresponding nature of factor demand.

Finally, my paper revisits a relatively old literature on derived factor demand and market structure through the lenses of the demand manifold framework. While the Marshall rules of derived factor demand have been conventionally developed and taught for the case of a competitive industry, back in the '70-80s a few papers extended this analysis to various settings of imperfect competition. Maurice and Ferguson (1973) and Foran (1976) to monopoly, Waterson (1980) and de Meza (1982) to oligopoly. While they already noted that the relevant elasticity is the elasticity of the marginal revenue curve, my paper extends their analyses along two dimensions. First, by extending the analysis beyond CES demand, I can express the elasticity of marginal revenue in terms of meaningful features of output demand. [3] Second, by applying the insights from the manifold, I can relate it to other comparative static predictions. More generally, I consider how the derived factor demand changes with a *Hicks-neutral* productivity shock, while the focus on the literature on derived factor demand has been on the effect of firm-specific or industry-wide changes in factor price(s).[4]

[3]Contrary, for example, to Maurice and Ferguson, according to which the elasticity of marginal revenue "*is unquestionably related to the elasticity of commodity demand... the relation is a tenuous one, and it cannot be stated explicitly in meaningful economic terms*" (p. 185)."

[4]As a matter of fact, Dobbs, Hill, and Waterson (1987) consider the effect of derived demand of factor-augmenting technical change, by looking at the cross-price elasticity of labor with respect to the price of capital. However, also their focus is restricted to CES demand, where the mechanism highlighted in my paper is not at play.

2 Theory

Why firms with market power do not necessarily increase their factor demand after a Hicks-neutral productivity shock? In this section, I characterize the theoretical conditions for this result to emerge both in terms of the features of output demand and market structure. In Section 2.1, I show that this result ultimately depends on the incomplete pass-through of productivity to output which is driven a firm's market power. To build the intuition, in Section 2.2 I present the main theoretical results under monopoly. In this setting, in Section 2.3 I describe how both the elasticity and curvature of output demand determine the responsiveness of input demand to productivity shocks and show that many widely-used demand functions lead to this result. Beyond monopoly, in Section 2.4 I prove that this result occurs also in standard models of monopolistic and oligopolistic competition and discuss how it influenced by the market structure and degree of competition across firms.

2.1 Derived factor demand and productivity shocks

To build the intuition, consider a generic profit-maximizing firm i that produces and sells a single product under the following assumptions.

Assumption A1 (**Input factor**). *The firm produces its output q_i according to a standard production function $q_i = f(x_i, \omega_i)$ where x_i is a single input factor, which is static and variable, and ω_i denotes its Hicks-neutral productivity level.*

Assumption A2 (**Technology**). *The productive technology of the firm is linear $q_i = x_i, \omega_i$ and exhibits constant returns to scale.*

Assumption A3 (**Input price**). *The firm is price-taker on the input market and $w > 0$ is the prevailing market price at which the firm can purchase the input x.*

I consider this stylized setting to highlight in a parsimonious way the key economic mechanism underlying the main result of the paper. However, in Section 2.5 I relax one by one all these assumptions and show that the main results of the paper hold in more general environments with multiple input factors, with non-constant returns to scale and with market power in the input market.

In such a setting, assume that this firm becomes more productive ($\uparrow \omega_i$). How would this firm adjust its derived factor demand, i.e. the amount of input used to maximize its profits? This is the key comparative static I focus on throughout the paper. Underlying this question, two opposing forces are at play. On one hand, a higher level of productivity means that the firm can produce the same level of output with a lower amount of input. On the other, a higher productivity level also means that its marginal costs are now lower. Therefore, the firm has the incentive to expand output in order to keep maximizing its profits. As a result, the firm will increase its input demand to meet its higher production targets. How these two effects balanced out remains ambiguous. In general, the net effect depends on the optimal rate of output expansion. To see this, start from the profit-maximizing levels of output and input,

$$q_i^* = x_i^* \, \omega_i \, ,$$

express it in logs and isolate the derived factor demand x^* so that

$$\log(q_i^*) = \log(x_i^*) + \log(\omega_i)$$
$$\log(x_i^*) = \log(q_i^*) - \log(\omega_i) \, .$$

By taking the total derivatives with respect to (log) productivity, the relationship between optimal input and output changes after a productivity shock can be expressed in terms of elasticities.[5] In particular,

$$\frac{d\log(x_i^*)}{d\log(\omega_i)} = \frac{d\log(q_i^*)}{d\log(\omega_i)} - \frac{d\log(\omega_i)}{d\log(\omega_i)} \tag{1}$$

$$\eta_{x_i^*,\,\omega_i} \;=\; \eta_{q_i^*,\,\omega_i} \;-\; 1 \;. \tag{2}$$

From this, it becomes apparent how the elasticity of input to productivity $\eta_{x^*,\,\omega}$ depends on the elasticity of optimal output to productivity $\eta_{q^*,\,\omega}$:

$$\eta_{x_i^*,\,\omega_i} < 0 \iff \eta_{q_i^*,\,\omega_i} < 1\,. \tag{3}$$

After a +1% productivity shock, **factor demand decreases** *if and only if* **the firm decides to increase output by less than 1%**. This incomplete adjustment of output is the key mechanism that leads a firm to reduce its input demand after a positive productivity shock. As this decision is essentially an equilibrium outcome, it is influenced by the structural features of the market in which the firm operates. Thus, to properly analyze the responsiveness of factor demand to productivity, it is necessary to consider the features of both output demand and market structure. In the next sessions, I analyze them in full generality.

2.2 Monopoly

To start analyzing the role of output demand in influencing the responsiveness of input demand to productivity shocks, I begin with the simplest setting of imperfect competition: monopoly. There is a single firm producing and selling the product.[6] The buyers' willingness to pay for the good produced by the monopolist is assume to be well-defined and downward-sloping with the following properties.

Assumption A4 (Demand). *The market demand for the output q is described by the (indirect) demand function $p(q)$, which is continuous, three-times differentiable, and strictly decreasing in output q, i.e. $p'(q) < 0$.*

To formalize how the features of output demand influence the responsiveness of input factor to productivity, I follow Mrázová and Neary (2017) in defining the following unit-free measures of the elasticity and curvature of demand:

$$\boldsymbol{\varepsilon}(q) \equiv -\frac{p(q)}{p'(q)\,q} \quad \text{and} \quad \boldsymbol{\rho}(q) \equiv -\frac{p''(q)\,q}{p'(q)}\,. \tag{4}$$

I intentionally express the elasticity and curvature as a function of the output level because, with the exception of CES demand, both of them vary along a demand curve.

Under the aforementioned assumptions on technology, costs and demand, a monopolist optimally chooses the output level to maximize its operating profits

$$\max_q \quad \pi = r(q) - mc\,q = (p(q) - mc)\;q\;,$$

where $r(q) = p(q)\;q$ denotes its revenue and $mc = \dfrac{w}{\omega}$ its (constant) marginal costs, which depends on the input price and its productivity level. Profit-maximization imposes restrictions on the possible values that $\boldsymbol{\varepsilon}$ and $\boldsymbol{\rho}$ can take at a profit-maximizing level of output q^*. From the first-order condition, a markup greater than one

[5] Throughout the paper, I use the notation $\eta_{g,\,y} \equiv \frac{\partial g(y,z)}{\partial y}\,\frac{y}{g(y,z)}$ to denote the elasticity of the function $g(y,z)$ with respect to y. The function $g(y,z)$ can have other arguments z, but they are not reported (unless it avoids possible confusion).

[6] In the presence of only one firm, I omit the subscript i in this section for notational convenience.

implies that the elasticity must be greater than one. From the second-order condition, the marginal revenue $mr(q) = (p + p'q)$ decreasing with output implies that the curvature must be strictly less than two. As a result,

$$\begin{aligned} \mu = \frac{p(q^*)}{mc} = \frac{\boldsymbol{\varepsilon}(q^*)}{\boldsymbol{\varepsilon}(q^*) - 1} > 1 \;\; &\Rightarrow \;\; \boldsymbol{\varepsilon}(q^*) > 1 \\ 2p'(q^*) \, + p''(q^*) \, q^* < 0 \;\; &\Rightarrow \;\; \boldsymbol{\rho}(q^*) < 2\,. \end{aligned} \tag{5}$$

The first-order condition in terms of x leads to the standard result that marginal revenue product of the input must be equated to its price:

$$\frac{\partial \pi}{\partial x} = \frac{\partial \pi}{\partial q}\frac{\partial q}{\partial x} = \left(p + p'\, q - \frac{w}{\omega}\right)\omega = 0 \quad \Leftrightarrow \quad \underbrace{(p + p'\, x\, \omega)\, \omega}_{\substack{mrp(x) \\ \text{marginal revenue} \\ \text{product}}} \;=\; w\,.$$

Based on this, the derived factor demand is

$$x^* = -\frac{p(q^*)}{p'(q^*)}\frac{1}{\omega} + \frac{w}{p'(q^*)\,\omega^2}\,. \tag{6}$$

After multiplying and dividing it by the price, the derived factor demand can be expressed in terms of the Lerner index:

$$x^* = \frac{1}{\omega}\,\frac{p(q^*)}{-p'(q^*)}\,\frac{p(q^*) - mc}{p(q^*)} = \frac{1}{\omega}\,\frac{p(q^*)}{-p'(q^*)}\,\frac{1}{\varepsilon(q^*)} \tag{7}$$

where $\varepsilon(q^*)$ is the value of the price elasticity of demand, evaluated at the profit-maximizing output level. By taking the derivative of x^* with respect to ω, it becomes evident that input demand does not necessarily increase after a productivity increase. It depends on how two opposing effects balance out, which is related to the market power exercised by the firm.

$$\frac{\partial\, x^*}{\partial\, \omega} = \underbrace{-\frac{1}{\omega^2}\,\frac{p(q^*)}{-p'(q^*)}\,\frac{1}{\varepsilon(q^*)}}_{<0} + \underbrace{\frac{1}{\omega}\frac{\partial}{\partial\, \omega}\left(\frac{p(q^*)}{-p'(q^*)}\,\frac{1}{\varepsilon(q^*)}\right)}_{>0} \gtreqless 0\,. \tag{8}$$

On one hand, a higher productivity level means that the firm can produce the same level of output with a lower amount of input. This is why the first term is always negative. On the other, a higher productivity level means that the firm marginal costs decrease. As a result, the firm has the incentive to expand output, demanding more input to do so. This is reflected in the second term being always positive (as $p'(q) < 0$).

However, whenever the price elasticity of demand decreases with output $\frac{\partial\, \varepsilon(q)}{\partial\, q} < 0$, this second term remains positive but decreases in productivity. This is because, as a firm becomes more productive and expands production, it moves to portion of demand with lower price elasticity of demand ($\downarrow \varepsilon(q^*)$). Facing a less elastic demand, it has a lower incentive to pass on this marginal cost reduction through higher output at lower price. Fundamentally, this incomplete pass-through of productivity to output is the mechanism through which market power can reduce the responsiveness of factor demand to productivity shock. In the following result, I show that this effect can be so strong at low levels of $\varepsilon(q)$ that the derived factor demand become completely unresponsive and even decline after a productivity shock, i.e. $\frac{\partial\, x^*}{\partial\, \omega} < 0$. Moreover, as the monopolist adjust along its marginal revenue curve (rather than the demand curve), I show that the elasticity of marginal revenue $\eta_{mr,\,q}(q) \equiv \frac{\partial mr(q)}{\partial q}\frac{q}{mr(q)}$ pins down this result.

Proposition 1. *Under the assumptions on (**A1-A2**) technology, (**A3**) input costs and (**A4**) demand, a monopolist reacts to a productivity shock by decreasing its input use if and only if either of these equivalent conditions holds at its profit-maximizing level of output q^*:*

$$\eta_{x^*,\,\omega} < 0 \iff \eta_{q*,\,\omega} < 1 \iff \eta_{mr,\,q}(q^*) < -1 \iff \boldsymbol{\varepsilon}(q^*) < 3 - \boldsymbol{\rho}(q^*)$$

(a) the elasticity of its optimal output with respect to productivity $\eta_{q*,\,\omega}$ is lower than 1;

(b) the elasticity of marginal revenue with respect to output $\eta_{mr,\,q}(q^*)$ is lower than -1.

(c) the price elasticity of demand is lower than 3 minus the value of the curvature of demand;

Proof reported in Appendix A.1.

Why a monopolist would expand its output less than proportionally to a +1% productivity shock, i.e. *(a)* $\eta_{q*,\,\omega} < 1$? The economics behind this decision is simple: because it starts experiencing strongly diminishing returns to increase its output. This is due to the fact that a declining price elasticity of demand implies that the marginal revenue generated by an additional unit of output decline too. As a result, if marginal revenue starts decreasing by more than 1% (i.e. *(b)* $\eta_{mr,\,q} < -1$), the firm takes its "foot off the gas" to prevent its margins from declining too rapidly and expands output by less than 1%. It does so precisely to keep maximizing its profits. This happens when demand becomes *nearly-satiated*, such that to convince buyers to purchase 1% more output, the price of the product must fall so much that marginal revenue starts decreasing by more than 1%.

To understand the last condition (i.e. *(c)* $\boldsymbol{\varepsilon}(q^*) < 3 - \boldsymbol{\rho}(q^*)$), note that the rate at which marginal revenue declines with output $\frac{\partial mr(q)}{\partial q} = 2p' + q\,p''$ depends not only on the first, but also on the second derivative of demand.[7] Therefore, both the slope of the demand curve (i.e. the elasticity) and the rate at which the slope decline with output (i.e. the curvature) determines whether and at which point $\eta_{mr,\,q}(q^*) < -1$. Since the marginal revenue depends of the features of demand, the shapes and determinants of these two curves are clearly linked.

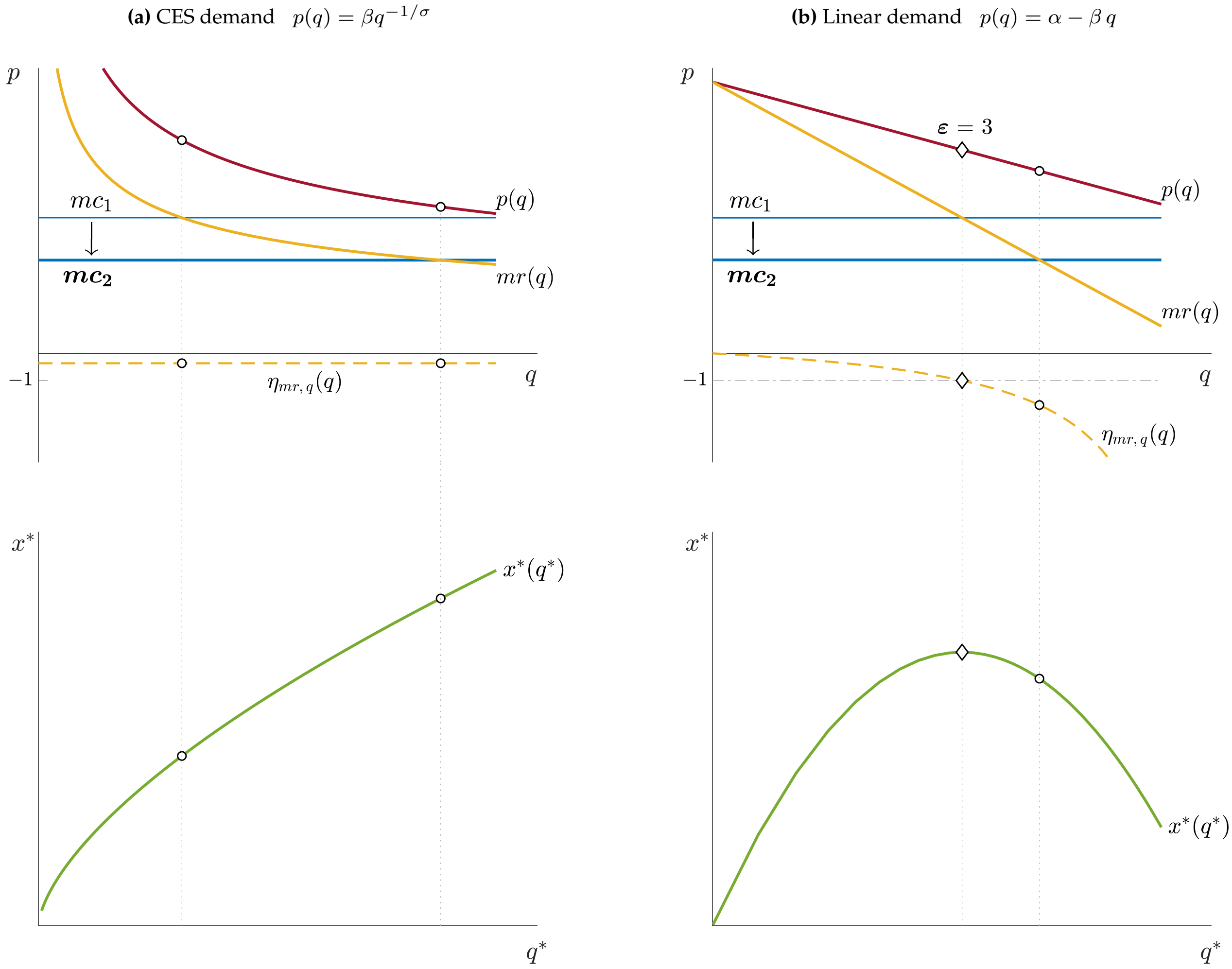
$$\eta_{mr,\,q}(q^*) \equiv \frac{(2p' + q^*\,p'')q^*}{p + q^*\,p'} = -\frac{2 - \boldsymbol{\rho}(q^*)}{\boldsymbol{\varepsilon}(q^*) - 1}\,. \tag{9}$$

This is why the value of the price elasticity of demand below which $\eta_{x^*,\,\omega} < 0$ depends also on the value of the curvature. Beyond this point, the higher productivity is enough to produce this less-than-proportional increase in output, so input demand starts decreasing and $\eta_{x^*,\,\omega} < 0$.

Illustration. To provide further intuition for this theoretical result, I illustrate in Figure 1 the comparative static of a productivity shock with the two commonly-used demand: CES and linear. Respectively defined as $p(q) = \beta q^{-1/\sigma}$ and $p(q) = \alpha - \beta\,q$ with $\sigma > 1$ and $\alpha,\ \beta > 0$. The price elasticity of CES demand is constant, while it declines with output in the case of linear. By comparing these two functional forms, it becomes clear that the derived factor demand does not always increase after a productivity shock, as commonly presumed in the literature. As productivity increases (from ω_1 to ω_2), marginal cost of the firm falls ($mc \downarrow$). The monopolist always reacts by expanding output ($q \uparrow$) and selling it at a lower price ($p \downarrow$). However, the firm adjusts in input use x^* (illustrated by the green line) very differently depending on the type of demand it faces: it always increases its input use if demand is CES, while with linear demand its derived factor demand halts and starts decreasing at a certain level of output.

[7]This is because it combines both the response of the price of the additional unit of output as well as the impact on revenue from the change in price on infra-marginal units.

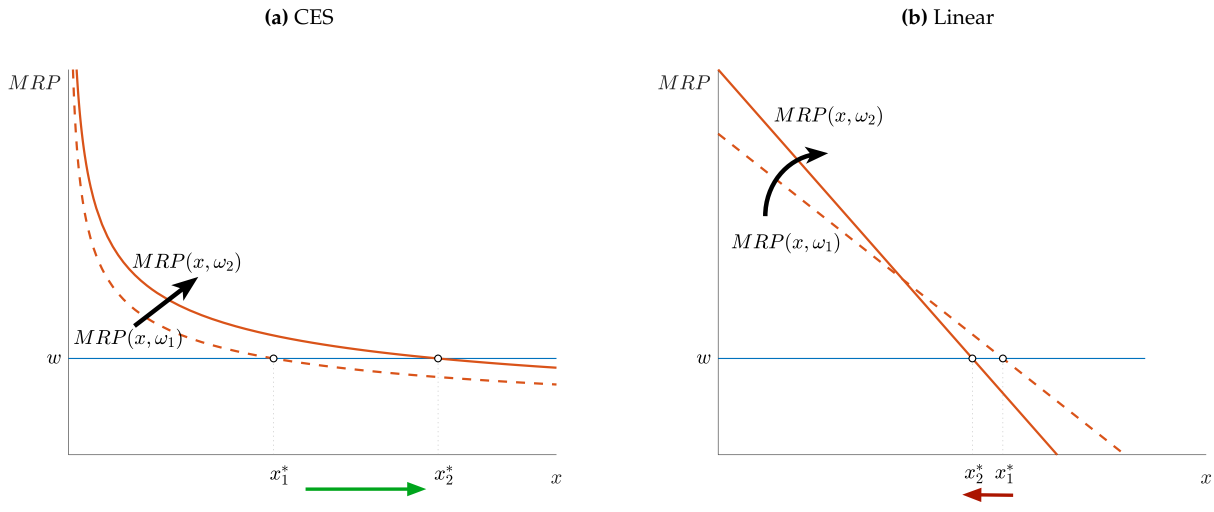
Figure 1. Reaction of a monopolist to a productivity shock $\uparrow \omega$.



All the conditions derived in Proposition 1 for this result to emerge show up. First, note how different is the increase in output q^* between the two types of demands. When demand is CES, the firm expands output more than proportionally to the productivity increase, in particular $\eta_{q^*,\omega} = \sigma - 1 > 1$. With linear demand, instead, the same productivity shock leads to a much smaller increase in output ($\eta_{q^*,\omega} \downarrow$ in q). This is reflected also by the value of the elasticity of marginal revenue (yellow dashed lines). In the case of CES, it remains constant at $\eta_{mr,q}(q) = -\frac{1}{\sigma}$ which is always greater than -1. With linear demand, instead, it decreases in q and falls below -1 exactly when input demand starts declining. Regarding condition *(c)*, in the case of linear demand this non-monotonicity kicks-in at $\varepsilon(q) = 3$ since the curvature $\boldsymbol{\rho}(q) = 0\ \forall q$. As a result, $\varepsilon(q^*) < 3 - 0$. Below this value of the price elasticity of demand, any further increase in productivity leads to a decrease in derived factor demand.

A complementary perspective on this result comes from visualizing the first-order condition in terms of the marginal revenue product (mrp), i.e. the amount of revenue a firm can generate by purchasing one additional unit of input. Figure 2 shows how the $mrp(x)$ reacts to an exogenous productivity shock. Since the firm is assumed to be *price-taker* on the input market, the intersection between its $mrp(x)$ curve and the factor price w determines the profit-maximizing level of input x^*. Contrary to common belief, however, a productivity shock does not necessarily lead to an *outward shift* of the mrp. While this is the case for CES demand, it is not when demand is linear. In fact, a productivity shock leads to a *rotation* of the marginal revenue product of the input, because of strongly diminishing marginal revenue gains from increasing input.

Figure 2. Reaction of $mrp(x)$ to a productivity shock ($\uparrow \omega$): outward shift *vs.* rotation.



2.3 Features of output demand

Beyond linear, in this section I show that many other commonly-used demand functions lead to a non-monotonic relationship between derived factor demand and productivity. I do that by bringing the insight from Proposition 1*(c)* into a general framework to compare demand functions based only on the relationship between the values of the elasticity and curvature. Moreover, I connect the responsiveness of input demand to productivity to predictions on firm-level markups and pass-through rates, since they are essentially driven by same characteristics of output demand.

Demand manifold framework. Mrázová and Neary (2017) show that any well-behaved demand function can be represented by its *demand manifold*, a smooth curve relating the value of the elasticity $\varepsilon(q)$ of demand to its convexity $\boldsymbol{\rho}(q)$.[8] This is the case for any demand that fulfills assumptions A4. As shown in Equation (5), the first- and second-order conditions restrict the possible values of $(\varepsilon, \boldsymbol{\rho})$ in which a profit-mazimizing monopolist with constant marginal costs would operate. The shaded area in Figure 3 (a) illustrates the resulting admissible region in the $(\varepsilon,\ \boldsymbol{\rho})$-space.[9]

To fix the ideas, the manifold of a linear demand $p(q) = \alpha - \beta q$ is illustrated in Figure 3 (b). Along any linear demand curve, the elasticity $\varepsilon(q)$ declines with output: $\varepsilon(q) = \frac{\alpha}{\beta q} - 1$. As p'' is zero $\forall q$, the curvature is always $\boldsymbol{\rho}(q) = 0$. This is why the corresponding manifold is a vertical line at $\boldsymbol{\rho} = 0$. When a firm expand production, it will face a lower point-elasticity of demand, which is represented by a downward movement along the manifold (as illustrated by the arrow from q_1 to q_2).

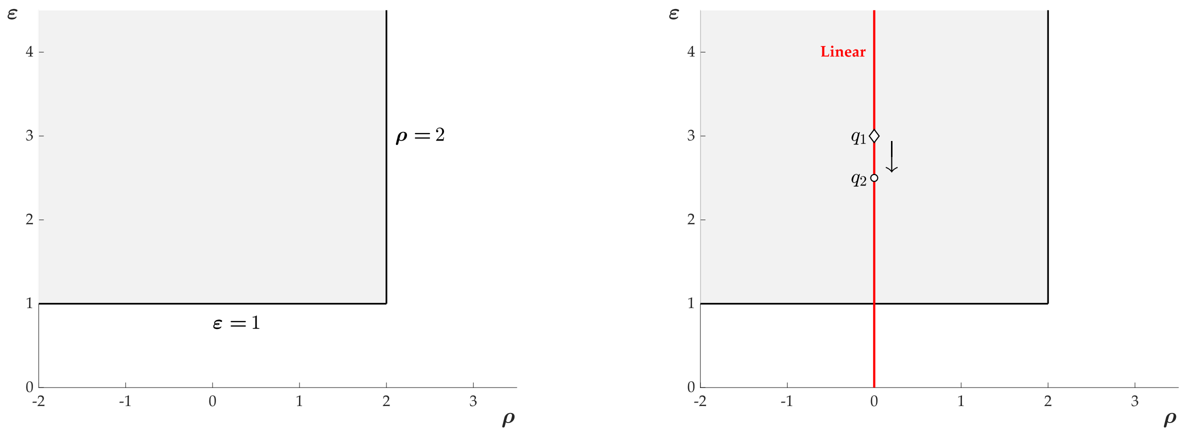
One of the advantage of working with demand manifolds rather than the demand functions themselves is that being located in $(\varepsilon, \boldsymbol{\rho})$ space, they reveal the implications of local demand feature at a higher level of generality. In the case of linear, for example, the exogenous parameters α and β shift the perceived demand curve, but they do not shift the corresponding demand manifold. Mrázová and Neary call this property "manifold invariance". When it holds, exogenous shocks lead only to movements along the manifold, not to shifts in it.

[8]For the sake of clarity, note that the definitions of $\varepsilon(q)$ and $\boldsymbol{\rho}(q)$ are not entirely consistent with each other. The measure of curvature $\boldsymbol{\rho}(q)$ equals the elasticity of the slope of inverse demand. Mrázová and Neary follow this standard practice and work throughout with the price elasticity of direct demand, given its greater intuitive appeal (at least in industrial organization) and its focus on the region of parameter space where comparative statics results are ambiguous.

[9]Consumers may be willing to consume outside this region, but such values of $(\varepsilon, \boldsymbol{\rho})$ cannot represent a profit-maximizing equilibrium.

Figure 3. Overview of the Convexity-Elasticity Space.

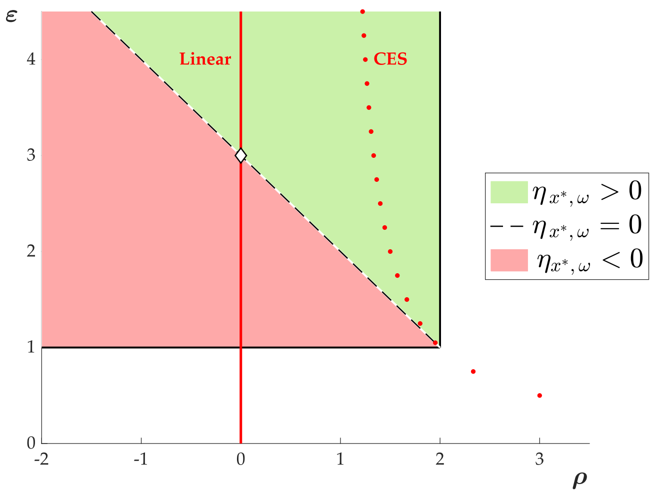
(a) Admissable region **(b)** Linear demand manifold



Notes: the admissible region is in fact $\{(\varepsilon, \rho) : 1 \leq \varepsilon < \infty$ and $-\infty < \rho < 2\}$. Following Mrázová and Neary (2017), I highlighted only a subset of the admissible region, which is where most interesting issues arise and is also consistent with available empirical evidence.

Equipped with this theoretical construct, I can bring the results from Proposition 1.*(c)* on the responsiveness of factor demand to productivity into the manifold framework. As shown before, a firm will decrease its input demand after a productivity shock if it faces a low price elasticity of demand. How low? It depends on the values of the curvature, in particular $\varepsilon(q) < 3 - \rho(q)$. The red region in Figure 4 represents the corresponding combination of elasticity and convexity values where $\eta_{x^*,\omega} < 0$, while the dashed line illustrates the values that lead exactly to $\eta_{x^*,\omega} = 0$.

Figure 4. Proposition 1*(c)* in the manifold space.



The demand manifold for CES demand is represented, instead, by a dotted line. Every point on this dotted curve corresponds to a CES demand function with a different value σ, because $\varepsilon(q) = \sigma \; \forall q$ and $\rho(q) = \frac{\sigma+1}{\sigma} \; \forall q$. Since $\sigma > 3 - \frac{\sigma+1}{\sigma}$, a profit-maximizing monopolist facing a CES demand will never be in the (red) region where $\eta_{x^*,\omega} = 0$. This is because a monopolist will face the same level of elasticity for any value of output.

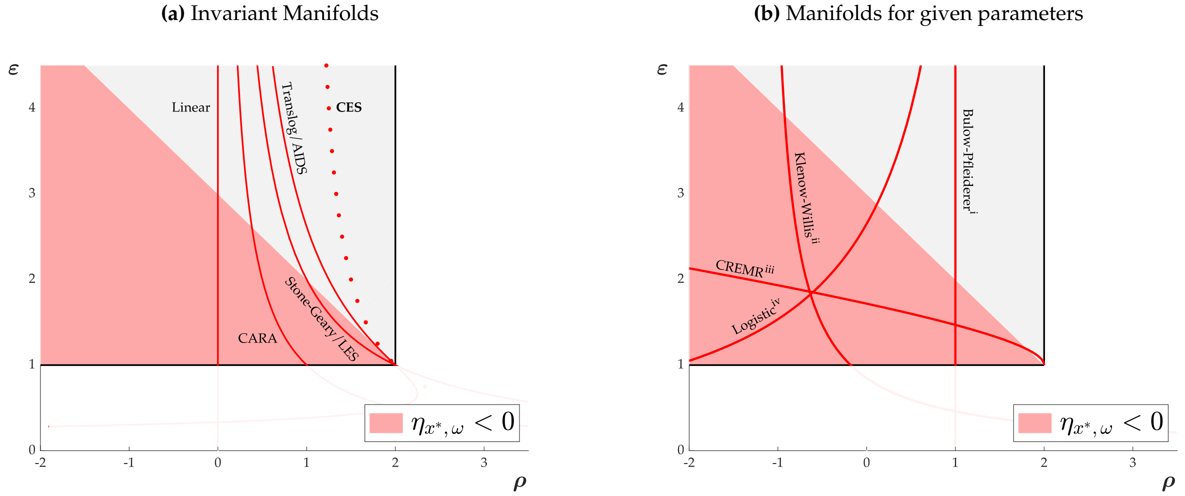
In this respect, the CES is very special and represents an important knive-edge case. Mrázová and Neary (2019) show that the CES manifold divides the admissible region in two: at an arbitrary point any demand will be either more or less convex at that point than a CES demand function with the same elasticity. Demand functions whose manifold are located to the right of CES are called ”superconvex”, while “subconvex” those to the left. The elasticity $\varepsilon(q)$ increases with output if a demand is superconvex, while it decreases with output if it is subconvex (e.g. Linear demand). This represents an important boundary for several comparative static predictions. Among other things, this determines the relationship between demand markups and output. This is the case also in the case of responsiveness of input demand to productivity shock.

Corollary 1. *The responsiveness of factor demand to productivity can become negative $\eta_{x^*,\omega} < 0$ only if output demand is sub-convex, i.e. its price elasticity of demand declines with output $\frac{\partial\varepsilon(q)}{\partial q} < 0$.*

Sub-convexity property is often called "Marshall's Second Law of Demand".[10]. Although many other terminologies are used in the literature: Arkolakis, Costinot, Donaldson, and Rodríguez-Clare (2019) describe such demands by being "log concave in log prices", Zhelobodko, Kokovin, Parenti, and Thisse (2012) denote them through the concept of "increasing relative love of variety", while Kimball (1995) defines this property as "positive super-elasticity of demand". However called, this property is considered theoretically more plausible since consumers are more responsive to price changes the greater their consumption and it is a sufficient condition for the existence (and sometimes uniqueness) of equilibrium in models of industrial organization Caplin and Nalebuff (1991).[11] Moreover, sub-convexity is also consistent with much of the available empirical evidence on incomplete pass-through and the fact that larger firms set higher markups. For these reasons, throughout the paper, I will focus on subconvex demand functions.

In Figure 5, I illustrate the manifolds for widely-used demand specifications that are sub-convex. In general, both $\varepsilon(q)$ and $\rho(q)$ vary with output.

Figure 5. Demand Manifolds for common sub-convex demand functions.



Notes: in **(b)** I consider certain demand functions in which the location of the manifold depends on specific parameter values. For illustration, I take these values from previous calibration in the literature. In (i) the Bulow and Pfleiderer (1983) demand the absolute pass-through from cost to price is 1 (i.e. dollar-for-dollar). In (ii) the value of super-elasticity is 2.18 based on the calibration by Edmond et al. (2018). In (iii) the parameter for CREmr is set to $\sigma = 1.11$ following Mrazova et al. (2021). In (iv) I follow the specification of Cowan (2016). For additional details, I refer the reader to the Appendix A.1.1.

This shows that a non-monotonic relationship between derived factor demand and productivity arises in most commonly-used demand specifications. In particular, input demand becomes unresponsive to productivity shocks in the cases of linear demand, the negative exponential or CARA ("constant absolute risk-aversion")

[10] As it was originally introduced by Marshall (1890) in his *Principles of Economics*, where he argued that "*the elasticity of demand is great for high prices, and great, or at least considerable, for medium prices; but it declines as the price falls; and gradually fades away if the fall goes so far that satiety level is reached. This rule appears to hold with regard to nearly all commodities and with regard to the demand of every class*" (Book III).

[11] It is a sufficient condition for a unique equilibrium to exist in common models of Cournot competition and differentiated products Bertrand competition.

system used by Behrens et al. (2020), the Logistic demand (Cowan, 2016), the Klenow and Willis (2016) specification of Kimball demand, and the CREmr demands ("Constant-Revenue-Elasticity-of-Marginal-Revenue") introduced by Mrazova et al. (2021). This will happen also in Bulow and Pfleiderer (1983) demand and Stone-Geary/LES ("linear expenditure system") at very low levels of elasticity. The only exception is Translog demand, which from the firm's perspective is consistent also with the Almost Ideal or "AIDS" model of Deaton and Muellbauer (1980). These are just a few examples of commonly used demand specifications. If interested in the manifolds of other demands, I refer the reader to Mrázová and Neary (2017) and their rich Appendix for additional material.[12]

2.3.1 Corresponding values of markups and pass-through

Another advantage of illustrating the result of Proposition 1 in terms of demand manifolds is the possibility to link these new insights on derived factor demand with other firm-level outcomes. This is particularly the case with markups and pass-through behaviors, as they are determined by the elasticity and curvature of demand too. The following result highlights what else we should expect in correspondence of $\eta_{x^*,\omega} = 0$.

Corollary 2. *The derived factor demand of a monopolist halts and starts declining after a productivity shock in correspondence of high levels markups and low (incomplete) pass-through rates.*

It is important to note that higher markups and incomplete pass-through rates represent a necessary and sufficient condition for the unresponsiveness of factor demand to productivity. To the best of my knowledge, this connection between the nature of factor demand and pricing behavior has remained overlooked in the literature so far.

Table 1. Correspondence between $\eta_{x^*,\omega} = 0$ and firm-level outcomes.

		(i) $\boldsymbol{\rho} = -1$	(ii) $\boldsymbol{\rho} = 0$	(iii) $\boldsymbol{\rho} = 1$
(a) Responsiveness on input to productivity	$\eta_{x^*,\omega} = \dfrac{\varepsilon + \boldsymbol{\rho} - 3}{2 - \boldsymbol{\rho}}$	0	0	0
(b) Markups	$\mu = \dfrac{\varepsilon}{\varepsilon - 1}$	1.33	1.5	2
(c) Cost-*to*-price pass-through *(€-to-€)*	$\dfrac{\partial p}{\partial mc} = \dfrac{1}{2 - \boldsymbol{\rho}}$	0.33€	0.5€	1€
(d) Cost-*to*-price pass-through rate *(in %)*	$\eta_{p,mc} = \dfrac{\varepsilon - 1}{\varepsilon(2 - \boldsymbol{\rho})}$	25%	33%	50%

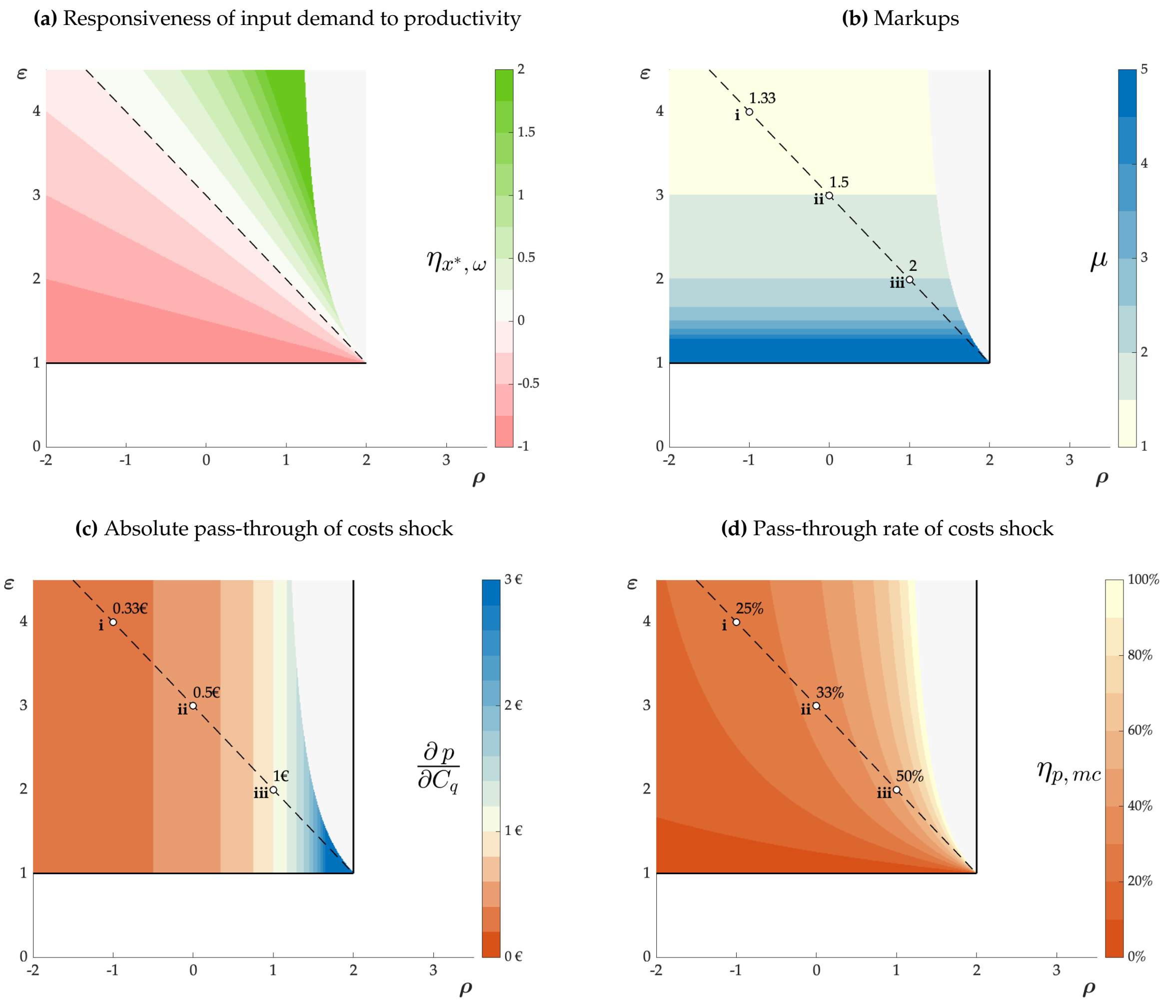
In the case of linear demand, the derived factor demand becomes unresponsive to productivity shock in correspondence of a markup $\mu = 1.5$ and a cost-to-price pass-through rate of 33%. In general, the actual values of markups and pass-through rates at which $\eta_{x^*,\omega} = 0$ depend on the convexity of demand. To illustrate this, I report in Table 1 the values of markups and two pass-through measures in correspondence of $\eta_{x^*,\omega} = 0$ for three different values of curvature, (i) $\boldsymbol{\rho} = -1$, (ii) $\boldsymbol{\rho} = 0$, (iii) $\boldsymbol{\rho} = +1$.[13]

[12] In this regard, Proposition 1 is fundamentally related to their result about the super-modularity of firms' profits in their own marginal cost and the iceberg transport cost they face in models with heterogeneous firms, where firms choose between two alternative ways of serving a market (such as the choice between exports and foreign direct investment). In light of this, any demand functions that is shown to be super-modular in their paper, it will also lead to $\eta_{x^*,\omega} < 0$.

[13] In industrial organization the focus is usually on the absolute pass-through (i.e. by how much a firm raises its price if marginal cost increases by one euro), while in macro/international economics the standard measure is the proportional pass-through.

For a given level of output, $\eta_{x^*,\omega} = 0$ will occur in correspondence of relatively lower markups and pass-through rates for more concave/less convex demands (i.e. lower $\boldsymbol{\rho}$). This is because in such demands a firm reaches sooner (at a relatively lower level of output) the point where demand become nearly-satiated, i.e. where an additional unit of output will cause a relatively large fall in price reducing the incentive for the firm to expand output and, in turns, the responsiveness of factor demand.
To put these values in context, in Figure 6 I illustrate within the demand manifold framework the overall correspondence between the values of $\eta_{x^*,\omega}$, markups and pass-through rates for sub-convex demands.

Figure 6. Values of responsiveness, markups and pass-through in the (ε, $\boldsymbol{\rho}$) space.



This correspondence between the (un)responsiveness to productivity shocks and markups level will prove very useful to detect $\eta_{x^*,\omega} \leq 0$ in the data in Section 3 because the levels of markups and pass-through rates can be estimated.

2.4 Role of market structure

Restricting the analysis to monopoly has been useful to highlight the features of output demand in driving the reactions of derived factor demand after a productivity shock. The main take-away is that the inputs become completely unresponsive to a positive productivity shock at lower levels of price elasticity of demand $\varepsilon(q)$. Beyond this archetypal model of imperfect competition, however, the relevant elasticity for firms' behavior is not the elasticity of market demand, but the elasticity of the residual demand faced by the each firm. For this reason, any factor that affects the latter will influence the responsiveness of factor demand, on top of the feature of output demand. In the following sections, I analyze in standard models of monopolistic competition and oligopoly competition in quantity the role of the degree of competition, the characteristics of competitors and the market structure. Overall, I find that the responsiveness of factor demand to productivity shocks declines with firm productivity and it is more likely to become negative in less competitive markets.

2.4.1 Monopolistic competition

A monopolistic competitive market is comprised of many small firms, each producing a different variety of a product. The assumptions on their technology (**A1-A2**) and input price (**A3**) remain the same to the monopoly setting, but firms have different levels of productivity ω_i. Mrázová and Neary (2017) already proved that the comparative statics predictions developed within the manifold framework for a monopolist can be translated into the canonical general-equilibrium model of monopolistic competition, which was formalized by Dixit and Stiglitz (1977), extended to firm heterogeneity by Melitz (2003) and then generalized to non-CES demands by Zhelobodko et al. (2012) and Dhingra and Morrow (2019). The only condition for this is that consumers' preferences need to be symmetric and that the elasticity of demand for a variety depends only on its level of consumption. Among others, this is the case for additively separable preferences, which are very common in the literature (see Mayer et al. (2021)). In this setting, the assumption on demand is fundamentally the same as (**A4**), but micro-founded from consumers' preferences.

Assumption A5 (Demand). *Let $i \in [0, N]$ be the continuum of horizontally differentiated varieties available to L consumers, whose preferences are represented by*

$$\mathcal{U} = U\left[\int_0^N u(q_i)di\right] \text{ with } U' > 0\,,$$

where $u(\cdot)$ is strictly increasing and concave, i.e. $u'(q_i) > 0$ and $u''(q_i) < 0$ for $q_i \geq 0$, and $u(0) = 0$. Given prices for each variety, each consumer maximizes utility subject to his/her budget constraint. This leads to a strictly decreasing residual (inverse) demand for each variety defined by

$$p_i(q_i) = \frac{u'(q_i)}{\lambda}\,,$$

where λ is the marginal utility of income of the consumer. The sub-utility function $u(\cdot)$ is such that the elasticity of residual demand declines with output.[14]

Being negligible to the market, each firm chooses its output level to maximize its operating profits by taking the residual demand function it faces as given. This includes the aggregate demand conditions described by λ. Given the assumption of separable preferences, λ is the unique endogenous aggregate demand shifter. In Appendix A.2 I describe the conditions that determines the value of λ at the unique free-entry equilibrium.

[14] The inverse demand inherits the properties of $u(\cdot)$. In particular, $p_i(q_i)$ is strictly decreasing because $u(\cdot)$ is strictly concave. Moreover, Zhelobodko et al. (2012) show that $\varepsilon(q_i) = -\frac{p(q_i)}{p'(q_i)q_i} = -\frac{u'(q_i)}{u''(q_i)q_i}$.

The standard prediction of this class of models is that the levels of profits, output, and revenues are monotonic increasing in firm productivity. With sub-convex demands, this is the case also for markups.

$$\frac{\partial \pi_i(\omega_i,\,\lambda)}{\partial \omega_i} > 0, \; \frac{\partial q_i(\omega_i,\,\lambda)}{\partial \omega_i} > 0, \; \frac{\partial r_i(\omega_i,\,\lambda)}{\partial \omega_i} > 0, \; \frac{\partial \mu_i(\omega_i,\,\lambda)}{\partial \omega_i} > 0 \;\text{ but }\; \frac{\partial x_i(\omega_i,\,\lambda)}{\partial \omega_i} \gtreqless 0\,.$$

The presumption in both the theoretical and empirical literature is that this prediction holds also for input demand, both in a cross-section and in changes. Illustrative and authoritative examples of this can be found in the most-cited review article on productivity by Syverson (2011), in the chapter on firm productivity by De Loecker and Syverson (2021) within the latest IO Handbook ("*higher-productivity producers are also larger even when measured by input use*") or in the introduction of the model of superstar firms by Autor et al. (2020) ("*more productive firms will have higher levels of factor inputs and greater output*").

In fact, input use is not always monotonic increasing in productivity and the results of Proposition 1 extend to monopolistic competition. This is because a monopolistic competitive firm behaves like a monopolist on the residual demand for its variety. Even if the residual demand and marginal revenue curves depend on λ, their elasticities are independent of it.[15] As a result, all the results on the features of output demand from Section 2.3 remains valid from the perspective of each firm.

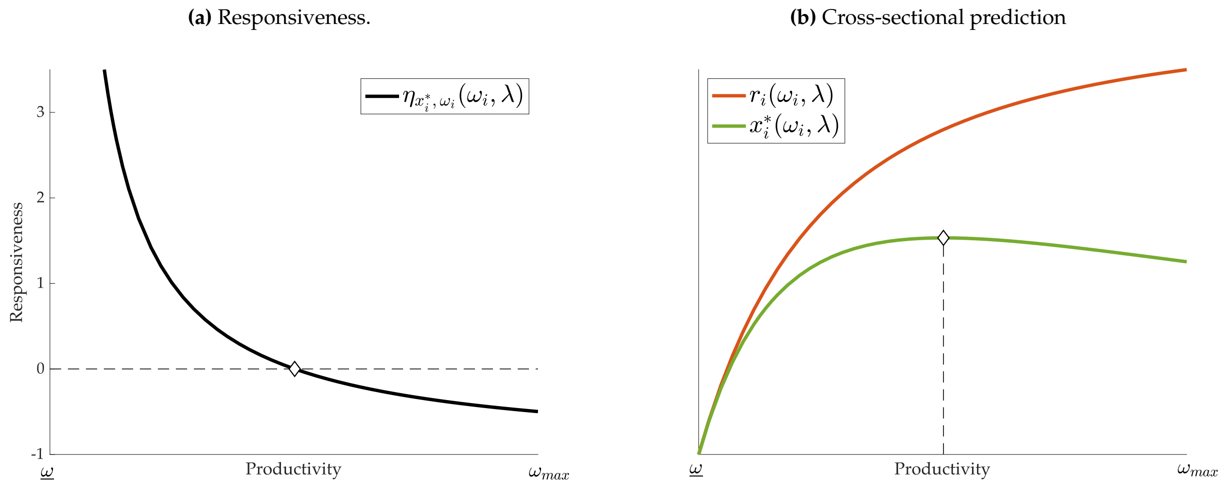
Proposition 2. *Under the assumptions on (A1-A2) technology, (A3) input costs and (A5) demand, a monopolistic competitive firm reacts to a productivity shock by decreasing its input use if and only if either of these equivalent conditions holds at its profit-maximizing level of output $q_i^*(\omega_i, \lambda)$:*

$$\eta_{x_i^*,\,\omega_i} < 0 \iff \eta_{q_i^*,\,\omega_i} < 1 \iff \eta_{mr_i,\,q_i}(q_i^*) < -1 \iff \boldsymbol{\varepsilon}(q_i^*) < 3 - \boldsymbol{\rho}(q_i^*)$$

Since $\boldsymbol{\varepsilon}(q_i(\omega_i,\,\lambda))$ declines with q_i, this implies that the most productive firms within a market:
(a) have a lower responsiveness of $\eta_{x_i^*,\,\omega_i}$ than less productive firms;
(b) are not necessarily the largest in terms of input.

In Figure 7, I plot both these predictions in an illustrative setting with linear demand.

Figure 7. Monopolistic competition equilibrium outcomes



Notes: equilibrium outcome with linear demand $p_i(q_i) = \alpha - \beta\, q_i$. The white point corresponds to the firm with productivity level where $\eta_{x_i^*,\,\omega_i} = 0$. The productivity values ranges from the minimum cut-off $\underline{\omega}$ to ω_{max}. See Appendix A.2 for details about the simulation.

[15] In particular, $mr_i(q_i,\,\lambda) = \frac{u'(q_i)+u''(q_i)q_i}{\lambda}$. Therefore, $\boldsymbol{\varepsilon}(q_i) = -\frac{p(q_i)}{p'(q_i)q_i} = -\frac{u'(q_i)}{u''(q_i)q_i}$ and $\boldsymbol{\rho}(q_i) = -\frac{p''(q_i)q_i}{p'(q_i)} = -\frac{u'''(q_i)q_i}{u''(q_i)}$ do not depend on λ, their relationship (i.e. the demand manifold) and the elasticity of marginal revenue $\eta_{mr_i,\,q_i}(q_i^*)$ are independent of λ too.

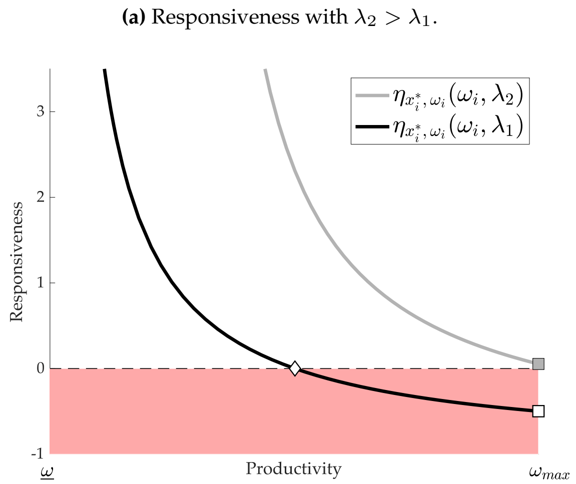
The cross-sectional prediction follows simply by considering two firms i and j with different productivity levels $\omega_j > \omega_i$. The more productive firm j will be larger in terms of input if and only if $\frac{log(q_j^*)-log(q_i^*)}{log(\omega_j)-log(\omega_i)} > 1$. This is the case if and only if at $q_i^*(\omega_i, \lambda)$ the conditions in Proposition 2 are met. Note, however, that all these firm-level outcomes do not depend only on the productivity level of each firm but also of the level of λ. In this setting, λ represents the degree of competition in the market and is a general equilibrium outcome that reflects all its structural features: the number of consumers, the entry costs, the distribution of firm productivity, etc. . Therefore, any change that affects the level of competition in a market will change the equilibrium responsiveness of factor demand $\eta_{x_i^*, \omega_i}$ of each firm, as formalized in the following result.

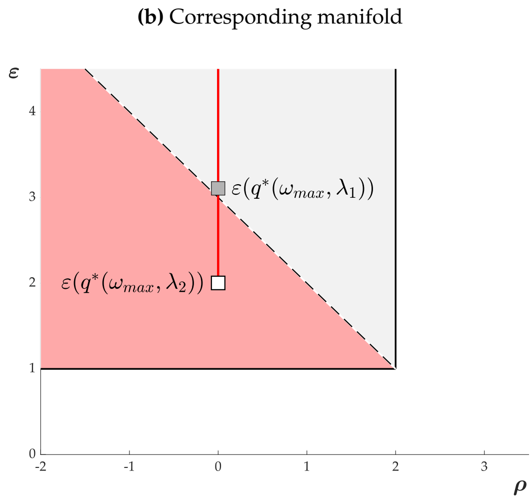
Corollary 3. *The responsiveness of factor demand to productivity of monopolistic competitive firms is positively related to the degree of competition in a market: less competitive pressures (lower λ) induce a lower responsiveness, while more competition (higher λ) increases $\eta_{x_i^*, \omega_i}$ for any firm i.*

The reason for this is that, whenever the equilibrium value of λ increases, the output of each firm $q_i^*(\omega_i, \lambda)$ will be lower since all residual demand curves shift outward. This decrease in output, in turn, increases the value of the elasticity of the residual demand $\varepsilon(q_i^*)$, which has a direct impact on firm-level markups and the responsiveness $\eta_{x_i^*, \omega_i}$. Mayer et al. (2021) show that the shape of marginal revenue curve determines the gradient of this change since $\eta_{q_i, \lambda} = \frac{1}{\eta_{mr_i, q_i}}$.

In Figure 8(a), I illustrate this result for two market equilibria with $\lambda_2 > \lambda_1$, keeping consumers' preferences and thus features of output demand fixed. In (b), I show the corresponding values of (ε, $\boldsymbol{\rho}$) for the most productive firm in both scenarios. As the elasticity $\varepsilon(q_i^*)$ is higher with more competition, this corresponds to an upward shift along the demand manifold. A more competitive environment can increase the responsiveness of this firm so much that it is "pushed" out of the (red) region where $\eta_{x_i^*, \omega_i} < 0$.

Figure 8. Impact of an increase in competition.





Notes: equilibrium outcome with quadratic preferences leading to linear demand $p_i(q_i) = \alpha - \beta q_i$. The squares denote the firm with the highest productivity level. The higher λ_2 is the result of a reduction in entry costs f_E which increases the equilibrium level of competition in the market. See Appendix A.2 for details about the simulation.

2.4.2 Oligopoly

In this section I investigate how strategic interactions across firms shape the responsiveness to productivity shocks of their input demand. To do so, I consider a market with a limited number of firms $i = 1, ..., N$ which produce a homogeneous good q_i. The assumptions on their technology (**A1-A2**) and input price (**A3**) remain the same as previous settings. This implies that their marginal costs are constant $mc_i = \dfrac{w}{\omega_i}$ and depend on (given) input factor prices w and the firm-specific productivity level ω_i. In this setting, the assumption on demand is fundamentally the same as in monopoly (**A4**), but expressed here in terms of total output.

Assumption A6 (**Demand**). *The market demand for the homogeneous good is described by the (indirect) demand function $p(Q)$, which is continuous, three-times differentiable, and strictly decreasing in total output $Q = \sum_{i=1}^{N} q_i$.*

The essential strategic interaction that I focus on is the extent to which a firm's quantity choice q_i affects other firms' profits through the aggregate quantity Q. In doing so, I follow the conduct parameter approach to model oligopolistic interactions and assume that this relationship is summarized by the parameter $\theta \equiv \frac{\partial Q}{\partial q_i}$. This assumption nests a number of well-known special cases. The standard Cournot oligopoly model corresponds to $\theta = 1$, in which each firm takes the quantities of the other firms as given, conjecturing that total output increases by the same amount as its own quantity. The case of perfect collusion corresponds to $\theta = N$, in which each firm conjectures that each rival will fully match a quantity increase. If $\theta \to 0$, this leads to perfect competition since each firm conjectures that the rivals contract their quantities in response to a change in its own quantity so that Q output remains constant.[16] In such a setting, the first-order condition of profit maximization for each firm is

$$\underbrace{p + \theta\, p'(Q)\, q_i}_{mr_i} = mc_i\,. \tag{10}$$

If we divide and multiply the left-hand side by Q, it can be expressed in terms of the elasticity of market demand $\varepsilon(Q)$ and the market shares of each firm $s_i = \dfrac{q_i}{Q}$.

$$p\left(1 - \frac{\theta\, s_i}{\varepsilon(Q)}\right) = mc_i\,.$$

Summing over the first-order conditions of all the competitors $j \neq i$, note that the market share of the firm s_i depends on the elasticity of market demand, the number and the average marginal costs of its competitors:

$$s_i = 1 - (N-1)\frac{\varepsilon(Q)}{\theta}\left(1 - \frac{\overline{mc}_j}{p}\right)\,. \tag{11}$$

Similarly to monopoly, profit-maximization imposes restrictions on the possible values that ε and ρ can take at a profit-maximizing equilibrium. From the first-order condition, a markup greater than one implies that the price elasticity of residual demand must be greater than one. From the second-order condition, the marginal revenue decreasing with its own output implies that the curvature of the residual demand must be strictly less then two. In terms of elasticity and curvature of market demand, this imply that for each active firm i it must be that

$$\varepsilon(Q) \geq \theta\, s_i \quad \text{and} \quad \rho(Q) < \frac{2}{\theta s_i}\,.$$

Following Seade (1980), an additional restriction on the curvature of market demand $\rho(Q) < \frac{N}{\theta} + 1$ is set by the stability criterion. Against this background, I show that the insights from Proposition 1 about the

[16] As discussed by Verboven and Van Dijk (2009), *"outside these special cases this framework has little game-theoretic appeal, since it aims to capture dynamic responses within a static model. It has, however, often been used in empirical work to estimate the conduct or average collusiveness of firms without having to specify a fully dynamic model"*.

responsiveness of derived factor demand to a productivity shock can be applied, with few adjustments, also to a setting with strategic interactions.

Proposition 3. *Under the assumptions on (**A1-A2**) technology, (**A3**) input price and (**A6**) demand, in a standard oligopolistic setting of competition in quantity a firm i reacts to an idiosyncratic shock to its productivity by decreasing its inputs if and only if either of these equivalent conditions holds at its optimal level of output q_i^*:*

(a) the elasticity of its optimal output with respect to productivity $\eta_{q_i^*,\,\omega_i}$ is lower than 1;

(b) the elasticity of its marginal revenue curve with respect to its output $\eta_{mr_i,\,q_i}$ is lower than -1.

(c) the price elasticity of its residual demand is lower than $3-\boldsymbol{\rho}(Q)\,\theta\,s_i$;

$$\eta_{x_i^*,\,\omega_i} < 0 \iff \eta_{q_i^*,\,\omega_i} < 1 \iff \eta_{mr_i,\,q_i} < -1 \iff \frac{\boldsymbol{\varepsilon}(Q)}{\theta\,s_i} < 3-\boldsymbol{\rho}(Q)\,\theta\,s_i$$

Proof reported in Appendix A.3.

As an oligopolist acts exactly as a monopolist on its *residual* demand, these comparative statics predictions are almost identical to Proposition 1 and the underlying economic mechanism leading to an incomplete pass-through of productivity to input demand is the same. At lower levels of elasticity, a firm that becomes more productive has a lower incentive to further expand production, because of strongly diminishing marginal revenue gains from doing so. However, the condition *(c)* is re-expressed in terms of the elasticity and curvature of the residual demand of each firm, which depends on its market share s_i and the conduct parameter θ. This shows how the market structure influences the responsiveness by affecting the elasticity of the residual demand curve of a firm.

Corollary 4. *The degree of competition across competitors and their number affect the responsiveness of factor demand to productivity changes: less competitive pressure (i.e. lower N and/or higher θ) leads to a lower elasticity of derived factor demand to its productivity shocks $\eta_{x_i^*,\,\omega_i}$.*

To see why, assume without loss of generality that firms are symmetric in their productivity and thus in their marginal costs. In such situation, market shares are $s_i = \frac{1}{N}\ \forall i$ and Equation (A3) implies that

$$\eta_{x_i^*,\,\omega_i} = -1 - \frac{1-\frac{\boldsymbol{\varepsilon}(Q)N}{\theta}}{2-\frac{\boldsymbol{\rho}(Q)\,\theta}{N}}$$

Ceteris paribus, fewer competitors ($\downarrow N$) and/or a higher degree of collusion ($\uparrow \theta$) reduce $\eta_{x_i^*,\,\omega_i}$. On the contrary, if $N \to \infty$ and/or $\theta \to 0 \Rightarrow \eta_{x_i^*,\,\omega_i} > 0\,\forall i$. This result highlights why in perfect competition the responsiveness is always positive.

Whenever firms are heterogeneous, the responsiveness is very different between large and small firms and the difference in their their productivity levels (and more generally in their cost structure) is another important determinant of $\eta_{x_i^*,\,\omega_i}$.

Corollary 5. *The largest firm in a market has the lowest responsiveness of its factor demand to an exogenous productivity shock and this is accentuated by its cost advantage relative to the competitors.*

Proof. Sum the first-conditions of all firms to obtain $p = \frac{\overline{mc}}{1-\frac{\theta}{N\,\boldsymbol{\varepsilon}}}$. Substitute it into Eq. (11) and, after a few simplifications, the market share of firm i can be expressed as

$$s_i = \frac{\boldsymbol{\varepsilon}}{\theta} - \frac{mc_i}{\overline{mc}}\left(\frac{\boldsymbol{\varepsilon}}{\theta} - \frac{1}{N}\right)$$

The ratio $r_i \equiv \dfrac{mc_i}{\overline{mc}}$ is an inverse measure of the cost advantage of firm i: the lower it is, the greater the cost advantage of firm i with respect to average marginal costs in the industry. Plug it into Eq. (A3), to obtain

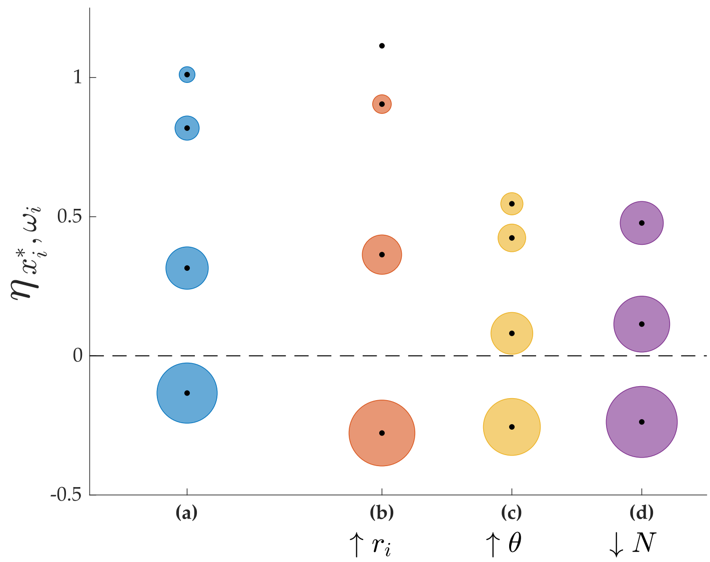
$$\eta_{x_i,\omega_i} = -1 + \frac{\frac{\boldsymbol{\varepsilon}}{\theta\left[\frac{\boldsymbol{\varepsilon}}{\theta} - r_i\left(\frac{\boldsymbol{\varepsilon}}{\theta} - \frac{1}{N}\right)\right]} - 1}{2 - \boldsymbol{\rho}\,\theta\left[\frac{\boldsymbol{\varepsilon}}{\theta} - r_i\left(\frac{\boldsymbol{\varepsilon}}{\theta} - \frac{1}{N}\right)\right]} = -1 + \frac{\frac{1}{1 - r_i\boldsymbol{\varepsilon} + r_i\frac{\theta}{\boldsymbol{\varepsilon}N}} - 1}{2 - \boldsymbol{\rho}\,\boldsymbol{\varepsilon}\left(1 - r_i + r_i\frac{\theta}{\boldsymbol{\varepsilon}N}\right)}.$$

Based on this, the smaller is r_i, the lower is η_{x_i,ω_i}. This is because a firm with a substantial cost-advantage is likely to control a large share of the market. If so, it faces a low price elasticity of its residual demand and has a lower incentive to increase its output after a productivity shock. ■

To illustrate these results, Figure 9 show the simulations of a market with $N = 4$ firms that compete à la Cournot ($\theta = 1$) and face a linear demand. Firms are ranked based on their productivity $i = 1, 2, 3, 4$. The size of the bubbles are proportional to the market share s_i and the values of $\eta_{x_i^*,\omega_i}$ for each firm are represented by the black dots.

In the baseline scenario **(a)**, it is clear that the least productive firm has the highest responsiveness, while the most productive firm (and largest in terms of market shares) has $\eta_{x_1^*,\omega_1} < 0$. In the scenario **(b)**, the largest firm has a higher productivity advantage which reduces its $\eta_{x_1^*,\omega_1} < 0$ ($\downarrow r_1$), while it increases $\eta_{x_i^*,\omega_i}$ for the others ($\uparrow r_{-1}$). In **(c)**, I consider a more collusive conduct $\theta > 1$. This decrease the responsiveness for all firms. In the last scenario **(d)**, the least productive firm does not operate anymore in the market ($N \to 3$). All the other firms face a lower competitive pressure which reduces their responsiveness.

Figure 9. The impact of market structure on the responsiveness.



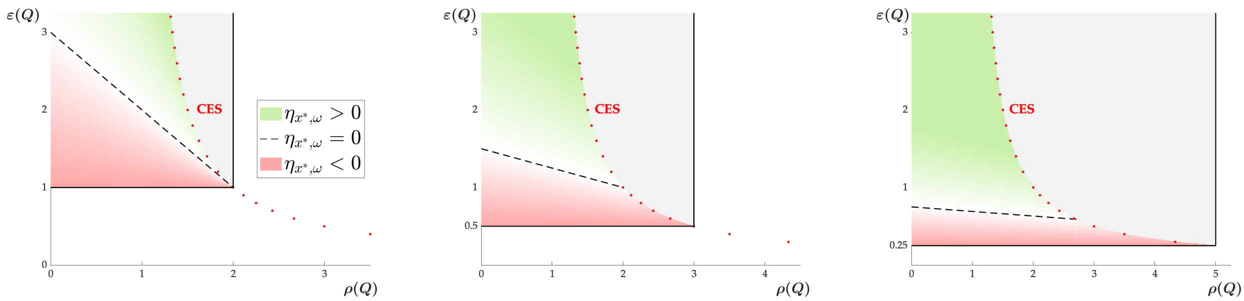
Notes: own simulation based on $w = 4$ and productivity values $\omega_i = [0.95; 0.97; 1.05; 1.2]$. In (b) the highest productivity value become $\omega_{max} = 1.3$ so that r_i changes. In (c) the conduct parameter becomes $\theta = 1.25$. In (d) the least productive exits and $N = 3$.

This example illustrates how a productivity shock can have very different effects in term of input demand depending on the market in which firms operate, the degree of competition and the competitive positioning of each firm.

Demand manifolds in oligopoly. Mrázová and Neary (2017) suggest that the demand manifold framework can also be applied to oligopoly, but they left it to future research. In deriving the results for this section, I extend their analysis to oligopolistic settings which is slightly more complex as the admissible region becomes endogenous to firms' market shares. I discuss about it in more details in Appendix A.3.1. The main take-away is that as the region of possible values of elasticity and curvature of market demand ($p(Q)$) changes, the values leading to $\eta_{x_i^*,\omega_i} \leq 0$ change too. I illustrate this result below in an oligopoly setting à la Cournot ($\theta = 1$) with symmetric firms.

Figure 10. Demand manifold regions in monopoly *vs.* oligopoly.

(a) $N = 1$ **(b)** $N = 2$ **(c)** $N = 4$



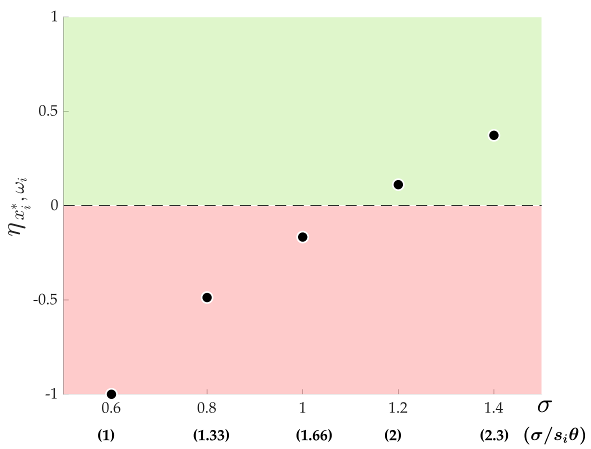
Essentially, this is because the elasticity of residual demand tends to be higher than the elasticity of market demand, i.e. $\frac{\varepsilon(Q)}{\theta\, s_i} \geq \varepsilon(Q)$. As a result, the elasticity of market demand can be lower than 1 in an oligopoly equilibrium, despite the price elasticity of the residual demand of each firm is still greater than 1.

Corollary 6. *A demand function can lead to $\eta_{x_i^*,\omega_i} \leq 0$ in oligopoly, even if this is not the case under monopoly or monopolistic competition. Notably, this is the case for CES demand.*

To highlight this result, in Fig. 11 I show the responsiveness of a firm with $s_i = 60\%$ in a oligopolistic market à la Cournot with CES demand. I consider different values of the elasticities of market demand σ. With less elastic market demand (lower σ), the responsiveness of this firm becomes negative. In the table, I report for different values of σ, the market shares and the corresponding values of markups above which $\eta_{x_i^*,\omega_i} \leq 0$. The lower is σ, the lower the market shares and markups above which the decoupling of factor demand to productivity occurs even with CES demand.

Figure 11. Responsiveness in Oligopoly à la Cournot with CES demand.

(a) Responsiveness for $s_i = 60\%$



(b) Thresholds above which $\eta_{x_i^*,\omega_i} \leq 0$

$\varepsilon(Q) = \sigma$	$s_i \geq \ldots$	$\mu_i \geq \ldots$
0.5	21%	1.7
0.6	26%	1.76
0.75	34%	1.83
0.8	37%	1.85
1	50%	2
1.2	70%	2.4
1.25	83%	3
1.4	-	-

2.5 Extensions

Until this point, I restricted the focus to a production function with a single input factor (**A1**), a constant returns to scale technology (**A2**) and price-taking behavior in the input market (**A3**). Under these assumptions, I show that the responsiveness of derived factor demand to productivity shocks declines with output and becomes negative at low levels of elasticity of output demand. Is it the case also in the presence of multiple input factors or if the technology does not feature constant returns to scale? What if a firm exercises some market power also over its suppliers of input? In this section, I show that this result holds also in these more general environments. While the key economic mechanism remains the same, the elasticity of factor demand to productivity $\eta_{x_i^*,\omega_i}$ gets enriched of other structural determinants. To shed light on each on them in the simplest way, throughout this section I restrict the focus to a profit-maximizing monopolist that produces and sells a single product.

2.5.1 Multiple input factors

Focusing on a single input factor, however simplistic, proved essential to shed light on the key mechanism through which market power can reduce the responsiveness of factor demand to productivity. In a less parsimonious way, I show that the same mechanism remains at play in the presence of multiple input factors. To keep the problem tractable, I consider a simple static maximization problem with only two variable inputs, labor (l) and material (m).[17] Their prices are given to the firm and are denoted by w_l and w_m, respectively.

$$\max_{l,m} \quad \pi = p(q)\, q - w_l\, l - w_m\, m\,.$$

Assumption A1Ext (Inputs). *The firm produces its output q according to a standard production function $q = f(\boldsymbol{x},\omega) = \phi(\boldsymbol{x})\,\omega$ where ω denotes its Hicks-neutral productivity level and $\boldsymbol{x} = l, m$ is a vector of variable input factors. The production function f is assumed to be continuous, increasing ($f_l > 0$, $f_m > 0$) and strictly quasi-concave ($f_{ll} \leq 0$, $f_{mm} \leq 0$) in each input.*

The prices and the marginal products of each input play an important role in determining the optimal combination of inputs for a given level of output and with that the values of $\eta_{l^*,\omega}$ and $\eta_{m^*,\omega}$. However, the (un)responsiveness of factor demand to a Hicks-neutral productivity shock is ultimately driven by the rate of expansion of output that maximizes the profit of a firm with market power. Whenever the firm optimally increases its output less than proportionally to a productivity shock, this has a negative effect across the board for all the variable inputs. In this regard, it is the revenue-maximizing side of the profit-maximization problem, rather than the cost-minimization one, that ultimately leads to $\eta_{l^*,\omega} = 0$ and $\eta_{m^*,\omega} = 0$

Proposition 4. *Under the assumptions on (**A1Ext**) technology, (**A3**) input prices and (**A4**) demand, a monopolist reacts to a Hicks-neutral productivity shock by decreasing the use of its input factors if and only if the elasticity of marginal revenue at its profit-maximizing level of output $\eta_{mr,q}(q^*)$ is lower than -1.*

$$\frac{\partial l^*}{\partial \omega} < 0 \quad \Leftrightarrow \quad \underbrace{\left(-\frac{f_l}{\omega} f_{mm} + \frac{f_m}{\omega} f_{lm}\right)}_{>0} (1 + \eta_{mr,q}) < 0 \quad \Leftrightarrow \quad \eta_{mr,q} < -1$$

$$\frac{\partial m^*}{\partial \omega} < 0 \quad \Leftrightarrow \quad \underbrace{\left(-\frac{f_m}{\omega} f_{ll} + \frac{f_l}{\omega} f_{ml}\right)}_{>0} (1 + \eta_{mr,q}) < 0 \quad \Leftrightarrow \quad \eta_{mr,q} < -1\,.$$

Proof reported in Appendix A.4.1.

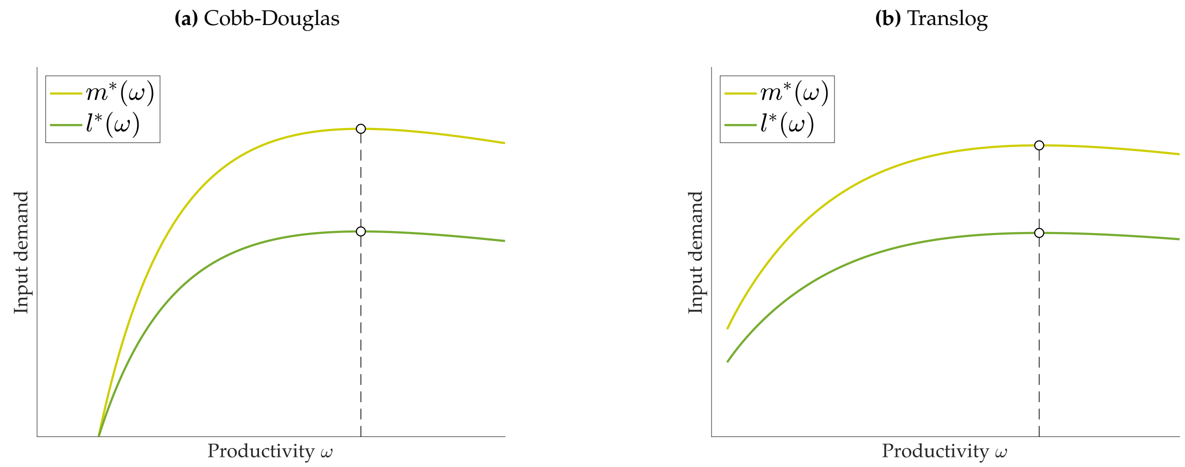
[17] While the theoretical results can be generalized to numerous factor inputs, the key insights can be gained by examining just two factors. Following De Loecker (2011), capital is not part of the analysis as it assumed to a fixed input determined in the previous period.

Having established that the mechanism highlighted with a composite input is at play also when multiple inputs are considered, a related question is what we can infer from the distribution of each of these factors.

Corollary 7. *For a given production function and relative factor prices, a non-monotonic relationship between derived factor demand and Hicks-neutral productivity is inherited by both inputs and occurs at the same level of* ω*.*

I illustrate this result for a Cobb-Douglas production function $q = l^{\beta_l} m^{\beta_m} \omega$, for which the condition in Proposition 4 simplifies to $(\beta_l + \beta_m)(1 + \eta_{mr,q})$, and for a more general Translog production function.

Figure 12. Non-monotonicity with multiple input factors.



Notes: own simulations based on a linear demand and identical factor costs $w_l = w_m$. In figure **(a)** I consider a Cobb-Douglas production function with $\beta_l = 0.4$ and $\beta_m = 0.6$, while in **(b)** a translog production function $log(q) = \beta_l log(l) + \beta_m log(m) + \beta_{ll} log(l)^2 + \beta_{mm} log(m)^2 + \beta_{ml} log(l) log(m) + log(\omega)$ with the same β_l and β_m, $\beta_{ll} = -0.02$, $\beta_{mm} = -0.03$ and $\beta_{ml} = 0.01$. In the Appendix A.4.1, I illustrate also the results for $\eta_{l^*,\omega}$ and $\eta_{m^*,\omega}$.

This result contrasts with the previous analyses by Levinsohn and Petrin (2003) and De Loecker (2011), who proved that the demand for a variable input is always monotonic increasing in productivity. This led to the idea of using changes in variable input factors to control for unobservable productivity shocks. This monotonicity, however, is a necessary condition to apply the control function approach in the estimation of production functions with unobservable productivity. While the proof in Levinsohn and Petrin works in a competitive environment where firms take output prices as given (so markups are zero), De Loecker considers a monopolistic competitive industry with CES demand where markups are constant and $\eta_{mr,q} > -1 \, \forall q$. Beyond these two settings, Proposition 4 proves that this relationship is not always monotonic, which poses a fundamental challenge to the control function approach.

2.5.2 Non-constant returns to scale

In this section, I consider the role of technology in shaping the responsiveness of derived factor demand to productivity shocks. To restrict the focus only on returns to scale, I keep considering a single input (**A1**) but allow for a more general production function.

Assumption A2Ext (Technology). *The productive technology of the firm is described by a homothetic production function* f *which is assumed to be continuous, increasing* $f' > 0$ *and strictly quasi-concave* $f'' \leq 0$ *for any* x*.*

Start from profit-maximizing output and input levels

$$q^* = f(x^*, \omega) = \phi(x^*)\, \omega \,.$$

Express it in logs and differentiate it with respect to ω_i to obtain

$$\frac{d\log(q^*)}{d\log(\omega)} = \frac{d\log(\phi)}{d\log(x^*)}\frac{d\log(x^*)}{d\log(\omega)} + \frac{d\log(\omega)}{d\log(\omega)}$$
$$\eta_{q^*,\,\omega} \;=\; \eta_{\phi,\,x^*} \quad \eta_{x^*,\,\omega} \;+\; 1\;.$$

Compared to Equation (1) with constant returns to scale, the responsiveness of input to productivity is now influenced also by the scale elasticity $\eta_{\phi,\,x^*}$. In particular,

$$\eta_{x^*,\,\omega} = \frac{\eta_{q^*,\,\omega} - 1}{\eta_{\phi,\,x^*}}\,. \tag{12}$$

In fact, also $\eta_{q^*,\,\omega}$ is different because the marginal costs of the firm changes whenever returns to scale are not constant and this influences the incentive to expand production after a productivity shock. The assumptions of homotheticity of the production function (and Hicks-neutral productivity) ensures that relative changes in cost can be decoupled into output and productivity effects. In particular, I follow Bakhtiari (2009) in denoting the two components of the (dual) cost function as $C(q,\omega,w) = c_1(q)c_2(\omega)\,w$ with $c_1' > 0$ and $c_2' < 0$. As shown in the proof for Proposition 5, this turns out to be useful because $\eta_{q^*,\,\omega} = \frac{\eta_{c_2,\,\omega}}{\eta_{mr,\,q}-\eta_{mc,\,q}}$ where $\eta_{c_2,\,\omega} = \frac{c_2'\omega}{c_2}$ is the elasticity of the component of the cost function directly related to productivity. Moreover, for homothetic production functions, the scale elasticity $\eta_{\phi,\,x^*}$ equals the returns to scale (RTS) of the production function and is equal to the inverse of the elasticity of the costs function with respect to quantity $\eta_{C,\,q} = \frac{C_q\,q}{C}$.[18]

This connection between the technology based definition of scale economies and the cost based definition turns out very useful to further unpack Equation (12).

Proposition 5. *Under the assumptions on (**A1-A2Ext**) technology, (**A3**) input prices and (**A4**) demand, the responsiveness of factor demand to productivity shock of a monopolist is influenced also by the features of its cost function:*

$$\eta_{x^*,\omega} = \eta_{c_2,\omega}\frac{\eta_{C,q} + \eta_{mr,q} - \eta_{mc,q}}{\eta_{mr,q} - \eta_{mc,q}}\,.$$

Proof reported in Appendix A.4.2.

This can also be expressed directly in terms of elasticity and convexity of the cost function since $\eta_{mc,q} \equiv \frac{C_{qq}\,q}{C_q} = \rho_{C,q}$

$$\eta_{x^*,\omega} = \eta_{c_2,\omega}\frac{\eta_{C,q} + \eta_{mr,q} - \rho_{C,q}}{\eta_{mr,q} - \rho_{C,q}}$$

In practice, this result has the following implication about the influence of returns to scale on the elasticity of derived factor demand.

Corollary 8. *Decreasing returns to scale reduces the responsiveness of a firm's derived factor demand to a positive productivity shock. Increasing returns have the opposite effect.*

If returns to scale are decreasing, indeed, marginal costs increase as a firm produces more. In turns, this reduces the incentive to further expand production after a productivity shock. The opposite is true with increasing

[18] As discussed by Panzar (1989) in a setting with multiple input factors, under mild regularity conditions the technology based definition of scale economies $\tilde{S} = \frac{\sum x_i f_i(\mathbf{x})}{f(\mathbf{x})}$ and the cost based definition $S = \frac{C(q,\omega,w)}{q\,C_q(q,\omega,w)}$ are equivalent. With a different terminology, Frisch (1965) show that the scale elasticity $\tilde{S}$ (called "passus coefficient") is the reciprocal of the elasticity of the costs function (referred as "substitumal cost flexibility"). This has been recently revisited also by Syverson (2019). In the context under analysis, this is the case since

$$\frac{1}{\eta_{C,\,q}} = \frac{C}{q\,C_q} = \frac{wx^*}{\phi(x)\omega\,\frac{w}{\omega\phi_x}} = \frac{\phi_x\,x^*}{\phi} = \eta_{\phi,\,x_i^*}\,.$$

This also implies that the scale elasticity $S(q) = \frac{C(q,\omega,w)}{q\,C_q(q,\omega,w)}$, which describes how the ratio of average to marginal costs varies with output, depends only on the output level.

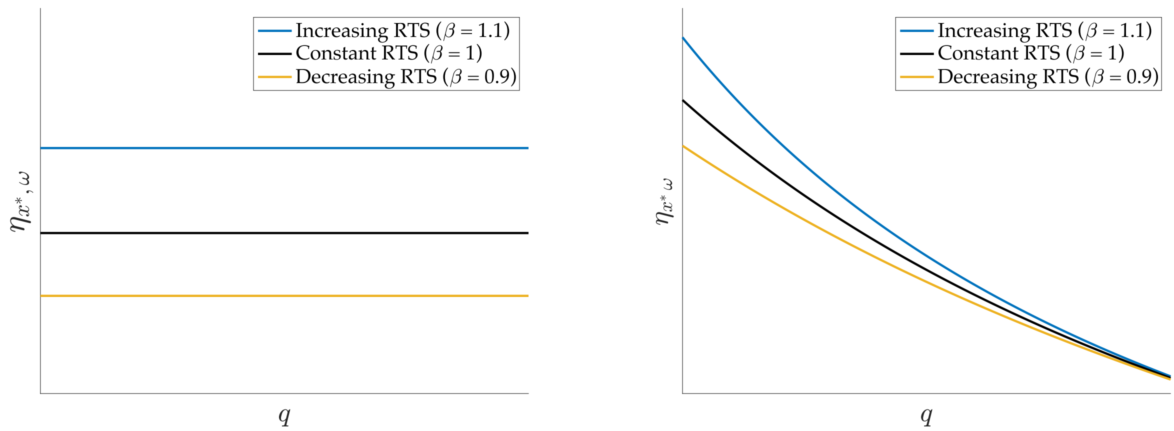
returns to scale. To understand this, consider a simple Cobb-Douglas production function, $q = x^{\beta}\omega$, where Proposition 5 leads to

$$\eta_{x^*,\omega} = -\frac{1}{\beta}\frac{\frac{1}{\beta} + \eta_{mr,q} - \left(\frac{1}{\beta} - 1\right)}{\eta_{mr,q} - \left(\frac{1}{\beta} - 1\right)} = \frac{1 + \eta_{mr,q}}{1 - \beta - \beta\eta_{mr,q}} .$$

This is illustrated below for different degrees of returns to scale (i.e. values of β) with CES and linear demand.

Figure 13. Influence of returns to scale on $\eta_{x^*,\omega}$.

(a) CES demand **(b)** Linear demand



This result implies that with Cobb-Douglas production function (and many other homogeneous production functions) the degree of scale economies enjoyed by the firm affects the level of $\eta_{x^*,\omega}$, but it does not change the level of output at which $\eta_{x^*,\omega} = 0$. In Appendix A.4.2, I discuss the implications for $\eta_{x^*,\omega}$ when returns to scale vary with output.

2.5.3 Monopsonistic power in input market

The last extension relates to the price paid by the firm to purchase/hire/rent the input factor x. If the input market is not competitive, but the firm exercises some market power also on its suppliers.[19]

Assumption A3Ext (Input price). *The firm faces an upward-sloping inverse supply curve for the input factor x, i.e. $w(x)$ with $w' > 0$.*

In such a setting, the resulting first-order condition for profit-maximization is

$$\frac{\partial \pi}{\partial x} = \frac{\partial \pi}{\partial q}\frac{\partial q}{\partial x} = 0 \quad \Leftrightarrow \quad \underbrace{(p + p'\, q)\,\omega}_{mrp} = \underbrace{w + w'(x)x}_{me} .$$

The marginal revenue product of the input (mrp) is set equal to the marginal expenditure (me). However, instead of just being the input price w, the me has now an extra term reflecting the fact that a monopsonist must raise the input price if it demands and purchases an additional unit of input. In a static framework, the degree of monopsony power is measured by the wedge between the marginal revenue product and the factor price

$$\frac{mrp}{w} = (1 + \eta_{w,x}) ,$$

[19] Since the seminal work of Robinson (1933), the concept of monopsony has been predominantly used when referring to market power in the labour markets. However, it can be applied to any factor market in which a firm manages to set a price below the marginal product of the input.

which ultimately depends on the elasticity of inverse supply $\eta_{w,x} \equiv \dfrac{w'(x)\,x}{w(x)}$. The higher it is, the higher will be the monopsony power exercised by the firm. With monopsony power, the derived factor demand of a monopolist is

$$x^* = \frac{w - p\,\omega}{p'\omega^2 - w'}\,. \tag{13}$$

Based on this, it is possible to extend the results of Proposition 1 to the presence of monopsony power in the following way.

Proposition 6. *Under the assumptions on (**A1-A2**) technology, (**A3Ext**) input prices and (**A4**) demand, the responsiveness of factor demand to productivity shock of a monopolist is influenced also by the elasticity of its marginal expenditure* $\eta_{me,x} \equiv \frac{me'(x)\,x}{me(x)}$. *In particular,*

$$\eta_{x^*,\omega} = -\frac{1+\eta_{mr,q}}{\eta_{mr,q} - \eta_{me,x}}\,.$$

Proof reported in Appendix A.4.3.

This result implies that, in the presence of monopsony power, productivity improvements will lead to a smaller increase in optimal output because the firm faces an additional trade-off between keeping its margins and increasing output. This mechanism is now at play on its cost-side. As a firm produces more, indeed, its marginal expenditures (me) increase as well due to monopsonistic pecuniary effects. As a result, monopsony power represents an additional factor that refrains a firm from expanding its output after a productivity shock. This can be seen by comparing output adjustment in the presence of monopsony or not:

$$\begin{aligned} \eta^{Monopsony}_{q^*,\omega} &\leq \eta^{Competitive}_{q^*,\omega} \\ \frac{\eta_{me,x}\,\eta_{x^*,\omega}}{\eta_{mr,q}} - \frac{1}{\eta_{mr,q}} &\leq -\frac{1}{\eta_{mr,\,q}}\,. \end{aligned}$$

As $\eta_{me,x} \geq 0$, there is an extra negative component on the left-hand side. Therefore, *ceteris paribus* $\eta_{q^*,\omega}$ tends to be lower if a firm can exercise some monopsonistic power on its suppliers, with direct consequence for derived factor demand.

Corollary 9. *The presence of monopsonistic power in the input market reduces the responsiveness of a firm's derived factor demand to productivity shocks, but it does not change the level of output at which* $\eta_{x^*,\,\omega} = 0$.

$$\eta_{x^*,\,\omega} = 0 \quad \Leftrightarrow \quad \eta_{mr,q} = -1$$

Proof reported in Appendix A.4.3, where I illustrate this result within the manifold framework (Figure A3). Similarly to the elasticity of marginal revenue, the elasticity of marginal expenditure ($\eta_{me,x}$) depends on the shape of the inverse supply curve. In particular, I find that it is jointly determined by its elasticity $\eta_{w,x}$ and convexity $\rho_{w,x}$ according to

$$\eta_{me,x} = -\frac{\eta_{w,x}\,(2+\rho_{w,x})}{1+\eta_{w,x}} \quad \text{where} \quad \eta_{w,x} \equiv \frac{w'(x)\,x}{w(x)} \quad \text{and} \quad \rho_{w,x} \equiv \frac{w''(x)\,x}{w'(x)}$$

Mirroring the demand function, a higher elasticity of inverse supply curve $\eta_{w,x}$ refrains a firm from getting even larger, as marginal factor costs increase due to monopsonistic pecuniary effects. On top of it, also the convexity of the inverse supply curve matters as it determines the rate at which these effects on marginals costs increase as a firm gets larger, moving along the inverse supply curve.

3 From theory to empirics: how relevant is it?

Having established the generality of this result from a theoretical point of view, the logical next step is to assess its empirical relevance. To do that, the ideal approach would be to observe a firm decreasing its input use while increasing its output after a well-identified positive Hicks-neutral productivity shock. In particular,

$$\begin{cases} \Delta x_i^* \ (\Delta\omega_i) < \ 0 \\ \Delta q_i^* \ (\Delta\omega_i) > \ 0 \end{cases} \tag{14}$$

would be the indication of a decoupling of derived factor demand from productivity growth, i.e. $\eta_{x_i^*,\omega_i} < 0$.

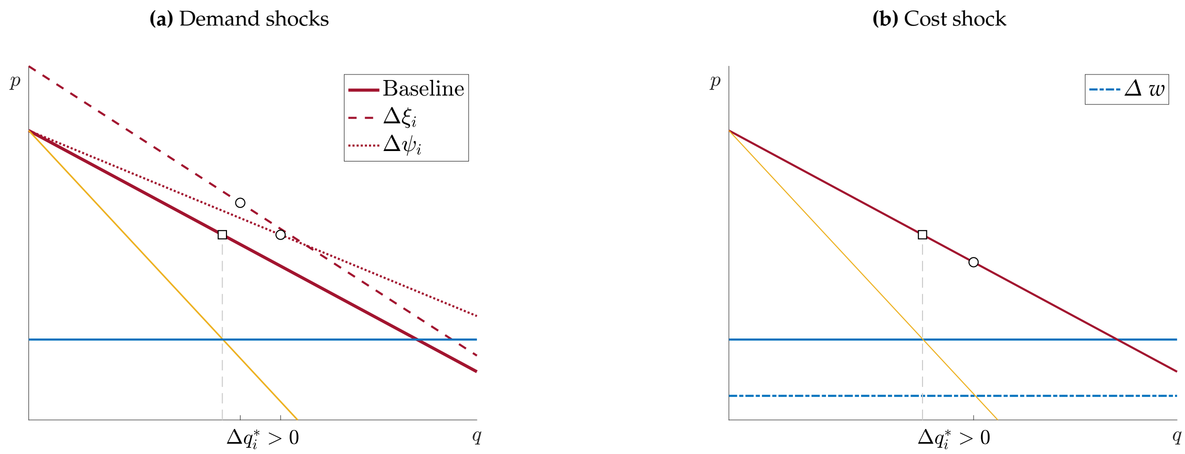
3.1 Challenges

When brought to the data, however, this simple prediction gets confronted by a number of issues which makes it more difficult to detect whether a firm is operating in the range of elasticity of demand where $\eta_{x_i^*,\omega_i} < 0$.

Other firm-level shocks. A first issue is fundamentally related to the fact that the prediction about $\eta_{x^*,\omega} \leq 0$ is a comparative statics in which by construction everything else is held constant (or abstracted away). In the presence of other contemporaneous shocks that lead a firm to increase both its input and output, the effect $\Delta x_i^* \ (\Delta\omega_i) < \ 0$ can be overshadowed.

This is the case, for example, for standard demand shocks and cost shocks. In general, the demand for a firm's product can increase between two periods because of changes in its appeal or because of changes in the market size. For a given functional form of demand $p(q_i)$, in the first case the price consumers are willing to pay changes by the same factor (denoted by ξ_i) for all quantities. The resulting demand for the firm is equal to $\xi_i \, p(q_i)$. In the case of market size changes, demand varies by the same factor ψ_i for any price so that $p(\frac{q_i}{\psi_i})$. This may due to more consumers being present in a market or the firm entering in another geographical market with the same features of demand. In terms of costs shocks, the price paid by a firm to purchase its input can change. In Appendix B, I prove that these demand shocks, differently from productivity, always lead to $\Delta x_i^* > 0$ and $\Delta q_i^* > 0$. This is the case also for a reduction in input prices. To fix the ideas, I illustrate below the comparative statics of these other shocks in the case of a monopolist facing a linear demand.

Figure 14. Reaction of $\Delta q_i^* > 0$ to other firm-level shocks.



Notes: comparative statics of demand and cost shocks for a monopolist facing a linear demand. In Appendix B, I show this also in terms of $mrp(x)$.

As a result, if a firm experiences a productivity shock and at the same time another of these shocks, the net effect on input demand may not be necessarily negative, even if $\Delta x_i^* \; (\Delta \omega_i) < \; 0$.

$$\begin{cases} \Delta x_i^* \; (\Delta \omega_i, \, \underbrace{\Delta \xi_i}_{+}, \, \underbrace{\Delta \psi_i}_{+}, \, \underbrace{-\Delta w}_{+}) \gtreqless 0 \\ \Delta q_i^* \; (\Delta \omega_i, \, \underbrace{\Delta \xi_i}_{+}, \, \underbrace{\Delta \psi_i}_{+} \, \underbrace{-\Delta w}_{+}) > 0 \, . \end{cases}$$

This raises a fundamental challenge in detecting $\eta_{x_i^*, \, \omega_i} < 0$.

Revenue *vs.* physical output. A second challenge is empirical and posed by the fact that in most firm-level datasets output is reported only in terms of revenue rather than physical quantities ($r_i = p_i \, q_i \; vs. \; q_i$). As a result, firm-level productivity estimated as a residual from a production function is likely to biased. This is a well-known issue in the literature on productivity and the standard solution is to use deflate revenues with a price index. While this should solve the problem under perfect competition, this is not the case with imperfect competition since deflated revenue reflects firm-level output price heterogeneity.[20] This problem is accentuated in the presence of additional sources of firm heterogeneity, such as firm-level demand or costs differences (in levels and changes).

[20] In fact, Klette and Griliches (1996) as well as De Loecker (2011) propose a solution that is valid only under monopolistic competition and if demand when CES. However, I can not adopt this solution since my focus is explicitly on contexts of imperfect competition with variable elasticity of demand. Despite progress has been made towards this direction (e.g. Kasahara and Sugita, 2021), we are currently not equipped (yet) to precisely estimate productivity under imperfect competition beyond CES.

3.2 Approaches to detect $\eta_{x^*,\omega} < 0$

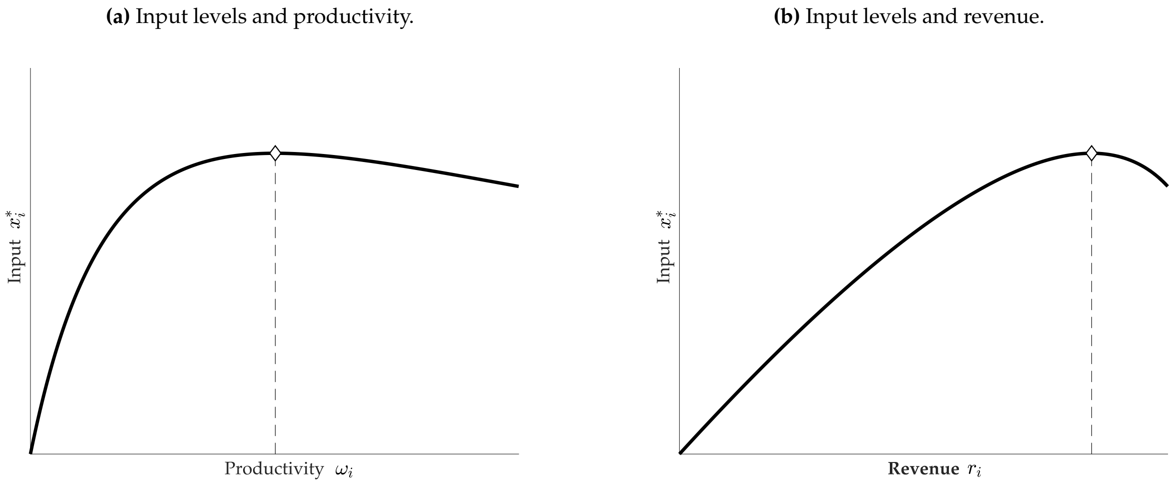
Despite these challenges, in this section I argue that under certain assumptions it is possible to infer directly from data on revenue and input whether firms in a given market react to a productivity shock by decreasing their input use. In particular, I propose two approaches, one based in levels and another one in changes, which can be adopted by researchers even if firm-level productivity can not be estimated and without taking a stance on the specific functional form of demand faced by the firms.

3.2.1 Detection test *in levels*

As shown in Figure 7 and discussed in Section 2.4.1, in a monopolistic competitive market the cross-sectional distribution of firms' input and productivity levels has to be non-monotonic if $\eta_{x_i^*,\omega_i} < 0$. On top of that, I find that this non-monotonicity gets reflected also in the relationship between derived factor demand and revenue. As a result, the distribution of revenue and input across firms can be used to infer whether larger firms are operating in the range of price elasticity of demand where $\eta_{x^*,\omega} < 0$. This *self-reflection* property is illustrated in the figure below. I provide a proof for this in Appendix B.

Prediction in *(levels)*: in a given cross-section of monopolistic competitive firms, $\eta_{x_i^*,\omega_i} < 0$ leads to non-monotonic relationship between their revenues and inputs.

Figure 15. Non-monotonic relationship between input demand and productivity.



Notes: equilibrium outcome with linear demand (details reported in Appendix A.2). The white diamond corresponds to $\eta_{x^*,\omega} = 0$.

However, this test is not robust if there are additional sources of firm heterogeneity, especially when correlated with productivity. For example if more productive firms have a higher demand for their products because of higher perceived quality/appeal by consumers, these firms will have both higher revenue and input in equilibrium. As a result, the cross-sectional relationship would look monotonic even if these firms face a price elasticity of demand such that $\eta_{x_i^*,\omega_i} < 0$.[21] If a researcher has reasons to believe that in the market under analysis there are other fundamental differences between firms in addition to their productivity, this approach in levels may fail to detect $\eta_{x_i^*,\omega_i} < 0$ even if it is actually there and deliver *de facto* many "false negatives".

[21] In numerical simulation I find that this test in levels remain valid in the presence of demand appeals heterogeneity if and only if the dispersion in ξ_i is not too large and if it is not positively correlated with productivity.

3.2.2 Detection test in *changes*

Looking at the relationship between changes in input and revenues can be informative even if there are multiple sources of firm heterogeneity and shocks. The logic for that is simple. If a profit-maximizing firm increases its revenue ($\Delta r_i^* > 0$) between two periods t and $t+1$, something must have changed in its fundamentals. It can be the result of a positive productivity change ($\uparrow \omega_i$), but also a positive demand shock ($\uparrow \xi_i$ or $\uparrow \psi_i$), a reduction in input prices ($\downarrow w$) or even a mix of them.[22] It is usually hard for a researcher to quantify the magnitude of each shocks or even distinguish which type of shock has hit a firm. Among these shocks, however, only a productivity shock can ultimately lead to $\Delta x_i^* \leq 0$ if a firm face a low elasticity of demand. As I show in Table 2, all the other shocks lead to higher input demand $\Delta x_i^* \geq 0$.

Table 2. Optimal reaction of output and input to various shocks.

	$\eta_{q_i^*,\ldots}$	$\eta_{r_i^*,\ldots}$	$\eta_{x_i^*,\ldots}$	
Demand appeal ($\uparrow \xi_i$)	>0	>0	>0	
Market size ($\uparrow \psi_i$)	>0	>0	>0	
Input costs ($\downarrow w$)	>0	>0	>0	
Productivity ($\uparrow \omega_i$)	>0	>0	>0 High $\varepsilon(q_i^*)$	<0 Low $\varepsilon(q_i^*)$

This sign restriction is the cornerstone for the detection test in changes, which is based on the ratio of input to revenue relative changes between two periods, t and $t+1$.

$$\text{Ratio} = \frac{\%\Delta x_i^*}{\%\Delta r_i^*} = \frac{log(x_{i,t+1}^*) - log(x_{i,t}^*)}{log(r_{i,t+1}^*) - log(r_{i,t}^*)}$$

The value of this ratio can be directly linked to the values of elasticity and curvature of demand. As I show in details in Appendix B, this Ratio becomes negative if and only if $\varepsilon < 3 - \rho$ so that

$$\text{Ratio} = \left.\frac{\%\Delta x_i^*}{\%\Delta r_i^*}\right|_{\Delta r_i^* > 0} < 0 \iff \varepsilon < 3 - \rho \iff \eta_{x_i^*,\omega_i} < 0\,.$$

For this reason, it provides direct guidance whether and to what extent this non-monotonicity has kicked-in (or not) in a given market. Conditioning on $\Delta r_i > 0$ restricts the focus on firms that effectively have increased their revenue and output between t and $t+1$. The behavior of this ratio across the distribution of firms' revenue within a market reveals whether some of these firms operate in the range of elasticities where $\eta_{x_i^*,\omega_i} < 0$. As the firms with higher revenues tend to face a lower elasticity of demand, the following cross-sectional prediction holds and is testable in the data.

Prediction *(in changes)*: within a market if $\eta_{x_i^*,\omega_i} < 0$, the Ratio $= \frac{\%\Delta x_i}{\%\Delta r_i}$ declines with revenue and is more likely to be negative among firms with higher revenues.

Even if a researcher does not know whether or to what extent other firm-level shock has driven this increase, the sign of the Ratio of revenue to input changes a researcher remains informative. This is because a productivity shock is the only one among these shocks that can lead to an increase in revenue and at the same

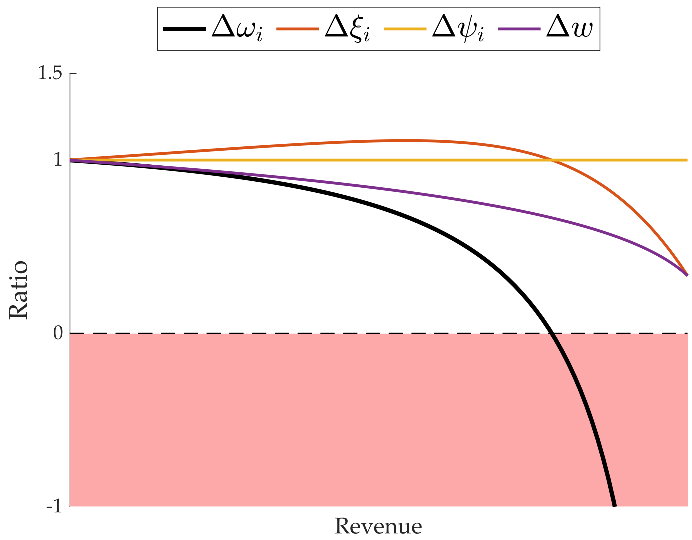
[22] While I focus on these canonical shocks, the list of other possible shocks is obviously not exhaustive.

time reduction in input use. Also the other shocks affect the Ratio but they keep it always above zero, since they lead to an increase in the input use.

$$\text{Ratio} = \frac{\%\Delta x_i}{\%\Delta r_i} = \frac{\eta_{x_i^*,\dots}}{\eta_{r_i^*,\dots}} = \frac{\eta_{x_i^*,\xi_i} + \eta_{x_i^*,\psi_i} + \eta_{x_i^*,-w} + \eta_{x_i^*,\omega_i}}{\eta_{r_i^*,\xi_i} + \eta_{r_i^*,\psi_i} + \eta_{r_i^*,-w} + \eta_{r_i^*,\omega_i}} = \underbrace{\frac{\eta_{x_i^*,xi_i} + \eta_{x_i^*,\psi_i} + \eta_{x_i^*,-w}}{\eta_{r_i^*,\dots}}}_{>0} + \underbrace{\frac{\eta_{x_i^*,\omega_i}}{\eta_{r_i^*,\dots}}}_{\gtreqless 0}$$

Therefore, a negative ratio is possible if and only if $\eta_{x_i^*,\omega_i} < 0$. However, the opposite is not true: a positive Ratio does not imply that $\eta_{x_i^*,\omega_i} > 0$ since the productivity shock might be overshadowed by other shocks. I illustrate the logic of this test below by plotting the values of Ratio $= \frac{\%\Delta x_i}{\%\Delta r_i} = \frac{\eta_{x_i^*,\dots}}{\eta_{r_i^*,\dots}}$ for different shocks in the case of linear demand.[23] In the presence of multiple shocks, the net effect on the Ratio will be a combination of these outcomes.

Figure 16. Values of Ratio for different shocks along the revenue distribution.



Notes: the revenue distribution is in the baseline year *t*.

[23]In Appendix B.2.1, I discuss the different predictions between sub-convex and super-convex demands.

4 Application to Chinese manufacturing firms

I apply these two detection tests to firm-level data the Census of Chinese manufacturing over the period 1998-2007, building on the work by Brandt et al. (2017).

Results will be available soon.

5 Conclusions

In this paper, I identify and unpack an overlooked mechanism through which market power can reduce and even mute the responsiveness of firms' factor demand to productivity shocks. This simple yet powerful result challenges the common presumption in the literature that firms always expand their factor demand when they become more productive. To my knowledge, this (dis)connection between product and factor markets has remained overlooked so far.

From a theoretical point of view, I am currently investigating the relevance of this result for the measurement of reallocation of productive resources within industries. If large firms are not necessarily the most productive, looking at the covariance between size and productivity is not very informative anymore. This result is also related to the highly-debated aggregation of firm-level outcomes, such as markups and labor share. If large firms are not necessarily the most productive firms, using cost *vs.* revenue shares is likely to deliver a different result. Finally, this new result may offer an alternative explanation to the secular decline responsiveness of labor demand to productivity shocks. While the literature has focused so far on the role of adjustment costs Decker et al. (2020), the role of firms' market power in the output and input markets has not received sufficient attention. Testing this alternative explanation on European firms is part of an ongoing research project with Inferrera, Mertens and Miranda. Methodologically, as discussed in Section 2.5.1, this non-monotonicity result poses a fundamental challenge to the control function approach. More work is certainly needed to properly analyze the actual impact on the estimates of the production function and the results that are based on them.

Despite the settings considered in the paper being intentionally simple, they are at the core of most theoretical and empirical analyses related to firm productivity, not only in industrial organization but also in international and macroeconomics. For this reason, the mechanism identified in this paper has wide-ranging implications, many of which still need to be unveiled.

References

Arkolakis, C., Costinot, A., Donaldson, D., & Rodríguez-Clare, A. (2019). The Elusive Pro-Competitive Effects of Trade. *Review of Economic Studies*, *86*(1), 46–80.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., & Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, *135*(2), 645–709.

Bakhtiari, S. (2009). *Quantifying Non-monotonicity in Productivity-Input Relations.* UNSW School of Economics Discussion Paper: 2009/11.

Bakhtiari, S. (2012). Markets and the non-monotonic relation between productivity and establishment size. *The Canadian Journal of Economics*, *45*(1), 345–372.

Behrens, K., Mion, G., Murata, Y., & Suedekum, J. (2020). Quantifying the gap between equilibrium and optimum under monopolistic competition. *Quarterly Journal of Economics*, *135*(4), 2299–2360.

Brandt, L., Van Biesebroeck, J., Wang, L., & Zhang, Y. (2017). Wto accession and performance of chinese manufacturing firms. *American Economic Review*, *107*(9), 2784–2820.

Bulow, J. I., & Pfleiderer, P. (1983). A Note on the Effect of Cost Changes on Prices. *Journal of Political Economy*, *91*(1), 182–185.

Caplin, A., & Nalebuff, B. (1991). Aggregation and imperfect competition: on the existence of equilibrium. *Econometrica*, *59*, 25–59.

Cowan, S. (2016). Welfare-increasing third-degree price discrimination. *RAND Journal of Economics*, *47*(2), 326–340.

De Loecker, J. (2011). Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica*, *79*(5), 1407–1451.

De Loecker, J., Eeckhout, J., & Mongey, S. (2021). *Quantifying Market Power and Business Dynamism in the Macroeconomy.* NBER Working Paper No. 28761.

De Loecker, J., & Syverson, C. (2021). An industrial organization perspective on productivity. In *Handbook of industrial organization* (4th ed., pp. 141–223). Elsevier.

Deaton, A., & Muellbauer, J. (1980). An Almost Ideal Demand System. *The American Economic Review*, *70*(3), 312–326.

Decker, Haltiwanger, Jarmin, & Miranda. (2020). Changing Business Dynamism and Productivity: Shocks versus Responsiveness. *American Economic Review*, *110*(12), 3952–3990.

de Meza, D. (1982). Generalized Oligopoly Derived Demand With an Application To Tax Induced Entry. *Bulletin of Economic Research*, *34*(1), 1–16.

Dhingra, S., & Morrow, J. (2019). Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity. *Journal of Political Economy*, *127*(1), 196–232.

Dixit, A. K., & Stiglitz, J. E. (1977). American Economic Association Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, *67*(3), 297–308.

Dobbs, I. . M. ., Hill, M. . B. ., & Waterson, M. . (1987). Industrial Structure and the Employment Consequences of Technical Change. *Oxford Economic Papers*, *39*(3), 552–567.

Edmond, C., Midrigan, V., & Xu, D. (2018). *How Costly Are Markups?* NBER Working Paper Series 24800.

Foran, T. G. . (1976). Market Structure and Derived Demand. *Economica*, *43*(169), 83–87.

Frisch, R. (1965). *Theory of production*. Dordrecht: Springer.

Ilut, C., Kehrig, M., & Schneider, M. (2018). Slow to hire, quick to fire: Employment dynamics with asymmetric responses to news. *Journal of Political Economy*, *126*(5), 2011–2071.

Kehrig, M., & Vincent, N. (2021). The micro-level anatomy of the labor share decline. *The Quarterly Journal of Economics*, *136*(2), 1031–1087.

Kimball, M. S. (1995). The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit, and Banking*, *27*(4), 1241.

Klenow, P. J., & Willis, J. L. (2016). Real Rigidities and Nominal Price Changes. *Economica*, *83*(331), 443–472.

Klette, T. J., & Griliches, Z. (1996). The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and. *Journal of Applied Econometrics*, *11*(4), 343–361.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, *70*(2), 317–341.

Matsuyama, K., & Ushchev, P. (2022). *Selection and sorting of heterogeneous firms through competitive pressures.* CEPR Discussion Paper No. DP17092.

Maurice, S. . C., & Ferguson, C. . E. . (1973). Factor Demand Elasticity under Monopoly and Monopsony. *Economica*, *40*(158), 180–186.

Mayer, T., Melitz, M. J., & Ottaviano, G. I. (2021). Product Mix and Firm Productivity Responses to Trade Competition. *The Review of Economics and Statistics*, *103*(5), 874–891.

Melitz, M. J. (2003). The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity. *Econometrica*, *71*(6), 1695–1725.

Melitz, M. J., & Ottaviano, G. I. (2008). Market Size, Trade, and Productivity. *Review of Economic Studies*, *75*(1), 295–316.

Mrázová, M., & Neary, J. P. (2017). Not So Demanding: Demand Structure and Firm Behavior. *American Economic Review*, *107*(12), 3835–3874.

Mrázová, M., & Neary, J. P. (2019). Selection Effects with Heterogenous firms. *Journal of the European Economic Association*, *17*(4), 1294–1334.

Mrázová, M., & Neary, J. P. (2020). IO for exports(s). *International Journal of Industrial Organization*, *70*.

Mrazova, M., Parenti, M., & Neary, J. P. (2021). Sales and markup dispersion: theory and empirics. *Econometrica*, *89*(4), 1753–1788.

Panzar, J. C. (1989). Technological determinants of firm and industry structure. In *Handbook of industrial organization* (Vol. 1, pp. 4–56). Elsevier.

Robinson, J. (1933). *The economics of imperfect competition*. London: MacMillan.

Seade, J. (1980). On the Effects of Entry. *Econometrica*, *48*(2), 479–489.

Syverson, C. (2011). What determines productivity. *Journal of Economic Literature*, *49*(2), 326–365.

Syverson, C. (2019). Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives*, *33*(3), 23–43.

Verboven, F., & Van Dijk, T. (2009). Cartel damages claims and the passing-on defense. *Journal of Industrial Economics*, *57*(3), 457–491.

Waterson, M. (1980). Oligopoly and Derived Demand. *Economics Letters*, *5*, 115–118.

Weyl, E. G., & Fabinger, M. (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. *Journal of Political Economy*, *121*(3), 528–583.

Zhelobodko, E., Kokovin, S., Parenti, M., & Thisse, J. (2012). Monopolistic Competition: Beyond the Constant Elasticity of Substitution. *Econometrica*, *80*(6), 2765–2784.

Appendices

A Additional theoretical results

A.1 Monopoly

Proof of Proposition 1.

Proof. (not reported in main text) Start from the derived factor demand of the monopolist x^* reported in Eq. (7) and take its derivative with respect to productivity:

$$\frac{\partial x^*}{\partial \omega} = \frac{1}{p'\,\omega^2}\,(p-2c) + \frac{p''}{p'^2\,\omega}\,\frac{\partial q^*}{\partial \omega}\,(p-c) - \frac{1}{\omega}\,\frac{\partial q^*}{\partial \omega}\ \gtreqless 0\,.$$

Express it in terms of elasticity by multiplying by $\dfrac{\omega}{x^*}$

$$\eta_{x^*,\omega} = \frac{\omega}{x^*}\,\frac{1}{p'\,\omega^2}\,(p-2c) + \frac{\omega}{x^*}\,\frac{p''}{p'^2\,\omega}\,\frac{\partial q^*}{\partial \omega}\,(p-c) - \frac{\omega}{x^*}\,\frac{1}{\omega}\,\frac{\partial q^*}{\partial \omega}\,.$$

The conditions for profit maximization imply that the marginal cost must be equal to marginal revenue, so that $(p-2mc) = -p - 2\,p'q$ and $(p-mc) = -p'q$. As a result,

$$\eta_{x^*,\omega} = \frac{1}{x^*}\,\frac{(-p-2\,p'q)}{p'\,\omega} + \frac{1}{x^*}\,\frac{p''}{p'^2}\,\frac{\partial q^*}{\partial \omega}\,(-p'q) - \frac{1}{x^*}\,\frac{\partial q^*}{\partial \omega}\,.$$

Substitute $x^*\omega = q^*$ and apply the definitions of elasticity and curvature from Eq. (4)

$$\eta_{x^*,\omega} = \frac{(-p-2\,p'q)}{p'q} - \frac{1}{x^*}\,\frac{p''\,q}{p'}\,\frac{\partial q^*}{\partial \omega} - \frac{1}{x^*}\,\frac{\partial q^*}{\partial \omega} = \boldsymbol{\varepsilon}(q^*) - 2 + (\boldsymbol{\rho}(q^*) - 1)\,\frac{1}{x^*}\,\frac{\partial q^*}{\partial \omega}\,. \tag{A1}$$

To get $\dfrac{\partial q^*}{\partial \omega}$, take the derivative of the first-order condition with respect to ω so that

$$\frac{\partial\,(p+p'q)}{\partial q}\,\frac{\partial q}{\partial \omega} = \frac{\partial\,mc}{\partial \omega} \quad\Rightarrow\quad \frac{\partial q}{\partial \omega} = \frac{\dfrac{\partial\,mc}{\partial \omega}}{\dfrac{\partial\,(p+p'q)}{\partial q}} = \frac{-\dfrac{w}{\omega^2}}{2p'+p''q} = -\frac{c}{\omega(2p'+p''q)}\,.$$

At q^*, it must be that $mc = p + p'q^*$ and $\dfrac{1}{\omega} = \dfrac{x^*}{q^*}$. Therefore,

$$\frac{\partial q^*}{\partial \omega} = -x^*\,\frac{p+p'q^*}{q^*(2p'+p''q^*)} = x^*\,\frac{-p\,(1+\frac{p'q^*}{p})}{q^*\,p'\,(2+\frac{p''q}{p'})} = x^*\,\frac{\boldsymbol{\varepsilon}(q^*)\left(1-\frac{1}{\boldsymbol{\varepsilon}(q^*)}\right)}{2-\boldsymbol{\rho}(q^*)} = x^*\,\frac{\boldsymbol{\varepsilon}(q^*)-1}{2-\boldsymbol{\rho}(q^*)} > 0\,.$$

Plug $\dfrac{\partial q^*}{\partial \omega} > 0$ back into Eq. (A1) to obtain

$$\eta_{x^*,\omega} = \boldsymbol{\varepsilon}(q^*) - 2 + (\boldsymbol{\rho}(q^*) - 1)\,\frac{\boldsymbol{\varepsilon}(q^*)-1}{2-\boldsymbol{\rho}(q^*)} = \frac{\boldsymbol{\varepsilon}(q^*)+\boldsymbol{\rho}(q^*)-3}{2-\boldsymbol{\rho}(q^*)}\,. \tag{A2}$$

Since the denominator is always positive for the second-order condition to hold (i.e. $\boldsymbol{\rho}(q^*) < 2$), condition *(c)* is proved:

$$\eta_{x^*,\omega} < 0 \overset{(c)}{\iff} \boldsymbol{\varepsilon}(q^*) < 3 - \boldsymbol{\rho}(q^*)\,.$$

Based on this, substitute $\dfrac{1}{x^*} = \dfrac{q^*}{\omega}$ into Eq. (A1) to obtain $\eta_{q^*,\omega} = \frac{\partial q^*}{\partial \omega}\frac{\omega}{q^*}$ and

$$\eta_{x^*,\omega} = \boldsymbol{\varepsilon}(q^*) - 2 + (\boldsymbol{\rho}(q^*) - 1)\,\eta_{q^*,\omega}\,.$$

Express $\eta_{x^*,\omega}$ in terms of demand primitives from Eq. (A2)

$$\begin{aligned}
(\boldsymbol{\rho}(q^*) - 1)\,\frac{\boldsymbol{\varepsilon}(q^*) - 1}{2 - \boldsymbol{\rho}(q^*)} &= (\boldsymbol{\rho}(q^*) - 1)\,\eta_{q^*,\omega} \\
\frac{\boldsymbol{\varepsilon}(q^*) - 1}{2 - \boldsymbol{\rho}(q^*)} &= \eta_{q^*,\omega} \\
\eta_{x^*,\omega} + 1 &= \eta_{q^*,\omega}\,.
\end{aligned}$$

It follows that

$$\eta_{x^*,\omega} < 0 \overset{(a)}{\iff} \eta_{q^*,\omega} < 1.$$

Similarly, condition *(b)* is proved by expressing $\eta_{mr,\,q}$ in terms of its demand primitives:

$$\eta_{x^*,\omega} < 0 \overset{(b)}{\iff} \eta_{mr,\,q}(q^*) = \frac{(2p' + q^*\,p'')q^*}{p + q^*\,p'} = -\frac{2 - \boldsymbol{\rho}(q^*)}{\boldsymbol{\varepsilon}(q^*) - 1} = -\frac{1}{\eta_{q^*,\omega}} < -1\,.$$

As marginal revenue is equal to marginal cost, the elasticity of marginal revenue with respect to output is indeed the inverse of the elasticity of output with respect to productivity (and more general marginal cost). Condition *(b)* can also be proved by reshuffling Eq. (A2) so that

$$\eta_{x^*,\omega} = \frac{\boldsymbol{\varepsilon}(q^*) + \boldsymbol{\rho}(q^*) - 3}{2 - \boldsymbol{\rho}(q^*)} = -\frac{\eta_{mr,\,q}(q^*) + 1}{\eta_{mr,\,q}(q^*)} < 0 \overset{(b)}{\iff} \eta_{mr,\,q}(q^*) < -1\,.$$

■

A.1.1 Demand Manifold

In this section, I review the functional forms and describe the parameter restrictions associated with condition *(c)* in Proposition 1.
Work in progress

A.2 Monopolistic competition

As explained by Mrázová and Neary (2019), the specification in (**A5**) is consistent also with a very broad class of demands that Pollak (1972) calls "generalized additive separability", such that the inverse demand for each good depends on its own quantity and on a single aggregate. In addition to (directly and indirectly) additive preferences, this class includes quasi-linear quadratic preferences as in Melitz and Ottaviano (2008), where λ equals the total sales of all firms; and the family of choke-price demands considered by Arkolakis et al. (2019), where λ is an aggregate price index.

Additional details about the setting. Prior to entry, firms face uncertainty about their productivity and entry requires a sunk cost f_E. Once the entry cost is paid, firms observe their productivity, which is drawn from a distribution $G(\omega)$ with support $[\omega_{min}, \omega_{max}]$. Last, after observing its type, each entrant decides to produce or not based on its operating profits:

$$\pi_i(\omega_i, \lambda) = \max_{q_i} \quad \left(p_i(q_i, \lambda) - \frac{w}{\omega_i} \right) L\, q_i \ .$$

As these ultimately depend on the productivity term, this implies that there will be the minimum level of productivity $\underline{\omega}$ to remain profitably active. This is determined by two conditions. First, a break-even condition that all producers make nonnegative operating profits. Second, a zero-expected-profit condition, which drives the entry decision and requires that entry occurs until the expected value of taking a productivity draw is zero. The unique free-entry equilibrium determines the productivity cut-off $\underline{\omega}$, the mass of firms N and the marginal utility of income λ, which can be interpreted as a measure of the degree of competition each firm faces.

Details about the simulation in Fig. 7 and Fig. 8 The equilibrium values of responsiveness and cross-sectional outcomes are obtained by assuming a mass of consumers $L = 100$ with quadratic preferences $u(q) = \alpha q - \frac{\beta}{2} q^2$ where $\alpha = 5$ and $\beta = 1$. The productivity distribution is assumed to be a bounded Pareto with $k = 5$ and $\omega_i \in [1, 4]$. The fixed entry cost is $f_E = 1$ and is reduced to $f_E = 0.1$ in the second scenario in Fig. 8. The resulting equilibrium value of λ moves from $\lambda_1 = 6.7$ to $\lambda_1 = 10.2$.

A.3 Oligopoly

Proof of Proposition 3.

Proof. Start from $x_i^* = \dfrac{q_i^*}{\omega_i}$ and take its derivative with respect to productivity, so that

$$\frac{\partial x_i^*}{\partial \omega_i} = \frac{1}{\omega_i} \frac{\partial q_i^*}{\partial mc_i} \frac{\partial mc_i}{\partial \omega_i} - \frac{q_i^*}{\omega_i^2} = -\frac{1}{\omega_i} \frac{\partial q_i^*}{\partial mc_i} \frac{w}{\omega_i^2} - \frac{x_i^*}{\omega_i} \ .$$

Expressing it in terms of elasticity, it follows that

$$\eta_{x_i^*, \omega_i} \equiv \frac{\partial x_i^*}{\partial \omega_i} \frac{\omega_i}{x_i^*} = -1 - \frac{1}{\underbrace{\omega_i\, x_i^*}_{q_i^*}} \underbrace{\frac{w}{\omega_i}}_{mc_i} \frac{\partial q_i^*}{\partial mc_i} = -1 - \frac{c_i}{q_i^*} \frac{\partial q_i^*}{\partial mc_i} = -1 - \eta_{q_i^*, mc_i} \ . \tag{A3}$$

Noting that $\eta_{q_i^*, mc_i} = -\eta_{q_i^*, \omega_i}$, condition *(a)* holds

$$\eta_{x_i^*, \omega_i} < 0 \iff \eta_{q_i^*, \omega_i} < 1 \ .$$

To analyze how a firm in oligopoly adjusts its output, derive its first-order condition with respect to mc_i

$$p' + \frac{\partial Q}{\partial q_i^*} \frac{\partial q_i^*}{\partial mc_i} + p''\, \theta\, q_i^* \frac{\partial Q}{\partial q_i^*} + p'\, \theta \frac{\partial q_i^*}{\partial mc_i} = 1 \ .$$

Isolate $\frac{\partial q_i^*}{\partial mc_i}$ and express it in terms of elasticity to obtain

$$\eta_{q_i^*, mc_i} \equiv \frac{\partial q_i^*}{\partial mc_i} \frac{mc_i}{q_i^*} = \frac{1}{2\, p'\, \theta + p''\, \theta^2\, q_i^*} \frac{p + p'\, \theta\, q_i^*}{q_i^*} = \ldots = \frac{1 - \frac{\boldsymbol{\varepsilon}(Q)}{\theta\, s_i}}{2 - \boldsymbol{\rho}(Q)\, \theta\, s_i} \ .$$

By plugging this back into Eq. (A3)

$$\eta_{x_i^*,\omega_i} = -1 - \eta_{q_i^*,mc_i} = -1 - \frac{1 - \frac{\varepsilon(Q)}{\theta\, s_i}}{2 - \rho(Q)\,\theta\, s_i} = \frac{-3 + \rho(Q)\,\theta\, s_i + \frac{\varepsilon(Q)}{\theta\, s_i}}{2 - \rho(Q)\,\theta\, s_i}\,, \tag{A4}$$

condition *(c)* holds since the denominator must be always positive

$$\eta_{x^*,\omega} < 0 \iff \frac{\varepsilon(Q)}{\theta\, s_i} + \rho(Q)\,\theta\, s_i < 3.$$

Condition *(b)* follows from deriving the mr_i with respect to q_i

$$\eta_{mr_i,q_i} \equiv \frac{\partial mr_i}{\partial q_i}\frac{q_i}{mr_i} = \left(p'\theta + \theta p''\frac{\partial Q}{\partial q_i}q_i + p'\theta\right)\frac{q_i}{mr_i} = \ldots = \frac{2 - \rho(Q)\,\theta\, s_i}{1 - \frac{\varepsilon(Q)}{\theta\, s_i}} = \frac{1}{\eta_{q_i^*,mc_i}}\,.$$

Thus,

$$\eta_{x_i^*,\omega_i} < 0 \iff \eta_{mr_i,q_i} < -1.$$

■

Proof of Corollary 5.

Proof. Sum the first-conditions of all firms to obtain $p = \frac{\bar{c}}{1-\frac{\theta}{N\,\varepsilon}}$. Substitute it into Eq. (11) and, after a few simplifications, the market share of firm i can be expressed as

$$s_i = \frac{\varepsilon}{\theta} - \frac{mc_i}{\overline{mc}}\left(\frac{\varepsilon}{\theta} - \frac{1}{N}\right)$$

The ratio $r_i \equiv \dfrac{mc_i}{\overline{mc}}$ is an inverse measure of the cost advantage of firm i: the lower it is, the greater the cost advantage of firm i with respect to average marginal costs in the industry. Plug it into Eq. (A3), to obtain

$$\eta_{x_i,\omega_i} = -1 + \frac{\frac{\varepsilon}{\theta\left[\frac{\varepsilon}{\theta} - r_i\left(\frac{\varepsilon}{\theta} - \frac{1}{N}\right)\right]} - 1}{2 - \rho\,\theta\left[\frac{\varepsilon}{\theta} - r_i\left(\frac{\varepsilon}{\theta} - \frac{1}{N}\right)\right]} = -1 + \frac{\frac{1}{1 - r_i\varepsilon + r_i\frac{\theta}{\varepsilon N}} - 1}{2 - \rho\,\varepsilon\left(1 - r_i + r_i\frac{\theta}{\varepsilon N}\right)}\,. \tag{A5}$$

Based on this, the smaller is r_i, the lower is η_{x_i,ω_i}. This is because a firm with a substantial cost-advantage is likely to control a large share of the market. If so, it faces a low price elasticity of its residual demand and has a lower incentive to increase its output after a productivity shock. ■

A.3.1 Demand manifold in oligopoly.

Work in progress.

A.4 Extensions

A.4.1 Multiple inputs

Proof of Proposition 4

Proof. Start from the first-order conditions for both inputs

$$\begin{cases} \dfrac{\partial \pi}{\partial l} = mr(q) f_l - w_l = 0 \\ \\ \dfrac{\partial \pi}{\partial m} = mr(q) f_m - w_m = 0 \,. \end{cases}$$

Differentiate them with respect to productivity to obtain the following system

$$\begin{bmatrix} mr\, f_{ll} + f_l^2 \frac{\partial mr}{\partial q} & mr\, f_{lm} + f_l\, f_m \frac{\partial mr}{\partial q} \\ mr\, f_{ml} + f_m\, f_l \frac{\partial mr}{\partial q} & mr\, f_{mm} + f_m^2 \frac{\partial mr}{\partial q} \end{bmatrix} \begin{bmatrix} \frac{\partial l^*}{\partial \omega} \\ \frac{\partial m^*}{\partial \omega} \end{bmatrix} = \begin{bmatrix} -mr\, f_{l\omega} - f_l\, f_\omega \frac{\partial mr}{\partial q} \\ -mr\, f_{m\omega} - f_m\, f_\omega \frac{\partial mr}{\partial q} \end{bmatrix} . \tag{A6}$$

Solve it by using Cramer's rule and collect mr in order to express it in terms of elasticity. The derived labor demand changes with ω according to

$$\frac{\partial l^*}{\partial \omega} = \frac{mr^2 \left[(-f_{l\omega} - \frac{f_l f_\omega}{q} \eta_{mr,q})(f_{mm} + \frac{f_m^2}{q} \eta_{mr,q}) - (f_{lm} + \frac{f_l f_m}{q} \eta_{mr,q})(f_{ml} + \frac{f_m f_l}{q} \eta_{mr,q}) \right]}{mr^2 \left[(f_{ll} + \frac{f_l^2}{q} \eta_{mr,q})(f_{mm} + \frac{f_m^2}{q} \eta_{mr,q}) - (f_{lm} + \frac{f_l f_m}{q} \eta_{mr,q})(f_{ml} + \frac{f_m f_l}{q} \eta_{mr,q}) \right]} \,.$$

As the denominator is always positive under the profit-maximizing conditions, the sign of $\frac{\partial l^*}{\partial \omega}$ ultimately depends on the numerator. After a few simplifications, this becomes

$$-f_{lm} f_{mm} - \frac{f_l f_\omega}{q} f_{mm} \eta_{mr,q} - f_{l\omega} \frac{f_m^2}{q} \eta_{mr,q} + f_{m\omega} f_{lm} + \frac{f_l f_m}{q} f_{m\omega} \eta_{mr,q} + f_{lm} \frac{f_m f_\omega}{q} \eta_{mr,q} \,.$$

This can be further simplified since the productivity term is assumed to be *Hicks-neutral*. In this case, indeed, $f_{l\omega} = \frac{f_l}{\omega}$, $f_{m\omega} = \frac{f_m}{\omega}$ and $f_\omega = \frac{q}{\omega}$. Thus, the sign of $\frac{\partial l^*}{\partial \omega}$ depends on the value of $\eta_{mr,q}$:

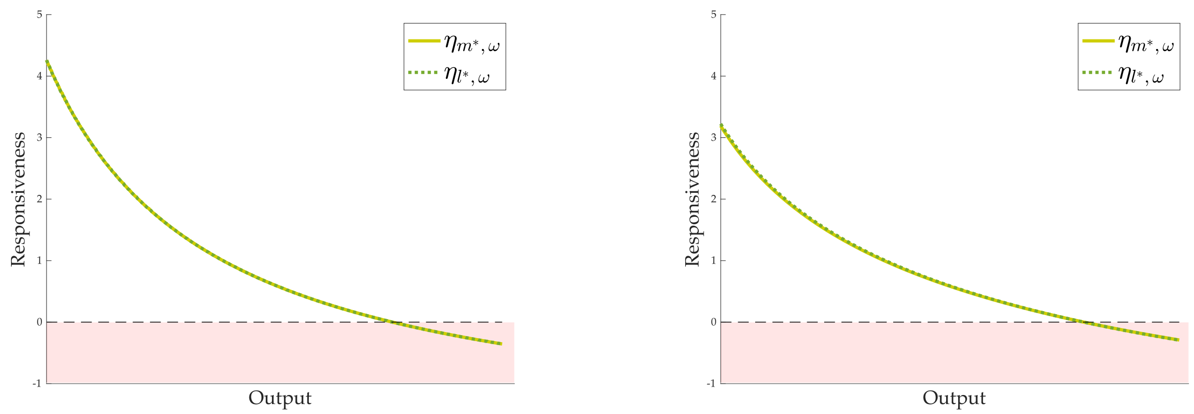
$$\frac{\partial l^*}{\partial \omega} < 0 \quad \Leftrightarrow \quad \underbrace{\left(-\frac{f_l}{\omega} f_{mm} + \frac{f_m}{\omega} f_{lm} \right)}_{>0} (1 + \eta_{mr,q}) < 0 \,. \tag{A7}$$

Following the same steps, a similar result holds for m^*. The equivalent conditions in terms of $\eta_{q^*,\omega} < 1$ and $\varepsilon(q^*) < 3 - \rho(q^*)$ follows directly from Proposition 1. ∎

Below I illustrate the values of $\eta_{l^*,\omega}$ and $\eta_{m^*,\omega}$ in the simulations reported in Fig. 12. In both cases, there is no difference between the responsiveness to productivity of the two inputs. This is an implications of considering Hicks-neutral productivity shocks.

Figure A1. Non-monotonicity with multiple input factors.

(a) Cobb-Douglas **(b)** Translog



A.4.2 Technological returns to scale

Proof of Proposition 5.
This is a reformulation of previous derivations by Bakhtiari (2009), which I extend in the last part.

Proof. Starting from the cost function $C(q, \omega, w) = c_1(q)c_2(\omega)\, w$, isolate the derived factor demand

$$x^* = \frac{C(q, \omega, w)}{w} = c_1(q)c_2(\omega)\,.$$

Take its derivative with respect to ω

$$\frac{\partial x^*}{\partial \omega} = c_1'(q)\frac{\partial q^*}{\partial \omega}c_2(\omega) + c_1(q)c_2'(\omega)\,,$$

then express $\dfrac{\partial x^*}{\partial \omega}$ in terms of elasticity

$$\eta_{x^*,\,\omega} \equiv \frac{\partial x^*}{\partial \omega}\frac{\omega}{x^*} = \frac{c_1'(q)c_2(\omega)\omega}{c_1(q)c_2(\omega)}\frac{\partial q^*}{\partial \omega} + \frac{c_1(q)c_2'(\omega)\omega}{c_1(q)c_2(\omega)} = \frac{c_1'(q)\omega}{c_1(q)}\frac{\partial q^*}{\partial \omega} + \frac{c_2'(\omega)\omega}{c_2(\omega)}\,. \tag{A8}$$

To obtain $\frac{\partial q^*}{\partial \omega}$, take the derivative of the first-order condition $R_q(q) - C_q(q, \omega, w) = 0$ with respect to ω so that

$$(R_{qq} - C_{qq})\frac{\partial q^*}{\partial \omega} = C_{q\omega} \quad \Rightarrow \quad \frac{\partial q^*}{\partial \omega} = \frac{C_{q\omega}}{R_{qq} - C_{qq}} = \frac{c_1'(q)\, c_2'(\omega)\, w}{R_{qq} - C_{qq}} = \frac{c_2'(\omega)}{c_2(\omega)}\frac{C_q}{R_{qq} - C_{qq}}\,.$$

Plug it back into Eq. (A8)

$$\eta_{x^*,\,\omega} = \frac{c_1'(q)\omega}{c_1(q)}\frac{c_2'(\omega)}{c_2(\omega)}\frac{C_q}{R_{qq} - C_{qq}} + \frac{c_2'(\omega)\omega}{c_2(\omega)}\,.$$

Rearrange, multiply and divide by q the first term so that

$$\eta_{x^*,\,\omega} = \frac{c_1'(q)q}{c_1(q)}\frac{c_2'(\omega)\omega}{c_2(\omega)}\frac{C_q}{q(R_{qq} - C_{qq})} + \frac{c_2'(\omega)\omega}{c_2(\omega)}\,.$$

By the first-order condition $C_q = R_q$, so this simplifies to

$$\eta_{x^*,\omega} = \eta_{c_1,q}\, \eta_{c_2,\omega} \frac{1}{\eta_{mr,q} - \eta_{mc,q}} + \eta_{c_2,\omega} = \eta_{c_2,\omega} \frac{\eta_{c_1,q} + \eta_{mr,q} - \eta_{mc,q}}{\eta_{mr,q} - \eta_{mc,q}} .$$

Since $\eta_{c_1,q} = \eta_{C,q}$ and $\eta_{C,q} - \eta_{mc,q} - 1 = \eta_{S,q}$, it holds that

$$\eta_{x^*,\omega} = \eta_{c_2,\omega} \frac{1 + \eta_{mr,q} + \eta_{S,q}}{\eta_{mr,q} - \eta_{mc,q}} .$$

■

Varying returns to scale. In the paper, I considered a Cobb-Douglas production function in which returns to scale take the same value for any level of output, either < 1 if RTS are decreasing or > 1 when RTS are increasing. This is the case also with CES production function and other homogeneous production functions with are usually applied in empirical studies to ease the estimation of output elasticities.

However, in theory returns to scale can take different values depending on the output level. For example, a U-shaped average cost curve requires the scale elasticity to vary, in particular to be larger than one at low output levels and to decrease below one as a firm produces more. In general, the rate at which the returns to scale vary with output is described by

$$\eta_{S,q} \equiv \frac{\partial S(q)}{\partial q} \frac{q}{S(q)} .$$

This leads to the following result.

Corollary 10. *If returns to scale declines with output, the responsiveness of a firm's derived factor demand to productivity shocks decreases. Moreover, the level of output at which $\eta_{x^*,\omega} = 0$ is lower, restricting the range of elasticities of output demand for a monotonic relationship.*

Proof. Note that the sign of $\eta_{x^*,\omega}$ in Proposition 5 depends on the sign of its numerator. This is because $\eta_{c_2,\omega} < 0$ and $(\eta_{mr,q} - \eta_{mc,q}) < 0$ under profit-maximization. As a result,

$$\eta_{x^*,\omega} \geq 0 \quad \Leftrightarrow \quad 1 + \eta_{mr,q} + \eta_{S,q} \geq 0 \quad \Leftrightarrow \quad \eta_{mr,q} \geq -1 - \eta_{S,q}$$

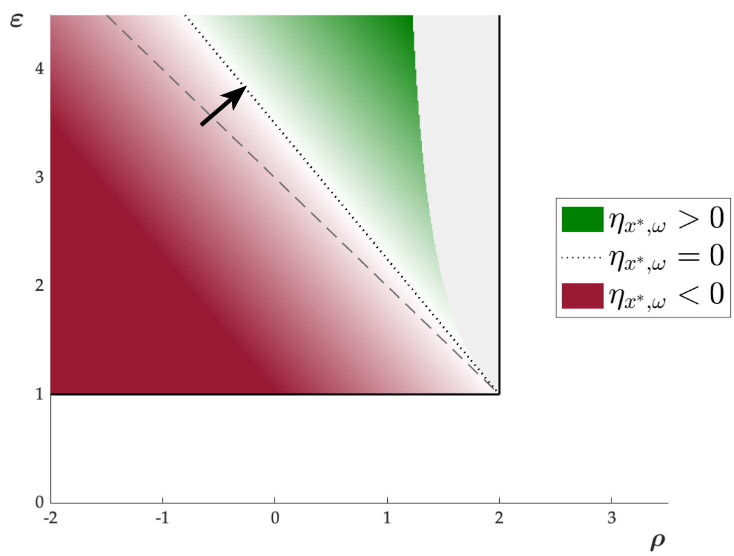
If $\eta_{S,q} < 0$, $\eta_{x^*,\omega} = 0$ must occur at a lower level of q. ■

As a result, the range of possibilities for a monotonic relationship is restricted whenever technological scale economies get exhausted at higher level of output. This is the case, for example, for production technologies that lead to a U-shaped average cost function, which is a common assumption in many settings. The prediction of Corollary 10 is clearly visible in the manifold framework since

$$\eta_{x^*,\omega} < 0 \quad \Leftrightarrow \quad -\frac{2 - \boldsymbol{\rho}(q)}{\boldsymbol{\varepsilon}(q) - 1} < -1 - \eta_{S,q}(q) \quad \Leftrightarrow \quad \boldsymbol{\varepsilon} < \frac{3 - \boldsymbol{\rho} + \eta_{S,q}}{1 + \eta_{S,q}} .$$

Figure A2 shows that the region of elasticity and convexity values at which $\eta_{x^*,\omega} < 0$ expands. This means that, whenever (technological) returns to scale declines with output, even higher values of elasticities of demand can lead to a decoupling of derived factor demand from productivity growth.

Figure A2. Monotonicity in the manifold with declining RTS ($\eta_{S,q} < 0$).



Notes: illustrative example with $\eta_{S,q} = -0.25$. The dashed grey line represents the threshold with constant RTS, i.e. $\eta_{S,q} = 0$ (see Figure 4).

A.4.3 Monopsonistic power in input market

Proof for Proposition 6

Proof. Following the same logic of Proposition 1, take the derivative of Eq. (13) with respect to ω

$$\frac{\partial x^*}{\partial \omega} = \frac{\frac{\partial (w - p\,\omega)}{\partial \omega}(p'\omega^2 - w') - (w - p\,\omega)\frac{\partial (p'\omega^2 - w')}{\partial \omega}}{(p'\omega^2 - w')^2} .$$

After expressing it in terms of elasticity and a few other manipulations, it becomes

$$(p'\omega^2 - 2w' - x^* w'')\,\eta_{x^*,\omega} = \eta_{q^*,\omega}\,(-p' - p''\,q)\omega^2 - 2p'\omega^2 - \frac{p\omega^2}{q} .$$

To know how optimal output q^* is adjusted, derive also the first-order condition with respect to ω. As a result,

$$mr + mr'\,\frac{\partial q^*}{\partial \omega}\omega = me'\,\frac{\partial x^*}{\partial \omega} .$$

Isolate $\dfrac{\partial q^*}{\partial \omega}$ and express it in terms of elasticity

$$\frac{\partial q^*}{\partial \omega}\frac{\omega}{q^*} = \frac{me'}{mr'\,q^*}\frac{\partial x^*}{\partial \omega} - \frac{mr}{mr'\,q^*} .$$

Then, multiply the second term by $\dfrac{\omega\,mr}{me}\dfrac{x^*\omega}{x^*\omega}$ and rearrange so that

$$\eta_{q^*,\omega} = \frac{\eta_{me,x}\,\eta_{x^*,\omega}}{\eta_{mr,q}} - \frac{1}{\eta_{mr,q}} . \tag{A9}$$

Then plug it back into the equation

$$(p'\omega^2 - 2w' - x^*w'')\,\eta_{x^*,\omega} = \frac{\eta_{me,x}\,\eta_{x^*,\omega}}{\eta_{mr,q}} - \frac{1}{\eta_{mr,q}}\,(-p' - p''\,q)\,\omega^2 - 2p'\omega^2 - \frac{p\,\omega^2}{q}$$

and multiply both sides by $\eta_{mr,\,q}$

$$\eta_{mr,q}\,(p'\omega^2 - me')\,\eta_{x^*,\omega} = \eta_{me,x}\,\eta_{x^*,\omega} - (-p'\,\omega^2 - p''\,q\,\omega^2) - \eta_{mr,q}\left(2p'\omega^2 + \frac{p\,\omega^2}{q}\right).$$

Finally, isolate $\eta_{x^*,\omega}$ and collect $(-p'\,\omega^2 - p''\,q\,\omega^2)$

$$\eta_{x^*,\omega} = \frac{(p'\,\omega^2 + p''\,q\,\omega^2) - \left(2p'\omega^2 + \frac{p\,\omega^2}{q}\right)\eta_{mr,q}}{\eta_{mr,q}(p'\omega^2 - me') - \eta_{me,x}(-p'\,\omega^2 - p''\,q\,\omega^2)}$$

$$\eta_{x^*,\omega} = \frac{-1 - \eta_{mr,q}\frac{\left(2p'\omega^2 + \frac{p\,\omega^2}{q}\right)}{(-p'\,\omega^2 - p''\,q\,\omega^2)}}{\eta_{mr,q}\frac{(p'\omega^2 - me')}{(p'\,\omega^2 + p''\,q\,\omega^2)} - \eta_{me,x}}\,.$$

Since the term highlighted in orange becomes

$$\frac{\left(2p'\omega^2 + \frac{p\,\omega^2}{q}\right)}{(-p'\,\omega^2 - p''\,q\,\omega^2)} = \frac{\frac{p\,\omega^2}{q}\left(\frac{2p'\omega^2 q}{p\,\omega^2} + 1\right)}{\frac{p\,\omega^2}{q}\left(\frac{-p'\,q}{p} - \frac{p''\,q^2}{p}\right)} = \frac{-\frac{2}{\varepsilon} + 1}{\frac{1}{\varepsilon} - \frac{\rho}{\varepsilon}} = -\frac{\varepsilon - 2}{\rho - 1}\,,$$

the numerator simplifies to

$$-1 + \frac{\varepsilon - 2}{\rho - 1}\,\eta_{mr,\,q}\,.$$

The term highlighted in blue, instead, simplifies to

$$\frac{mr'\,q}{mr}\,\frac{mr}{q}\,\frac{\left(\frac{p'\,\omega^2\,q}{mr} - \frac{me'\,x\,\omega^2}{me}\right)}{p'\,\omega^2\left(-1 - \frac{p''\,q\,\omega^2}{p'\,\omega^2}\right)} = mr'\,\frac{\left(\frac{p'\,q}{mr} - \frac{me'\,x}{me}\right)}{p'\left(-1 - \frac{p''\,q}{p'}\right)} = \frac{mr'}{p'}\,\frac{\left(\frac{p'\,q}{mr} - \eta_{me,\,x}\right)}{(-1 + \rho)}\,.$$

Therefore, the denominator becomes

$$= \frac{mr'}{p'}\,\frac{\left(\frac{p'\,q}{mr} - \eta_{me,\,x}\right) - (\rho - 1)\eta_{me,\,x}}{(\rho - 1)} = \frac{\eta_{mr,\,q} - \frac{2p' + p''q}{p'}\,\eta_{me,\,x} - \eta_{me,\,x} - (\rho - 1)\eta_{me,\,x}}{(\rho - 1)}$$
$$= \frac{\eta_{mr,\,q} - (2 - \rho)\,\eta_{me,\,x} - \eta_{me,\,x} - (\rho - 1)\eta_{me,\,x}}{(\rho - 1)} = \frac{\eta_{mr,\,q} - \eta_{me,\,x}}{\rho - 1}\,.$$

Bringing them back together, we obtain

$$\eta_{x^*,\omega} = \frac{-(\rho - 1) + (\varepsilon - 2)\,\eta_{mr,\,q}}{\eta_{mr,\,q} - \eta_{me,\,x}}\,.$$

Since $\eta_{mr,\,q} = -\dfrac{2 - \rho}{\varepsilon - 1}$, the numerator can be further simplified to

$$-(\rho - 1) - (\varepsilon - 2)\frac{2 - \rho}{\varepsilon - 1} = \frac{-\rho(\varepsilon - 1) + \varepsilon - 1 - (\varepsilon - 2)(2 - \rho)}{\varepsilon - 1} = -\frac{\rho + \varepsilon - 3}{\varepsilon - 1} = -(1 + \eta_{mr,q})\,.$$

As a result, it holds that

$$\eta_{x^*,\omega} = -\frac{1 + \eta_{mr,q}}{\eta_{mr,\,q} - \eta_{me,\,x}}\,.$$

■

Proof of Corollary 9

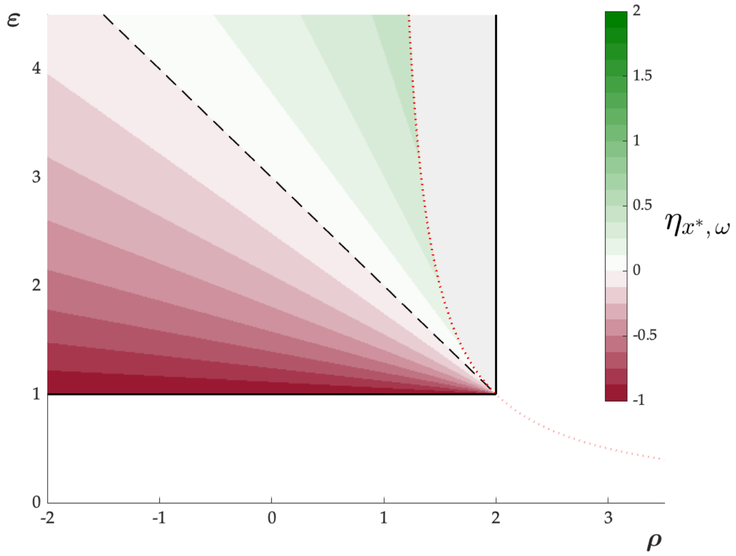
Proof. This follows from the fact that the denominator in Proposition 6 has to be negative for the second-order condition to hold. Alternatively, by simply rewriting Equation (A9) in terms of $\eta_{x^*,\,\omega}$, it holds that

$$\eta_{x^*,\,\omega} = \frac{\eta_{q^*,\,\omega}\,\eta_{mr,\,q}}{\eta_{me,\,x}} + \frac{1}{\eta_{me,\,x}} = 0 \quad \Leftrightarrow \quad \eta_{q^*,\,\omega} = -\frac{1}{\eta_{mr,\,q}}\,.$$

■

To illustrate Corollary 9 in the manifold space, I consider the simplest case of an isoleastic inverse supply curve $w(x) = g\,x^m$, which leads to a constant $\eta_{me,x} = m \geq 0$. In comparison to Figure 5, the lower responsiveness of derived demand to productivity shocks in the presence of monopsony power becomes apparent.

Figure A3. Lower values of $\eta_{x^*,\,\omega}$ with monopsony power.



Notes: this is illustrated for a constant $\eta_{me,x} = m = 1.5$.

B Different impacts of shocks: ω_i *vs.* cost *vs.* demand

B.1 Comparative statics of demand and cost shocks.

To illustrate the difference impact on derived factor demand between a productivity shock (see Figure 2) demand and cost shocks, in Figure A4 I plot the first-order conditions before and after a multiplicative demand shock ξ_i. This represents a standard demand shifter due, for example, to an increased appeal for the firm's product. In Figure A4 I show the comparative statics of market size shock ψ_i. Differently from a productivity shock, these demand shocks always lead to an increase in derived factor demand, also for linear demand. It is worth emphasizing that comparative statics of both marginal costs and demand shocks for factor demand remains the same. Always, outward shift.

Figure A4. Reaction of $mrp(x)$ to a demand shock $\uparrow \xi$.

(a) CES **(b)** Linear

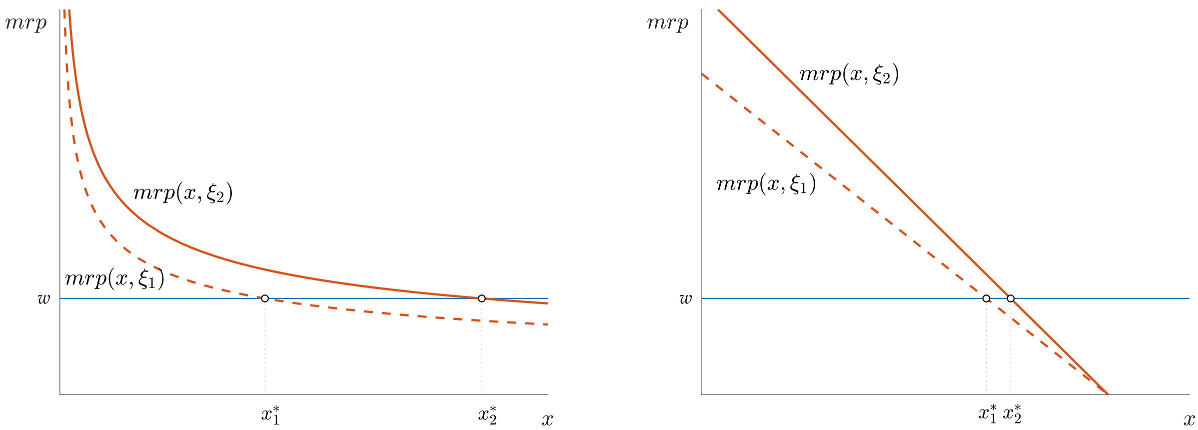
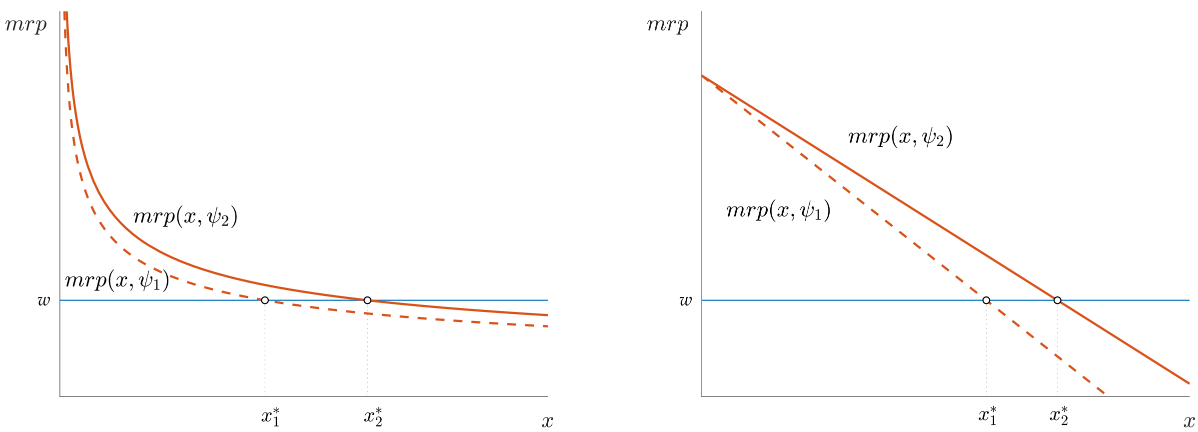


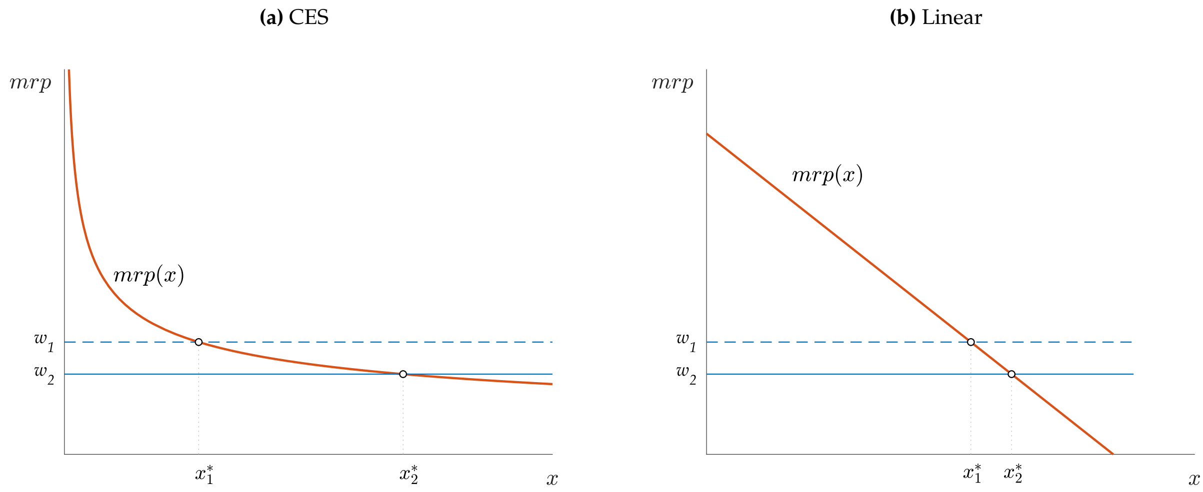
Figure A5. Reaction of $mrp(x)$ to a demand shock $\uparrow \psi$.

(a) CES **(b)** Linear



As can be seen below, this is the case also for exogenous shocks that reduce factor price *w*. The optimal reaction of a firm is indeed to increase its derived factor demand $x_2^* > x_1^*$.

Figure A6. Reaction of $mrp(x)$ to a cost shock $\downarrow w$.



These illustrative examples show that the comparative static predictions of derived factor demand after exogenous productivity/demand/costs shocks are not the same, as generally presumed so far. Productivity shocks are fundamentally different because they can lead to a reduction in the in the amount of input optimally used by a firm (i.e. $\Delta x_i^* < 0$).

B.2 Determinants and values of the $Ratio = \frac{\%\Delta x_i}{\%\Delta r_i}$.

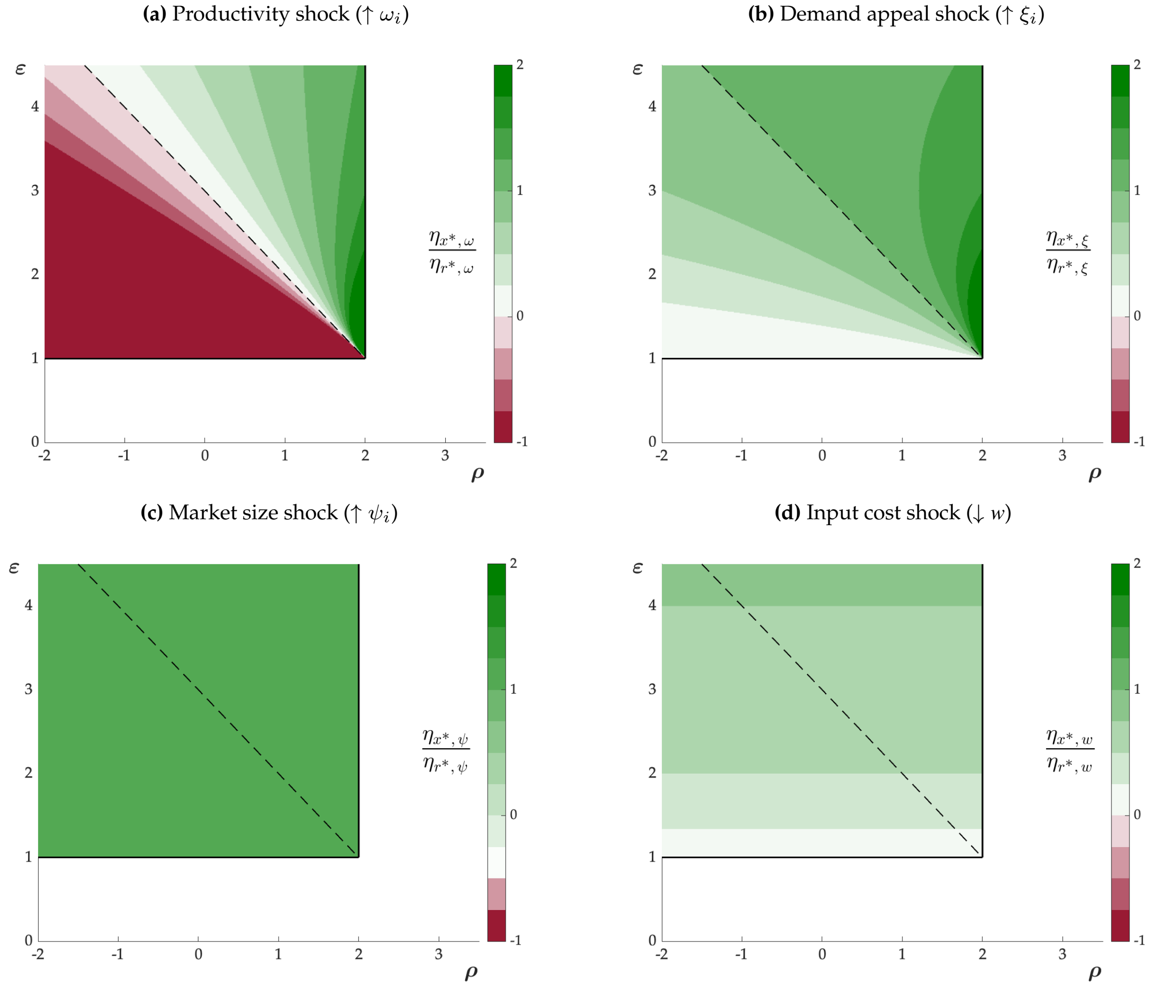
From the first-order conditions, I derive the elasticity of derived factor demand and revenue with respect to productivity, demand and cost shocks in the case of a monopolist. As reported in the table, all these elasticity are a function of the values of elasticity and curvature of demand.

Table A1. Structural determinants of the $Ratio$.

	$\eta_{x_i^*,\dots}$	$\eta_{r_i^*,\dots}$	$Ratio = \frac{\eta_{x_i^*,\dots}}{\eta_{r_i^*,\dots}}$
Productivity (ω_i)	$\frac{\boldsymbol{\varepsilon}+\boldsymbol{\rho}-3}{2-\boldsymbol{\rho}}$	$\frac{2-\boldsymbol{\rho}}{\boldsymbol{\varepsilon}}$	$\frac{(\boldsymbol{\varepsilon}+\boldsymbol{\rho}-3)\boldsymbol{\varepsilon}}{(\boldsymbol{\varepsilon}-1)^2}$
Demand appeal (ξ_i)	$\frac{\boldsymbol{\varepsilon}-1}{2-\boldsymbol{\rho}}$	$\frac{(\boldsymbol{\varepsilon}-1)^2}{\boldsymbol{\varepsilon}(2-\boldsymbol{\rho})}+1$	$\frac{\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}-1)}{(\boldsymbol{\varepsilon}-1)^2+2-\boldsymbol{\rho}}$
Market size (ψ_i)	1	1	1
Input costs (*w*)	$-\frac{\boldsymbol{\varepsilon}-1}{2-\boldsymbol{\rho}}$	$-\frac{(\boldsymbol{\varepsilon}-1)^2}{\boldsymbol{\varepsilon}(2-\boldsymbol{\rho})}$	$\frac{\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}-1}$

I illustrate below the values of the $Ratio$ for different productivity, demand and costs shocks in the manifold space.

Figure A7. Values of Ratio $= \frac{\%\Delta x_i}{\%\Delta r_i} = \frac{\eta_{x_i^*, \cdots}}{\eta_{r_i^*, \cdots}}$ in the manifold space.

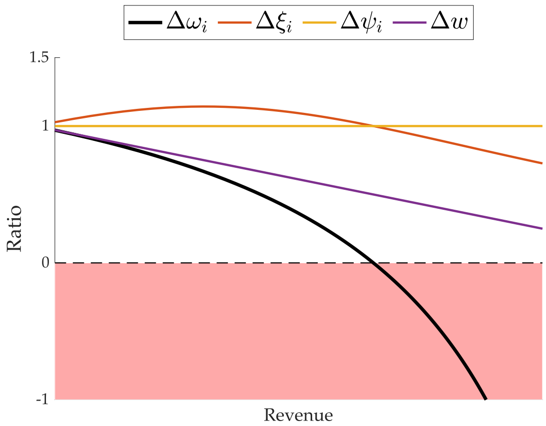


B.2.1 $Ratio$ for sub-convex and super-convex demands.

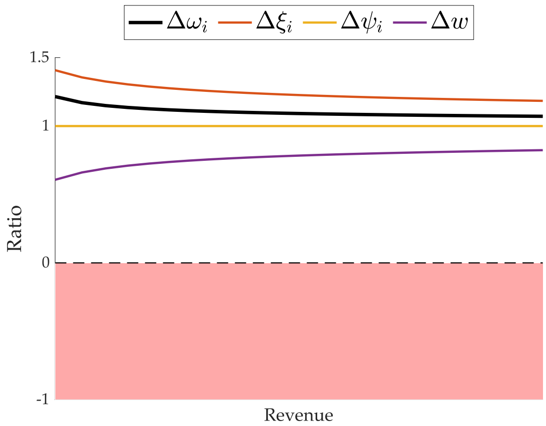
This happen only if the revenue and markups are positively correlated. I plot the values of $Ratio = \frac{\%\Delta x_i}{\%\Delta r_i} = \frac{\eta_{x_i^*,\dots}}{\eta_{r_i^*,\dots}}$ for different shocks along the revenue distribution in the case of a sub-convex demand and of a super-convex one.

Figure A8. Values of $Ratio$ for different shocks along the revenue distribution.

(a) Sub-convex demands.



(b) Super-convex demands.



Notes: Constant Proportional Pass-Through (CPPT) demands with (a) 50% and (b) 150% Pass-through rates.