FIRM ORGANIZATION AND INFORMATION TECHNOLOGY: MICRO AND MACRO IMPLICATIONS*

Asier Mariscal **

February 15, 2023

Abstract

I develop a theory of how information and communication technology (IT) capital affects firms' organization of production, consistent with multiple micro and macro facts. Firms, organized as hierarchies, choose at every layer capital, labor, and worker knowledge. To produce, production workers encounter and solve problems, whereas only managers use IT and specialize in problem solving. There are four key results. First, as in the data, an IT price reduction reallocates within-firm knowledge and wages away from production and towards managerial layers. Second, despite the model featuring capital-labor complementarity for both managers and production workers, there is an *endogenous* IT capital-to-production-labor substitution: a lower IT price makes problem solving at managerial layers cheaper, raising factor demands at these layers. Third, firms' IT capital and total knowledge are endogenously complementary and increase firms' measured TFP, as in the empirical literature. Fourth, counterfactual exercises show that a decrease in the demand elasticity generates the observed value-added concentration in large firms and the decline of the aggregate labor share of GDP. A decline in the IT price explains the decline of the real wage and wage bill share of production workers, and the increase of both for managers.

JEL codes: D21, D24, J3, L2, L16, E25.

---

*I am very grateful to Thomas Chaney and Sam Kortum for their detailed comments. I also thank Pol Antràs, Antonio Cabrales, Lorenzo Caliendo, Antonia Díaz, William Fuchs, Luis Garicano, Matthias Kredler, David Lagakos, Ben Moll, Ferdinando Monte, Ezra Oberfield, Gianmarco Ottaviano, Fidel Perez, Climent Quintana, Natalia Ramondo, Veronica Rappoport, Devesh Raval, B. Ravikumar, Diego Restuccia, Esteban Rossi-Hansberg, Raul Santaeulalia-Llopis, Anna Sanz de Galdeano, Dominik Thaler, Eric Verhoogen, Felix Wellschmied for very useful conversations, and participants at the SED 2018, RIDGE 2018, CEMFI (REDg2018), Global Macro at CERGE-EI for their comments. I gratefully acknowledge financial help from UC3M and U. Alicante, under project ECO2015-70540-P (MINECO/FEDER). The usual disclaimer applies.

**U. Rovira i Virgili, Department of Economics. Campus Bellissens, Av. Universitat 1, 43204, Reus, Tarragona; email: asier.mariscal@urv.cat, phone: +34 977 758903.
Word count (excluding Appendix): 18,606.

I. INTRODUCTION

During the past 40 years, there has been a rapid adoption of information and communication technology (IT) capital, strongly impacting the labor market. In the United States, Wilson (2006) reports computers alone represent 30% of firms' investment in 1998, and the IT price today is a third of the 1980 price (Eden and Gaggl 2018). At the same time, wage inequality in the US is rising with the surprising feature that wages for the less skilled workers are falling, despite aggregate productivity growth (see Autor 2014). Changes in the distribution of income have also manifested in a lower share of national income for low-skill workers and a larger share for high-skill workers, with an overall decline in the labor share of GDP (IMF 2017). Many of these labor market trends are not unique to the United States but are shared by European countries (see Autor 2010; Karabarbounis and Neiman 2013). Similarly widespread is the slowing down of measured aggregate TFP growth during this period (Syverson 2017).

This paper rationalizes these and several other micro and macro facts within a unified framework that opens the black box of firms' organizational choices. In the past 40 years, the reorganization of the typical white-collar workplace has been notably visible as firms have introduced computers, email, and other internal communication technologies. Simultaneously, production workers have seen their autonomy, understood as independent problem solving, narrowed.

An Amazon warehouse illustrates this new organization of work. Production workers serving an order must deal with four relevant questions, which I next describe together with the organization of their solutions. First, what is the right packaging? This problem is not solved by the worker but by a computerized information system. Second, in which shelves are the ordered products located? Actually, workers do not need to find the ordered products in the warehouse, because shelves glide to them through the floor while workers, standing in their cubicles, simply place products inside the boxes. Third, how many products should a worker pick at once so that none falls and breaks? Suppose the order requires nineteen units of an item. The worker's so-called "Jennifer headset" might request him to sequentially pick five, then five, then five, and, finally, four items. Finally, where is the order delivered? In fact, the address is automatically stamped on the exit belt, as the package leaves the premises. In this example, the autonomy of production workers is reduced to an extreme degree through a rational design of IT systems that complement managers. As a consequence, Amazon reduces its costs by paying lower wages than otherwise to the numerous production workers. More generally, Appendix Section A documents how truck drivers, low-level lawyers, and customer services have been reorganized by reducing production workers' autonomy and wages, while managers fine-tune firms' information systems, suggesting these trends occur across industries.

The knowledge-based hierarchy literature is a natural theory for thinking about optimal problem solving within organizations (Garicano 2000; Garicano and Rossi-Hansberg 2006; Caliendo and Rossi-Hansberg 2012). In this paper, firms' organizational decision involves choosing capital, labor, and worker knowledge at each layer of the hierarchy, as well as the number of layers, as suggested in Lucas (1978). As in the Amazon example, production workers use their time and non-IT capital to generate *potential* production, or problems. Production workers have knowledge that allows them to solve some of the problems, which *realizes* output. Problems unsolved at the production layer are sent to managers in the first layer, who use IT capital to solve them. These managers also solve the problems they know and have the recourse of the next-level management, in a process that continues up to the CEO. Hence, realized production is the result of potential production that materializes due to worker knowledge across all layers of the organization. Since managers' time is relatively expensive, using it optimally implies they deal with infrequent, nonroutine problems. In this theory, firms' IT capital is a new technology through which managers and the CEO leverage their available time to solve problems. On the one hand, adding a layer involves extra employees and capital, an expense that acts like a fixed cost to the firm since, without knowledge, these factors do not produce output. However, it allows a more efficient knowledge allocation across layers, which reduces the marginal cost. Hence, the theory endogenizes the menu of fixed and marginal costs and a firm with a sufficiently large demand, chooses more managerial layers.

I contribute to this literature by adding endogenous physical capital adoption, which allows me to quantitatively evaluate the model using standard firm production functions. Hence, I can address the main questions that information technology raises: What are the implications of IT for the wages of unskilled workers? Why are IT and unskilled workers substitutes? How does IT affect firm productivity? And what are the macroeconomic implications of this new technology that complements managers? To answer these questions, the theory characterizes firms' optimal reorganization as IT prices decline, and speaks to macro questions using micro discipline.

The first key result of the paper is that cheaper IT reallocates problem solving within the firm, reducing knowledge and wages of production workers and increasing those of managers. Lower IT prices alters the relative price of problem solving across layers: promotes the use of managers and IT, which become cheaper, instead of production workers, who become relatively expensive. This theory is consistent with the observed decline in production worker wages, despite intense and widespread adoption of a new general-purpose technology, IT.

The second key result is that IT capital and production labor are *indirect* substitutes, despite an external calibration with capital-labor complementarity, more for managerial than production layers as in the capital-skill complementarity hypothesis by Griliches (1969). This substitution is indirect for three reasons. First, the model does not impose any explicit constraint to the substitution or complementarity possibilities between IT capital and production workers. Hence, IT capital and production labor substitution is a consequence of *optimal* firm reorganization when IT prices fall. The mechanism is connected to knowledge reallocation, the first key result above: lower IT prices, reduce the relative price of problem solving at managerial layers, and the firm optimally solves fewer problems at the production layer. To deal with more problems at the managerial layers, the firm increases the demand for IT and managers relative to production workers. Second, it is indirect because it is in contrast to the classic approach in macroeconomics, and empirical work, where a constant parameter is assumed to govern factor's elasticities. Third, quantitatively, it is the mean of the distribution of problems that matters for this result, and not the within-layer capital-labor elasticity parameters. The IT capital-to-plant-labor elasticity is a quantitative result (as opposed to a proposition) and a long-run elasticity, because I focus on 1980-2015 changes, consistent with an analysis across two steady states.

The third key result is that firms' total worker knowledge and IT are endogenously complementary because lower IT prices make it cheaper to deal with problems, so the marginal cost of production falls, and the firm expands. This increase in production tightens the CEO time constraint, which is optimally relieved through more knowledge in all layers, including the CEO. Hence, firms' total knowledge and IT increase simultaneously, a complementarity that has ample empirical support, as I review in the next section. Moreover, this theoretical result has implications for TFP measurement: ignoring the endogenous interaction between worker knowledge, firm organization, and IT is likely to result in biased firm and aggregate TFP estimates.

The fourth key result is the model's ability to match a wide range of micro and macro facts. The model is calibrated using microdata moments on firm organization, and uses two sources of variation to explain the aggregates: (1) an IT price decline from the Bureau of Economic Analysis (BEA) data and (2) a demand elasticity reduction such that markups increase more for larger firms, consistent with de Loecker, Eeckhout, and Unger (2020, hereafter DEU). Regarding micro facts, the quantitative model captures the positive association in the firm cross-section between value-added and (1) wages, (2) the capital-labor ratio, and (3) the inverse of the labor share. Within a firm, over time, (4) wages decline at low layers and increase at top layers, (5) the employment share of managers increases, (6) the wage bill share of managers increases and that of production workers declines, (7) the IT capital-labor ratio increases, and (8) the labor share of value-added of large firms declines (Autor et al. 2020, hereafter ADKPV). Regarding macro facts, over time, (9) the share of labor compensation going to managers increases and (10) that of production workers decreases, (11) the labor share of GDP declines (Karabarbounis and Neiman 2013), (12) value-added concentration rises (ADKPV), and (13) the aggregate markup increases (DEU).

The empirical success of the model makes the framework valuable for counterfactual exercises. The model rationalizes observed factor demands (i) in the firm cross-section and (ii) as IT prices decline, providing the necessary production-side discipline to demand-based explanations for macro trends. In fact, both the measurement of the firm markup and the mechanism for sales reallocation across firms depend crucially on the assumed production function. My experiments show that IT is the key driver of the above inequality facts, including the diverging macro trends for managers and production workers. On the contrary, lower demand elasticity produces value-added concentration in large firms and the decline of the aggregate labor share where, for both, organizational choices have a quantitatively large dampening effect.

Related Literature. This paper relates to knowledge-based theories of the firm (see Garicano 2000). The model is closer to that of Caliendo and Rossi-Hansberg (2012), who study how firms that are heterogeneous in demand organize and how their productivity responds to international trade. Garicano and Rossi-Hansberg (2006) study inequality and argue that there are two types of IT when it comes to their wage inequality consequences:[1] IT that decreases knowledge access costs, like computers, and

[1] See also Antràs, Garicano, and Rossi-Hansberg (2006) and Antràs, Garicano, and Rossi-Hansberg (2008), who study the implications of offshoring on wage inequality within cross-country organizations.

IT that decreases communication costs. I use a more coarse concept of IT, and I dispense with their elegant assignment model between managers and workers, which allows me to contribute in the following ways. First, to this theoretical literature I add optimal capital choices to allow a natural quantification and mapping to micro and macro facts. Second, because capital is a choice, I can speak to the literature that studies the shape of the microeconomic production function, with an emphasis on how new technologies and different types of labor may complement or substitute for each other (see Goldin and Katz 1998). Third, I fully characterize firms' optimal organization in the cross-section and their reorganization as IT prices decline, and discipline determinants of macro trends.

The empirical evidence on firm organization provides credibility to the mechanisms in my model. Caliendo, Monte, and Rossi-Hansberg (2015) provide evidence based on French linked employer-employee data that is consistent with the predictions in Caliendo and Rossi-Hansberg (2012); for example, given the number of layers, wages rise with firm value-added, a result also obtained in this paper. Bloom et al. (2014) show that IT reduces production workers' autonomy as measured by the capacity of independent decision making, as obtained in this theory. Further, Caliendo et al. (2015) find that, when a firm expands its production by adding layers, its price and marginal cost falls, whereas the opposite occurs to both when the number of layers is fixed. Quantitatively, I obtain marginal and average costs functions with the same properties, extending the previous results to a theory with optimal capital choices.

The theory is also consistent with the empirical literature that studies the impact of IT. In the model, computers are effectively substituting routine tasks/workers, as in Autor, Levy, and Murnane (2003). Autor, Katz, and Krueger (1998) show in United States data that computer adoption and the skilled wage bill share are positively related, as in the model. The theory also rationalizes the evidence of complementarity between IT adoption, skills, and firm reorganization provided by Bresnahan, Brynjolfsson, and Hitt (2002), Caroli and Van Reenen (2001), and Bloom, Sadun, and Van Reenen (2012). Moreover, the theory is consistent with the declining routine wages due to adoption of broadband Internet in Norway (Akerman, Gaarder, and Mogstad

While focusing on different implications, Gumpert (2018) and Gumpert, Steimer, and Antoni (2018) study the role of communication costs on multiplant organization. Santamaria (2018) studies how spatial sorting affects inequality through two channels: spatial differences in technology and endogenous organization of production. Sforza (2017) studies the impact of trade and credit shocks on firm organization. I also relate to Eeckhout and Kircher (2018) in that our models allow input quality and quantity choices in the firm cross-section, though our questions are fundamentally different.

2015) and due to more IT investments induced by tax deductions in the UK (Gaggl and Wright 2017, hereafter GW); also, for the United States in the past 40 years, with the decline of all percentiles below the median of the real wage distribution of *only* the largest firms (see Song et al. 2019). For the United States, ADKPV and Kehrigy and Vincent (2018) provide rich evidence related to the decline in the labor share and, specifically, show the importance of reallocation of sales to "superstar firms", with a low labor share, a result that this theory matches. Lashkari, Bauer, and Boussard (2018) show a positive relation between firm size and its IT capital-labor ratio in French data, and estimate a production function featuring a nonhomothetic demand for IT.[2] I complement these papers by developing a production organization theory consistent with multiple organizational facts, including inequality, and quantitatively analyze the micro and macro effects of declining IT prices and rising markups.[3]

This paper also relates to a macro literature. Oberfield and Raval (2014), with a micro estimate of capital-labor elasticity lower than one, show that firms' technical change is key to understanding the decline in the labor share. Unlike them, this paper microfounds firms' production function to explicitly capture organizational choices with IT, and their endogenous response to cheaper IT, providing valuable macroeconomic insights as Antràs and Rossi-Hansberg (2009) view to be desirable. There is also related literature that uses aggregate production functions. Acemoglu and Restrepo (2018) and Hemous and Olsen (2016) study the impact of automation from a directed technical change perspective. Eden and Gaggl (2018) study the macro effects of lower IT prices using a CES production function with routine and nonroutine labor, IT and non-IT capital. Unlike them, this theory focuses on firms' decisions and features capital-labor complementarity for both unskilled and skilled workers, though more for the latter, which Goldin and Katz (1998) consider valid for technologies in the post-WWII period.

This paper is structured as follows. In Section II, I introduce the model, solve it, and define the equilibrium. In Section III, describes the micro implications of the model, and in Section IV, I calibrate the model using micro moments on firm organization for the United States and evaluate the quantitative performance of the model for micro and macro facts. In Section V, I study two counterfactual scenarios: (1) lower IT prices

[2] I review these papers in more detail when mapping the model to the data.

[3] See also Hsieh and Rossi-Hansberg (2019) and Aghion et al. (2019) for theories connecting sales concentration and how IT has affected market entry costs of firms. See Koh, Santaeulàlia-Llopis, and Zheng (2018) for how the decline in the labor share can be explained by the capitalization of intellectual property products.

and (2) lower elasticity of demand. Section VI concludes.

II. MODEL

II.A. Model Setup

Preferences and Occupational Choice. Agents maximize total consumption, Q,

$$\int_{\omega\in\Omega} \alpha(\omega)^{1/\rho(\omega)} \left(\frac{q^c(\omega)}{Q}\right)^{\frac{\rho(\omega)-1}{\rho(\omega)}} d\omega = 1 \tag{1}$$

where for each variety $\omega \in \Omega$, $q^c(\omega)$ denotes the quantity consumed, $\alpha(\omega)$ is the product-appeal shifter, and $\rho(\omega) > 1$ is the elasticity of substitution. Equation 1 implicitly defines the constant relative elasticity of income and substitution (CREIS) utility function due to Lashkari and Mestieri (2017), which generalizes the constant elasticity of substitution (CES) utility. These preferences allow $\rho(\omega)$, the elasticity of demand for good ω, to be nonconstant across ω. Still, its implied inverse demand for ω has the familiar form

$$p(\omega) = \left(\frac{q^c(\omega)}{\alpha(\omega)R}\right)^{-1/\rho(\omega)} D(\omega) \tag{2}$$

where $D(\omega) \equiv P^{\frac{\rho(\omega)-1}{\rho(\omega)}} \frac{M}{m(\omega)}$, $m(\omega) \equiv \frac{\rho(\omega)}{\rho(\omega)-1}$, R is aggregate spending, and the ideal price and markup indexes, P and M, respectively, are defined implicitly by

$$\begin{cases} \int_{\omega\in\Omega} \alpha(\omega) \left(\frac{m(\omega)}{M}\right)^{-\rho(\omega)} \left(\frac{p(\omega)}{P}\right)^{-(\rho(\omega)-1)} d\omega = 1 \\ \int_{\omega\in\Omega} \alpha(\omega) \left(\frac{m(\omega)}{M}\frac{p(\omega)}{P}\right)^{-(\rho(\omega)-1)} d\omega = 1 \end{cases} \tag{3}$$

Note the inverse demand in Equation 2 is very similar to that obtained under CES. In fact, when $\rho(\omega) = \rho$, $\forall\omega$, CREIS collapses to CES. I use CREIS instead of CES, so that the quantitative analysis is, in a simple way, consistent with the evidence on markup heterogeneity in DEU, thereby focusing instead on a theory about production organization. An important feature of the demand function is that, a variety ω with a higher $\alpha(\omega)$ delivers a larger utility, and hence, ceteris paribus, its demanded quantity is larger. Each product ω corresponds to one firm only and has a different $\alpha(\omega)$, which is the source of exogenous firm/product heterogeneity in this model.

The economy is populated by N agents who inelastically supply one unit of labor.

They are ex-ante identical, but acquire different knowledge ex-post, and with their net wage they consume. To consume, agents use their available unit of time in the labor market where they obtain a wage, w, and also receive their equal share in aggregate profits and factor income. Agents can work either in the production sector or in the education sector. Agents that acquire knowledge z to produce, pay a training cost of wcz, because learning requires teachers in the education sector who are paid w, and c is the amount of teaching time required per unit of knowledge. Firms exactly compensate for the cost of acquiring knowledge, leaving all agents indifferent across sectors, firms, and occupations with a net wage of w in equilibrium. In the next section, I introduce the benefits of knowledge, which arise from firms' production choices.

Firms. A unit mass of firms obtain a draw for α, which determines the level of demand for their product variety ω, from a cumulative distribution function (CDF) denoted $G(\alpha)$, assumed to be log-normal distribution with parameters (μ_α, ξ_α). Organizing production implies choosing optimally the number of managerial layers, $2 \leq L \leq 4$,[4] and the amount of each of three factors at each layer l, namely, capital, k_l, labor, n_l, and knowledge, z_l.[5] This adds capital choices to the model by Caliendo and Rossi-Hansberg (2012), to which I relate at the end of this section, after describing firm technology in this model.

Within an organization, agents are broadly grouped into production workers, at layer 1, and managers, at layers $l > 1$. Workers are dedicated to production and problem solving and send their unsolved problems to the managers. As in the Amazon example, to capture that managers use and control firms' information systems, I assume that, at layer 1 production workers, n_1, use production, non-IT, capital, k_1, while managers, n_l $\forall l > 1$, use IT capital, k_l $\forall l > 1$.

At the production layer, capital and labor are combined in an input bundle $y_1 \equiv \left(k_1^{\frac{\sigma_1-1}{\sigma_1}} + n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}}$. Per unit of input bundle, one problem arises that, if successfully dealt with, yields A produced quantity. Accordingly, total *potential* output is Ay_1. Each unit of the production input bundle y_1 is associated with a problem drawn from a CDF, $F(x) = 1 - \exp(-\lambda x)$. For a problem z drawn from $F(.)$ to become *realized* output,

[4] I set $2 \leq L \leq 4$ following the empirical mapping of occupations to layers in business hierarchies in France by Caliendo, Monte, and Rossi-Hansberg (2015) available in Table XII in Appendix Section C.2.2. I exclude self-employment as in Caliendo and Rossi-Hansberg (2012).

[5] For notational simplicity, I omit dependence of choices on L unless necessary for clarity.

the production worker must have a knowledge set that includes such problem. Workers can deal with all of the y_1 problems, but can solve only those they know. Suppose workers learn the interval $[0, z_1]$, then $AF(z_1)y_1$ is produced with their knowledge, and all problems above z_1 are sent to the managers at the layer immediately above. Assume for exposition that such managers exist.

In layer 2, managers, using IT capital, deal with unsolved problems from $l = 1$, and the required managerial input bundle $y_l \equiv \left(k_l^{\frac{\sigma_l - 1}{\sigma_l}} + n_l^{\frac{\sigma_l - 1}{\sigma_l}}\right)^{\frac{\sigma_l}{\sigma_l - 1}}$ at $l = 2$ needs to satisfy the restriction:

$$y_2 = y_1(1 - F(z_1)) \tag{4}$$

The left-hand side (LHS) of Equation 4 is the managerial input at $l = 2$, composed of managers and IT, whereas the right-hand side (RHS) is the amount of problems unsolved at layer 1, which have been passed on to layer 2. To deal with such a volume of problems in layer 2, the firm optimally chooses the number of managers, their knowledge, and IT capital. Assume managers learn the interval $[z_1, z_1 + z_2]$, the amount of problems passed on to layer 3 is $y_1(1 - F(z_1 + z_2))$, that is, the unsolved problems at all previous layers, those above $z_1 + z_2$.

More generally, at layer l, $1 < l < L$, managers learn $[Z_{l-1}, Z_l]$ with $Z_l \equiv \sum_{j=1}^{l} z_j$, and satisfy $y_l = y_1(1 - F(Z_{l-1}))$. The process continues with all managerial layers $l > 1$ using IT capital k_l, until it reaches the top layer, L, that is, the CEO. In this layer, I assume a single CEO, whose time is also fixed at one unit, and uses IT capital, k_L, to attempt $y_L \equiv 1 + Bk_L^{\beta_L}$ problems, where B is a TFP shifter for CEO capital, and $0 < \beta_L < 1$ captures decreasing returns to CEO's IT capital. This layer must also satisfy $y_L = y_1(1 - F(Z_{L-1}))$.

As layers and knowledge are added, more of the potential production is realized, such that for a firm with total knowledge Z_L, realized production q is

$$q = AF(Z_L)y_1 \tag{5}$$

With a decreasing density $F(.)$, economizing on knowledge implies that workers at lower layers learn the more common problems, whereas those at higher layers learn the more exceptional problems. Next, Section II.B describes the optimization problem and its solution for an arbitrary firm with its full set of constraints.

In Caliendo and Rossi-Hansberg (2012)'s theory, $y_1 = n_1$, and $y_l = \frac{1}{h}n_l$ for $l > 1$, where $h < 1$ is a parameter capturing the time per problem required by a manager in layer l to find the answer, and communicating it to those implementing the solution. In this theory, I endogenize h modeling it as firms' IT capital choices for optimal problem solving, an aspect that was absent in the knowledge-based hierarchy literature. The way I introduce capital in the organization problem is sensible. The restriction at the CEO layer assumes total time is fixed at one unit and allow the CEO to leverage his unit of time through IT capital, k_L, although its use is subject to decreasing returns to scale, $0 < \beta_L < 1$. These technology assumptions are a realistic way to capture corporations in the real world, where CEO time is limited, and information systems relieve the CEO from time-consuming activities.

The fixed CEO time assumption is shared with the knowledge-based hierarchy literature Garicano (2000); Garicano and Rossi-Hansberg (2006); Caliendo and Rossi-Hansberg (2012). It is also empirically relevant, since, as I show in the next section, fixing CEO time in the model makes knowledge and wages grow with firm size conditional on L as in the data (Caliendo, Monte, and Rossi-Hansberg 2015). In addition, IT capital at the top layer allows CEOs to increase their span (see e.g. Guadalupe, Li, and Wulf 2014). For all other layers, I use the simplest form that allows for capital-labor complementarity for both managers and production workers, as in the capital-skill complementarity hypothesis by Griliches (1969), which Goldin and Katz (1998) find plausible for the post-WWII era. To discipline the results, I impose $\sigma_l = \sigma_2$ for $1 < l < L$, and assume that IT capital at any layer has price p_2, whereas non-IT capital has price p_1.

II.B. Model Solution

The maximization problem of the firm can be split into two parts: optimization given L and the choice of L, which I next describe sequentially.

Firm Choices Given L. The problem of a firm with demand α, given L, is

$$\max_{q,\{z_l,n_l,k_l\}_{l=1}^{L}} \pi_L(\alpha) = p(q)q - \sum_{l=1}^{L} (n_l w(cz_l + 1) + p_l k_l) \tag{6}$$

subject to:

$$p(q) = \left(\frac{q}{\alpha R}\right)^{-1/\rho(\alpha)} D(\alpha) \tag{7}$$

$$q = A[1 - \exp(-\lambda Z_L)]y_1(k_1, n_1) \tag{8}$$

$$y_l(k_l, n_l) = y_1(k_1, n_1)\exp(-\lambda Z_{l-1}), 1 < l < L \tag{9}$$

$$y_L(k_L) = y_1(k_1, n_1)\exp(-\lambda Z_{L-1}) \tag{10}$$

$$n_L = 1 \tag{11}$$

$$z_l > 0, \forall l \geq 1 \tag{12}$$

In this problem, profit, $\pi_L(\alpha)$, is the objective function, composed of revenue minus total cost. Total cost is composed of the wage bill and capital spending where, at each layer l, p_l is the price of capital and $w_l \equiv w(cz_l + 1)$ is the cost of a unit of labor. The first constraint is the demand of the firm for its α variety. The second is the production function. The rest are the managerial input constraints introduced in the previous section.

To understand the solution to the maximization problem, I next describe the solution to the associated cost minimization problem for a given q, denoted $C_L(q)$.[6] The first-order condition for z_1 is

$$wcn_1 - \sum_{l=2}^{L} \psi_l y_1(k_1, n_1)\lambda \exp(-\lambda Z_{l-1}) = \phi A\lambda \exp(-\lambda Z_L)y_1(k_1, n_1) \tag{13}$$

where ψ_l is the multiplier associated to the layer l managerial input constraint, Equation 9, and ϕ is the multiplier associated to the output constraint, Equation 8. Note ψ_l is the effect on total cost of an extra unit of layer l input bundle, and ϕ is the cost effect of an extra unit of output, that is, the marginal cost. So the LHS is the marginal cost of providing a unit of knowledge to all production workers, minus the reduction in cost due to fewer problems arriving to the upper layers, whereas $RHS = \frac{\partial C_L(q)}{\partial q}\frac{\partial q}{\partial z_1} = \frac{\partial C_L(q)}{\partial z_1}$ is the reduction in total cost that an extra unit of knowledge z_1 generates. At the optimum, both are equal, an intuition that is similar for all other layers.

[6] The interested reader can find the detailed derivations in the Appendix Section B.2.

By combining the first-order conditions for k_l and n_l at every layer $l < L$:

$$\frac{k_l}{n_l} = \left(\frac{p_l}{w(cz_l + 1)}\right)^{-\sigma_l} \tag{14}$$

This is the familiar condition arising in two-factor CES production functions with elasticity parameter σ_l with a crucial difference: in this model, knowledge is a production factor that is chosen optimally and, hence gross wages are not exogenous to the firm.

Using the problem constraints and Equation 14, I obtain for each $l < L$

$$n_l = \frac{\exp(-\lambda Z_{l-1})}{A[1 - \exp(-\lambda Z_L)]} \left(\frac{(w(cz_l + 1))}{P_l}\right)^{-\sigma_l} q, \tag{15a}$$

$$k_l = \frac{\exp(-\lambda Z_{l-1})}{A[1 - \exp(-\lambda Z_L)]} \left(\frac{p_l}{P_l}\right)^{-\sigma_l} q, \tag{15b}$$

where $P_l \equiv \left(p_l^{1-\sigma_l} + (w(cz_l + 1))^{1-\sigma_l}\right)^{1/(1-\sigma_l)}$. Equation 15 are conditional factor demands and knowledge z_l is also a production factor in this theory and is determined implicitly by its first-order condition. Equation 15 has a familiar CES form, in which knowledge appears in three places: as wages, as the denominator, and as a shifter in the numerator, inside the exponential term. The reason for its appearance in wages is the usual one: to capture the negative relation between the factor price and its demand. The intuition for the latter two is also simple. Total knowledge Z_L in the denominator acts as a Hicks-neutral TFP, similar to the role of A in that a larger value delivers more production, reflecting the role that $F(Z_L)$ plays in realizing potential production. Finally, the numerator contains an exponential term that captures the trade-off between knowledge and the other two factors, labor and capital; the more knowledge at layers lower than l, the less factors at layers $l' \geq l$ are needed to produce q. At the top layer L

$$1 + Bk_L^{\beta_L} = \frac{\exp(-\lambda Z_{L-1})}{A[1 - \exp(-\lambda Z_L)]} q \tag{16}$$

that is, the CEO, with IT capital, deals with all the unsolved problems at the previous layer.

Using Equations 13 and 15-16 in total cost delivers $C_L(q)$, the cost function given q and L. In Proposition 1, I characterize how factor demands change with output given L, using comparative statics of the first-order conditions of the cost minimization problem.[7]

[7]When formally solving the cost minimization problem, I allow the firm to choose zero knowledge, but solutions of that sort will not occur. In Proposition 5 in Appendix Section B.2.1, I show that a firm optimally choosing L will never have intermediate layers with zero knowledge. The reason is that intermediate layers are costly but do not add anything to production if their knowledge is zero. Section B.1 in the Appendix provides the parameter assumptions I impose throughout and discusses their intuition. The proofs for all propositions in the main text can be found in Appendix Section B.2.

Proposition 1. *Given the number of layers L,*

(1) the knowledge of all employees, z_l $\forall l$,

(2) the number of employees, n_l $\forall l < L$,

(3) capital, k_l $\forall l$, and

(4) capital-labor ratios, k_l/n_l $\forall l$,

all increase with q.

Proof. See Appendix, Section B.2.2.

Proposition 1 states that factor demands given L are increasing in scale (except CEO hours). Intuitively, given L, as the firm expands its scale, the CEO time constraint becomes tighter. Using more IT capital allows the firm to leverage CEO time, as does increasing knowledge at all other layers. As knowledge increases, so does the relative price of labor to capital at each layer, which induces the firm to increase all within-layer capital-labor ratios, and the firm-level capital-labor ratio. Capital and labor at each layer also increase, as long as the substitution effect due to rising knowledge and w_l is smaller than the scale effect caused by increasing output and $F(Z_L)$; to be consistent with the empirical evidence, I assume this to be the case through sufficient assumptions on parameters on Appendix Section B.1.

As a consequence of this result, the following properties of the marginal and average costs given L, $MC_L \equiv \frac{\partial C_L(q)}{\partial q}$ and $AC_L \equiv \frac{C_L(q)}{q}$, respectively, can be proven:

Proposition 2. *Given L:*

(1) The marginal cost is positive and increasing in q with $\lim_{q\to\infty} MC_L = \infty$.

(2) The average cost satisfies: (i) $\lim_{q\to 0} AC_L = \lim_{q\to\infty} AC_L = \infty$. (ii) AC_L decreases at $q \to 0$ and increases at $q \to \infty$. (iii) AC_L has a U-shape with a unique minimum at q^*, the minimum efficient scale.

Proof. See Appendix, Section B.2.3.

That the MC_L is increasing with scale is a consequence of the CEO time being fixed. To understand why, it is useful to compare the benchmark model and two alternatives: (i) a model with constant returns to scale at all layers, denoted CRS, and (ii) a model without capital at the CEO level, denoted NOk_L. Only the constraints at layer L are different relative to the benchmark and are given by

$$\begin{cases} \left(k_L^{\frac{\sigma_L-1}{\sigma_L}} + n_L^{\frac{\sigma_L-1}{\sigma_L}}\right)^{\frac{\sigma_L}{\sigma_L-1}} = \left(k_1^{\frac{\sigma_1-1}{\sigma_1}} + n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}} \exp(-\lambda Z_{L-1}) & (CRS) \\ 1 = \left(k_1^{\frac{\sigma_1-1}{\sigma_1}} + n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}} \exp(-\lambda Z_{L-1}) & (NOk_L) \end{cases} \tag{17}$$

In the CRS model, the associated cost function is constant returns to scale in capital and labor, optimal knowledge is independent of scale, and the marginal cost is constant. On the contrary, under NOk_L, CEO time is fixed and $k_L = 0$, so increasing production given L necessarily involves increasing knowledge at all layers, which makes the marginal cost increase.

In the benchmark model, to increase q, the firm can increase k_L, which allows a lower marginal cost for any Z_L relative to NOk_L. Moreover, as long as k_L is not too advantageous relative to z_L, both of these factors are used to increase production, all z_l rise (Proposition 1) and the marginal cost increases with output. This outcome is guaranteed under Assumption 3 in the Appendix Section B.1, which bounds the returns to IT. If the assumption was not satisfied due to, for example, too large increasing returns to capital, $\beta_L > 1$, then, to increase output, the firm could find it optimal to reduce knowledge at all layers and increase capital with potentially lower marginal costs. I conclude describing (1) in Proposition 2: since $\frac{\partial MC_L}{\partial q} > 0$, MC_L tends to infinity as output rises.

Proposition 2 also deals with the average cost function properties. AC_L goes to infinity because zero knowledge at all layers still implies a positive fixed cost of compensating workers for their time (w times the number of employees) and capital expenses, but delivers no production, hence the label fixed cost, despite it is endogenously determined and varies with q. Note that, at the very least, the CEO must be compensated for his time. As knowledge increases (optimally) with scale, the fixed part of the cost of a layer can be spread over more units, initially lowering the AC_L until reaching the minimum efficient scale (MES), from which point on the AC_L increases with output due to the increasing marginal cost.[8] Hence, AC_L has U-shape.

[8]Propositions 1 and 2, which build on Proposition 5 and Lemma 1 in the Appendix Section B.2.1, extend the content of Proposition 1, B1, and Lemma 1 in Caliendo and Rossi-Hansberg (2012) to this theory with capital choices.

To connect cost minimization and profit maximization, note that in the latter, the first-order condition for q delivers the familiar condition that price equals a constant markup over marginal cost,

$$p(\alpha) = \frac{\rho(\alpha)}{\rho(\alpha) - 1} MC_L(q(\alpha)) \tag{18}$$

which together with Equation 2 determines the quantity produced. Firms are heterogeneous in their optimal scale, and those with a larger demand shifter α produce more. This is shown combining the implicit expression for $q(\alpha)$ with Proposition 2.

Firm Organization Choice, L. A firm with demand α chooses among three organizations by maximizing profits across L, that is

$$\max_{2 \leq L \leq 4} \pi_L(\alpha) = \max_q \; p(\alpha)q - C(q) \tag{19}$$

where $C(q) \equiv \min\limits_{2 \leq L \leq 4} C_L(q)$ is the cost function, $p(\alpha)$ is given by Equation 2.

To understand the layer number choice, I next graphically illustrate average and marginal cost for $L = 2, 3$ but the same arguments apply to any $L + 1$ vs L choice: choosing an extra layer implies a trade-off between a lower marginal cost but a larger fixed cost.

Figure I shows, on the LHS, average costs for given L have a U-shape as q changes for a given firm (Proposition 2). An extra layer, even with zero knowledge for its managers, implies a larger fixed labor cost due to the cost of employees' time (w for each worker) and capital spending. So, given a low production scale, $AC_3 > AC_2$ holds. However, a firm with an extra layer also features a lower marginal cost for any output, shown in the RHS of Figure I; the $L = 3$ organization economizes on knowledge of the lower layers by optimally choosing knowledge at the extra layer. For this reason, as output increases, the fixed part of the cost of the extra layer can be spread out and, for a large enough output, the $L = 3$ organization features both a lower average and marginal cost for a given q. Since firms with a larger α maximize profits with a larger q, these arguments also imply a larger L is chosen by a firm with a large enough α. I solve the L choice and the firm problems given α numerically, determining at the same time the menu of endogenous fixed and variable costs of each organization.[9]

[9] A numerical determination of a menu of fixed and variable costs also occurs in Caliendo and

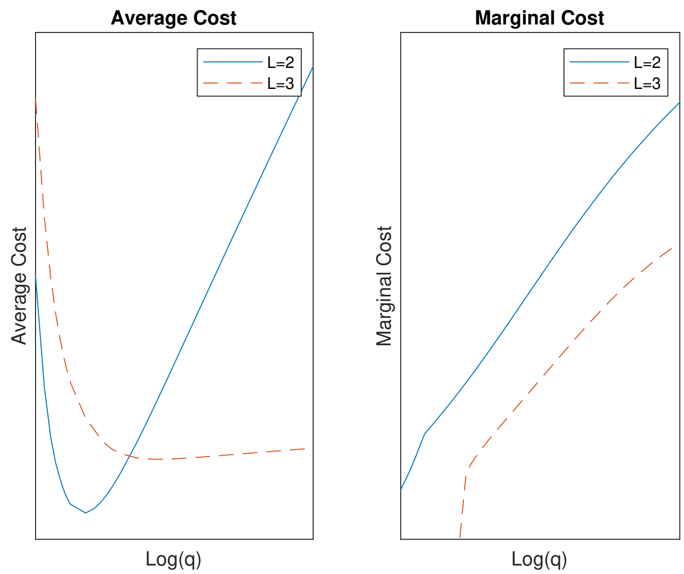

FIGURE I: Average and marginal costs as a function of q conditional on $L = 2, 3$.

II.C. Equilibrium

To close the model, I assume the existence of a large enough "outside" sector with a constant share of total income of $(1 - \eta)$, absorbing any excess factor supplies and determining factor prices, (w, p_1, p_2), similar, for example, to Antràs, Fort, and Tintelnot (2017). Income from the exogenously given factor supplies as well as firms' aggregate profits are evenly distributed to all consumers, which implies the latter's aggregate spending in consumption equals national income, denoted I.

Given factor prices, (w, p_1, p_2), the general equilibrium in this economy is given by an output price index, P, and a markup index, M, for a firm producing variety ω, its output $q(\omega)$, factor demands $\{n_l(\omega), k_l(\omega), z_l(\omega)\}_{l=1}^{L(\omega)}$ and organization choice $L(\omega) \in \{2, 3, 4\}$, and for consumers, for each variety ω, the quantity consumed, $q^c(\omega)$, such that: (i) consumers maximize utility subject to their budget constraint, $\int_\Omega p(\omega) q^c(\omega)\, d\omega = \eta I = R$, as described in Section II (ii) firms maximize profits, as described in Section II, and (iii) labor, both types of capital, and goods markets clear.

Rossi-Hansberg (2012), and in an international trade context, for example, Antràs, Fort, and Tintelnot (2017).

III. MICRO IMPLICATIONS

In this section, I obtain the firm-level implications of the theory through comparative statics using propositions and numerical results with the calibration in Section IV. Specifically, in Section III.A, I describe the cross-sectional implications, and in Section III.B, I characterize reorganization when IT prices fall; in both sections, I compare the theoretical predictions to the existing evidence. Finally, I discuss the next two key results of the paper: the indirect IT capital-to-production-labor elasticity and the TFP measurement implications of IT adoption, in Sections III.C and III.D, respectively.

III.A. The Implications for the Firm Cross-Section

To map the predictions of the model in the α cross-section to the data, I first describe the profit maximizing choices given L and then add the L choice. Henceforth, when I describe the model predictions, I treat sales and value-added interchangeably, as in ADKPV.

Proposition 1 proves that factor demands and the capital-labor ratio rise with q given L, and, by profit maximization, larger α implies more q. Hence, the content of Proposition 1 also extends to the observed firm cross-section given L. Additionally, Figure II shows the labor cost both as a share of cost and value-added for $L = 2$ in the α cross-section, using the calibrated parameters from Section IV, and $\rho = 3.8, \forall \alpha$; the qualitative behavior of other organizations' labor share's is the same.

The labor share declines for low α and then rises slightly. The labor share of cost, denoted Λ^L, changes with q according to

$$\frac{\partial \Lambda^L}{\partial q} = \frac{\sum_{l=1}^{L} \left(\frac{\partial w_l}{\partial q} n_l + \frac{\partial n_l}{\partial q} w_l \right) - MC_L \Lambda^L}{C_L} \tag{20}$$

Note that the sign depends on the numerator, whose first term is positive due to Proposition 1 and, in the second term, marginal cost behaves as in Proposition 2. For low output, the labor share of cost falls because w_l and n_l are small (Proposition 1), whereas the second term is large: for any L, as q approaches zero, the time compensation of the CEO must be incurred making Λ^L large, but marginal cost approaches a constant. As α and output increases, the labor share of cost is lower, while the first term becomes larger larger and dominates making Equation 20 positive.[10] A similar equation and arguments can be used for the labor share of value added. Next, I show this theory has predictions for the elements in Equation 20, which are observed in the data, providing production-side discipline to the behavior of the firm-level labor share, a key element to understand the aggregate labor share decline, and other macro trends.

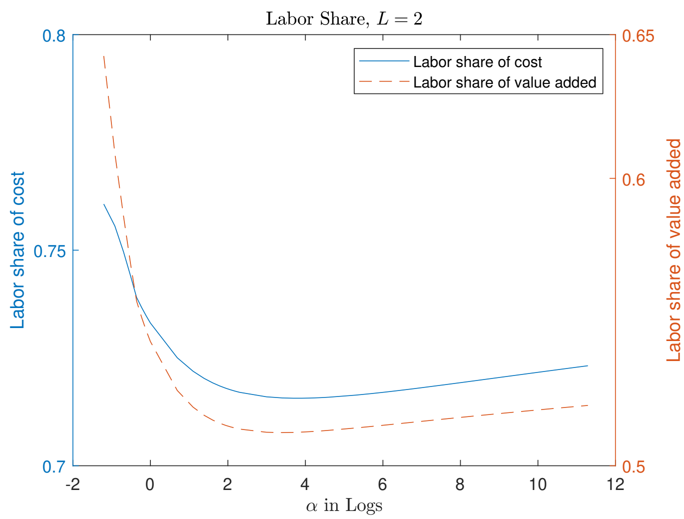

FIGURE II: Labor share of cost as a function of α, conditional on $L = 2$.

Table I shows comparative statics of key moments with respect to firm size unconditional on L, to map the model to the observed firm cross-section. To construct the table, I use calibrated parameters from Section IV, and $\rho = 3.8, \forall \alpha$, and the $*$ symbol indicates the sign is obtained as a proposition for given L. The model-implied regression slopes unconditional on L have the same sign as the correlations obtained for given L (Propositions 1 and 2). To understand the results, next I discuss how the L choice affects each moment on Table I, for the firm with an α such that it is indifferent between adding an extra layer or not.

An extra layer reduces the marginal cost because, at the margin, an extra layer allows the firm to economize on knowledge of all preexisting layers, and their wages

[10]Compare this theory to a Cobb-Douglas production function with L labor types. Note that Equation 20 can be written in elasticity form as $\frac{\partial \log \Lambda^L}{\partial \log q} = \sum_{l=1}^{L} \left(\varepsilon_{w_l,q} + \varepsilon_{n_l,q}\right) s_{l,wb} - \frac{MC_L(q)}{AC_L(q)}$, where $\varepsilon_{x_l,q} \equiv \frac{\partial \log(x_l)}{\partial \log(q)}$ for $x_l = \{w_l, n_l\}$ and $s_{l,wb} \equiv \frac{w_l n_l}{\sum_{l=1}^{L} w_l n_l}$. For such production function, $\varepsilon_{n_l,q}$ is the inverse of the output elasticity of factor n_l, is constant and equal to $1/s_{l,wb}$, and the labor share of cost does not change with output if, for example, (1) the firm is a price taker in factor markets, hence $\varepsilon_{w_l,q} = 0$, and (2) the marginal cost equals the average cost.

fall; this result is the same as in Caliendo and Rossi-Hansberg (2012), which is not surprising, since this theory converges to theirs in the limit when capital prices go to infinity. However, the rise of w_l and MC_L with α given L dominates (Proposition 1) and a positive correlation is predicted for w_l and $MC(q) \equiv \frac{\partial C(q)}{\partial q}$ and sales in the cross-section.

TABLE I: COMPARATIVE STATICS: FIRM CROSS-SECTION IMPLICATIONS

	Dependent Variable								
	w_1	w_2	w_3	w_4	$\frac{\sum_{l=2}^{L} k_l}{\sum_{l=2}^{L} n_l}$	$\frac{\sum_{l=1}^{L} k_l}{\sum_{l=1}^{L} n_l}$	$\frac{\sum_{l=1}^{L} w_l n_l}{C(q)}$	$\frac{\sum_{l=1}^{L} w_l n_l}{p(q)q}$	$\frac{\partial C(q)}{\partial q}$
$p(\alpha)q(\alpha)$	$+^*$	$+^*$	$+^*$	$+^*$	$+$	$+^*$	-	-	$+^*$

Notes. Model implied regression slopes of firm-level (1) wages in each layer, (2) IT capital-labor ratio, (3) capital-labor ratio, (4) labor share of cost and (5) of value-added, (6) marginal cost on firm revenue in the firm cross-section, unconditional to L. The $*$ symbol indicates the result holds as a Proposition for given L. Signs unconditional on L are numerical results using the calibration in Section IV and $\rho = 3.8$, $\forall \alpha$.

The capital-labor ratio can be written as $\frac{\sum_{l=1}^{L} k_l}{\sum_{l=1}^{L} n_l} = \sum_{l=1}^{L} \frac{k_l}{n_l} \frac{n_l}{\sum_{l=1}^{L} n_l}$. An extra layer reduces w_l at every preexisting l, lowering $\frac{k_l}{n_l} = \left(\frac{p_l}{w_l}\right)^{-\sigma_l}$ in those layers and contributing to a lower firm-level capital-labor ratio. But two effects more must be accounted for. First, the extra layer lowers the weight of preexisting layers ($\frac{n_l}{\sum_{l=1}^{L} n_l}$) in the firm-level capital-labor ratio, lowering the latter. Second, because quantitatively wages satisfy $w_l > w_{l-1}$, the extra management layer features a higher capital-labor ratio than preexisting managerial layers; however, this extra management layer has a smaller weight due to the hierarchical organization, which implies $y_l < y_{l-1}$ (and numerically $n_l < n_{l-1}$). Quantitatively, when a layer is added $\frac{\sum_{l=1}^{L} k_l}{\sum_{l=1}^{L} n_l}$ falls; however, the rise with α given L dominates (Proposition 1) and determines the positive correlation in the firm cross-section. With a similar argument, a positive correlation between IT capital-labor ratio and sales in the cross-section is predicted.

The firm labor share of cost of an organization L, denoted Λ^L, can be written as $\Lambda^L = \sum_{l=1}^{L} \Lambda_l^L s_{l,c}^L$, where $\Lambda_l^L \equiv \frac{w_l^L n_l^L}{w_l^L n_l^L + p_l k_l^L}$ is the labor share of cost at layer l, $s_{l,c}^L \equiv \frac{w_l^L n_l^L + p_l k_l^L}{C_L(q)}$ is the layer l share in total cost, and superscripts highlight choices of the L organization. Accordingly, the total change in the firm labor share across organizations L and $L+1$ can be decomposed using a shift-share approach (Baily, Hulten, and Campbell 1992) as

$$\Delta\Lambda = \underbrace{\sum_{l\in\Gamma} \Delta\Lambda_l s_{l,c}^{L+1}}_{\text{within-layer change}} + \underbrace{\sum_{l\in\Gamma} \Lambda_l^L \Delta s_{l,c}}_{\text{cost share reallocation}} + \underbrace{\Lambda_L^{L+1} s_{L,c}^{L+1}}_{\text{extra layer}} \tag{21}$$

where Γ is the set of common layers across both organizations and $\Delta x \equiv x^{L+1} - x^L$ for any variable x. The first term captures within-layer labor share changes, whereas the second is the reallocation of cost across layers, and both sum across all the common layers; the third is the contribution of the extra layer, L, in the $L+1$ organization. Regarding the first term, for $l < L$ the sign of $\Delta\Lambda_l$ due to an extra layer can be inferred from expression $\frac{w_l^L n_l^L}{p_l k_l^L} = \left(\frac{w_l^L}{p_l}\right)^{1-\sigma_l}$: an extra layer reduces w_l^L at every preexisting layer, an effect which tends to reduce the firm labor share of cost, due to within-layer capital-labor complementarity. Contrary, the labor to capital expense rises at the CEO layer because a larger share of problems are solved at layers below, fewer problems arrive to the CEO, and hence $k_{L+1}^{L+1} < k_L^L$; this contributes to a positive $\Delta\Lambda$, but its weight is small due to the hierarchical nature of the firm, which quantitatively implies $s_{l+1,c}^L < s_{l,c}^L \forall L$. The second term in Equation 21 captures changes in preexisting layer's cost shares (reallocation). An extra layer allows economizing on cost of all preexisting layers, that is, negative $\Delta s_{l,c} \forall l \in \Gamma$, lowering $\Delta\Lambda$. Finally, the third term, the extra (intermediate) layer of the $L+1$ relative to L organization, contributes to an increase in $\Delta\Lambda$, but its weight is small relative to the lower layers ($s_{L,c}^{L+1} < s_{l,c}^{L+1}, \forall l < L$). Quantitatively, the two effects that lower $\Delta\Lambda$ dominate and the labor share of cost declines when increasing L, which together with Figure II explains the negative correlation between value-added and labor share of cost in Table I.

To conclude, the labor share of value-added can be written as

$$\frac{\sum_{l=1}^{L} w_l^L n_l^L}{p(q^L) q^L} = \Lambda^L \frac{AVC_L}{MC_L} \frac{MC_L}{p(q^L)} \tag{22}$$

Using the optimal price decision of the firm, we have $\frac{MC_L(q^L)}{p(q^L)} = \frac{1}{m}$, so the third term does not depend on L. Regarding the second term, an extra layer increases fixed costs and reduces marginal cost, expanding output; quantitatively, an extra layer increases $\frac{AVC_L}{MC_L}$, so the LHS increases with L. However, the negative correlation between firm size and labor share of value-added given L (Figure II) dominates and is responsible for the sign on Table I.[11]

In summary, Table I shows that firms with larger sales (a higher α) are associated with (1) higher wages at every layer, (2) a higher IT capital-labor ratio, (3) a higher capital-labor ratio, (4) a lower labor share relative to both cost and value-added, and (5) a higher marginal cost. Table I slopes are calculated with sales as independent variable, but output or value-added deliver the same signs.

Next, I show the predictions of the theory for the cross-section are observed in the data. This provides production-side discipline to the calibration of ρ (for which I use Equation 22 in Section IV), which has to be also consistent with aggregate trends, like the decline in the labor share and sales reallocation to large firms.

The Evidence on the Firm Cross-Section. Empirically, wages rising with output and sales in the cross-section firm sales is the well-known size-wage premium (see e.g., Brown and Medoff 1989; Abowd, Kramarz, and Margolis 1999). This robust stylized fact is rationalized by this organization theory where capital choices are optimal, extending the results in Caliendo and Rossi-Hansberg (2012)'s only labor theory. Moreover, Caliendo, Monte, and Rossi-Hansberg (2015) use French data and a map from occupations to hierarchical layers and test the predictions in Caliendo and Rossi-Hansberg (2012). They find that, when firms growth without changing the number of layers, wages at all layers rise.

In the theory, as firms try to produce more, they optimally choose more knowledge, making labor at all layers relatively expensive. As a consequence, the firm raises its capital-labor ratios at all layers, and also at the firm level. Evidence that capital-labor ratios increase in the firm value-added cross-section is found among others by Raval (2010) in United States manufacturing, and by Abowd, Kramarz, and Margolis (1999) in French data for several industries. Similarly, Lashkari, Bauer, and Boussard (2018) show a positive association between firm size and the IT capital-labor ratio, which in this theory is a consequence of firms substituting towards capital as managers' knowledge, and their wages, increase with size.[12] Raval (2019) shows that the factor cost ratio,

[11]In the calibration, Section IV, I return to Equation 22 as I introduce heterogeneity in demand elasticity in the firm cross-section and there will be an extra source of variation. As I will show in that section, heterogeneity in $\rho(\alpha)$ only reinforces the negative correlation on Table I between labor share of value-added and sales, since markups rise with firm size in the data and the model.

[12]Their paper uses a nonhomothetic production function, which they estimate, and also perform general equilibrium counterfactual experiments.

the ratio of capital cost to labor cost, is positively associated to firm value-added (so negatively to the labor share of cost) in US manufacturing. Similarly, ADKPV report the ratio of payroll to sales to be negatively related to sales, using manufacturing and service sectors microdata from the United States Economic Census; Autor et al. (2017) explain their observation with a model featuring a fixed cost of overhead labor, which in this theory is the fixed part of the labor cost of layers.

Turning to the properties of the marginal and average cost function (Figure I) have the same properties as in Caliendo et al. (2015), who study how organizational choices affect firm productivity. Using Portuguese employer-employee matched data and firms' production quantity and input data in the manufacturing sector, they study how productivity responds to reorganization as measured by a change in the number of management layers. They find that an exogenous demand increase that makes the firm add a management layer raises quantity-based productivity and revenue-based productivity falls. The difference between these two productivity measures implies that prices decline by adding an extra layer. Together, this suggests that firms reduce their marginal cost by adding an extra layer as in the theory in Caliendo and Rossi-Hansberg (2012) and this paper, which extends their results by adding optimal capital choices.[13]

Data and model agree qualitatively, specifically on the micro responses that determine the change of the labor share of cost in the firm cross-section. Moreover, the macro implications of the *quantitative* micro facts are potentially large: the correlation between factor cost and value-added implies "an increase of about 35% to 50% [in the factor cost ratio] between plants with value-added at the industry mean and the largest plants in the industry" (Raval 2019, 6); understanding its structural determinants requires a quantitative theory.

III.B. The Implications of the Organizational Channel of IT

Before turning to how knowledge changes across layers as a result of falling IT prices (reorganization), I describe the trade-off across layers when choosing knowledge (organization). Combining first-order conditions for knowledge in layers 1 and 2, and using a convenient expression for the multiplier in layer 2, delivers, at the optimum of an

[13]Interestingly, Almunia et al. (2018) also find evidence of increasing marginal costs, as in these organizational theories. Almunia et al. (2018) show that an exogenous shock to domestic demand, reduced firms' marginal cost allowing them to expand their sales in foreign markets through exports.

$L > 2$ organization:[14]

$$\underbrace{wcn_1}_{\text{MC of } z_1} - \underbrace{\overbrace{\frac{p_2}{\frac{\partial y_2(k_2,n_2)}{\partial k_2}}}^{\text{Cost per problem}} \overbrace{y_1(k_1,n_1)\lambda\exp(-\lambda z_1)}^{\text{Change in Mass of Problems}}}_{\text{Savings in Managerial Inputs}} = \underbrace{wcn_2}_{\text{MC of } z_2} \tag{23}$$

On the LHS, an extra unit of knowledge in layer 1 has marginal cost of wcn_1 (first term) but allows for less spending in communication costs to layer 2 (second term). This second term captures cost savings from the extra unit of z_1: the cost per problem (price of IT over mass of problems solved by the last unit of IT) times the reduction in the amount of problems received in layer 2 (number of problems times change in density sent to layer 2). At the optimum, the marginal cost of knowledge in layer 1 net of the change in communication costs (LHS) is equal to the marginal cost of knowledge in layer 2 (RHS).

Equation 23 also delivers the intuition for how knowledge reallocation occurs when the IT price, p_2, falls: savings in communication costs due to layer 1 knowledge are lower, which makes z_1 relatively expensive. This tends to increase knowledge at layer 2 and decrease it at layer 1. Thus, firm reorganization of knowledge across layers generates wage inequality within the firm, as is formally proven in the next two propositions. Proposition 3 describes how factors and marginal cost respond to declining IT prices holding q constant, that is, along an isoquant as in classic analysis (see Hicks 1932), building the intuition for the indirect IT capital-production-labor elasticity, in Section III.C, and for Proposition 4, where I allow output to adjust to map theory to the data.

Proposition 3. *Given q, the firm response to a decline in the IT price on*

(1) knowledge is such that

- *(i) for any L, $\frac{dz_1}{dp_2} > 0$,*
- *(ii) for any L, $\sum_{l=1}^{L} \frac{dz_l}{dp_2} = \frac{dZ_L}{dp_2} > 0$,*
- *(iii) for $L = 2$, $\frac{dk_L}{dp_2} < 0$, and $\frac{dz_L}{dp_2} < 0$; and*

(2) cost is such that for any L, $\frac{dMC_L}{dp_2} > 0$.

[14] These intuitions hold for any L, however.

Proof. See Appendix, Section B.2.4.

The knowledge result (i) formally shows the aforementioned intuition that when IT prices decline, the firm optimally reallocates knowledge away from the lowest layer (i.e., $\frac{dz_1}{dp_2} > 0$). The complete reorganization rationale is as follows. The number of problems, $y_1 \equiv y_1(k_1, n_1)$, and total knowledge, Z_L, are choices that act like substitutes to generate a given output, $q = AF(Z_L)y_1$. Since lower IT prices reduces the cost of managers dealing with problems, the firm reorganizes by increasing the number of problems, y_1, and reducing the now relatively expensive factor, Z_L (i.e., result [ii] in Proposition 3). In particular, this is achieved with a lower z_1 since this increases the volume of unsolved problems that is passed on to higher layers.[15]

The result (iii) examines the simpler case $L = 2$: the decline in IT prices makes IT capital increase, to attempt the solution of the increased volume of problems. CEO knowledge also increases because of the knowledge reallocation mechanism described in the previous paragraph. Quantitatively, (iii) holds $\forall L$. Finally, result (2) states that marginal costs fall when IT becomes cheaper.

Relative to Proposition 3, mapping the theory to the data requires that q optimally adjusts as p_2 declines, and is the subject of Proposition 4.

Proposition 4. *Firm responses to a decline in the IT price, for profit maximizing output q, have the same characteristics as in Proposition 3, except, for any L,*

(1) $\frac{dq}{dp_2} < 0$.

(2) $\sum_{l=1}^{L} \frac{dz_l}{dp_2} = \frac{dZ_L}{dp_2} < 0$.

Proof. See Appendix, Section B.2.5.

Proposition 4 shows that firms increase output when the price of IT declines, due to a reduction in marginal cost. As a consequence of the output increase, Z_L also increases because, as shown in Proposition 1, firm total knowledge increases with output.

The first key result of the paper is that, despite (a) production featuring capital-skill complementarity and (b) with IT capital deepening, Proposition 4 implies *lower*

[15] As Section III.C shows, the indirect IT capital-to-production-labor elasticity substitution hinges on $\frac{dz_1}{dp_2} > 0$ for given q.

wages for production workers, a result that Autor (2015) finds "hard to reconcile" in the context of Krusell et al. (2000). In fact, while Krusell et al. (2000) shares features (a-b), their framework "should [...] raise real wages of low-skilled workers" (Autor 2015, 3), a result that is counterfactual with the United States experience (Autor 2014). Of course, the proposition deals with nominal wages, but in Section IV, I show that real wages respond the same way in general equilibrium and, moreover, as in firm-level empirical studies of IT adoption.

Table II shows firm reorganization predicted by the model when the IT price falls allowing an optimal q, together with the available evidence. In the LHS of the table, I use the calibration in Section IV to obtain the response of an $L = 4$ organization to a decline in the IT price[16] and the $*$ symbol indicates the result is a Proposition for any L. The RHS of the table reports the implied empirical results in GW, who empirically study a tax incentive to IT investments in the UK. I first describe the theory predictions and then present the evidence.

The LHS of Table II, Panel A shows comparative statics for layer-level variables. The knowledge reallocation mechanism implies that w_l at $l = 1, 2$ falls, whereas the wage in managerial layers 3 and 4 rises. Because managerial layers' capital becomes cheaper and managers are complementary to IT, the managerial layers' employment relative to that in the production layer rises. While total employment rises, the CEO has fixed time, so his share in total firm labor falls. These two results imply that while the CEO span of control rises, that of intermediate managers declines. The wage bill of $l = 2, 3$ in the total wage bill rise, because relative employment rises at both layers and wages rise at $l = 3$, and while w_2 falls, this is small relative to the employment response. The relative wage bill of the production layer falls because both its relative employment and wage declines. Finally, because the CEO span of control rises, his share of the wage bill declines, despite CEO wage also rising.

[16] I use the median α given L to construct the table but the predictions are robust across α. Also, the model predictions are all the same for $L = 3$, except one needs to dispense with the $l = 2$ column and relabel $l = 3$ in the table with $l = 2$. Having done that you realize the behavior of z_2 is the only prediction that is different across $L = 3$ and $L = 4$: it increases in the former but falls in the latter organization.

TABLE II: COMPARATIVE STATICS: FIRM RESPONSE TO LOWER IT PRICES

Panel A. Firm-Layer Level							
	Model				Data		
	Layer				Layer		
	1	2	3	CEO	1	2	3
Moments							
w_l	-*	-	+	+	-	+	+
$\frac{n_l}{\sum_{j=1}^{L} n_j}$	-	+	+	-	-	+	+
$\frac{w_l n_l}{\sum_{j=1}^{L} w_j n_j}$	-	+	+	-	-	+	+
$\frac{k_l}{n_l}$	-*	+	+	+			
Panel B. Firm Level							
	Model				Data		
Moments							
$\frac{\sum_{l=2}^{L} k_l}{\sum_{l=1}^{L} n_l}$	+				+		
Z_L	+*				+		
$\frac{\sum_{l=1}^{L} w_l n_l}{p(q)q}$ and $\frac{\sum_{l=1}^{L} w_l n_l}{C(q)}$	+						

Notes. The first column reports the effects of lower p_2 in the model: + denotes an increase in the moment and − denotes a decrease. The $*$ symbol indicates the result is a Proposition for any L, whereas the rest are numerical results for the firm with median α given $L = 4$ using the calibration in Section IV, where only IT prices change between steady states. The numerical predictions are all the same for $L = 3$, except you need to dispense with the $l = 2$ column and relabel $l = 3$ in the table as $l = 2$. The RHS shows the results implied by GW, who empirically study a tax incentive to IT investments in the UK.

The last element in Panel A is the within-layer capital-labor ratio, which is affected by the endogenous response of knowledge, as it alters the relative price of labor to capital at each layer. On the production layer, w_1 falls but p_1 is constant, so more labor to capital is used. At $l = 2$, both p_2 and w_2 decline, but the former by more than the latter, so the capital-labor ratio increases. At $l > 2$, w_l rises, which also increases the capital-labor ratio, including at the CEO layer, whose time is fixed.

The LHS of Table II, Panel B reports the firm-level predictions of the theory. The IT capital-to-labor ratio increases because it does at all layers that use IT. As in Proposition 4, firm total knowledge increases, $\forall L$, due to the expansion of output.

Recall from Section II.B, that a firm's labor share of cost, Λ, can be written as $\Lambda = \sum_{l=1}^{L} \Lambda_l s_{l,c}$, where $\Lambda_l \equiv \frac{w_l n_l}{w_l n_l + p_l k_l}$ and $s_{l,c} \equiv \frac{w_l n_l + p_l k_l}{C_L(q)}$, where I omit the dependence of all choices on L because, in the following analysis, I fix $L = 4$. The total change in the firm labor share between periods t (high p_2) and $t+1$ (low p_2) can be decomposed as

$$\Delta\Lambda = \underbrace{\sum_{l=1}^{L} \Delta\Lambda_l s_{l,c}^{t+1}}_{\text{within-layer change}} + \underbrace{\sum_{l=1}^{L} \Lambda_l^t \Delta s_{l,c}}_{\text{cost share reallocation}} \tag{24}$$

where $\Delta x \equiv x^{t+1} - x^t$ for any variable x. The first term captures within-layer changes in labor share and the second is reallocation of cost shares across-layers. Regarding the first term, for $l < L$, the sign of the layer-level labor share change ($\Delta\Lambda_l$) due to lower p_2 comes from expression $\frac{w_l n_l}{p_l k_l} = \left(\frac{w_l}{p_l}\right)^{1-\sigma_l}$. Given the capital-labor complementarity assumption, and that $\frac{w_l}{p_l}$ rises for layers 2 and 3, their layer-level labor share rises, while the opposite occurs in layer 1. The CEO layer labor share falls due to a race between rising k_L and z_L, quantitatively won by the former, now a cheaper factor. Each of these changes affect the firm labor share according to each layer's weight in total cost ($s_{l,c}^{t+1}$), which satisfy $s_{l,c}^t < s_{l-1,c}^t, \forall t$, due to the hierarchical nature of the firm. Overall, due to this first term in Equation 24, the firm labor share increases. The second term in Equation 24 captures changes in layer-level cost shares ($\Delta s_{l,c}$). When p_2 falls, the cost share of managerial layer's increase, while in the production layer it decreases, and each of them contribute to the change in the firm labor share according to each layer's labor share (Λ_l^t). Layers satisfy $\Lambda_l^t < \Lambda_{l-1}^t$ and overall this channel contributes to a decline in the firm labor share. Quantitatively, the rise in managerial layer's labor share is larger than the opposing effects and the firm labor share *rises*. To infer changes in the labor share of value-added for $L = 4$, Equation 22 can be used: because the inverse markup is constant, and AVC_L/MC_L plays a small role, it rises like Λ;[17] hence, this result and Equation 22 imply that larger markups are required for a reduction in the labor share of value-added over time.

Figure III shows how heterogeneous firms optimally choose L and how their choices change with the IT price. Firms with larger α sell more, so they find it optimal to have a larger L since the corresponding "fixed" cost of the a layer is more than compensated

[17]On the contrary, for an $L = 2$ the labor share falls, since this organization does not feature intermediate managerial layers plus its k_L increases substantially. I numerically decompose the labor share of value-added across L into each contributing element in Equation 22 in Table VII in the calibration Section IV.

by the larger revenue caused by lower marginal cost. Also, with a lower p_2, the marginal cost falls for all organizations, as does the cost of an extra layer, and firms with a high α given L find it optimal to add an extra layer.

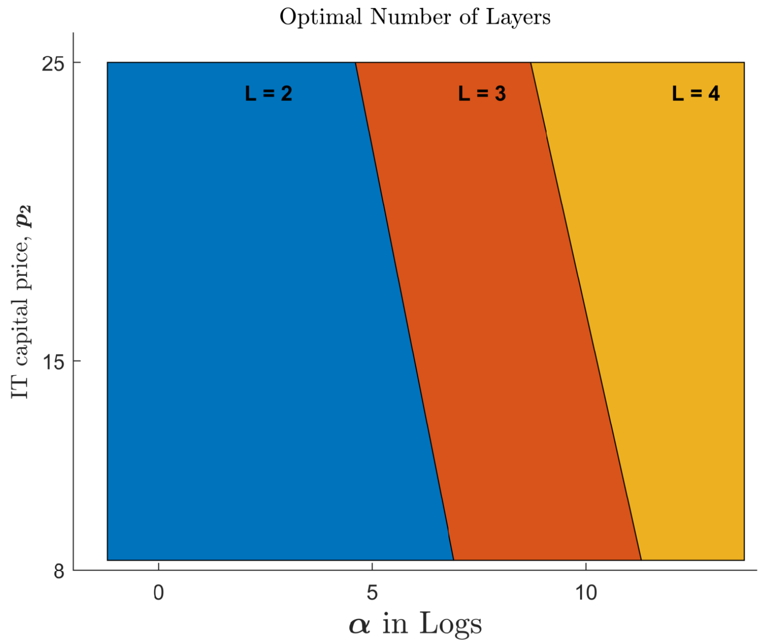

FIGURE III: Optimal number of layers as function of (α, p_2).

In the next section, I match these theoretical predictions to the existing evidence. To understand why the match matters, note that Equation 24 in elasticity form is

$$\frac{\partial \log \Lambda}{\partial \log p_2} = \sum_{l=1}^{L} \left(\varepsilon_{w_l,p_2} + \varepsilon_{n_l,p_2}\right) s_{l,wb} - \varepsilon_{C,p_2} \tag{25}$$

where $\varepsilon_{x_l,p_2} \equiv \frac{\partial \log(x_l)}{\partial \log(p_2)}$ for $x_l = \{w_l, n_l\}$, $s_{l,wb} \equiv \frac{w_l n_l}{\sum_{l=1}^{L} w_l n_l}$, and $\varepsilon_{C,p_2} \equiv \frac{\partial \log(C(q))}{\partial \log(p_2)}$. Next, the evidence is consistent with this theory's characterization of how firms' wages, factor demands and factor spending shares change when IT prices fall and, specifically, on ε_{w_l,p_2}, and, $\forall l$, $s_{l,wb}$ and ε_{n_l,p_2}. Regarding ε_{C,p_2}, as is always the case, one can only infer it indirectly from the elements on Table II. This provides production-side discipline to determine how changes in parameters p_2 and ρ over time have affected the observed micro labor share and also aggregate trends, like the decline in the labor share and sales reallocation to large firms.

The Evidence on the Organizational Channel of IT. I proceed in two parts: first, I use GW to construct the RHS of Table II, and then I review all other related evidence.

GW study a policy experiment in the UK in which small firms (those with fewer than 50 employees) were granted a 100% first year tax allowance on IT investments (implicitly, an IT price reduction). Using a regression discontinuity approach, they obtain effects of the policy for weekly earnings and weekly hours per worker for four task-based worker categories, that is, routine manual, routine cognitive, nonroutine manual, and nonroutine cognitive, and firm-level effects for investment by capital type. Appendix Section C.2.3 describes how I construct the RHS of Table II, including how to map their worker-level results, which use a task-based labor classification, to the labor types in this knowledge-based theory using occupational classification descriptions.[18]

Layer-level results are in Table II, Panel A. As a consequence of the policy, in GW, Table 4 wages in layer 1 fall and they increase in managerial layers 2 and 3, as in $L = 4$ firms in the model, except at layer 2. The data and the model agree with the predictions that the share of production labor hours declines, whereas that of both managerial layers increases. The empirical response of wage bill shares also supports the theory: the production labor wage bill over the total wage bill falls, whereas the wage bills of the two managerial layers rises.

Firm-level results are in Table II, Panel B. IT over total labor hours rises and total knowledge also rises, both as in the model. The latter result is particularly reassuring given the proof for Z_L in Proposition 4 is not an immediate result and is a unique prediction from the theory; moreover, the empirical result is likely to be downward biased for a $L = 4$ organization, since I ignore the z_L response (due to lack of specific empirical results for the CEO).

The empirical results in Table II have the advantage of coming from a single study. Moreover, they are aligned with the results obtained in other contexts, which I review next, roughly following the structure of the table from top to bottom.

In the late 1980s, a period of rapid adoption of computers in the United States, Krueger (1993) infers that workers using computers increased their wages using industry-level data. Akerman, Gaarder, and Mogstad (2015) study the effects of the introduc-

[18] One caveat is that GW do not report results for CEOs, so in constructing in Table II, I do not use nor report results for CEOs specifically.

tion of broadband internet in Norway, which under appropriate controls is a quasi-experiment, providing causal evidence in availability. Using worker data, they show declines (increases) in hourly wages of unskilled (skilled), where (un)skilled labor is defined as employees with(out) college education. Moreover, their results are robust to using the task interpretation of labor market outcomes (Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011): hourly wages of routine tasks decline and the increase for nonroutine abstract tasks. The latter results are also an empirical validation of the key theoretical prediction $\frac{\partial z_1}{\partial p_2} > 0$, since, as argued above for GW, the routine-versus-nonroutine categorization has a natural map to the knowledge-based hierarchy literature.

Bloom et al. (2014) study the effects of several technologies on the autonomy of workers and managers, using a firm hierarchy perspective and data from companies based in Europe and the United States. For the autonomy measures that capture the length of the knowledge interval of the employee, they use a management survey.[19] They find that production worker and plant manager knowledge falls with intranet, and plant manager knowledge increases with enterprise resource planning software (ERP), again consistent with the model in this paper. Importantly, their results are also causal, and their OLS estimates are confirmed by an IV approach. These results are based on nonwage measures of knowledge, and hence provide complementary evidence to studies measuring the impact of IT on knowledge using wages.

Song et al. (2019) use the Social Security employer-employee data for the United States, and find that, from 1980 through 2015, in firms with 10,000 or more employees, the median real wage has declined. This is a surprising fact, and the model in this paper suggests an explanation: large firms use IT more intensively, and as a consequence, when IT prices fall, they reallocate more problem solving away from the production workers towards managerial layers. Moreover, Song et al. (2019) find the decline in the median real wage *only* occurs in firms with 10,000 or more employees but rises in the smaller ones, with between 100 and 1,000 employees. I report the quantitative fit of the model

[19] Autonomy is defined following Bresnahan, Brynjolfsson, and Hitt (2002). Workers' autonomy is a dummy taking value one whenever decisions on both pace of work and allocation of production tasks are mostly taken by workers (i.e., both variables take values higher than three). Plant-manager autonomy is defined in four ways: (i) how much capital investment a plant manager could undertake without prior authorization from the corporate headquarters and (ii) where decisions were effectively made in three other dimensions: (a) hiring a new full-time permanent shop-floor employee, (b) the introduction of a new product, and (c) sales and marketing decisions. They interpret these definitions of autonomy in the data as the length of the knowledge interval (i.e., to how many problems are solved) of each employee type in a knowledge-based hierarchy model.

to the latter two facts in Table VIII of untargeted moments in Section IV, where I take care of the output price index.

Hence, the first key result, that knowledge on layer 1 falls when IT prices decline ($\frac{\partial z_1}{\partial p_2} > 0$ in Proposition 4) has ample empirical support as it comes from studies measuring knowledge with wages and also survey questions, and from different contexts with multiple variation sources (Akerman, Gaarder, and Mogstad 2015; Bloom et al. 2014; Gaggl and Wright 2017; Song et al. 2019). Moreover, this finding also supports the second key result, the indirect IT capital-to-production-labor substitution, as I show next on Section III.C.

Turning to employment shares, the seminal Autor, Levy, and Murnane (2003) provides early evidence on the impact of IT on labor organization. The paper argues that computers substitute for workers in performing routine cognitive and manual tasks, whereas they complement workers in performing nonroutine problem solving and complex communications tasks. Using data on task input for 1960 to 1998 for the United States, a context of rapid computer adoption, they provide evidence consistent with the hypothesis. Moreover, in the faster computerizing industries, labor shifted favoring nonroutine and against routine tasks. Bresnahan, Brynjolfsson, and Hitt (2002) provide firm-level evidence for the United States that IT is positively related to firms using a lower unskilled employment share, and a larger managerial employment share.[20] Guadalupe, Li, and Wulf (2014) use a panel of large United States firms and show the CEO span of control rises when firms invest in IT.

Relatedly, Autor, Katz, and Krueger (1998) show that computer adoption and the skilled wage bill share are positively related within manufacturing and nonmanufacturing industries in the United States. They view the spread of computers as altering the organization of work and, more generally, raising the relative demand for skilled workers. Using firm-level data for the UK, Caroli and Van Reenen (2001) show that organizational change is related to a lower unskilled manual wage-bill share.

Akerman, Gaarder, and Mogstad (2015) show causal firm-level evidence on the response of factor use. Following a Levinson-Petrin firm production function estimation approach, they find a decline (rise) in the output elasticity of unskilled (skilled) labor due to the availability of broadband. Moreover, their OLS results are robust to using firms' DSL adoption instrumented with availability. GW study mechanisms for

[20] They define IT as (i) percent of workers using general purpose computing; (ii) percent of workers using email; (iii) computerization of work; and (iv) computing power.

their empirical results on firm's factor reorganization and show that firms introduce "Advanced Management Techniques" and change "Organizational Structure".

To conclude, the evidence supports the model's qualitative prediction, specifically on ε_{w_l,p_2} and ε_{n_l,p_2}, $\forall l$, and indirectly on ε_{C,p_2}, which determine the change of the labor share of cost with p_2. Moreover, the *quantitative* micro facts are large: for example, between 1980 and 2015, the median real wage in firms with 10,000 or more employees has *declined* by 7% (Song et al. 2019), a large effect which requires a quantitative theory to understand its structural determinants and its macro implications. I return to the connection with aggregates in the calibration, Section IV.

III.C. The Indirect IT capital-to-Production-Labor Elasticity of Substitution

The theory speaks to the literature that studies the shape of the microeconomic production function, with an emphasis on how new technologies and different types of labor may complement or substitute for each other (see Goldin and Katz 1998). This theory does not assume any explicit constraint to the substitution or complementarity possibilities between IT capital and production workers. As a consequence, how IT and production labor relate to each other is not immediately apparent.

To study this issue, I focus on an IT capital-to-production-labor elasticity defined as

$$\varepsilon_{k_l,n_1} \equiv -\left.\frac{d\log\left(\frac{k_l}{n_1}\right)}{d\log\left(\frac{p_2}{w_1}\right)}\right|_{q=\overline{q}} \tag{26}$$

for any layer $l \geq 2$, for fixed output $\overline{q}$, when all factors are allowed to adjust. This definition of the elasticity of substitution is an intuitive way to capture reorganization effects, and follows the classic analysis for two factors along an isoquant in Hicks (1932), and its multiple factor generalization, the Morishima elasticity, ME henceforth (see Morishima 1967; Blackorby and Russell 1989). ε_{k_l,n_1} is the percent response of the k_l-to-n_1 ratio to a percent change in the p_2-to-w_1 ratio, fixing q and allowing all other factors to also adjust. Note that the specific mechanisms in the model imply that there is not one perfect mapping to any elasticity used in the literature; usually factor prices in the elasticity definition are exogenous to the firm, but in this model changes to p_2 affect w_1 through firms' first-order conditions that determine all production factors.

Absent this endogenous response of $w_l, \forall l$, this would be the ME defined for an IT capital price change.[21]

The firm-level IT capital-to-production-labor elasticity satisfies:

$$\varepsilon_{IT,n_1} \equiv -\frac{d\log\left(\frac{\sum_{l>1}^{L} k_l}{n_1}\right)}{d\log\left(\frac{p_2}{w_1}\right)} = \sum_{l>1}^{L} \varepsilon_{k_l,n_1} s_{l,k}$$

where $s_{l,k} \equiv \frac{k_l}{\sum_{j>1}^{L} k_j}$. Note that there is no perfect way to compare the elasticity across different organizations due to the different number of layers; more complex firms have more layers, and hence more IT capital. However, due to the hierarchical nature of firms, it is easy to see from Equation 9 that input bundles across layers satisfy $y_l < y_{l-1}$, for any l. Moreover, the same inequality sequence holds quantitatively for capital and labor across layers, that is, $n_l < n_{l-1}$ and $k_l < k_{l-1}$, for any l. As a consequence the layer 2 share of total IT capital is largest, and ε_{k_2,n_1} is quantitatively more important than $\varepsilon_{n_1,k_l}, \forall l > 2$, in determining the quantitative value of $\varepsilon_{n_1,IT}$. For this reason, from here onward, I focus on ε_{k_2,n_1}, noting that ε_{IT,n_1} is also larger than one.

Figure IV shows quantitative results for ε_{k_2,n_1}, when only parameter p_2 changes but $w_l, \forall l$, optimally respond, using the calibrated parameter values from Section IV. Like Karabarbounis and Neiman (2013), I focus on a long-run analysis, from an initial to a final steady state, using the IT price change between 1980 and 2015 from the BEA as the only difference across periods. The calibration uses off-the-shelf long-run capital-labor elasticities (σ_l), such that there is capital-labor complementarity for both production workers and managers (i.e., $\sigma_l < 1, \forall l < L$), though more so for the latter (i.e., $\sigma_1 < \sigma_l = \sigma_2 < 1$), consistent with the capital-skill complementarity hypothesis by Griliches (1969), a view that Goldin and Katz (1998) and Autor (2015) consider plausible for the post-WWII era (see Hamermesh 1993, for a survey).

In Figure IV, ε_{k_2,n_1} is larger than one in the firm cross-section, implying k_2 to n_1 *indirect* substitution, despite capital and labor being complementary for managers and production workers. Figure IV shows substantial differences in the elasticity across organizations: in the simpler $L = 2$ firms, layer 2 capital and production labor are very substitutable. To understand why, note that the numerator of ε_{k_2,n_1} is just the

[21] For a more detailed discussion on measuring capital-labor elasticities, see Appendix Section B.3.

difference in growth rate of k_2 and n_1, whereas the denominator is the percent change of $\frac{w_1}{p_2}$. The interpretation of the latter is common to all organizations as the elasticity of w_1 to p_2, denoted by $\varepsilon_{w_1,p_2} \equiv \frac{d\log(w_1)}{d\log(p_2)}$. Table III shows ε_{w_1,p_2} is quantitatively small and similar across organizations, relative to the values in Figure IV. So the heterogeneity in ε_{k_2,n_1} across organizations comes from the numerator to which I turn next.

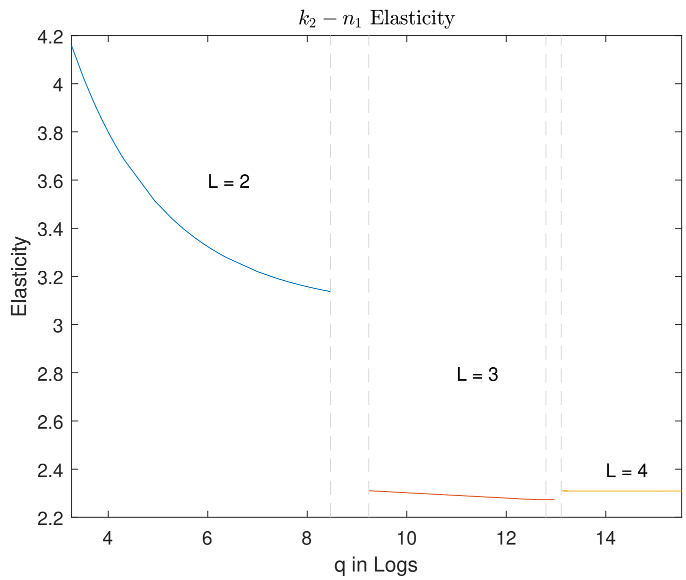

FIGURE IV: Layer 2 IT Capital-to-Production-Labor Elasticity.

For $L = 2$, k_2 is the only IT capital the firm uses so it is very elastic. The most complex firms with $L = 4$ have three layers of management whose capital becomes cheaper when p_2 declines. Moreover, more complex firms also increase n_2, unlike $L = 2$ firms whose CEO time is fixed, so k_2 responds less in $L = 4$ firms. Finally, n_1 rises when p_2 falls due to $\frac{\partial z_1}{\partial p_2} > 0$, an effect which dampens ε_{k_2,n_1}, and by Table III more for $L = 4$ firms than $L = 2$ organizations.

TABLE III: PRODUCTION WAGE RESPONSE TO IT PRICE

	Organization		
	L=2	L=3	L=4
ε_{w_1,p_2}	0.06	0.08	0.09

Notes. Percent response of the production-wage to a percent change in the IT price, that is, $\varepsilon_{w_1,p_2} \equiv \frac{d\log(w_1)}{d\log(p_2)}$, for the median firm given L, for all organizations, using the calibration in Section IV with the IT price as only source of variation across periods.

A quantitative decomposition, focusing only on optimal knowledge decisions, highlights the specific organizational channels determining ε_{k_2,n_1}. Next, I focus on the intuition for $L = 3, 4$ organizations and provide the full quantitative analysis for all L in Appendix Section C.3. When p_2 changes, three terms matter for ε_{k_2,n_1}: (i) two within-layer price index changes for layers $l = 1, 2$, and the (ii) reorganization of problem solving across layers, which is a function of λ. The first two are standard, except for endogenous wage responses, capture substitution across factors within a layer and directly depend on σ_l. At layer 1, w_1 falls, n_1 rises, so this term contributes to a lower ε_{k_2,n_1}. At layer 2, both factors' prices (w_2 and p_2) change so its sign is undetermined a priori. Finally, knowledge reorganization implies $\frac{\partial z_1}{\partial \log(p_2)} > 0$ (Proposition 3), and more problems are sent to the upper layers, raising factor demands at layer 2 and contributing to more IT capital-to-production-labor substitution. This latter term contributes around 95% of ε_{k_2,n_1}, with the remaining terms less than 5% each in absolute value.

The second key result is that ε_{k_2,n_1} is quantitatively determined by the endogenous response of z_1 to p_2 together with λ, not by the standard parameters governing substitution, σ_l. Hence, this *indirect* substitution mechanism would not be a relevant implication of the theory if empirically we found the opposite (i.e., if $\frac{\partial z_1}{\partial \log(p_2)} < 0$). In Section III.B, I provide ample evidence supporting that production worker knowledge decline with IT adoption, suggesting the indirect IT capital-to-production-labor substitution is at work. This is in contrast to the standard approach in the macro literature, where the parameters that govern the substitution pattern between capital and unskilled labor in the production function are calibrated/estimated (see e.g., Krusell et al. 2000; Acemoglu and Restrepo 2018; Hemous and Olsen 2016; Eden and Gaggl 2018), hence the label indirect for ε_{IT,n_1}.

III.D. TFP, Firm Reorganization, and IT

This section discusses the third key result of the paper: the TFP effects of IT and its connection to firm organization. The production function is $q = AF(Z_L)y_1(k_1, n_1)$. In the productivity estimation literature, TFP is usually isolated, assuming a Cobb-Douglas production function that aggregates each labor and capital into its respective input quantity index. A naive measure of the Solow Residual in this model would be $TFP_m \equiv AF(Z_L)$, whereas true TFP is actually A. From this perspective, the behavior of the naively measured TFP is associated with that of $F(Z_L)$, so I first describe how Z_L responds to cheaper IT using the earlier propositions and then compare the model to the empirical literature.

Proposition 3, which studies reorganization for fixed q, and 4, reorganization for optimal q, imply opposite results for Z_L, and combining both is illustrative for studying TFP measurement. Denote by γ_x, the growth rate of x when p_{IT} falls, $\bar{q}$ output before p_{IT} changes, and q^* optimal output after p_{IT} changes, then

$$\gamma_{TFP_m}\big|_{q=q^*} = \underbrace{\gamma_{TFP_m}\big|_{q=\bar{q}}}_{\substack{\text{Solow Paradox} \\ <0}} - \rho \underbrace{\Psi \gamma_{MC_L}\big|_{q=q^*}}_{\substack{\text{Scale Expansion} \\ <0}} > 0 \tag{27}$$

where $\Psi > 0$.[22] Equation 27 shows that measured TFP growth, when q adjusts as in the data, can be decomposed as the difference between two effects: (1) measured TFP growth, when q is fixed, minus (2) a term proportional to MC_L growth.

The first term can be labeled a Solow Paradox effect: since by Proposition 3, holding q constant, $\frac{dZ_L}{dp_2}$ is positive, cheaper IT makes measured TFP growth lower; as Robert Solow stated, "You can see the computer age everywhere but in the productivity statistics." To the extent that worker knowledge is ignored in productivity estimation, as is often done, TFP_m is downward biased. The second term captures the effect on knowledge of the output expansion due to lower marginal costs; it contributes positively to $\gamma_{TFP}|_{q=q^*}$ because Z_L increases with q, as shown in Proposition 1. The overall effect of both terms is positive: naively measured TFP and IT capital adoption are positively correlated, or as a different observational statement, endogenous reorganization makes total firm knowledge and IT adoption positively correlated.

This is the third key result of the paper. It is a precise mechanism for the empirical results found on (1) the complementarity between firm knowledge and IT and (2) positive effects of IT on the naively measured TFP. Specifically, in Proposition 4, I show a joint increase of k_L and Z_L as IT prices decline, for $L = 2$. For more complex firms, I obtain the same results quantitatively. This discussion shows that unless worker knowledge, firm organization, and IT are jointly studied, measurement of true TFP, A, is biased.

The Evidence on TFP measurement and IT. Evidence on the existence of the Solow Paradox is found by Acemoglu et al. (2014), who study the connection between productivity growth and IT capital intensity at the industry level in the United States,

[22] Note Ψ is obtained in the proof of Proposition 4 in Appendix Section B.2.5.

and conclude that IT usage has little impact on productivity. The model suggests this is due to use of aggregate data, where it is not possible to control for firms' scale. Scale matters because Proposition 1 states that Z_L rises with output. If, at the industry level, IT adoption is concentrated in expanding (contracting) firms, industry-level measured productivity growth will tend to be positive (negative), which provides a rationale for these empirical results supporting the Solow Paradox.

Using firm-level data, Bloom, Sadun, and Van Reenen (2012) study the productivity effects of IT and show that European affiliates of American firms are better managed and are more IT capital-labor intensive than are European companies. European affiliates also have higher productivity effects of IT capital, which is suggested to be due to larger organizational capital of United States parents being transplanted to their European affiliates. In fact, an index of "people management" practices accounts for most of the differential output elasticity of IT capital across these firm types. Their results are aligned with this theory's predictions that TFP, Z_L and IT capital are all pairwise positively correlated. Fox and Smeets (2011) find that controlling for labor quality reduces the TFP dispersion in the firm cross-section, consistent with a need for a richer view of how labor is dealt with when estimating firm productivity.

Bresnahan, Brynjolfsson, and Hitt (2002) use detailed firm-level data, and show evidence of complementarity among IT and skill in the workforce in factor demand and productivity regressions. They interpret their results as IT inducing an organizational redesign to achieve efficiency gains. Their results suggest a mechanism whereby, lower IT prices increase firms' IT capital-labor ratios, which in turn raises their relative demand of skilled to unskilled workers, as in the model in this paper. Incidentally, this mechanism is also confirmed through qualitative interviews with the firm's managers.

Similarly, Brynjolfsson and Hitt (2003) show that TFP, measured using standard growth accounting, is positively associated with computer investments in a panel of large United States firms. Using five different instrumental variables for computer investments confirms the mentioned OLS results. GW use a regression discontinuity design around the eligibility cut-off of 50 employees, for a policy that provided a 100% first-year tax allowance on IT investments. They find positive labor productivity effects of IT as well as positive effects on organizational change measured as an "implementation of advanced management techniques" and "implementation of major changes to your organizational structure".

IV. CALIBRATION

I calibrate the model for two steady states, 1980 and 2015, with common parameters except time-varying (1) IT capital prices and (2) the elasticity of demand in the firm cross-section. Except for the latter parameters, which require data for 2015, I use moments for 1980 or the closest available year. In the quantitative implementation, I use a grid for α and solve each firm problem given L, and compare profits across L. I find the equilibrium using a procedure similar to the inner-loop in the algorithm in Burstein and Vogel (2017); more details on the solution method can be found in Appendix Section C.1.

TABLE IV: CALIBRATED PARAMETERS

Parameter	Value	Description	Source/Target
Panel A. Calibrated Externally			
$\hat{\Delta}p_1$	0	Change in capital price at layer $l = 1$	Eden and Gaggl (2018, Figure 3B)
$\hat{\Delta}p_l$, $l > 1$	-2/3	Change in capital price at layers l $\forall l > 1$	Eden and Gaggl (2018, Figure 3B)
σ_1	0.87	Capital-labor elasticity at layer $l = 1$	Raval (2010, Table 5)
σ_l, $l > 1$	0.36	Capital-labor elasticity at layers $1 < l < L$	Raval (2010, Table 5)
η	0.13	Sector spending share	Eaton and Kortum (2002)
Panel B. Calibrated Internally			
ρ	See Table V	Demand elasticity, distribution	Labor share of VA, top firms' average, 1980&2015 Slope(log(Markup),log(TFP)), 1980&2015
β_L	0.51	Capital exponent at layer L	Slope(log(FC),log(VA))
B	5	TFP of CEO IT capital	P75/P25 FC distribution
c	0.1	Employee training cost	Labor share of cost, average
λ	1	Mean of problem CDF, $F(.)$	Slope(log(Capital-Labor Ratio),log(VA))
A	5	TFP	Employment share by firm size (7 bins)
μ_α	5	Mean of α	
ξ_α	2.4	Standard deviation of α	
w	2	Wage rate	
p_1	0.001	Production-level capital price	
p_2	25	IT capital price	

Notes. Calibrated parameters in the baseline. FC is the factor-cost ratio, that is, the ratio of capital cost to the wage bill, and VA is value-added. Slope(Y,X) indicates the regression coefficient of Y on X. Px indicates x-th percentile of a distribution. $\hat{\Delta}x \equiv \frac{x_t - x_{t-1}}{x_{t-1}}$ for a variable x between periods $t-1$ and t. Parameter values are constant across periods except ρ and p_2.

In Table IV, there are two sets of parameters according to whether they are calibrated externally or internally. I start describing the former. IT capital prices for the United States come from Eden and Gaggl (2018, Figure 3B), who use data from the BEA detailed fixed accounts; for more details on the data aggregation, see Appendix Section C.2.1. Between pre-1980 and 2015, real rental rates decline by two-thirds for IT capital and are constant for non-IT capital, values I use for the change in p_l, $\forall l > 2$ and p_1, respectively. I follow Eaton and Kortum (2002) in setting the income share of the outside sector to $1 - \eta = 0.87$, using it to compute the aggregate price index with a Cobb-Douglas function.

The calibration uses capital-labor complementarity, $\sigma_l < 1, \forall l < L$. I follow the capital-skill complementarity hypothesis as in Griliches (1969), which proposed that capital is complementary to *both* unskilled and skilled labor, though more to the latter, that is, $\sigma_1 < \sigma_2 < 1$. Goldin and Katz (1998) and Autor (2015) consider this view plausible for the post-WWII era and, in fact, Raval (2010) obtains estimates consistent with it, using data from the United States Census of Manufactures and local labor market wage variation. Hence, I use $0 < \sigma_l = \sigma_2 < 1$, for $2 \leq l < L$, and values within the interval of those reported in Raval (2010, Table 5) for production and nonproduction employees, which have a close mapping to production workers and managers, respectively.[23] These estimates are long-run elasticities which fit well with my analysis across the two steady states, 1980 and 2015.[24]

Table IV, Panel B contains ten internally calibrated parameters, (β_L, B, c, λ, A, μ_α, ξ_α, w, p_1, p_2), and the cross-section of ρ in both periods, which I obtain by searching over the parameter space using as a loss function the norm of the percentage deviation difference between fifteen moments from the model and those in the data, similar to Caliendo and Rossi-Hansberg (2012). I next describe how each parameter is informative of each simulated moment.

A higher CEO capital exponent, β_L, implies lower decreasing returns, allowing to economize on labor costs for all firms, but particularly more for larger firms, which are more restricted by a fixed CEO time. The TFP of CEO IT capital, B, is relevant for the dispersion of the labor share of cost in the firm cross-section. A higher B reduces the relevance of knowledge as it allows firms to expand through dealing with more problems instead. This reduces the relative importance of the wage bill relative to

[23]This is not surprising. It is well known that the vast majority of estimates of the capital and labor elasticity are lower than one (for surveys see Chirinko 2008; Raval 2017). As a consequence macroeconomists have studied questions related to technical change, the labor share, and growth while respecting the estimated capital-labor complementarity (see e.g., Acemoglu 2003; Antràs 2004; Lawrence 2015; Grossman et al. 2017a,b). Moreover, capital and skilled labor being complementary has been emphasized by many, with authors defining "skill" differently (e.g., Autor, Katz, and Krueger 1998; Krusell et al. 2000). My paper follows this stream of papers, and my calibration features IT capital-to-managerial-labor complementarity through $0 < \sigma_l = \sigma_2 < 1$, for $2 \leq l < L$.

[24]In robustness exercises, I have also experimented with other values for σ_l, and my quantitative results are not meaningfully affected. In particular, as I argued in Section III.C, the IT capital-to-production-labor substitution is quantitatively driven by λ, the problem distribution parameter, not σ_l.

capital spending (the inverse of the factor cost ratio) in the firm cross-section. I match β_L to the slope of a regression of the factor-cost ratio on value-added and B to the 75/25 percentile ratio of the factor-cost ratio distribution; both data moments are for the United States Census of Manufacturing in 1987 from Raval (2019).

Parameters c and λ are both related to worker knowledge. Reducing the cost per unit of knowledge, c, makes knowledge cheap, which raises the share of labor in total cost. Hence, I match it to an average labor share of cost of 0.7, which I impute. Regarding λ, the mean of the problem distribution, higher values imply more density at low z values and less at high z values. For a given c, this raises the marginal benefit of knowledge at low values of the support of $F(.)$ and lowers it for high levels, which reduces the capital-labor ratio of large firms relative to small firms. I match λ to the firm cross-sectional correlation between value-added and the capital-labor ratio for United States manufacturing in 1987 from Raval (2010). In my calibration, the ratio $\frac{c}{\lambda} = 0.1$ is similar to the 0.25 value in Caliendo and Rossi-Hansberg (2012).

The set of parameters (A, μ_α, ξ_α, w, p_1, p_2) matches the firm size distribution. First, note that average firm scale is determined by a combination of true TFP, A, and the demand shifter α, whose distribution is log-normal with parameters (μ_α, ξ_α). I turn next to factor prices. A higher w makes more complex organizations more expensive thereby limiting their size. The price of IT, p_2, similarly affects more the larger organizations, which use this capital type in more managerial layers. Finally, a lower production capital price lowers the scale of all firms but comparatively more that of the simpler organizations. I match these six parameters to the United States aggregate employment shares for seven firm size-classes reported by the Bureau of Labor Statistics (BLS) for 1993.[25]

Finally, the demand elasticity parameter, ρ, is assumed to be heterogeneous in the cross-section of firms to be consistent with the evidence of increasing markups in a simple way. I identify changes in the demand elasticity as a residual, building on DEU and ADKPV, Section IV. G. Equation 28 shows how the model provides production-side discipline to the inference about $\rho(\alpha)$, since using the definition of markup as output price over marginal cost, $m \equiv \frac{p}{MC}$, the firm-level labor share of value-added can be

[25] I calibrate factor price levels because factor markets clear at those prices, according to the general equilibrium definition in Section II.C.

expressed as

$$\underbrace{\frac{\sum_{l=1}^{L} w_l n_l}{p(q)q}}_{\text{Labor share of VA}} = \underbrace{\frac{\sum_{l=1}^{L} w_l n_l}{C(q)}}_{\text{Labor share of cost}} \underbrace{\frac{C(q)}{qMC(q)}}_{AVC/MC} \underbrace{\frac{1}{m}}_{\text{Inverse of markup}} \tag{28}$$

The expression applies to any model, since only definitions are required. For the particular case of the Cobb-Douglas production function $q = AN^{\beta}K^{1-\beta}$ with N denoting labor and K capital, the labor share of cost is β, which equals the output elasticity of labor, $\frac{\partial q}{\partial N}\frac{N}{q}$, and, assuming no fixed costs, $\frac{AVC}{MC} = 1$. This implies that movements in m are reflected one-to-one in the firm-level labor share of value-added. However, in general, not only m matters, but all three terms determine the LHS.

Production-side discipline is needed to understand the micro and macro implications of changes in the elasticity of demand and changes in IT capital prices, the two exogenous changes in parameters. Given a change in a structural parameter, the response of firms' conditional factor demands, as well as the cost function matter, since, for general production functions, they change with scale and factor prices, thereby affecting the labor share of cost. This theory has the necessary micro discipline as it is consistent with the evidence on eight cross-sectional facts (wages by layer, the IT and total capital-labor ratio, and the labor share of cost and value-added in Table I in Section III.A) and also with ten facts on firm reorganization as IT prices change (by layer: wages at layer l, the hour share of layer l in total hours, and share of the wage bill of layer l in the total wage bill; and, at the firm-level, the IT capital-labor ratio, and total knowledge in Table II in Section III.B); as seen in those sections, these responses determine changes to the labor share of cost in any theory, which is a key element in Equation 28. Moreover, how the elasticity of demand changes also needs to be consistent with sales reallocation towards large firms; I detail how firm organization matters for this result in Section V.

For each period, 1980 and 2015, I obtain the cross-section of ρ imposing it to be a linear relation of α; for each period, I need two data moments: the level of ρ is pinned down by the largest firms' labor share of value-added, whereas the slope in the α cross-section is determined by the correlation between TFP and markup across firms. A lower ρ, by raising markups, tends to lower the firm labor share of value-added; the corresponding data moment, for both 1980 and 2015, comes from ADKPV, Figure

A.18. It is calculated as the ratio of the wage bill to EBITDA (earnings before interest, tax, depreciation and amortization) for top firms in Compustat that are "nonglobally engaged", defined as firms ranked in the top 500 based on sales and with a low share of foreign sales in total sales relative to their 2 digit industry median.[26] For the second moment, the firm cross-sectional correlation between firm TFP and markup, in 1980 is obtained from a log-log regression in de Loecker and Warzynski (2012, 2463), who find a slope of 0.3 using a production function approach and Slovenian manufacturing firm-level data for the 1990s.[27] For 2015, I impute the TFP and m correlation to ten times the 1980 value in order to be consistent with: (i) a positive cross-sectional correlation between firm size and markups (ADKPV, Figure A.4), (ii) the increase in the upper tail of the markup distribution between 1980 and 2015 (DEU) and (iii) over time, an increasing correlation between size and low labor share (Kehrigy and Vincent 2018). I construct the model counterpart using the "naively measured" firm TFP, $AF(Z_L)$, and m.

Table V shows key moments of the sales-weighted distribution for the calibrated ρ and markups. The largest firm in 1980 has a calibrated demand elasticity of $\rho = 3.81$, close to 3.85 in Antràs, Fort, and Tintelnot (2017), and 3.8 in Caliendo and Rossi-Hansberg (2012), whereas, in 2015, it has a lower ρ of 2.47. Overall, the estimated ρ distributions are sensible and within the bounds of the estimates in Broda and Weinstein (2006), who report a mean of 4 and a median of 2.2. The table also compares my implied markups to those in DEU. The largest firm markups have risen by 23% between both periods, much less than 66.66% in DEU, whereas my calibrated sales-weighted median has risen slightly at about 10% and is flat at 6% when unweighted, quantitatively like ADKPV, Figure 10, Panels A and B. As in DEU, the calibration also implies that the

[26] I map it to these firms because my model omits international trade and FDI, but reassuringly, firms they define (complementarily) as "globally engaged" have very similar labor share levels (and changes over time). Moreover, (Kehrigy and Vincent 2014) report 50% to be the firm-level labor share change over time for "hyper-productive plants" in United States manufacturing, defined as those with large value-added or low labor shares, mapping well to large firms in this theory. Equation 28 shows that within-firm changes in the labor share provide crucial identification information on markup changes. As ADKPV argue, the firm-level labor share decline in large firms is the empirically relevant case in all sectors, including services, the aggregate labor share also falls in many services sectors (see for example, Figure A.8), and sales concentration is also present both in services and manufacturing sectors (see for example, Figure A.1). Also, DEU, Figures 16.1-16.2 show that aggregate markups rise with roles for both within and between firm in many industries, including manufacturing, retail trade, and finance, insurance and real state. For these reasons, I believe the model correctly captures multiple, manufacturing and nonmanufacturing, sectors of the economy.

[27] I am unaware of *quantitative* evidence of this type for the United States; see ADKPV, Figure A.4 for qualitatively similar evidence for the United States, as well as DEU.

markups for higher percentiles rise by more than for the lower percentiles.

TABLE V: EMPIRICAL AND CALIBRATED MARKUPS

	DEU			Calibration				
	Markup			Markup			Elasticity, ρ	
	1980	2014	% Change	1980	2015	% Change	1980	2015
Min				1.33	1.4	5.00	3.81	2.47
P50	1.2	1.2	0.00	1.33	1.48	10.84	3.97	3.04
P75	1.28	1.50	17.18	1.34	1.51	12.67	3.94	3.15
P90	1.5	2.5	66.66	1.35	1.54	14.44	3.95	3.26
Max				1.36	1.68	23.81	4.01	3.5

Notes. Markups from DEU and markup, $m \equiv \frac{\rho}{\rho-1}$, implied by the calibrated elasticity of demand in the model, ρ. Px denotes x-th percentile of the respective sales-weighted distribution.

I conclude the calibration discussion comparing the fit of the model and data moments. Table VI shows the source for the targeted moments and their values, as well as those implied by the model which are close. First, the average labor share of value-added at top firms in both 1980 and 2015 is close in the model and the data. This is important since, as I have argued, it is a crucial identifying moment.

TABLE VI: TARGETED MOMENTS

Moment	Model	Data	Data Source
Labor share of VA, top firms' average, 1980	0.52	0.51	ADKPV, Figure A.18
Labor share of VA, top firms' average, 2015	0.43	0.43	ADKPV, Figure A.18
Slope(log(Markup),log(TFP)), 1980	0.37	0.3	de Loecker and Warzynski (2012, 2463)
Slope(log(Markup),log(TFP)), 2015	3.76	3	Imputed
Slope(log(FC),log(VA))	0.07	0.02	Raval (2019, Table 3)
P75/P25 of factor cost distribution	1.20	2.1	Raval (2019, Table 1)
Labor share of cost, average	0.73	0.7	Imputed
Slope(log(Capital-labor ratio),log(VA))	0.07	0.15	Raval (2010, Table 9)
Employment share by firm size (7 bins)	R^2= 0.76		BLS

Notes. Data and simulated moments and their source. Top firms are defined as in the top 500 in Compustat based on the sales ranking. FC is the firm factor-cost ratio, that is, the ratio of capital cost to the wage bill, and VA is firm value-added. Slope(Y,X) indicates the regression coefficient of Y on X. Px indicates x-th percentile of a distribution. Data on levels are from 1980 and data on changes use the 1980-2015 period; whenever data on any of those years is not available, I use the most proximate years available in the cited papers.

In 1980, the correlation between markups and TFP in data and model is close, whereas it is larger than my imputed number for 2015. The correlation between factor cost and value-added is close, which is also important since heterogeneity is large and matters in a context where sales reallocate to large firms due to changes in structural parameters.[28] The model has lower factor cost dispersion, measured as the 75/25 percentile ratio, than the data. The numbers for the average labor share of cost, and the correlation between the firm capital-labor ratio and value-added in the cross-section for the model and the data are both close. Finally, the simulated employment share distribution has a large explained variance.

Quantitative Micro and Macro Implications, 1980 to 2015. Figure V shows the firm cross-section of the labor share of value-added and of cost, as a function of α both in both periods.

FIGURE V: Labor share as function of α.

As value-added increases, the labor share decreases. Given L, larger α lowers the labor share as the wage bill is spread out over more production. In the RHS of the figure, because markups increase by more for firms with larger α, their labor share of value-added declines by more between the two periods. Adding a layer causes minor

[28]The data moment implies "an increase of about 35% to 50% [in the factor cost ratio] between plants with value-added at the industry mean and the largest plants in the industry" (Raval 2019, 6).

discrete jumps in both figures:[29] the labor share of cost falls, and the labor share of value-added rises. Regarding the former, an extra layer allows the firm to reorganize knowledge and pay less to labor as a share of cost (see Section III.A); in turn, as a share of value-added, the role of AVC/MC must be factored in (Equation 28): the term rises because the fixed cost of the layer increases AVC whereas the MC decreases.

Table VII, reports the percent contribution of each term in Equation 28 relative to the total change in the labor share of value-added for each organizations' median α firm. Across organizations, most of the decline in the labor share of value-added comes from an increase in the markup, with the remaining terms having small quantitative roles. For $L = 4$ firms, the labor share of cost tends to increase as p_2 falls, but the opposite holds for $L = 2$ firms. For simpler organizations, CEO capital responds intensely, reducing these firms' labor share of cost. On the other hand, for $L = 4$ firms, the CEO layer does not have a large quantitative impact, whereas intermediate managerial layers' relative cost of labor rises and so does their layer-level labor share, raising these firms' labor share.

TABLE VII: DECOMPOSITION OF THE CHANGE IN THE MICRO LABOR SHARE

	Organization		
	$L = 2$	$L = 3$	$L = 4$
$\frac{\sum_{l=1}^{L} w_l n_l}{C(q)}$	10	-4	-2
$\frac{AVC(q)}{MC(q)}$	-2	0	0
$\frac{1}{m}$	92	104	102

Notes. Decomposition of the change in the labor share of value-added across organizations into the contributions of the three components in Equation 28. Values in percentages relative to the total change, rounded to the nearest integer for the median α given L.

Table VIII shows the untargeted moments, both at the firm level and aggregate. So far, I have not taken a stand on the specific sector I study because many sectors behave qualitatively in the same way (see ADKPV and Footnote 26). The aggregate moments in Table VIII refer to manufacturing unless otherwise stated. The model delivers moments that are close to the data for a variety of outcomes. In Table VIII,

[29] Note that along the x-axis both α and $\rho(\alpha)$ change, hence there is not a one-to-one map to the description of the labor share in the cross-section in Section III.A.

Panel A revenue concentration increases as measured by the sales share of the four largest firms, and revenue concentration correlates with the decline in the sectoral labor share in the model and data (ADKPV). The labor share in the model declines, explaining 35% of the decline in the manufacturing sector. This is a conservative result because the calibration targets the labor share of top firms in Compustat *across sectors*, which declines *only* by 16% (ADKPV). An alternative calibration that targets the labor share of "hyperproductive" manufacturing firms, with a 50% decline in their labor share (Kehrigy and Vincent 2014), produces very similar Tables V and VI, and explains 50% of the aggregate labor share decline in United States manufacturing.[30] The within-firm decline in the labor share contributes to the decline of the aggregate labor share, but sales reallocation to large firms also plays a large quantitative role for the aggregate decline, as evidenced in ADKPV.

TABLE VIII: UNTARGETED MOMENTS, MODEL AND DATA

Moment	Model	Data	Data Source
Panel A.Aggregate moments			
Revenue concentration, CR4 % change	28	10.25	ADKPV, Figure 4
Slope(log(CR20),log(Aggregate labor share))	-0.43	-0.9	ADKPV, Figure 6
Aggregate labor share, % change	-10.06	-28.69	Kehrigy and Vincent (2014, Table 1)
Aggregate markup, % change	10.99	16.13	DEU, Figure 6
Routine aggregate share of wage bill, % change	-0.21	-30.36	Eden and Gaggl (2018, Figure 4)
Nonroutine aggregate share of wage bill, % change	16.32	10.83	Eden and Gaggl (2018, Figure 4)
Panel B. Firm-level moments			
P50 Real wage, % change in firms with 100 to 1000 employees	4.86	31	Song et al. (2019, Figure VI)
P50 Real wage, % change in firms with 10,000+ employees	-2.81	-7	Song et al. (2019, Figure VI)
P75 Real wage, % change in firms with 10,000+ employees	17.29	64	Song et al. (2019, Figure VI)
Highest real wage, % change in firms with 10,000+ employees	12.02	137	Song et al. (2019, Figure VI)

Notes. Data on levels are from 1980 and data on changes use the 1980-2015 period, and alternatively, I use the most proximate years available in the cited papers. CRx denotes the revenue concentration of the x largest firms. Slope(Y,X) indicates the regression coefficient of Y on X. Px indicates average of the x-th percentile of the within-firm wage distribution across the corresponding sample of firms.

The wage bill share of managers increases and that of production workers falls. This is a consequence of the within-firm changes due to IT: wages of managers increase and so does their relative employment share, whereas the opposite happens for production workers. The change in the wage bill share by occupation type for the United States economy are from Eden and Gaggl (2018), who assign occupations to routine or non-routine labor based on Acemoglu and Autor (2011); in the model, their classification

[30]The corresponding labor share levels reported on Kehrigy and Vincent (2014) for those firms are 0.6 and 0.3 in 1980 and 2014, respectively.

corresponds to nonmanagers and managers, respectively.[31]

The aggregate markup in the model increases by 10.99% less than the benchmark 16.13% in DEU, Figure 6, for the United States Census of Manufactures, using labor as the variable input and revenue as weights, whereas it is about 66% using also intermediates as variable input (Figure 17.1). When using value-added weights it ranges between 18.32% and 68.75%, depending on the methodology used (ADKPV, Figure 10).

Table VIII, Panel B reports firm-level moments. The surprising fact that the median real wage in 10,000+ employee firms in the United States has declined by 7% during the past 40 years (Song et al. 2019) is closely matched by the model; I am not aware of any other theory rationalizing of this fact. Moreover, it is only in these largest firms that the median wage has declined, whereas the same moment for the subsample of 100 to 1000 employee firms has risen by 31% (Song et al. 2019, Figure VI, Panel A); in the model, I obtain a rise of 4.86% for the latter sample of firms. Finally, the model also predicts a rise in the real wage 75-th percentile and the largest paid employee for firms with 10,000+ employees, though smaller than that in the data. In the theory, a firm with $L = 4$ has more layers that use IT, and hence its reorganization of knowledge is quantitatively stronger than that of simpler organizations; in particular, for the same IT price decline, knowledge at the production layer shows a steeper decline for more complex organizations (see Table III, in Section III.C).

The quantitative evidence lends detailed credibility to the mechanisms in the model, in addition to the qualitative evidence on the firm cross-section and as IT prices fall (Sections III.B and III.A, respectively), providing production-side discipline to the counterfactual analysis in the next section.

V. COUNTERFACTUAL EXERCISES

In this section, I use the model to evaluate the separate role that changes to (1) IT prices and (2) the elasticity of demand have on income inequality and macroeconomic aggregates.

Table IX shows the results of the counterfactual experiments. The values on the table are relative to the baseline and in percentage terms. Table IX, Panel A shows

[31] See Table XII in Appendix Section C.2 for the mapping, which is based on occupational descriptions.

that declining IT prices increases the labor share of value-added for large firms in 2015 relative to the baseline moment, because $L = 4$ firms increase intermediate managers' labor share (Table II in Section III.B). On the contrary, lowering the elasticity of demand delivers a labor share as in the baseline. Together these show that changes in ρ for large firms are identified out of labor share changes over time. In Table IX, Panel B for the IT experiment, the within-firm labor share increases and value-added reallocation to large firms is too small, and hence the aggregate labor share does not decline; the latter is because the reduction in marginal cost does not produce a large enough output expansion. Moreover, the IT experiment cannot generate the correlation between the decline in the aggregate labor share and value-added concentration either. On the contrary, Table IX, panel B shows that a lower demand elasticity delivers sales reallocation towards large firms and an aggregate labor share decline essentially as in the baseline.[32]

TABLE IX: BASELINE VS COUNTERFACTUAL MOMENTS

Moment	IT Price	Demand Elasticity
Panel A. Targeted moments		
Labor share of VA, top firms' average, 2015	121	99
Panel B. Untargeted moments: Aggregate		
Revenue concentration, CR4 % change	14	94
Slope(log(CR20),log(Aggregate labor share))	-11	106
Aggregate labor share, % change	-2	100
Routine aggregate share of wage bill, % change	211	-177
Nonroutine aggregate share of wage bill, % change	211	-177
Panel C. Untargeted moments: Firm-level		
P50 Real wage, % change in firms with 100 to 1000 employees	-74	174
P50 Real wage, % change in firms with 10,000+ employees	212	-112
P75 Real wage, % change in firms with 10,000+ employees	43	71
Highest real wage, % change in firms with 10,000+ employees	42	49

Notes. Numbers in table are the ratio of counterfactual to baseline values in percentage. Column title refers to an experiment where only the label changes as in the baseline, with the rest of parameters in 2015 fixed at their 1980 values. Rx denotes the revenue concentration of the x largest firms. Px indicates average of the x-th percentile of the within-firm wage distribution across the corresponding sample of firms. Values are rounded to closest integer.

Table IX, Panel B shows that lower IT prices explain the fall of production work-

[32] I omit moments related to markups in the interest of space: they are driven by the demand elasticity experiment.

ers' wage due to within-firm knowledge reallocation across layers; IT capital increases, raising firms' demand for managers and, on the aggregate, their share in total labor payments increases, with the opposite happening for production workers. None of these changes can be explained by the demand elasticity, which shows the importance of IT in generating the observed inequality trends. In the demand elasticity experiment, firms expand, which implies firms demand more knowledge at all layers: production workers' wages increase very slightly, unlike in the IT experiment, and their employment grows. Managers' wages and employment increase but both by less than when IT prices change, since unlike in the IT experiment, their relative productivity has not increased. This muted reaction of both managerial factors (relative to the IT and baseline experiments) make their share in labor payments to fall, whereas that of production workers increases.

Table IX, Panel C reports firm-level moments. For firms with 100 to 1,000 employees, the decline in the IT price lowers the real median wage, while it increases in the baseline experiment. On the contrary, the lower elasticity of demand correctly predicts the sign because firms expand for which they need to increase knowledge at all layers. Together these two results show that both lower IT prices and demand elasticity are necessary to obtain enough revenue expansion to push these wages up in the baseline results. For firms with 10,000 or more employees, the IT experiment generates a much larger decline in the median real wage relative to the baseline results, whereas the demand elasticity experiment incorrectly predicts the sign. Finally, for the higher percentiles in the 10,000+ employee firms, both experiments correctly predict the signs relative to the baseline, because they both imply firms increase output, and hence knowledge.

In summary, the fourth key result of the paper is that inequality is the result of the firms reorganization due to the IT price decline; on the contrary, the aggregate labor share declines due to within-firm and value-added reallocation effects, which are consequence of lower demand elasticity. Next, I show how organizational choices and the demand elasticity interact to generate sales reallocation.

Firm Organization and Sales Reallocation. That a lower demand elasticity produces sales concentration in large firms is the opposite of what a pure Melitz (2003) model predicts, that is, a larger demand elasticity. Why this difference? A simple mathematical argument clarifies this point. I can approximate the response of revenues,

r, to changes in the demand elasticity with[33]

$$\varepsilon_{r,\rho} \equiv \frac{dr}{d\rho}\frac{\rho}{r} = \left(\frac{\rho}{\rho-1}MC\right)^{-2\rho}\left[1-(\rho-1)\ln\left(\frac{\rho}{\rho-1}MC\right)\right] \tag{29}$$

The expression can be positive or negative. When the marginal cost is sufficiently large, the expression is negative. In Melitz (2003), small firms have large marginal costs, and an increase in demand elasticity reallocates revenues away from small firms and towards large firms. On the contrary, in this production organization theory, the behavior of the marginal cost of a firm is more nuanced: it grows with q given L and jumps down when the firm increases L, because it can allocate knowledge more efficiently across layers. Quantitatively, larger firms have larger marginal costs but also a lower demand elasticity in the baseline calibration, so next I decompose the baseline percent revenue response, $\frac{\Delta r}{r}$, as

$$\frac{\Delta r}{r} = \varepsilon_{r,\rho}\frac{\Delta\rho}{\rho} \tag{30}$$

which highlights that only $\varepsilon_{r,\rho}$ depends on organization choices, through the marginal cost, MC, while the second term is simply the percent ρ change. Table X shows $\varepsilon_{r,\rho}$ for the calibrated model. First, the response is negative, meaning that all firms expand with a lower ρ, the opposite of Melitz (2003). Second, $\varepsilon_{r,\rho}$ falls with L so it is the larger absolute decline in ρ calibrated for larger firms that explains sales reallocation to them.

TABLE X: $\varepsilon_{r,\rho}$ ACROSS ORGANIZATIONS

	Organization		
	$L=2$	$L=3$	$L=4$
$\varepsilon_{r,\rho}$	-5.7	-4.4	-3.5

Notes. Percent revenue response to a percent ρ change across organizations for the median α given L in the baseline calibration.

A falling $|\varepsilon_{r,\rho}|$ with L is the result of both demand (ρ) and supply (MC), since in the baseline calibration both change with α. To determine the role of each, I use Equation 29 evaluated at the baseline MC and ρ, and compare it to Equation 29 with either MC or ρ constant in the cross-section, for which I use the smallest firm as reference and denote its marginal cost and demand elasticity by $\overline{MC}$ and $\overline{\rho}$, respectively. In the first row of Table XI, I report Equation 29 evaluated at $\left(\rho(\alpha), \overline{MC}\right)$ relative to

[33] It is an approximation because I abstract from equilibrium effects, and I focus on locally constant marginal costs.

the equation baseline value, that is, with $(\rho(\alpha), MC(\alpha))$. This shows how the cross-section of marginal cost affects $\varepsilon_{r,\rho}$. Conversely, in the second row of the table, I report Equation 29 evaluated at $(MC(\alpha), \overline{\rho})$ relative to its baseline value. This shows how heterogeneous ρ in the α cross-section affects $\varepsilon_{r,\rho}$.

Table XI reports the results for the median α given L, for all organizations. Focus on the $L = 4$, the last column. When its marginal cost is $\overline{MC}$, so lower than in the baseline, $\varepsilon_{r,\rho}$ for this firm is almost 300% larger; the increase in the marginal cost with value-added in the cross-section dampens sales growth in the baseline, and such implication of the production theory is quantitatively important for sales reallocation. In the second row, when the elasticity of demand is $\overline{\rho}$, that is, higher than in the baseline, $\varepsilon_{r,\rho}$ is almost 60% lower for the median $L = 4$ firm: the firm has a more elastic demand than in the baseline, which dampens it's sales growth. The numbers for other L in Table XI tell a similar story: the cross-sectional implications of the theory for the marginal cost have a large quantitative effect on sales reallocation.

TABLE XI: MARGINAL COST VERSUS DEMAND ELASTICITY IN EQUATION 29

	Organization		
	$L = 2$	$L = 3$	$L = 4$
$\frac{\varepsilon_{r,\rho}\left(\overline{MC},\rho\right)}{\varepsilon_{r,\rho}(MC,\rho)}$	275	297	292
$\frac{\varepsilon_{r,\rho}(MC,\overline{\rho})}{\varepsilon_{r,\rho}(MC,\rho)}$	78	61	57

Notes. The denominator in each row is Equation 29 using the cross-sectional values of MC and ρ as implied by the baseline calibration, that is, where the former rises with α whereas the latter falls with α. The numerator in the top row evaluates Equation 29 by fixing the marginal cost at $\overline{MC}$ and allowing ρ to change with α as in the baseline. The bottom row's numerator evaluates Equation 29 using the baseline $MC(\alpha)$ but fixing the demand elasticity at $\overline{\rho}$. The reported values are in percentages and for the median α given L.

Theoretically, multiple mechanisms can deliver reallocation of sales to larger firms. In this theory, demand and marginal cost functions interact, and a lower elasticity produces reallocation; this is consistent with a number of other micro and macro facts.

VI. CONCLUSION

During the past 40 years, there has been a rapid adoption of IT capital. This new technology has raised several questions: What are the implications of IT for the wages of unskilled workers? Why are IT and unskilled workers substitutes? How does IT affect firm productivity? And, what are the macroeconomic implications of this general purpose technology? To answer these questions, I develop a firm theory of production organization, where IT capital complements managers, consistent with multiple micro and macro facts. In the theory, firms, organized as hierarchies, choose capital, labor, and worker knowledge, at every layer, as well as a discrete number of layers.

The theory builds on the knowledge-based hierarchy literature by Garicano (2000), Garicano and Rossi-Hansberg (2006), and, specifically, adapts Caliendo and Rossi-Hansberg (2012). I study optimal physical capital adoption, which is absent in these previous theories, with the following benefits. First, I characterize the firm's conditional factor demands. Second, I show how within-firm factor demands and their relative spending respond to a decline in the price of IT capital.

Third, I impose capital-labor complementarity at every layer though more for managers than for production workers Griliches (1969), a view Goldin and Katz (1998) consider plausible for the post-WWII era. Despite this, I obtain that IT capital and production workers are indirect substitutes, due to *endogenous* firm reorganization when IT prices fall: firms optimally reallocate problem solving to the, now cheaper, managerial layers, and away from the production layer. Importantly, all these theoretical predictions are consistent with the evidence in the literature.

On the demand side, firms are heterogeneous in the size of their product appeal and compete monopolistically, charging a price that is a markup over their marginal cost. Because this theory provides empirically grounded conditional factor demands and cost functions, I can perform credible counterfactuals to determine the role that changes in firms' demand elasticity and IT prices have had in recent macroeconomic trends.

I find lower IT prices are responsible for within-firm (1) wage declines at the production layers and increase at managerial layers, (2) employment share of managers increase, (3) IT capital-labor ratio increases, and, at the aggregate level (4) an increase in the managerial share of labor spending, and (5) a decrease in that of production workers' share. On the other hand, (6) the decline in the labor share of value-added of large firms, (7) the decline in the aggregate labor share, and (8) reallocation of sales to larger firms are due to a decline in the elasticity of demand. For the latter two results, the marginal cost increase with scale substantially dampens sales reallocation to the large firms, highlighting the interaction between demand and production organization.

Finally, the theory has implications for the measurement of the TFP effects of IT capital. First, firm-level data is needed to control for scale adjustment when estimating factor's output elasticities. Second, the Solow residual is affected by the optimal knowledge responses to changing IT prices, demand size and true TFP changes. As a consequence, estimating true TFP when IT capital is adopted requires controlling for worker knowledge, organization, and firm scale choices of firms, which opens an avenue for future work.

Universitat Rovira i Virgili

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67(2): 251–333. 23

Acemoglu, Daron. 2003. "Labor- And Capital-Augmenting Technical Change." *Journal of the European Economic Association* 1(1): 1–37. 41

Acemoglu, Daron, and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of labor economics*, eds. David Card and Orley Ashenfelter, vol. 4, chap. 12, 1043–1171. Amsterdam: Elsevier-North. 32, 48, 98

Acemoglu, Daron, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. 2014. "Return of the Solow Paradox? IT, Productivity, and Employment in US Manufacturing." *American Economic Review Papers and Proceedings* 104(5): 394–99. 38

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race Between Man and Machine: Implications of Technology for Growth, Factor Shares and Employment." *American Economic Review* 108(6): 1488–1542. 7, 37

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li. 2019. "A Theory of Falling Growth and Rising Rents." *NBER Working Paper No. 26448*. 7

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. 2015. "The Skill Complementarity Of Broadband Internet." *Quarterly Journal of Economics* 130(4): 1781–1824. 6, 31, 33

Almunia, Miguel, Pol Antràs, David Lopez-Rodriguez, and Eduardo Morales. 2018. "Venting Out: Exports during a Domestic Slump." *NBER Working Paper No. 25372* December. 24

Antràs, Pol. 2004. "Is the U.S. Aggregate Production Function Cobb-Douglas? New Estimates of the Elasticity of Substitution." *The B.E. Journal of Macroeconomics* 4(1): 1–36. 41

Antràs, Pol, Teresa C. Fort, and Felix Tintelnot. 2017. "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms." *American Economic Review* 107(9): 2514–64. 18, 44

Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg. 2006. "Offshoring in a Knowledge Economy." *Quarterly Journal of Economics* 121(1): 31–77. 5

Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg. 2008. "Organizing Offshoring: Middle Managers and Communication Costs." In *The Organization of Firms in a Global Economy.*, eds. Elhanan Helpman, Dalia Marin, and Thierry Verdier, chap. 10, 311–39. Cambridge, MA: Harvard University Press. 5

Antràs, Pol, and Esteban Rossi-Hansberg. 2009. "Organizations and Trade." *Annual Review of Economics* 1: 43–64. 7

Autor, David. 2015. "Market Structure, Organizational Structure and Wage Structure." Lecture notes 14.662, MIT. 27, 35, 41

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. "The Fall of the Labor Share and the Rise of Superstar Firms." *NBER Working Paper No. 23396* May. 24

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135(2): 645–709. 5, 7, 19, 24, 42, 43, 44, 47, 48, 49

Autor, David, Lawrence F. Katz, and Alan B. Krueger. 1998. "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 113(4): 1169–1214. 6, 33, 41

Autor, David H. 2010. "The Polarization of Job Opportunities in the U.S. Labor Market: Implications for Employment and Earnings." *Center for American Progress and The Hamilton Project* April. 2

Autor, David H. 2014. "Skills, Education, and the Rise of Earnings Inequality Among the "Other 99 Percent"." *Science* 344(6186): 843–851. 2, 27

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118(4): 1279–1333. 6, 32, 33

Baily, Martin N., Charles Hulten, and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity* 187–267. 21

Blackorby, Charles, and R. Robert Russell. 1981. "The Morishima Elasticity of Substitution; Symmetry, Constancy, Separability, and its Relationship to the Hicks and Allen Elasticities." *The Review of Economic Studies* 48(1): 147–158. 94

Blackorby, Charles, and R. Robert Russell. 1989. "Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of the Allen/Uzawa and Morishima Elasticities)." *The American Economic Review* 79(4): 882–888. 34, 94

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. 2014. "The Distinct Effects of Information Technology and Communication Technology on Firm Organization." *Management Science* 60(12): 2859–2885. 6, 32, 33

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "Americans Do IT Better: US Multinationals and the Productivity Miracle." *American Economic Review* 102(1): 167–201. 6, 39

BLS. 1998. "Competition Drives The Trucking Industry." *Monthly Labor Review, Bureau of Labor Statistics* April. 64

Bresnahan, Timothy F., Eric Brynjolfsson, and Lorin M. Hitt. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics* 117(1): 339–376. 6, 32, 33, 39

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains From Variety." *The Quarterly Journ al of Economics* 121(2): 541–585. 44

Brown, Charles, and James Medoff. 1989. "The Employer Size-Wage Effect." *Journal of Political Economy* 97(5): 1027–1059. 23

Brynjolfsson, Eric, and Lorin M. Hitt. 2003. "Computing Productivity: Firm-Level Evidence." *Review of Economics and Statistics* 85(4): 793–808. 39

Brynjolfsson, Erik, and Eric McAfee. 2011. *Race Agains the Machine.* Lexington, Massachusetts: Digital Frontier Press. 65

Burstein, Ariel, and Jonathan Vogel. 2017. "International Trade, Technology, and the Skill Premium." *Journal of Political Economy* 125(5): 1356–1412. 40, 94

Caliendo, Lorenzo, Giordano Mion, Luca David Opromolla, and Esteban Rossi-Hansberg. 2015. "Productivity and Organization in Portuguese Firms." *NBER Working Paper No. 21811*. 6, 24

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg. 2015. "The Anatomy of French Production Hierarchies." *Journal of Political Economy* 123(4): 809–852. 6, 9, 11, 23, 96, 98

Caliendo, Lorenzo, and Esteban Rossi-Hansberg. 2012. "The Impact of Trade on Organization and Productivity." *Quarterly Journal of Economics* 127(3): 1393–1467. 3, 5, 6, 9, 10, 11, 16, 17, 21, 23, 24, 41, 42, 44, 54, 66, 67, 71

Caroli, Eve, and John Van Reenen. 2001. "Skill Biased Organizational Change? Evidence from a Panel of British and French Establishments." *Quarterly Journal of Economics* 116(4): 1449–1492. 6, 33

Chirinko, Robert S. 2008. "σ: The Long and Short of It." *Journal of Macroeconomics* 30(2): 671–686. 41

de Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. "The Rise of Market Power and the Macroeconomic Implications." *The Quarterly Journal of Economics*. 5, 8, 42, 44, 45, 49

de Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review* 102(6): 2437–2471. 44

Eaton, Jonathan, and Sam Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70(5): 1741–1779. 40

Eden, Maya, and Paul Gaggl. 2018. "On the welfare implications of automation." *Review of Economic Dynamics* 29: 15–43. 2, 7, 37, 40, 48, 95

Eeckhout, Jan, and Philipp Kircher. 2018. "Assortative Matching With Large Firms." *Econometrica* 86(1): 85–132. 6

Fox, Jeremy T., and Valérie Smeets. 2011. "Does Input Quality Drive Measured Differences In Firm Productivity?" *International Economic Review* 52(4): 961–989. 39

Gaggl, Paul, and Greg C. Wright. 2017. "A Short-Run View of What Computers Do: Evidence from a UK Tax Incentive." *American Economic Journal: Applied Economics* 9(3): 262–94. 7, 27, 28, 31, 32, 33, 39, 98, 99

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108(5): 874–904. 3, 5, 11, 54

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121(4): 1383–1435. 3, 5, 11, 54

Goldin, Claudia, and Lawrence F. Katz. 1998. "The Origins of Technology-Skill Complementarity." *Quarterly Journal of Economics* 113(3): 693–732. 6, 7, 11, 34, 35, 41, 54

Griliches, Zvi. 1969. "Capital-Skill Complementarity." *The Review of Economics and Statistics* 51(4): 465–68. 4, 11, 35, 41, 54

Grossman, Gene M., Elhanan Helpman, Ezra Oberfield, and Thomas Sampson. 2017a. "Balanced Growth Despite Uzawa." *American Economic Review* 107(4): 1293–1312. 41

Grossman, Gene M., Elhanan Helpman, Ezra Oberfield, and Thomas Sampson. 2017b. "The Productivity Slowdown and the Declining Labor Share: A Neoclassical Exploration." *CEPR Discussion Paper 12342*. 41

Guadalupe, Maria, Hongyi Li, and Julie Wulf. 2014. "Who Lives in the C-Suite? Organizational Structure and the Division of Labor in Top Management." *Management Science* 6(4): 824–844. 11, 33

Gumpert, Anna. 2018. "The organization of knowledge in multinational firms." *Journal of the European Economic Association* 16(6): 1929–1976. 6

Gumpert, Anna, Henrike Steimer, and Manfred Antoni. 2018. "Firm Organization with Multiple Establishments." *CESifo Working Paper Series 7435*. 6

Hamermesh, Daniel. 1993. *Labor Demand*. Princeton, New Jersey: Princeton University Press. 35

Harford, Tim. 2017. "50 things that made the modern economy: Robots." *BBC News* 8 March. 66

Hemous, David, and Morten Olsen. 2016. "The Rise of the Machines: Automation, Horizontal Innovation and Income Inequality." *Mimeo U. Zurich*. 7, 37

Hicks, John R. 1932. *Theory of Wages*. London, Macmillan and Co., Ltd. 25, 34, 93

Hsieh, Chang-Tai, and Esteban Rossi-Hansberg. 2019. "The Industrial Revolution in Services." *NBER Working Paper No. 25968*. 7

IMF. 2017. "Understanding the downward trend in labor income shares." *World Economic Outlook* April. 2

Karabarbounis, L., and B. Neiman. 2013. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129(1): 61–103. 2, 5, 35

Kehrigy, Matthias, and Nicolas Vincent. 2014. "Growing Productivity without Growing Wages: The Micro-Level Anatomy of the Aggregate Labor Share Decline." *Economic Research Initiatives at Duke (ERID) Working Paper No. 244*. 44, 48

Kehrigy, Matthias, and Nicolas Vincent. 2018. "The Micro-Level Anatomy of the Labor Share Decline." *NBER Working Paper No. 25275*. 7, 44

Knight, Will. 2015. "Inside Amazon's Warehouse, Human-Robot Symbiosis." *MIT Technology Review* July 7. 66

Koh, Dongya, Raül Santaeulàlia-Llopis, and Yu Zheng. 2018. "Labor Share Decline and Intellectual Property Products Capital." *BGSE Working Paper 927*. 7, 95

Krueger, Alan B. 1993. "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984–1989." *Quarterly Journal of Economics* 108(1): 33–60. 31

Krusell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull, and Giovanni L. Violante. 2000. "Capital-skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica* 68(5): 1029–1053. 27, 37, 41

Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard. 2018. "Information Technology and Returns to Scale." *Mimeo Boston College*. 7, 23

Lashkari, Danial, and Martí Mestieri. 2017. "Gains from Trade with Heterogeneous Income and Price Elasticities." *Mimeo Harvard*. 8

Lawrence, Robert Z. 2015. "Recent Declines in Labor's Share in US Income: A Preliminary Neoclassical Account." *NBER Working Paper No. 21296*. 41

Lucas, Robert E. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics* 9(2): 508–523. 3

Melitz, Marc J. 2003. "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71(6): 1695–1725. 51, 52

Morishima, Michio. 1967. "A Few Suggestions On The Theory Of Elasticity." *Keizai Hyoron (Economic Review)* 16: 144–150. 34

Oberfield, Ezra, and Devesh Raval. 2014. "Micro Data and Macro Technology." *NBER Working Paper No. 20452*. 7

Premack, Rachel. 2018. "Truck driver salaries have fallen by as much as 50 percent." *Business Insider* 26 Sep. 65

Raval, Devesh. 2017. "What's Wrong with Capital in the 21st Century's Model?" In *After Piketty: The Inequality Agenda post-Capital in the 21st Century*, eds. Heather Boushey, J. Bradford DeLong, and Marshall Steinbaum, chap. 4. Cambridge, MA: Harvard University Press. 41

Raval, Devesh R. 2010. "Beyond Cobb-Douglas: Estimation of a CES Production Function with Factor Augmenting Technology." *Mimeo The University of Chicago* November. 23, 41, 42

Raval, Devesh R. 2019. "The micro elasticity of substitution and non-neutral technology." *The RAND Journal of Economics* 50(1): 147–167. 23, 24, 42, 46

Robinson, Joan V. 1933. *The economics of imperfect competition.* London, Macmillan and Co., Ltd. 94

Santamaria, Clara. 2018. "Small Teams in Big Cities: Inequality, City Size, and the Organization of Production." *Mimeo Princeton University*. 6

Sforza, Alessandro. 2017. "Shocks and the organization of the firm: Who pays the bill?" *Mimeo LSE*. 6

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics* 134(1): 1–50. 7, 32, 33, 34, 49

Syverson, Chad. 2017. "Challenges to Mismeasurement Explanations for the US Productivity Slowdown." *Journal of Economic Perspectives* 31(2): 165–186. 2

Wilson, Daniel J. 2006. "Investment Behavior of U.S. Firms over Heterogeneous Capital Goods: A Snapshot." *Federal Reserve Bank of San Francisco Working Paper 2004-21* March. 2

APPENDIX

A. Case Studies: Firm Organization, Information Technology and Wages

As described in the main text, an Amazon warehouse is a good illustration of how firms reorganize when adopting IT. This theory contributes to the knowledge-based hierarchy literature by endogenizing the adoption of capital, in particular, IT capital, as a complementary way to leverage managerial time, since this type of capital is, arguably more than anything else, a problem solving tool. Hence, in the model, the adoption of IT capital allows managers to solve more problems, including some that were previously done by employees in the lower layers, which lowers the knowledge and wages of low-layer employees. Reorganizing in this way minimizes cost, since it lowers compensation to the numerous workers at the production level, as they now require less skill for their job.

This reorganization mechanism is ubiquitous in the real world. Trucking jobs have seen similar trends. As early as 1998, a BLS analysis of the sector studied the many new technologies that had been adopted (BLS 1998). Examples listed were electronic data interchange, new vehicle location detection systems, and voice and data communication services. The leading provider of such technology in the transportation sector is Qualcomm, which has been marketing state-of-the-art satellite-based mobile communication systems, as well as decision support tools since 1985. One particularly widespread technology is its "OmniTRACS" system. It involves an "in cab" communication device, the Automatic Vehicle Location (AVL) unit, which allows the driver to communicate with his dispatcher, who usually informs the driver of their pick-up and drop-off locations. If the AVL unit is connected to a mobile data terminal or a computer it also allows the driver to input the information from a bill of lading into a simple dot matrix display screen. The driver inputs the information, using a keyboard, into an automated system of preformatted messages known as macros. There are macros for each stage of the loading and unloading process, such as "loaded and leaving shipper" and "arrived at the final destination". Moreover, the system also enables companies to monitor extremely detailed statistics, such as vehicle location, mileage traveled on a specific vehicle, direction, fuel efficiency, speed, gear optimization, and the best fueling locations. As in the Amazon example, the extent to which autonomy has been removed from these workers' set of problems is extreme and, consequently, wages have been lowered in this occupation. In fact, BLS data shows that median wages for truck drivers have decreased 21% on average since 1980 (Premack 2018).

Customer services are another case in point. As documented by Brynjolfsson and McAfee (2011), in 2011, the translation services company Lionbridge announced the launch of Geofluent, a new language translation technology developed in partnership with IBM. Geofluent is based on statistical machine translation software and is used by large high-tech companies for conversations with clients and other parties. It develops a memory that is the specific to the translation content, making it particularly accurate and fast. The technology takes words written in one language from, say, an online chat message from a customer seeking help with a problem, and translates them accurately and immediately into another language, such as that of the customer service representative in another country. Customer representatives in firms with this technology are definitely likely to earn less, at least relative to a similar job in which language skills are necessary.

Skilled jobs are not immune to this deskilling trend. In a 2011 *New York Times* article titled "Armies of Expensive Lawyers, Replaced by Cheaper Software" John Markoff explains how computers' pattern recognition abilities are being exploited by the legal industry. Preparing for litigation in big cases requires the evaluation of large numbers of documents, the cost of which can be immense, as the required legal staff hours are potentially very large. Thanks to advances in artificial intelligence, a new "e-discovery" software can analyze documents in a fraction of the time for a fraction of the cost. For example, the company Blackstone Discovery of Palo Alto, California, analyzed 1.5 million documents for less than $100,000 (Brynjolfsson and McAfee 2011). The type of software they used goes beyond finding documents with relevant terms because it can extract the relevant concepts, even in the absence of specific search terms, or infer patterns that would have eluded lawyers examining millions of documents. Naturally, the consequence of this reorganization of law firms is a substitution of entry-level lawyers for clerical workers who scan and organize documents for a fraction of the knowledge and wages of the former employees. Even without occupational substitution, it is hard to imagine this technology not affecting negatively the wages of entry-level lawyers.

These examples illustrate that cost-minimizing firms have an incentive to adopt IT capital and deskill the job content in layers with numerous workers. The model in the main text formalizes this mechanism.[34]

B. Mathematical Appendix

B.1. Assumptions

The following assumptions are maintained throughout.

Assumption 1. *Parameters* $(\lambda,\ c,\ \sigma_l)$ *are such that* $c/\lambda < \frac{1}{\sigma_1}$, $\forall l$.

Assumption 2. *Parameters* $(p_1, p_2, \sigma_1, \sigma_2, w, c, \lambda)$ *are such that*

(i) $\left(p_1^{1-\sigma_1} + w^{1-\sigma_1}\right)^{1/(1-\sigma_1)} + \frac{wc}{\lambda}\left(1 + \frac{p_1^{1-\sigma_1}}{w^{1-\sigma_1}}\right)^{\sigma_1/(1-\sigma_1)} > \frac{wc}{\lambda}$

(ii) $\forall l > 1$: $\frac{\lambda}{wc}\left(p_l^{1-\sigma_l} + w^{1-\sigma_l}\right)^{1/(1-\sigma_l)} + \left(1 + \frac{p_l^{1-\sigma_l}}{w^{1-\sigma_l}}\right)^{\sigma_l/(1-\sigma_l)} > \left(1 + \frac{p_{l-1}^{1-\sigma_{l-1}}}{w^{1-\sigma_{l-1}}}\right)^{\sigma_{l-1}/(1-\sigma_{l-1})}$

Assumption 3. *Parameters* (β_L, p_L, w, c, B) *are such that*

(i) $\frac{1}{2} < \beta_L < 1$, *and*

(ii) $2^{2\beta_L}\left(\frac{\lambda p_L(1-\beta_L)}{wc}\right)^{\beta_L}(2\beta_L - 1)^{1-2\beta_L} > B.$

Next, I provide the intuition of the role of assumptions. Under the conditions on parameters in Assumptions 1 and 2, I show that $k_L > 0$ and that $z_l > 0$, $\forall 1 < l \leq L$ and $L > 2$, in a form analogous to Caliendo and Rossi-Hansberg (2012).

First, to have $k_L > 0$ for any production scale, it is enough that the cost of learning, c/λ, is low enough relative to the inverse of the capital-labor elasticity at layer 1, $c/\lambda - \frac{1}{\sigma_1} < 0$. This condition is reminiscent of that in Caliendo and Rossi-Hansberg (2012), where the cost of learning knowledge relative to the communication cost is required to be low, specifically, they require that $c/\lambda - \frac{h}{1-h} \leq 0$.

Solutions where $z_1 > 0$ but $z_L = 0$ are ruled out by a parameter restriction where the cost of the CEO is lower than the cost of layer 1 when $z_1 = 0$

$$\left(p_1^{1-\sigma_1} + w^{1-\sigma_1}\right)^{1/(1-\sigma_1)} + \frac{wc}{\lambda}\left(1 + \frac{p_1^{1-\sigma_1}}{w^{1-\sigma_1}}\right)^{\sigma_1/(1-\sigma_1)} > \frac{wc}{\lambda}$$

[34] The Amazon example is based on Harford (2017) and Knight (2015). For the rest, I use the materials cited at my own discretion.

The first element on the LHS is the unit cost of layer 1 when $z_1 = 0$, a sort of "fixed" cost of using that layer. The second term on the left is the marginal cost of knowledge in layer 1 at $z_1 = 0$, adjusted for the density of problems at that z_1. Intuitively, this second term captures how fast the costs of knowledge changes relative to its return: when there are many problems near $z_1 = 0$ (i.e., a high λ) the second term matters less because a large mass of problems exist around $z_1 = 0$. The term on the RHS is the CEO learning cost of knowledge adjusted for the mass of problems at zero knowledge. When the cost of an extra unit of layer 1 is large enough relative to CEO knowledge, zero CEO knowledge is not optimal. For $L > 2$, the second condition in Assumption 2 is sufficient to make knowledge at layer $L - 1$ optimally zero at a higher output level than for any layer below $L - 1$; this is the subject of Lemma 1), which establishes that an L organization has positive knowledge at the lower $L - 2$ layers. Note also that, intuitively, setting knowledge at layer $L - 1$ to zero can never be optimal, since such a layer does not produce anything but it is costly (see Proposition 5). Together, these results determine positive knowledge at all layers for $L > 2$. In principle, $z_1 = 0$ is possible for $L = 2$, but the next condition rules out that case for large enough output, similar to Caliendo and Rossi-Hansberg (2012).

Assumption 1, $c/\lambda < \frac{1}{\sigma_1}$, is also enough to guarantee that z_l increases with scale $\forall l < L$, though a weaker assumption would suffice. Finally, to obtain the natural result that k_L, and z_L are increasing with scale, the two parametric conditions in Assumption 3 are enough. The first involves not-too-strong decreasing returns to scale on k_L (i.e., $1/2 < \beta_L < 1$). The second requires that B, the TFP of the CEO IT, is not too large,

$$2^{2\beta_L} \left(\frac{\lambda p_L (1 - \beta_L)}{wc} \right)^{\beta_L} (2\beta_L - 1)^{1-2\beta_L} > B$$

B.2. Proofs

The cost minimization problem of a firm given L, denoted CMP, is

$$\min_{\{n_l, z_l, k_l\}_{l=1}^{L}} \sum_{l=1}^{L-1} \left(n_l w (c z_l + 1) + p_l k_l \right) + w_L + p_L k_L$$

Subject to constraints:

$$\begin{cases} A[1-\exp(-\lambda Z_L)]\left(k_1^{\frac{\sigma_1-1}{\sigma_1}}+n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}}-q=0 & :\phi \\ \left(k_l^{\frac{\sigma_1-1}{\sigma_1}}+n_l^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}}=\left(k_1^{\frac{\sigma_1-1}{\sigma_1}}+n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}}\exp(-\lambda Z_{l-1}), L>l>1 & :\psi_l \\ 1+Bk_L^{\beta_L}=\left(k_1^{\frac{\sigma_1-1}{\sigma_1}}+n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}}\exp(-\lambda Z_{L-1}) & :\psi_L \\ z_l>0,\forall l & :\theta_l \\ k_L>0 & :\chi \end{cases}$$

Choice variables are $\{n_l, z_l, k_l\}_{l=1}^{L}$. All the rightmost terms in the constraints are multipliers. It will become apparent that $\{n_l, k_l\}_{l=1}^{L-1}$ are all positive due to the standard argument that the marginal product of each factor is infinite whereas its price is a bounded value. Showing that $z_l > 0$, $\forall l$, is addressed in Proposition 5. Note that ϕ because it is attached to the produced quantity q is interpreted as marginal cost given L, denoted MC_L. The first-order conditions with respect to labor and capital at $l = 1$ are, respectively,

$$\begin{aligned} &w(cz_1+1)-\phi A[1-\exp(-\lambda Z_L)]\left(k_1^{\frac{\sigma_1-1}{\sigma_1}}+n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}-1}n_1^{-1/\sigma_1}+\cdots \\ &+\sum_{l=2}^{L}\psi_l\left(k_1^{\frac{\sigma_1-1}{\sigma_1}}+n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}-1}n_1^{-1/\sigma_1}\exp(-\lambda Z_{l-1})=0 \end{aligned} \tag{31}$$

and

$$\begin{aligned} &p_1-\phi A[1-\exp(-\lambda Z_L)]\left(k_1^{\frac{\sigma_1-1}{\sigma_1}}+n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}-1}k_1^{-1/\sigma_1}+\cdots \\ &\sum_{l=2}^{L}\psi_l\left(k_1^{\frac{\sigma_1-1}{\sigma_1}}+n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}-1}k_1^{-1/\sigma_1}\exp(-\lambda Z_{l-1})=0 \end{aligned} \tag{32}$$

Proceeding similarly for $l \neq L$, and combining both, one obtains

$$\frac{k_l}{n_l}=\left(\frac{p_l}{w_l}\right)^{-\sigma_l} \tag{33}$$

Note that using the optimal capital-labor ratio at layer 1 together with layer L constraint, $\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) = \left(k_1^{\frac{\sigma_1-1}{\sigma_1}} + n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}}$, I obtain the demand of labor, not as function of q but of k_L,

$$n_1 = P_1^{\sigma_1} w_1^{-\sigma_1}\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) \tag{34}$$

and hence

$$k_1 = P_1^{\sigma_1} p_1^{-\sigma_1}\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) \tag{35}$$

Combining both Equations 34 and 35, total cost of layer 1 inherits the CES structure:

$$\begin{aligned} w_1 n_1 + p_1 k_1 &= w_1 P_1^{\sigma_1} w_1^{-\sigma_1}\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) + p_1 P_1^{\sigma_1} p_1^{-\sigma_1}\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) = \\ &P_1\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) = P_1 Q_1 \end{aligned} \tag{36}$$

One can similarly show that, for any $1 < l < L$, Equation 33 along with layer l constraint delivers

$$n_l = P_l^{\sigma_l} w_l^{-\sigma_l}\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda \sum_{k=l}^{L-1} z_k)$$

and

$$k_l = P_l^{\sigma_l} p_l^{-\sigma_l}\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda \sum_{k=l}^{L-1} z_k)$$

so that,

$$w_l n_l + p_l k_l = P_l Q_l \tag{37}$$

with $Q_l \equiv \left(1 + Bk_L^{\beta_L}\right)\exp(\lambda \sum_{k=l}^{L-1} z_k)$ and $P_l \equiv \left(p_l^{1-\sigma_l} + (w(cz_l + 1))^{1-\sigma_l}\right)^{1/(1-\sigma_l)}$.

The above steps imply that I can rewrite CMP in a condensed form as

$$\begin{aligned} \mathcal{L}_L = &\sum_{l=1}^{L-1} P_l Q_l + w_L + p_L k_L - \phi\left(A[1 - \exp(-\lambda Z_L)]\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) - q\right) + \cdots \\ &- \sum_{l=1}^{L} \theta_l z_l - \chi k_L \end{aligned} \tag{38}$$

Note that Equation 38 is already in Lagrangian form with choice variables z_l, $\forall l$, and k_L. This form is different from the representation in the main text, where constraints are functions of output q, and is useful when proving Propositions 1 and 2 in the main text.

For notational simplicity, define $M_l \equiv \frac{P_l}{w_l}$, with $w_l \equiv w(cz_l + 1)$. After some algebra, the first-order conditions of $\mathcal{L}_L$ are, for k_L,

$$\frac{1}{\left(1 + Bk_L^{\beta_L}\right)} + \frac{\lambda}{wc}\frac{p_L}{B\beta_L k_L^{\beta_L - 1}} = \exp(\lambda z_{L-1}) M_{L-1}^{\sigma_{L-1}} \tag{39}$$

for z_L,

$$\exp(-\lambda z_L) = \frac{wc}{\phi A \lambda \left(1 + Bk_L^{\beta_L}\right)} \tag{40}$$

for z_1,

$$\frac{\lambda}{wc} P_1 + M_1^{\sigma_1} = \frac{\lambda}{wc}\phi A \tag{41}$$

and for z_l, $1 < l < L$,

$$\frac{\lambda}{wc} P_l + M_l^{\sigma_l} = \exp(\lambda z_{l-1}) M_{l-1}^{\sigma_{l-1}} \tag{42}$$

In this system, I have set all nonnegativity multipliers to zero, but later I obtain sufficient parameter conditions.

Using Assumptions 1 to 3, I am now ready to proof two intermediate results, Lemma 1 and Proposition 5, which I use in the proof of Propositions 1 and 2.

B.2.1 Auxiliary Results

Lemma 1. *Given the number of layers L, under Assumptions 1 and 2, as the multiplier on the output constraint, ϕ, decreases, the constraints over $z_l \geq 0, \forall l$ bind such that the first to hit zero is z_{L-1}, then z_{L-2}, and so on until z_1 and finally z_L. Moreover, $k_L > 0$ always holds.*

Proof. First, I show that $k_L > 0$ always holds. On the k_L first-order condition (Equation 39), if the term

$$\exp(\lambda z_{L-1}) M_{L-1}^{\sigma_{L-1}} - \frac{1}{\left(1 + Bk_L^{\beta_L}\right)} \tag{43}$$

is positive, then $k_L > 0$. Note that the equation increases as k_L increases. Hence, if $\exp(\lambda z_{L-1})M_{L-1}^{\sigma_{L-1}} - 1$ is positive, then Equation 43 is positive. I next focus on the latter simpler term to obtain a sufficient parameter condition for $k_L > 0$.

Two intermediate expressions that are used in several steps below are $\frac{dP_l}{dz_l} = wcM_l^{\sigma_l}$, and $\frac{dM_l}{dz_l} = \frac{wc}{w_l}(M_l^{\sigma_l} - M_l) = -\frac{wcp_l^{1-\sigma_l}}{w_l^2}P_l^{\sigma_l}$. Take derivatives with respect to z_{L-1} on Equation 43:

$$\frac{d\left(\exp(\lambda z_{L-1})M_{L-1}^{\sigma_{L-1}} - 1\right)}{dz_{L-1}} = \lambda \exp(\lambda z_{L-1})M_{L-1}^{\sigma_{L-1}}\left(1 - \frac{\sigma_{L-1}wc}{\lambda w_{L-1}}\left(\frac{p_{L-1}}{P_{L-1}}\right)^{1-\sigma_{L-1}}\right) \quad (44)$$

which is positive as long as $\lambda - \sigma_{L-1}\frac{wc}{wcz_{L-1}+w} > 0$; hence, it is enough that $\lambda - \sigma_{L-1}c > 0$ (Assumptions 1). In that case, Equation 43 increases in z_{L-1}, and as long as at its minimum

$$\exp(\lambda z_{L-1})P_{L-1}^{\sigma_{L-1}}w_{L-1}^{-\sigma_{L-1}}\Big|_{z_{L-1}=0} - 1$$

is positive, or equivalently $\left(w^{1-\sigma_{L-1}} + p^{1-\sigma_{L-1}}\right)^{\sigma_{L-1}/(1-\sigma_{L-1})} w^{-\sigma_{L-1}} > 1$, a condition that trivially holds, then Equation 43 is positive and $k_L > 0$.

Focus now on the z_L first-order condition, Equation 40 and since $k_L > 0$,

$$\exp(-\lambda z_L) = \frac{wc}{\phi A\left(1 + Bk_L^{\beta_L}\right)\lambda} < \frac{wc}{\phi A\lambda}$$

Define ϕ_L^* as that which makes $z_L = 0$, that is, $\phi_L^* \equiv \frac{wc}{A\left(1+Bk_L^{\beta_L}\right)\lambda} < \frac{wc}{A\lambda}$. Then, $z_L > 0$ for $\phi_L' \equiv \frac{wc}{A\lambda} > \phi_L^*$. Hence, as in Caliendo and Rossi-Hansberg (2012), it is enough that $\phi > \frac{wc}{A\lambda}$ for $z_L > 0$.

From here on, the proof requires understanding how the remaining first-order conditions behave as ϕ increases. Note that the LHS of Equations 41 and 42 has common form $\frac{\lambda}{wc}P_l + M_l^{\sigma_l}$. Taking derivatives with respect to z_l,

$$\frac{d\left(\frac{\lambda}{wc}P_l + M_l^{\sigma_l}\right)}{dz_l} = \lambda M_l^{\sigma_l} + \sigma_l M_l^{\sigma_l - 1}\left(-\frac{wcp_l^{1-\sigma_l}}{w_l^2}P_l^{\sigma_0}\right) = \lambda M_l^{\sigma_l}\left(1 - \frac{\sigma_l wc}{\lambda w_l}\left(\frac{p_l}{P_l}\right)^{1-\sigma_l}\right) \quad (45)$$

For the expression to be positive, Assumption 1 is enough. With this result, focus first on Equation 41, the z_1 first-order condition. The LHS is increasing in z_1 and the RHS is increasing in ϕ. Then at the lowest value of the LHS a unique ϕ_1^* exists such that,

$$\left(\frac{\lambda}{wc}P_1 + M_1^{\sigma_1}\right)\Big|_{z_{L-1}=0} = \frac{\lambda}{wc}\phi_1^* A \tag{46}$$

The multiplier ranking in Lemma 1 is $\phi_1^* > \phi_L' \equiv \frac{wc}{A\lambda} > \phi_L^*$, which holds if

$$\left(p_1^{1-\sigma_1} + w^{1-\sigma_1}\right)^{1/(1-\sigma_1)} + \frac{wc}{\lambda}\left(1 + \frac{p_1^{1-\sigma_1}}{w^{1-\sigma_1}}\right)^{\sigma_1/(1-\sigma_1)} > \frac{wc}{\lambda} \tag{47}$$

a condition in Assumption 2.

I turn now to Equation 42, the z_l first-order condition, shown again here for completeness:

$$\frac{\lambda}{wc}P_l + M_l^{\sigma_l} = \exp(\lambda z_{l-1})M_{l-1}^{\sigma_{l-1}}$$

Denote its right and left-hand sides as RHS_l and LHS_l for any $l > 1$. As shown above, the LHS_l is increasing in z_l under Assumption 1. The dependence of the RHS_l on ϕ requires taking derivatives,

$$\frac{d\left(\exp(\lambda z_l)M_l^{\sigma_l}\right)}{dz_l}\frac{dz_l}{d\phi} = \lambda \exp(\lambda z_l)M_l^{\sigma_l}\left(1 - \frac{\sigma_l wc}{\lambda w_l}\left(\frac{p_l}{P_l}\right)^{1-\sigma_l}\right)\frac{dz_l}{d\phi} \tag{48}$$

In Proposition 1, I show that $\frac{dz_l}{d\phi} > 0$, and the remaining term is positive due to Assumption 1. As a consequence RHS_l is increasing in ϕ. Now, similar to other layers, define ϕ_l^* implicitly as $LHS_l|_{z_l=0} = RHS_l|_{\phi_l^*}$. Since RHS_l increases with ϕ, requiring $\phi_l^* > \phi_{l-1}^*$ is the same as $RHS_l(\phi_l^*) > RHS_l(\phi_{l-1}^*)$. Also note that

$$LHS_l|_{z_l=0} = RHS_l|_{\phi_l^*} > RHS_l|_{\phi_{l-1}^*} = \left(\exp(\lambda z_{l-1})M_{l-1}^{\sigma_{l-1}}\right)\big|_{z_{l-1}=0} \tag{49}$$

Combining both extremes of the previous equation, I obtain

$$\frac{\lambda}{wc}\left(p_l^{1-\sigma_l} + w^{1-\sigma_l}\right)^{1/(1-\sigma_l)} + \left(1 + \frac{p_l^{1-\sigma_l}}{w^{1-\sigma_l}}\right)^{\sigma_l/(1-\sigma_l)} > \left(1 + \frac{p_{l-1}^{1-\sigma_{l-1}}}{w^{1-\sigma_{l-1}}}\right)^{\sigma_{l-1}/(1-\sigma_{l-1})} \tag{50}$$

which is the expression in Assumption 2. This concludes the proof that $\phi_{L-1}^* > ...\phi_2^* > \phi_1^* > \phi_L^*$ and $k_L > 0$ under Assumptions 1 and 2.

□

Proposition 5. *Under Assumptions 1 and 2, for any $L > 2$, and any production q, knowledge of agents at any layer is positive, $z_l > 0$, $\forall l$.*

Proof. Take any $L \geq 2$, I compare the cost minimization problem for the following two organizations:

1. A firm with L layers and unconstrained, with Lagrangian:

$$\begin{aligned}\mathcal{L}_U \equiv &\sum_{l=1}^{L-1} P_l Q_l + w_L + p_L k_L + \cdots \\ &- \phi\left(A[1-\exp(-\lambda Z_L)]\left(1 + Bk_L^{\beta_L}\right)\exp(\lambda Z_{L-1}) - q\right) - \sum_{l=1}^{L}\theta_l z_l\end{aligned} \tag{51}$$

2. A firm with $L+1$ layers and constrained to $z_L = 0$, with Lagrangian

$$\begin{aligned}\mathcal{L}_C \equiv &\sum_{l=1}^{L} P_l Q_l + w_{L+1} + p_{L+1} k_{L+1} + \cdots \\ &- \phi\left(A[1-\exp(-\lambda Z_{L+1})]\left(1 + Bk_{L+1}^{\beta_L}\right)\exp(\lambda Z_{L}) - q\right) - \sum_{l=1}^{L+1}\theta_l z_l\end{aligned} \tag{52}$$

Whenever needed, I will denote choice variables with superscripts U and C for problems $\mathcal{L}_U$ and $\mathcal{L}_C$, respectively. Specialize the constrained case by imposing $z_L^C = 0$ to get

$$\begin{aligned}\mathcal{L}_C = &\sum_{l=0}^{L-1} P_l Q_l + P_L|_{z_L=0}\, Q_L + w_{L+1} + p_{L+1} k_{L+1} + \cdots \\ &- \phi\left(A[1-\exp(-\lambda\left(Z_{L-1} + z_{L+1}\right))]\left(1 + Bk_{L+1}^{\beta_L}\right)\exp(\lambda Z_{L-1}) - q\right) - \sum_{\forall l>1, l\neq L}^{L+1}\theta_l z_l\end{aligned} \tag{53}$$

Next exploit that k_{L+1}^C in $\mathcal{L}_C$ is symmetric to k_L^U in $\mathcal{L}_U$, and similarly with z_{L+1}^C and z_L^U, and without loss of generality, relabel (p_L, k_L, z_L) in $\mathcal{L}_U$ as $(p_{L+1}, k_{L+1}, z_{L+1})$ to obtain:

$$\begin{aligned}\mathcal{L}_U =& \sum_{l=1}^{L-1} P_l Q_l + w_{L+1} + p_{L+1}k_{L+1} + \cdots \\ & - \phi\left(A[1-\exp(-\lambda\left(Z_{L-1}+z_{L+1}\right))]\left(1+Bk_{L+1}^{\beta_L}\right)\exp(\lambda Z_{L-1}) - q\right) + \cdots \\ & - \sum_{l=1}^{L-1}\theta_l z_l + \theta_{L+1}z_{L+1}\end{aligned} \tag{54}$$

Further recall $Q_l \equiv \left(1+Bk_L^{\beta_L}\right)\exp(\lambda \sum_{k=l}^{L-1} z_k)$ for a firm with L layers so, for the $\mathcal{L}_C$ firm with $L+1$ layers and $z_L^C = 0$, $Q_l^C \equiv \left(1+Bk_L^{\beta_L}\right)\exp(\lambda \sum_{k=l}^{L-1} z_k)$ for and $Q_L^C = \left(1+Bk_L^{\beta_L}\right)\exp(\lambda z_L)|_{z_L=0}$, its problem becomes

$$\begin{aligned}\mathcal{L}_C =& \sum_{l=1}^{L-1} P_l\left(1+Bk_L^{\beta_L}\right)\exp(\lambda\sum_{k=l}^{L-1} z_k) + P_L|_{z_L=0}\left(1+Bk_L^{\beta_L}\right) + w_{L+1} + p_{L+1}k_{L+1} + \cdots \\ & - \phi\left(A[1-\exp(-\lambda\left(Z_{L-1}+z_{L+1}\right))]\left(1+Bk_{L+1}^{\beta_L}\right)\exp(\lambda Z_{L-1}) - q\right) - \sum_{\forall l>1, l\neq L}^{L+1}\theta_l z_l\end{aligned} \tag{55}$$

Similarly, plug Q_l^U into $\mathcal{L}_U$,

$$\begin{aligned}\mathcal{L}_U =& \sum_{l=1}^{L-1} P_l\left(1+Bk_L^{\beta_L}\right)\exp(\lambda\sum_{k=l}^{L-1} z_k) + w_{L+1} + p_{L+1}k_{L+1} + \cdots \\ & - \phi\left(A[1-\exp(-\lambda\left(Z_{L-1}+z_{L+1}\right))]\left(1+Bk_{L+1}^{\beta_L}\right)\exp(\lambda Z_{L-1}) - q\right) + \cdots \\ & - \sum_{l=1}^{L-1}\theta_l z_l + \theta_{L+1}z_{L+1}\end{aligned} \tag{56}$$

Note that Lagrangians in Equations 55 and 56 are the same problem, except the former includes the extra term, $P_L|_{z_L=0}\left(1+Bk_L^{\beta_L}\right)$. Such term is the extra cost in $\mathcal{L}_C$ due to the extra layer L with zero knowledge. The extra layer generates no cost advantage elsewhere since the rest of the Lagrangian is the same as $\mathcal{L}_U$. Intuitively, unless knowledge is positive, an extra layer serves no purpose as it does not generate more output, nor reduce inputs at other layers. What is the effect of the extra layer on cost? By the Envelope Theorem,

$$\frac{d\mathcal{L}_C}{d\ P_L\big|_{z_L=0}} = 1 + Bk_L^{\beta_L}\Big|_{k_L=k_L^*} > 0$$

where k_L^* is optimal k_L^C (i.e., k_L under the constrained problem). The impact on the cost is strictly positive, as if the extra layer were a tax without benefits to the firm. Thus a firm will never choose to have a $L+1$ organization with zero knowledge on the extra layer L. If $z_L = 0$ is never optimal for the firm with $L+1$ layers, then, by Lemma 1, $z_l > 0$, $\forall l < L$, and $z_{L+1} > 0$ holds for such organization.

□

B.2.2 Proof of Proposition 1

The proof involves doing comparative statics, that is, obtaining $\frac{dz_l}{dq} = \frac{dz_l}{d\phi}\frac{d\phi}{dq}$, $\forall l$, and $\frac{dk_L}{dq} = \frac{dk_L}{d\phi}\frac{d\phi}{dq}$. The common term $\frac{d\phi}{dq}$ is derived at the end of the proof, and I start with the remaining terms.

Start with the z_1 first-order condition:

$$\frac{d\left(\frac{\lambda}{wc}P_1 + M_1^{\sigma_1}\right)}{dz_1} = \lambda M_1^{\sigma_1}\left(1 - \frac{\sigma_1 wc}{\lambda w_1}\left(\frac{p_1}{P_1}\right)^{1-\sigma_1}\right) = \frac{\lambda A}{wc}\frac{d\phi}{dz_1}$$

so

$$\frac{dz_1}{d\phi} = \frac{A}{wcM_1^{\sigma_1}\left(1 - \frac{\sigma_1 wc}{\lambda w_1}\left(\frac{p_1}{P_1}\right)^{1-\sigma_1}\right)}$$

which is positive iff $\left(1 - \frac{\sigma_1 wc}{\lambda w_1}\left(\frac{p_1}{P_1}\right)^{1-\sigma_1}\right) > 0$, a condition implied by $\left(1 - \frac{\sigma_1 c}{\lambda}\right) > 0$, that is, by Assumption 1.

Next the z_2 first-order condition. First, I rewrite its RHS using introductory results:

$$d\left(\exp(\lambda z_1)M_1^{\sigma_1}\right) = \lambda\exp(\lambda z_1)M_1^{\sigma_1}\left(1 - \frac{\sigma_1 wc}{\lambda w_1}\left(\frac{p_1}{P_1}\right)^{1-\sigma_1}\right)dz_1 \tag{57}$$

From totally differentiating the z_1 first-order condition:

$$M_1^{\sigma_1}\left(1 - \frac{\sigma_1 wc}{\lambda w_1}\left(\frac{p_1}{P_1}\right)^{1-\sigma_1}\right)dz_1 = \frac{A}{wc}d\phi$$

so Equation 58 can be written as

$$d\left(\exp(\lambda z_1)M_1^{\sigma_1}\right) = \lambda\exp(\lambda z_1)\frac{A}{wc}d\phi$$

Then differentiate the LHS of the z_2 first-order condition:

$$d\left(\frac{\lambda}{wc}P_2 + M_2^{\sigma_2}\right) = \lambda M_2^{\sigma_2}\left(1 - \frac{\sigma_2 wc}{\lambda w_2}\left(\frac{p_2}{P_2}\right)^{1-\sigma_2}\right)dz_2$$

Combining the two:

$$\lambda M_2^{\sigma_2}\left(1 - \frac{\sigma_2 wc}{\lambda w_2}\left(\frac{p_2}{P_2}\right)^{1-\sigma_2}\right)dz_2 = \lambda\exp(\lambda z_1)\frac{A}{wc}d\phi$$

so

$$\frac{dz_2}{d\phi} = \frac{\exp(\lambda z_1)\frac{A}{wc}}{M_2^{\sigma_2}\left(1 - \frac{\sigma_2 wc}{\lambda w_2}\left(\frac{p_2}{P_2}\right)^{1-\sigma_2}\right)}$$

which is positive iff $\left(1 - \frac{\sigma_2 wc}{\lambda w_2}\left(\frac{p_2}{P_2}\right)^{1-\sigma_2}\right) > 0$, a condition implied by $\left(1 - \frac{\sigma_2 c}{\lambda}\right) > 0$, that is, by Assumption 1.

Proceeding analogously gives $\frac{dz_l}{d\phi}$ for any $l > 2$. For any layer $l > 2$, $\frac{dz_l}{d\phi}$ involves taking the ratio of two derivatives. The denominator is the derivative of the LHS of the z_l first-order condition with respect to z_l. The numerator uses again the intermediate result that:

$$d\left(\exp(\lambda z_{l-1})M_{l-1}^{\sigma_{l-1}}\right) = \lambda\exp(\lambda z_{l-1})M_{l-1}^{\sigma_{l-1}}\left(1 - \frac{\sigma_{l-1} wc}{\lambda w_{l-1}}\left(\frac{p_{l-1}}{P_{l-1}}\right)^{1-\sigma_{l-1}}\right)dz_{l-1} \quad (58)$$

which is the derivative of the RHS of the first-order condition of z_l. Using it together with $\frac{dz_{l-1}}{d\phi}$, obtained in the previous step, gives the numerator. So for $l > 2$:

$$\frac{dz_l}{d\phi} = \frac{\lambda\exp\left(\lambda Z_{l-1}\right)\frac{A}{wc}}{\lambda M_l^{\sigma_l}\left(1 - \frac{\sigma_l wc}{\lambda w_l}\left(\frac{p_l}{P_l}\right)^{1-\sigma_l}\right)} \quad (59)$$

which is positive iff $\left(1 - \frac{\sigma_l wc}{\lambda w_l}\left(\frac{p_l}{P_l}\right)^{1-\sigma_l}\right) > 0$, a condition implied by Assumption $\left(1 - \frac{\sigma_l c}{\lambda}\right) > 0$.

Focus next on k_L. Total differentiation of the RHS of its first-order condition simply requires using results for the RHS of layer l applied to $L-1$, as just described in the immediately previous derivation. The total differentiation of the LHS of the k_L first-order condition delivers

$$d\left(\frac{1}{\left(1+Bk_L^{\beta_L}\right)}+\frac{\lambda}{wc}\frac{p_L}{B\beta_L k_L^{\beta_L-1}}\right)=\left(-\frac{Bk_L^{\beta_L-1}}{\left(1+Bk_L^{\beta_L}\right)^2}+\frac{\lambda}{wc}\frac{p_L}{B\beta_L}(1-\beta_L)k_L^{-\beta_L}\right)dk_L$$

Combining both and rewriting:

$$\frac{dk_L}{d\phi}=k_L^{\beta_L}(1+Bk_L^{\beta_L})^2\frac{\lambda\exp\left(\lambda Z_{L-1}\right)\frac{A}{wc}}{\left(\frac{\lambda}{wc}\frac{p_L}{B\beta_L}(1-\beta_L)\left(1+Bk_L^{\beta_L}\right)^2-\beta_L Bk_L^{2\beta_L-1}\right)}$$

which is weakly positive as long as $\left(\frac{\lambda}{wc}\frac{p_L}{B\beta_L}(1-\beta_L)\left(1+Bk_L^{\beta_L}\right)^2-\beta_L Bk_L^{2\beta_L-1}\right)>0$, a condition that is satisfied under Assumption 3 and is strictly positive due to Lemma 1.

I conclude with z_L. Take logs on its first-order condition to obtain

$$-\lambda z_L=\log(wc)-\log\left(\phi A\lambda\left(1+Bk_L^{\beta_L}\right)\right)$$

Totally differentiating and slightly rewriting:

$$-\lambda\frac{dz_L}{d\phi}+\frac{B\beta_L k_L^{\beta_L-1}}{1+Bk_L^{\beta_L}}\frac{dk_L}{d\phi}=-\frac{1}{\phi}$$

Using $\frac{dk_L}{d\phi}$ just obtained,

$$\frac{dz_L}{d\phi}=\frac{1}{\lambda\phi}+\frac{B\beta_L k_L^{2\beta_L-1}}{\left(1+Bk_L^{\beta_L}\right)}\frac{\exp\left(\lambda\left(Z_{L-1}\right)\right)\frac{A}{wc}}{\left(\frac{\lambda}{wc}\frac{p_L}{B\beta_L}(1-\beta_L)\left(1+Bk_L^{\beta_L}\right)^2-Bk_L^{2\beta_L-1}\right)}$$

which is also positive under Assumption 3.

The common term $\frac{d\phi}{dq}$ is obtained by implicit differentiation of the production function. First note using the expression for the optimal input bundle at layer 1, the production function can be written as

$$q = A[\exp(\lambda Z_{L-1}) - \exp(-\lambda z_L)] \left(1 + Bk_L^{\beta_L}\right)$$

and applying the Implicit Function Theorem,

$$\frac{d\phi}{dq} = \frac{1}{TT_1 + TT_2 + TT_3}$$

where

$$\begin{aligned} TT_1 &\equiv A[\sum_{l=0}^{L-1} \lambda \frac{dz_l}{d\phi} \exp(\lambda Z_{L-1}) \left(1 + Bk_L^{\beta_L}\right) \\ TT_2 &\equiv A\lambda \frac{dz_L}{d\phi} \exp(-\lambda z_L)] \left(1 + Bk_L^{\beta_L}\right) \\ TT_3 &\equiv A[\exp(\lambda Z_{L-1}) - \exp(-\lambda z_L)] B\beta_L k_L^{\beta_L - 1} \frac{dk_L}{d\phi} \end{aligned}$$

Since $\frac{dz_l}{d\phi} > 0$, $\forall l$, and $\frac{dk_L}{d\phi} > 0$, then $\frac{d\phi}{dq} > 0$, and also $\frac{dz_l}{dq} > 0, \forall l$, and $\frac{dk_L}{dq} > 0$.

All other factor demands are given by Equations 35 and 34. Taking derivatives of Equation 34 with respect to q,

$$\begin{aligned} \frac{dn_l}{dq} = P_l^{\sigma_l}(-\sigma_l) w_l^{-\sigma_l - 1} wc \frac{dz_l}{dq} \left(1 + Bk_L^{\beta_L}\right) \exp(\lambda \sum_{j=l}^{L-1} z_j) + \cdots \\ P_l^{\sigma_l} w_l^{-\sigma_l} \left(1 + Bk_L^{\beta_L}\right) \lambda \sum_{j=l}^{L-1} \frac{dz_j}{dq} \exp(\lambda \sum_{j=l}^{L-1} z_j) + \cdots \\ \frac{dP_l^{\sigma_l}}{dq} w_l^{-\sigma_l} \left(1 + Bk_L^{\beta_L}\right) \exp(\lambda \sum_{j=l}^{L-1} z_j) + \cdots \\ P_l^{\sigma_l} w_l^{-\sigma_l} \frac{d\left(1 + Bk_L^{\beta_L}\right)}{dq} \exp(\lambda \sum_{j=l}^{L-1} z_j) \end{aligned} \quad (60)$$

Note all the terms are positive except the first. Grouping terms by convenience,

$$\frac{dn_l}{dq} = P_l^{\sigma_l}(-\sigma_l)w_l^{-\sigma_l}\frac{wc}{w_l}\frac{dz_l}{dq}\left(1+Bk_L^{\beta_L}\right)\exp(\lambda\sum_{j=l}^{L-1}z_j) + \cdots$$
$$P_l^{\sigma_l}w_l^{-\sigma_l}\left(1+Bk_L^{\beta_L}\right)\lambda\frac{dz_l}{dq}\exp(\lambda\sum_{j=l}^{L-1}z_j) + OT \quad (61)$$

where

$$OT \equiv P_l^{\sigma_l}w_l^{-\sigma_l}\left(1+Bk_L^{\beta_L}\right)\lambda\sum_{j=l+1}^{L-1}\frac{dz_j}{dq}\exp(\lambda\sum_{j=l}^{L-1}z_j) + \cdots \quad (62)$$
$$\frac{dP_l^{\sigma_l}}{dq}w_l^{-\sigma_l}\left(1+Bk_L^{\beta_L}\right)\exp(\lambda\sum_{j=l}^{L-1}z_j) + \cdots \quad (63)$$
$$P_l^{\sigma_l}w_l^{-\sigma_l}\frac{d\left(1+Bk_L^{\beta_L}\right)}{dq}\exp(\lambda\sum_{j=l}^{L-1}z_j) \quad (64)$$

And with further collecting,

$$\frac{dn_l}{dq} = P_l^{\sigma_l}w_l^{-\sigma_l}\left(1+Bk_L^{\beta_L}\right)\exp(\lambda\sum_{j=l}^{L-1}z_j)\left[(-\sigma_l)\frac{wc}{w+wcz_l}+\lambda\right]\frac{dz_l}{dq} + OT \quad (65)$$

A sufficient condition for $\frac{dn_l}{dq} > 0$ is that the term in brackets,

$$\left[\lambda - \sigma_l c\frac{1}{1+cz_l}\right] \quad (66)$$

is positive. Suppose $z_l = 0$, then $[\lambda - \sigma_l c] > 0$ due to Assumption 1. For any $0 < z_l < \infty$, the negative term will be even smaller, and hence $\left[\lambda - \sigma_l c\frac{1}{1+cz_l}\right] > 0$, which completes the proof for $n_l, \forall l < L$. The proof for $k_l, \forall l > L$, is analogous and, in fact, simpler as its price p_l does not depend on z_l.

□

B.2.3 Proof of Proposition 2

I first show the properties of MC_L. Recall

$$MC_L \equiv \phi = \frac{wc}{\exp(-\lambda z_L) A\lambda \left(1 + Bk_L^{\beta_L}\right)} = \frac{[\exp(\lambda Z_L) - 1]wc}{\lambda q} \tag{67}$$

where the last equality uses the production function definition and the L layer constraint.

$MC_L > 0$ is trivial from Equation 67, and $\frac{dMC_L}{dq} > 0$ is a direct implication from Proposition 1. Regarding its behavior at the limits, note

$$\lim_{q\to 0} MC_L = \lim_{q\to 0} \frac{\exp(\lambda z_L)wc}{A\lambda \left(1 + Bk_L^{\beta_L}\right)} = \frac{wc}{A\lambda}$$

where the last equality comes from the monotonicity of choice variables with respect to q shown in Proposition 1. Using the second form of MC_L on Equation 67,

$$\lim_{q\to\infty} \frac{[\exp(\lambda Z_L) - 1]wc}{\lambda q} = \frac{\infty}{\infty} = \lim_{q\to\infty} wc\frac{\exp(\lambda Z_L)\lambda \sum_{l=1}^{L} \frac{dz_l}{dq}}{\lambda} = \infty$$

where the second equality uses L'Hopital's rule. This concludes the properties of MC_L.

I now show the properties of AC_L. Before turning to the stated properties of AC_L, note that a general property of cost functions is that

$$\frac{\partial AC_L}{\partial q} = \frac{\partial \frac{C_L(q)}{q}}{\partial q} = \frac{\frac{\partial C_L(q)}{\partial q} q - C_L(q)}{q^2} = -\frac{AC_L(q) - MC_L(q)}{q}$$

Hence, whether AC_L is increasing or decreasing depends on whether $AC_L > MC_L$ or vice versa, respectively. Denoting explicitly the dependence on q,

$$AC_L = \frac{\sum_{l=1}^{L-1} P_l(q)Q_l(q) + w_L(q) + p_L k_L(q)}{q} \tag{68}$$

and using the expressions for Q_l and q,

$$AC_L = \frac{\sum_{l=1}^{L-1} P_l(q)\exp\left(\lambda \sum_{k=l}^{L-1} z_k(q)\right)}{A[1-\exp(-\lambda Z_L(q))]\exp(\lambda Z_{L-1}(q))} + \cdots \\ + \frac{w_L(q) + p_L k_L(q)}{A[1-\exp(-\lambda Z_L(q))]\,(1+Bk_L(q)^{\beta_L})\exp(\lambda Z_{L-1}(q))} \tag{69}$$

Turn to the stated properties. First compute limits as $q \to 0$:

$$\lim_{q\to 0} AC_L = \lim_{q\to 0} \frac{\sum_{l=1}^{L-1} P_l(q)\exp\left(\lambda \sum_{k=l}^{L-1} z_k(q)\right) + w_L(q)}{A[1-\exp(-\lambda Z_L(q))]\exp(\lambda Z_{L-1}(q))} + \cdots \\ + \lim_{q\to 0} \frac{p_L k_L(q)}{A[1-\exp(-\lambda Z_L(q))]\,(1+Bk_L(q)^{\beta_L})\exp(\lambda Z_{L-1}(q))} \tag{70}$$

The first limit is

$$\frac{\sum_{l=1}^{L-1} P_l|_{z_l=0}}{0} + \frac{w}{0} = \infty$$

The second term is an indeterminacy $\frac{0}{0}$, and the only way to overturn the effect of the first limit would be that it is $-\infty$. Such is never the case due to the monotonicity properties in Proposition 1. Hence, $\lim_{q\to 0} AC_L = \infty$; moreover, for $q \to 0$, $AC_L > MC_L$ and AC_L is decreasing.

Now compute the limit $q \to \infty$:

$$\lim_{q\to \infty} AC_L = \lim_{q\to \infty} \frac{\sum_{l=1}^{L-1} P_l(q)\exp\left(\lambda \sum_{k=l}^{L-1} z_k(q)\right) + w_L(q)}{A[1-\exp(-\lambda Z_L(q))]\exp(\lambda Z_{L-1}(q))} + \cdots \\ + \lim_{q\to \infty} \frac{p_L k_L(q)}{A[1-\exp(-\lambda Z_L(q))]\,(1+Bk_L(q)^{\beta_L})\exp(\lambda Z_{L-1}(q))} \tag{71}$$

Note that the layer 1 component of the first limit is

$$\lim_{q\to\infty} \frac{P_1(q)}{A[1-\exp(-\lambda Z_L(q))]} = \infty$$

because the denominator converges to A and the numerator goes to ∞, both due to Proposition 1. While a priori the remaining terms are indeterminate of $\frac{\infty}{\infty}$ type, they are positive: applying L'Hopital and the monotonicity of policy functions rules out their convergence to $-\infty$. The other remaining terms are similarly analyzed which implies $\lim_{q\to\infty} AC_L = \infty$.

Showing that $AC_L > MC_L$ and AC_L is increasing as $q \to \infty$ requires more work because both AC_L and MC_L go to ∞. I next show that AC_L increases. Note that total cost given L, C_L, can be written as the fixed-plus-variable cost (i.e., fixed costs plus the integral of all marginal costs):

$$\begin{aligned}&\lim_{q\to\infty}\left[AC_L(q) - MC_L(q)\right] = \lim_{q\to\infty}\left[\frac{C_L(q)}{q} - MC_L(q)\right] = \\ &\lim_{q\to\infty}\left[\frac{F_L + \int_0^q MC_L(x)dx}{q} - MC_L(q)\right] = \lim_{q\to\infty}\left[\frac{\int_0^q MC_L(x)dx}{q} - MC_L(q)\right]\end{aligned} \tag{72}$$

where F_L is the fixed cost term, $F_L \equiv \lim_{q\to 0} C_L(q)$, at the very least including the CEO time compensation (w). It is now apparent that the limit as $q \to \infty$ depends on how MC behaves. Write it as a ratio so that calculating the limit is easier:

$$\lim_{q\to\infty}\frac{AC_L(q)}{MC_L(q)} = \lim_{q\to\infty}\frac{\int_0^q MC_L(x)dx}{qMC_L(q)}$$

Note that

$$\lim_{q\to\infty}\frac{\int_0^q MC_L(x)dx}{qMC_L(q)} = \lim_{q\to\infty}\int_0^q \frac{MC_L(x)}{qMC_L(q)}dx = \lim_{q\to\infty}\int_0^q h_1(q,x)dx$$

with $h_1(q,x) \equiv \frac{MC_L(x)}{qMC_L(q)}$. Properties of $h_1(q,x)$ include:

- $0 < h_1(q,x) < 1$ for any q and $x > 0$ and $h_1(q,0) = 0$.
- $\lim_{q\to\infty} h_1(q,x) = \lim_{q\to\infty}\frac{MC_L(x)}{qMC_L(q)} = 0$.

To exchange limit and integration in $\lim_{q\to\infty}\int_0^q h_1(q,x)dx$, I apply the dominated convergence theorem. Its application requires finding an integrable function g that bounds $h_1(q,x)$, that is, for all q, $h_1(q,x) \le g$. Setting $g = 1$ satisfies the requirement, so I exchange the operator order:

$$\lim_{q\to\infty}\frac{\int_0^q MC_L(x)dx}{qMC_L(q)} = \int_0^q \lim_{q\to\infty} h_1(q,x)dx = 0$$

Hence, $MC_L > AC_L$ as $q \to \infty$ and AC_L increases in that limit.

Finally, Proposition 2 states that the q^* that satisfies

$$\frac{F_L + \int_0^{q^*} MC_L(x)dx}{q^*} = MC_L(q^*)$$

is a unique minimum. To show it is unique, rewrite the condition that defines q^*:

$$\frac{F_L + \int_0^{q^*} \frac{[\exp(\lambda Z_L(q))-1]wc}{\lambda q} dq}{q^*} = \frac{[\exp(\lambda Z_L(q^*)) - 1]wc}{\lambda q^*}$$

$$F_L + \int_0^{q^*} \frac{[\exp(\lambda Z_L(q)) - 1]wc}{\lambda q} dq = [\exp(\lambda Z_L(q^*)) - 1]\frac{wc}{\lambda}$$

$$F_L + \int_0^{q^*} h_2(q)dq = h_2(q^*)q^*$$

where $h_2(q) \equiv \frac{[\exp(\lambda Z_L(q))-1]wc}{\lambda q}$. Finally, to show that such q^* is unique, I study the behavior of each side. The LHS,

$$\frac{\partial LHS}{\partial q^*} = h_2(q^*) > 0$$

and the right-hand side,

$$\frac{\partial RHS}{\partial q^*} = h_2(q^*) + h_2'(q^*)q^* > 0 \tag{73}$$

where $h_2' = wc\exp(\lambda Z_L)\sum_{l=1}^{L} \frac{dz_l}{dq} > 0$. Note that $0 < \frac{\partial LHS}{\partial q^*} < \frac{\partial RHS}{\partial q^*}$. This inequality together with the fact that total cost at $q^* = 0$, the LHS of Equation 73, is strictly positive, whereas its RHS is zero because $Z_L|_{q^*=0} = 0$, I conclude that q^* is unique.

To determine that q^* is a minimum, take the second-order derivative of AC_L,

$$\frac{\partial^2 AC_L(q)}{\partial q^2} = -\frac{\left(\frac{\partial AC_L}{\partial q} - \frac{\partial MC_L}{\partial q}\right) q - (AC_L - MC_L)}{q^2}$$

At q^* the expression simplifies due to $\frac{\partial AC_L}{\partial q} = 0$ and $AC_L - MC_L = 0$:

$$\frac{\partial^2 AC_L(q)}{\partial q^2} = \left.\frac{\frac{\partial MC_L}{\partial q}}{q}\right|_{q=q^*} > 0$$

since $\frac{\partial MC_L}{\partial q} > 0$ as long as $\frac{dz_l}{dq} > 0$ and $\frac{dk_L}{dq} > 0$.

□

B.2.4 Proof of Proposition 3

First, do case $L = 4$. The system of first-order conditions is:

$$\exp(-\lambda z_4) = \frac{wc}{\phi A\lambda\left(1 + Bk_L^{\beta_L}\right)}$$

$$\lambda P_1 + P_1^{\sigma_1}\left(w(cz_1+1)\right)^{-\sigma_1} wc = \phi A\lambda$$

$$\lambda P_2 + P_2^{\sigma_2}\left(w(cz_2+1)\right)^{-\sigma_2} wc = \exp(\lambda z_1)P_1^{\sigma_1}\left(w(cz_1+1)\right)^{-\sigma_1} wc$$

$$\lambda P_3 + P_3^{\sigma_3}\left(w(cz_3+1)\right)^{-\sigma_3} wc = \exp(\lambda z_2)P_2^{\sigma_2}\left(w(cz_2+1)\right)^{-\sigma_2} wc$$

$$\left(\exp(\lambda z_3)P_3^{\sigma_3}\left(w(cz_3+1)\right)^{-\sigma_3}\frac{wc}{\lambda} - \frac{wc}{\lambda\left(1+Bk_L^{\beta_L}\right)}\right)B\beta_L k_L^{\beta_L-1} = p_L$$

To prove Proposition 3, I first find compact expressions for some relevant derivatives. Recall that $P_l \equiv \left(p_l^{1-\sigma_l} + (w(cz_l+1))^{1-\sigma_l}\right)^{1/(1-\sigma_l)}$, so:

$$\frac{dP_l}{dp_2} = wc\left(\frac{P_l}{w_l}\right)^{\sigma_l}\left[\frac{1}{wc}\left(\frac{w_l}{p_l}\right)^{\sigma_l}\frac{dp_l}{dp_2} + \frac{dz_l}{dp_2}\right] \tag{74}$$

Notice that for $l = 2,3,4$, $\dfrac{dp_l}{dp_2} = 1$, but $\dfrac{dp_1}{dp_2} = 0$. Another relevant term is $P_l^{\sigma_l}(w(cz_l + 1))^{-\sigma_l}$, for which I obtain

$$\frac{d}{dp_2}\left(P_l^{\sigma_l}(w(cz_l+1))^{-\sigma_l}\right) = \sigma_l\left(\frac{P_l}{w_l}\right)^{\sigma_l}\left[\frac{1}{P_l}\frac{dP_l}{dp_2} - \frac{wc}{w_l}\frac{dz_l}{dp_2}\right]$$

Replacing the term $\frac{dP_l}{dp_2}$ with Equation 74, I get:

$$\frac{d}{dp_2}\left(P_l^{\sigma_l}(w(cz_l+1))^{-\sigma_l}\right) = \left(\frac{P_l}{w_l}\right)^{\sigma_l}\left[\frac{\sigma_l}{P_l}\left(\frac{P_l}{w_l}\right)^{\sigma_l}\frac{dp_l}{dp_2} + \left(\frac{\sigma_l wc}{P_l}\left(\frac{P_l}{w_l}\right)^{\sigma_l} - \frac{\sigma_l wc}{w_l}\right)\frac{dz_l}{dp_2}\right]$$

Finally, the term $e^{\lambda z_l} P_l^{\sigma_l}(w(cz_l+1))^{-\sigma_l}$ has the following derivative:

$$\begin{aligned}&\frac{d}{dp_2}\left(e^{\lambda z_l} P_l^{\sigma_l}(w(cz_l+1))^{-\sigma_l}\right) = \\ &e^{\lambda z_l}\left(\frac{P_l}{w_l}\right)^{\sigma_l}\left[\frac{\sigma_l}{P_l}\left(\frac{P_l}{p_l}\right)^{\sigma_l}\frac{dp_l}{dp_2}+\left(\lambda+\frac{\sigma_l wc}{P_l}\left(\frac{P_l}{p_l}\right)^{\sigma_l}-\frac{\sigma_l wc}{w_l}\right)\frac{dz_l}{dp_2}\right]\end{aligned} \tag{75}$$

With these expressions at hand, I obtain the derivatives of the system. Consider the z_4 first-order condition. First, apply the natural logarithm and rewrite it as:

$$-\lambda z_4 = \ln\left(\frac{wc}{A\lambda}\right) - \ln(\phi) - \ln\left(1+Bk_L^{\beta_L}\right)$$

Now by differentiating and simplifying the above:

$$\lambda\frac{dz_4}{dp_2} = \frac{1}{\phi}\frac{d\phi}{dp_2} + \frac{B\beta_L k_L^{\beta_L-1}}{1+Bk_L^{\beta_L}}\frac{dk_L}{dp_2}$$

Next, differentiating the z_1 first-order condition, and using Equations 74 and 75, and that $\frac{dp_1}{dp_2}=0$,

$$wc\left(\frac{P_1}{w_1}\right)^{\sigma_1}\left[\lambda+\frac{\sigma_1 wc}{P_1}\left(\frac{P_1}{w_1}\right)^{\sigma_1}-\frac{wc\sigma_1}{w_1}\right]\frac{dz_1}{dp_2} = A\lambda\frac{d\phi}{dp_2}$$

Similarly, differentiating the z_2 first-order condition, using Equations 74 and 75,

$$\begin{aligned}&wc\left(\frac{P_2}{w_2}\right)^{\sigma_2}\left[\frac{\lambda}{wc}\left(\frac{w_2}{p_2}\right)^{\sigma_2}+\frac{\sigma_2}{P_2}\left(\frac{P_2}{p_2}\right)^{\sigma_2}+\left(\lambda+\frac{\sigma_2 wc}{P_2}\left(\frac{P_2}{w_2}\right)^{\sigma_2}-\frac{\sigma_2 wc}{w_2}\right)\frac{dz_2}{dp_2}\right] = \\ &wce^{\lambda z_1}\left(\frac{P_1}{w_1}\right)^{\sigma_1}\left[\lambda+\frac{\sigma_1 wc}{P_1}\left(\frac{P_1}{w_1}\right)^{\sigma_1}-\frac{\sigma_1 wc}{w_1}\right]\frac{dz_1}{dp_2}\end{aligned}$$

Differentiating the z_3 first-order condition:

$$\begin{aligned}&wc\left(\frac{P_3}{w_3}\right)^{\sigma_3}\left[\frac{\lambda}{wc}\left(\frac{w_3}{p_3}\right)^{\sigma_3}+\frac{\sigma_3}{P_3}\left(\frac{P_3}{p_3}\right)^{\sigma_3}+\left(\lambda+\frac{\sigma_3 wc}{P_3}\left(\frac{P_3}{w_3}\right)^{\sigma_3}-\frac{\sigma_3 wc}{w_3}\right)\frac{dz_3}{dp_2}\right] = \\ &wce^{\lambda z_2}\left(\frac{P_2}{w_2}\right)^{\sigma_2}\left[\frac{\sigma_2}{P_2}\left(\frac{P_2}{p_2}\right)^{\sigma_2}+\left(\lambda+\frac{\sigma_2 wc}{P_2}\left(\frac{P_2}{w_2}\right)^{\sigma_2}-\frac{\sigma_2 wc}{w_2}\right)\frac{dz_2}{dp_2}\right]\end{aligned}$$

To differentiate the k_L first-order condition, I rewrite it as

$$e^{\lambda z_3}P_3^{\sigma_3}(w(cz_3+1))^{-\sigma_3} - \frac{1}{1+Bk_L^{\beta_L}} = \frac{\lambda}{B\beta_L wc}p_L k_L^{1-\beta_L}$$

and differentiate it to obtain

$$e^{\lambda z_3}\left(\frac{P_3}{w_3}\right)^{\sigma_3}\left[\frac{\sigma_3}{P_3}\left(\frac{P_3}{p_3}\right)^{\sigma_3}+\left(\lambda+\frac{\sigma_3 wc}{P_3}\left(\frac{P_3}{w_3}\right)^{\sigma_3}-\frac{\sigma_3 wc}{w_3}\right)\frac{dz_3}{dp_2}\right]+\frac{B\beta_L k_L^{\beta_L-1}}{(1+Bk_L^{\beta_L})^2}\frac{dk_L}{dp_2}=$$
$$\frac{\lambda k_L^{-\beta_L}}{B\beta_L wc}\left(k_L+p_L(1-\beta_L)\frac{dk_L}{dp_2}\right)$$

Finally, before differentiating the production function, I rewrite it as $\frac{q}{A}\frac{1}{1+Bk_L^{\beta_L}}=e^{\lambda(z_1+z_2+z_3)}-e^{-\lambda z_4}$, and hence

$$-\frac{q}{A}\frac{B\beta_L k_L^{\beta_L-1}}{\left(1+Bk_L^{\beta_L}\right)^2}\frac{dk_L}{dp_2}=\lambda\left[e^{\lambda(z_1+z_2+z_3)}\left(\frac{dz_1}{dp_2}+\frac{dz_2}{dp_2}+\frac{dz_3}{dp_2}\right)+e^{-\lambda z_4}\frac{dz_4}{dp_2}\right]$$

Before solving the system of differential equations, I define some new terms for compactness:

$$\overline{C_l}\equiv\left(\frac{P_l}{w_l}\right)^{\sigma_l}\left(\lambda+\frac{\sigma_l wc}{P_l}\left(\frac{P_l}{w_l}\right)^{\sigma_l}-\frac{\sigma_l wc}{w_l}\right)$$
$$D_l\equiv\frac{\sigma_l}{P_l}\left(\frac{P_l^2}{p_l w_l}\right)^{\sigma_l}$$
$$E_l\equiv\frac{\lambda}{wc}\left(\frac{P_l}{p_l}\right)^{\sigma_l}$$
$$X\equiv 1+Bk_L^{\beta_L}$$
$$X_k\equiv B\beta_L k_L^{\beta_L-1}$$
$$H\equiv\frac{\lambda}{wc}\frac{p_L}{X_k k_L}(1-\beta_L)-\frac{X_k}{G^2}$$

All the above terms are positive. This is straightforward for all the definitions except for $\overline{C_l}$ and H, so I only show the sign for these latter two. Under Assumption 1, $\frac{\lambda}{c}>\sigma_l,\forall l$. Therefore,

$$\overline{C_l}=\left(\frac{P_l}{w_l}\right)^{\sigma_l}\left[\frac{\sigma_l wc}{P_l}\left(\frac{P_l}{w_l}\right)^{\sigma_l}+c\left(\frac{\lambda}{c}-\sigma_l\frac{w}{w_l}\right)\right]>0$$

The above is true since $w_l>w$, $\forall l$. On the other hand, under Assumption 3,

$$\frac{\lambda}{wc}\frac{p_L}{B\beta_L}(1-\beta_L)\left(1+Bk_L^{\beta_L}\right)^2-B\beta_L k_L^{2\beta_L-1}>0$$

which can be written as

$$k_L^{\beta_L} X^2 \left[\frac{\lambda}{wc} \frac{p_L}{X_k k_L} (1-\beta_L) - \frac{X_k}{X^2} \right] = k_L^{\beta_L} X^2 H > 0$$

And since k_L and X are both greater than zero, then $H > 0$. With these definitions at hand, the system for the comparative statics becomes:

$$\lambda \frac{dz_4}{dp_2} = \frac{1}{\phi} \frac{d\phi}{dp_2} + \frac{X_k}{X} \frac{dk_L}{dp_2} \tag{76}$$

$$\overline{C_1} \frac{dz_1}{dp_2} = \frac{A\lambda}{wc} \frac{d\phi}{dp_2} \tag{77}$$

$$E_2 + D_2 + \overline{C_2} \frac{dz_2}{dp_2} = e^{\lambda z_1} \overline{C_1} \frac{dz_1}{dp_2} \tag{78}$$

$$E_3 + D_3 + \overline{C_3} \frac{dz_3}{dp_2} = e^{\lambda z_2} \left(D_2 + \overline{C_2} \frac{dz_2}{dp_2} \right) \tag{79}$$

$$e^{\lambda z_3} \left(D_3 + \overline{C_3} \frac{dz_3}{dp_2} \right) - H \frac{dk_L}{dp_2} = \frac{\lambda}{wcX_k} \tag{80}$$

$$-\frac{q}{A\lambda} \frac{X_k}{X^2} \frac{dk_L}{dp_2} = e^{\lambda(z_1+z_2+z_3)} \left(\frac{dz_1}{dp_2} + \frac{dz_2}{dp_2} + \frac{dz_3}{dp_2} \right) + e^{-\lambda z_4} \frac{dz_4}{dp_2} \tag{81}$$

I turn now to solving the system. From Equation 77, I can solve for $\frac{d\phi}{dp_2}$:

$$\frac{d\phi}{dp_2} = \frac{\overline{C_1} wc}{A\lambda} \frac{dz_1}{dp_2}$$

From Equation 76 and using the expression for $\frac{d\phi}{dp_2}$:

$$\frac{dz_4}{dp_2} = \frac{\overline{C_1} wc}{A\phi\lambda^2} \frac{dz_1}{dp_2} + \frac{X_k}{X\lambda} \frac{dk_L}{dp_2} \tag{82}$$

With Equation 78 and solving for $\frac{dz_2}{dp_2}$:

$$\frac{dz_2}{dp_2} = e^{\lambda z_1} \frac{\overline{C_1}}{\overline{C_2}} \frac{dz_1}{dp_2} - \frac{(E_2 + D_2)}{\overline{C_2}} \tag{83}$$

Using Equation 79, solving for $\dfrac{dz_3}{dp_2}$, and substituting the expression for $\frac{dz_2}{dp_2}$,

$$\frac{dz_3}{dp_2} = e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{C_3}}\frac{dz_1}{dp_2} - e^{\lambda z_2}\frac{E_2}{\overline{C_3}} - \frac{(E_3+D_3)}{\overline{C_3}} \tag{84}$$

Next, use Equation 80, solve for $\frac{dk_L}{dp_2}$ and substitute for $\frac{dz_3}{dp_2}$ expressed as in Equation 84:

$$\frac{dk_L}{dp_2} = e^{\lambda(z_1+z_2+z_3)}\frac{\overline{C_1}}{H}\frac{dz_1}{dp_2} - e^{\lambda(z_2+z_3)}\frac{E_2}{H} - e^{\lambda z_3}\frac{E_3}{H} - \frac{\lambda}{wcX_kH} \tag{85}$$

I now use $\frac{dz_2}{dp_2}$, $\frac{dz_3}{dp_2}$, and $\frac{dz_4}{dp_2}$ as just obtained in Equation 81. I multiply both sides by $e^{-\lambda(z_1+z_2+z_3)}$ to finally group alike terms to obtain

$$\begin{gathered}\left(1+e^{\lambda z_1}\frac{\overline{C_1}}{\overline{C_2}}+e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{C_3}}+e^{-\lambda Z_4}\frac{\overline{C_1}wc}{A\phi\lambda^2}\right)\frac{dz_1}{dp_2}+\left(e^{-\lambda Z_4}\frac{X_k}{X\lambda}+e^{-\lambda Z_3}\frac{q}{A\lambda}\frac{X_k}{X^2}\right)\frac{dk_L}{dp_2} = \\ \frac{(E_2+D_2)}{\overline{C_2}}+e^{\lambda z_2}\frac{E_2}{\overline{C_3}}+\frac{(E_3+D_3)}{\overline{C_3}}\end{gathered} \tag{86}$$

Note that the term multiplying $\dfrac{dk_L}{dp_2}$ can be simplified since, from the production function I have $e^{-\lambda Z_3}\frac{q}{AX} = 1-e^{-\lambda Z_4}$, so the term multiplying $\dfrac{dk_L}{dp_2}$ can be simplified by noticing,

$$e^{-\lambda Z_4}\frac{X_k}{X\lambda}+e^{-\lambda Z_3}\frac{q}{A\lambda}\frac{X_k}{X^2} = e^{-\lambda Z_4}\frac{X_k}{X\lambda}+(1-e^{-\lambda Z_4})\frac{X_k}{X\lambda} = \frac{X_k}{X\lambda}$$

Equation 86 can then be written as

$$\begin{gathered}\left(1+e^{\lambda z_1}\frac{\overline{C_1}}{\overline{C_2}}+e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{C_3}}+e^{-\lambda Z_4}\frac{\overline{C_1}wc}{A\phi\lambda^2}\right)\frac{dz_1}{dp_2}+\frac{X_k}{X\lambda}\frac{dk_L}{dp_2} = \\ \frac{(E_2+D_2)}{\overline{C_2}}+e^{\lambda z_2}\frac{E_2}{\overline{C_3}}+\frac{(E_3+D_3)}{\overline{C_3}}\end{gathered}$$

And if I substitute for $\frac{dk_L}{dp_2}$ as expressed in Equation 85, I obtain

$$\frac{dz_1}{dp_2} = \frac{\frac{(E_2+D_2)}{\overline{C_2}}+e^{\lambda z_2}\frac{E_2}{\overline{C_3}}+\frac{(E_3+D_3)}{\overline{C_3}}+e^{\lambda(z_2+z_3)}\frac{X_kE_2}{XH\lambda}+e^{\lambda z_3}\frac{X_kE_3}{XH\lambda}+\frac{1}{wcXH}}{1+e^{\lambda z_1}\frac{\overline{C_1}}{\overline{C_2}}+e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{C_3}}+e^{-\lambda Z_4}\frac{\overline{C_1}wc}{A\phi\lambda^2}+e^{\lambda(z_1+z_2+z_3)}\frac{X_k\overline{C_1}}{XH\lambda}} > 0 \tag{87}$$

From this result, Equation 77 also tells us that $\frac{d\phi}{dp_2}$ is positive. Obtaining expressions for the other derivatives simply involves plugging the solution for $\dfrac{dz_1}{dp_2}$, I just obtained. For example, for z_2, Equation 83 together with Equation 87 delivers the following:

$$\frac{dz_2}{dp_2} = e^{\lambda z_1} \frac{\left[\frac{(E_2+D_2)}{\overline{C_2}} + e^{\lambda z_2}\frac{E_2}{\overline{C_3}} + \frac{(E_3+D_3)}{\overline{C_3}} + e^{\lambda(z_2+z_3)}\frac{X_k E_2}{XH\lambda} + e^{\lambda z_3}\frac{X_k E_3}{XH\lambda} + \frac{1}{wcXH}\right]}{\left[\frac{\overline{C_2}}{\overline{C_1}} + e^{\lambda z_1} + e^{\lambda(z_1+z_2)}\frac{\overline{C_2}}{\overline{C_3}} + e^{-\lambda Z_4}\frac{\overline{C_2}wc}{A\phi\lambda^2} + e^{\lambda(z_1+z_2+z_3)}\frac{X_k\overline{C_2}}{XH\lambda}\right]} - \frac{(E_2+D_2)}{\overline{C_2}}$$

Similarly for $\dfrac{dz_3}{dp_2}$,

$$\frac{dz_3}{dp_2} = e^{\lambda(z_1+z_2)} \frac{\left[\frac{(E_2+D_2)}{\overline{C_2}} + e^{\lambda z_2}\frac{E_2}{\overline{C_3}} + \frac{(E_3+D_3)}{\overline{C_3}} + e^{\lambda(z_2+z_3)}\frac{X_k E_2}{XH\lambda} + e^{\lambda z_3}\frac{X_k E_3}{XH\lambda} + \frac{1}{wcXH}\right]}{\left[\frac{\overline{C_3}}{\overline{C_1}} + e^{\lambda z_1}\frac{\overline{C_3}}{\overline{C_2}} + e^{\lambda(z_1+z_2)} + e^{-\lambda Z_4}\frac{\overline{C_3}wc}{A\phi\lambda^2} + e^{\lambda(z_1+z_2+z_3)}\frac{X_k\overline{C_3}}{XH\lambda}\right]} + \cdots$$
$$- e^{\lambda z_2}\frac{E_2}{\overline{C_3}} - \frac{(E_3+D_3)}{\overline{C_3}}$$

and for $\dfrac{dz_4}{dp_2}$, which involves using also Equation 85,

$$\frac{dz_4}{dp_2} =$$
$$\left(\frac{\overline{C_1}wc}{A\phi\lambda^2} + e^{\lambda(z_1+z_2+z_3)}\frac{\overline{C_1}X_k}{XH\lambda}\right) \frac{\left[\frac{(E_2+D_2)}{\overline{C_2}} + e^{\lambda z_2}\frac{E_2}{\overline{C_3}} + \frac{(E_3+D_3)}{\overline{C_3}} + e^{\lambda(z_2+z_3)}\frac{X_k E_2}{XH\lambda} + e^{\lambda z_3}\frac{X_k E_3}{XH\lambda} + \frac{1}{wcXH}\right]}{\left[1 + e^{\lambda z_1}\frac{\overline{C_1}}{\overline{C_2}} + e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{C_3}} + e^{-\lambda Z_4}\frac{\overline{C_1}wc}{A\phi\lambda^2} + e^{\lambda(z_1+z_2+z_3)}\frac{X_k\overline{C_1}}{XH\lambda}\right]} \cdots$$
$$- \frac{X_k}{X\lambda}\left(e^{\lambda(z_2+z_3)}\frac{E_2}{H} + e^{\lambda z_3}\frac{E_3}{H} + \frac{\lambda}{wcX_kH}\right)$$

Note $\dfrac{dz_l}{dp_2}$, for $l > 1$ have indeterminate signs in this case $L = 4$. Finally, for k_L, from Equation 86 and plugging the solution to $\dfrac{dz_1}{dp_2}$,

$$\frac{dk_L}{dp_2} = \frac{X\lambda}{X_k}\left[\frac{(E_2+D_2)}{\overline{C_2}} + e^{\lambda z_2}\frac{E_2}{\overline{C_3}} + \frac{(E_3+D_3)}{\overline{C_3}}\right]\left(1 - \frac{a_2}{a_1}\right) - \cdots$$
$$\frac{X\lambda}{X_k}\left(e^{\lambda(z_2+z_3)}\frac{X_k E_2}{XH\lambda} + e^{\lambda z_3}\frac{X_k E_3}{XH\lambda} + \frac{1}{wcXH}\right)\frac{a_2}{a_1}$$

where

$$a_1 \equiv 1 + e^{\lambda z_1}\frac{\overline{C_1}}{\overline{\overline{C_2}}} + e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{\overline{C_3}}} + e^{-\lambda Z_4}\frac{\overline{C_1}wc}{A\phi\lambda^2} + e^{\lambda(z_1+z_2+z_3)}\frac{X_k\overline{C_1}}{XH\lambda}$$

and

$$a_2 \equiv 1 + e^{\lambda z_1}\frac{\overline{C_1}}{\overline{\overline{C_2}}} + e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{\overline{C_3}}} + e^{-\lambda Z_4}\frac{\overline{C_1}wc}{A\phi\lambda^2}$$

so $\frac{dk_L}{dp_2}$ also has an undetermined sign.

I am ready to show that $\sum_{l=1}^{L}\frac{dz_l}{dp_2} > 0$. The proof uses expressions for $\frac{dz_l}{dp_2}, \forall l > 1$, as functions of only $\frac{dz_1}{dp_2}$. I can express Equation 82 as

$$\frac{dz_4}{dp_2} = \frac{\overline{C_1}wc}{A\phi\lambda^2}\frac{dz_1}{dp_2} + \frac{X_k}{X\lambda}\left(e^{\lambda(z_1+z_2+z_3)}\frac{\overline{C_1}}{H}\frac{dz_1}{dp_2} - e^{\lambda(z_2+z_3)}\frac{E_2}{H} - e^{\lambda z_3}\frac{E_3}{H} - \frac{\lambda}{wcX_kH}\right) \quad (88)$$

Using Equations 83, 84, and 88 and grouping alike terms as well as substituting for Equation 87,

$$\sum_{l=1}^{L}\frac{dz_l}{dp_2} = \left(\frac{a_3}{a_1} - 1\right)b_1 > 0$$

because $a_1 < a_3$ where

$$a_3 \equiv 1 + e^{\lambda z_1}\frac{\overline{C_1}}{\overline{\overline{C_2}}} + e^{\lambda(z_1+z_2)}\frac{\overline{C_1}}{\overline{\overline{C_3}}} + \frac{\overline{C_1}wc}{A\phi\lambda^2} + e^{\lambda(z_1+z_2+z_3)}\frac{X_k\overline{C_1}}{XH\lambda}$$

and

$$b_1 \equiv \frac{(E_2 + D_2)}{\overline{C_2}} + e^{\lambda z_2}\frac{E_2}{\overline{\overline{C_3}}} + \frac{(E_3 + D_3)}{\overline{C_3}} + e^{\lambda(z_2+z_3)}\frac{X_kE_2}{XH\lambda} + e^{\lambda z_3}\frac{X_kE_3}{XH\lambda} + \frac{1}{wcXH}$$

Moreover, since $\frac{dz_1}{dp_2} = \frac{b_1}{a_1}$, then,

$$\sum_{l=1}^{L}\frac{dz_l}{dp_2} = \frac{\overline{C_1}wc}{A\phi\lambda^2}\left(1 - e^{-\lambda Z_4}\right)\frac{dz_1}{dp_2} > 0$$

Now the proof for $L = 2$. The system of first-order conditions is:

$$\exp(-\lambda z_2) = \frac{wc}{\phi A\left(1 + Bk_L^{\beta_L}\right)\lambda}$$

$$\lambda P_1 + P_1^{\sigma_1}\left(w(cz_1+1)\right)^{-\sigma_1} wc - \phi A\lambda = 0$$

$$\left(\exp(\lambda z_1)P_1^{\sigma_1}\left(w(cz_1+1)\right)^{-\sigma_1} wc - \frac{wc}{\left(1+Bk_L^{\beta_L}\right)}\right)\frac{\beta_L B k_L^{\beta_L-1}}{\lambda} = p_L$$

Note that the first two equations are the same as for $L = 4$ and the last equation is analogous. Accordingly, differentiating with respect to p_2 delivers, and in compact form, (i.e., written analogously to the $L = 4$ case, expressed as function of $\overline{C_l}$ and D_l),

$$\overline{C_1}\frac{dz_1}{dp_2} = \frac{A\lambda}{wc}\frac{d\phi}{dp_2} \tag{89}$$

$$\lambda\frac{dz_2}{dp_2} = \frac{1}{\phi}\frac{d\phi}{dp_2} + \frac{X_k}{X}\frac{dk_L}{dp_2}$$

$$e^{\lambda z_1}\overline{C_1}\frac{dz_1}{dp_2} + \left(\frac{X_k}{X^2} + \frac{\lambda}{wc}\frac{p_L}{X_k k_L}(\beta_L - 1)\right)\frac{dk_L}{dp_2} = \frac{\lambda}{wcX_k}$$

I will write the above equation simply as,

$$e^{\lambda z_1}\overline{C_1}\frac{dz_1}{dp_2} - H\frac{dk_L}{dp_2} = \frac{\lambda}{wcX_k}$$

and the constraint,

$$-\frac{q}{A\lambda}\frac{X_k}{X^2}\frac{dk_L}{dp_2} = e^{\lambda z_1}\frac{dz_1}{dp_2} + e^{-\lambda z_2}\frac{dz_2}{dp_2}$$

Following the same steps as for $L = 4$, I obtain,

$$\frac{dz_1}{dp_2} = \frac{1}{wcXH}\frac{1}{\left(1 + e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2} + e^{\lambda z_1}\frac{\overline{C_1}}{H}\frac{X_k}{X\lambda}\right)} > 0$$

and

$$\frac{dz_2}{dp_2} = \frac{1}{wcXH}\frac{-1 + \left(1 - e^{-\lambda Z_2}\right)\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2}}{\left(1 + e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2} + e^{\lambda z_1}\frac{\overline{C_1}}{H}\frac{X_k}{X\lambda}\right)}$$

Note that $\frac{dz_1}{dp_2} > 0$ implies by Equation 89 that $\frac{d\phi}{dp_2}$: marginal costs are decreasing in p_2.

Moreover, similarly to the $L = 4$ case, using the relationships implied by the production function, one can obtain,

$$-\frac{X_k}{X\lambda}\frac{dk_L}{dp_2} = \left(1 + e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2}\right)\frac{dz_1}{dp_2} < 0$$

This equation shows that knowledge at the lower layer and IT capital are substitutes, as suggested by the intuition in the main text on the knowledge trade-off, involving Equation 23. Moreover, rewriting the last equation as $\frac{dz_1}{dp_2} = \frac{-1}{\left(1+e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2}\right)}\frac{X_k}{X\lambda}\frac{dk_L}{dp_2}$, I obtain, after some manipulations,

$$\frac{dz_2}{dp_2} = \frac{X_k}{X\lambda}\left[\frac{e^{-\lambda Z_2}}{\left(1 + e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2}\right)}\right]\frac{dk_L}{dp_2} < 0$$

Overall, IT capital induces reallocation of knowledge across layers: there is a decrease in production knowledge, whereas CEO knowledge increases. Note that while $\frac{dz_1}{dp_2}$ moves in opposite direction to IT capital, $\frac{dz_2}{dp_2}$ moves in the same direction. Put differently, the increase in IT capital use as a consequence of its price decline, makes low layer wages decline and CEO knowledge increase.

I conclude with the response of total knowledge to the price decline, $\sum_{l=1}^{L}\frac{dz_l}{dp_2}$, which implies following similar steps to the above case, $L = 4$:

$$\sum_{l=1}^{2}\frac{dz_l}{dp_2} = \frac{1}{wcXH}\frac{\left[1 - e^{-\lambda Z_2}\right]\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2}}{1 + e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2} + e^{\lambda z_1}\frac{\overline{C_1}}{H}\frac{X_k}{X\lambda}} > 0$$

which shows that total knowledge falls as a consequence of reorganization responses to IT price declines. What is the intuition for what happens to total knowledge? Summing $\frac{dz_1}{dp_2}$ and $\frac{dz_2}{dp_2}$, as functions of $\frac{dk_L}{dp_2}$,

$$\sum_{l=1}^{2}\frac{dz_l}{dp_2} = \left(\frac{-1}{\left(1 + e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2}\right)} + \frac{e^{-\lambda Z_2}}{\left(1 + e^{-\lambda Z_2}\frac{1}{\phi}\frac{wc\overline{C_1}}{A\lambda^2}\right)}\right)\frac{X_k}{X\lambda}\frac{dk_L}{dp_2} =$$

which shows that the knowledge decline at the bottom is larger (in absolute value) than the CEO knowledge increase, and total knowledge decreases:

$$= \frac{e^{-\lambda Z_2} - 1}{1 + e^{-\lambda Z_2} \frac{1}{\phi} \frac{wc\overline{C_1}}{A\lambda^2}} \frac{X_k}{X\lambda} \frac{dk_L}{dp_2} > 0$$

□

B.2.5 Proof of Proposition 4

An analogous process to that for Proposition 3, together with the first-order condition associated with q, yields the results in the main text. Defining $q \equiv \bar{q}$, output before p_2 changes and $q \equiv q^*$ optimal output after p_2 changes, the combination of both propositions delivers Equation in the main text:

$$\frac{1}{F(Z_L)} \frac{dF(Z_L)}{dp_2}\Big|_{q=\bar{q}} = \frac{1}{F(Z_L)} \frac{dF(Z_L)}{dp_2}\Big|_{q=q^*} - \frac{\rho}{MC_L} \frac{dMC_L}{dp_2}\Big|_{q=q^*} \tag{90}$$

where $\Psi \equiv \exp(-\lambda Z_L) \left[\frac{\frac{C_1 wc}{A\phi\lambda} + e^{\lambda Z_3} \frac{X_k C_1}{XH} + \lambda\left(1 + e^{\lambda z_1} \frac{C_1}{C_2} + e^{\lambda(z_1+z_2)} \frac{C_1}{C_3}\right)}{e^{\lambda Z_3} \frac{X_k C_1}{XH} + \lambda\left(1 + e^{\lambda z_1} \frac{C_1}{C_2} + e^{\lambda(z_1+z_2)} \frac{C_1}{C_3}\right) + e^{-\lambda Z_L} \frac{C_1 wc}{A\phi\lambda}} \right]$. □

B.3. A Brief Introduction to Capital-Labor Elasticity Definitions

The elasticity of substitution was originally introduced by Hicks (1932) for the purpose of analyzing changes in the income shares of labor and capital. Hicks's key insight was that the effect of changes in the capital-labor ratio on the distribution of income, for a given output, can be completely characterized by a scalar measure of curvature of the isoquant. This measure is the two-variable elasticity of substitution.

To extend the Hicksian elasticity concept to multiple production inputs in some set I, there are different schools of thought on the appropriate elasticity. The simplest measure is the direct elasticity of substitution, which assumes that the other factors' quantities in the production function are fixed. Another measure, probably the most popular one, is the Allen (partial) elasticity of substitution (AES),

$$\sigma_{ij}^A \equiv -C(p, y) \frac{\frac{dlogx_i}{dlogp_j}}{x_i x_j} \tag{91}$$

where $C(p, y)$ is the unit cost function given output y and input-price vector p, with p_i for input i, and x_i is the conditional "input i" demand. A drawback of this measure is that it does not have a straightforward interpretation, except in its relation to the input demand elasticities. A third measure is the Morishima elasticity of substitution (ME),

$$ME_{ij} = \sigma_{ij}^{M} \equiv -\frac{dlog\left(\frac{x_j}{x_i}\right)}{dlog\left(\frac{p_j}{p_i}\right)} \tag{92}$$

Blackorby and Russell (1989) argue this is the "most sensible generalization of the Hicks elasticity of substitution because: (i) it is a measure of ease of substitution, (ii) is a sufficient statistic for assessing, quantitatively as well as qualitatively, the effects of changes in price ratios on relative factor shares, and (iii) is a logarithmic derivative of a quantity ratio with respect to a price ratio"; this latter interpretation of an elasticity of substitution was originally proposed by Robinson (1933) for production functions with two inputs, and the ME is the multiple input equivalent. The ME fixes output, but all inputs are allowed to adjust. Importantly, in its simplest form, requires that only the j-th price, in the ratio p_j/p_i, varies.[35] This implies the ME is naturally not symmetric, $\sigma_{ij} \neq \sigma_{ji}$.

In the case of two factors, the AES and ME coincide. In particular, for a two factor CES production function, $AES = ME = \sigma$, where σ is the CES parameter. On the other hand, for multiple inputs, the AES and ME are, in general, different and have properties as described above. ε_{IT,n_1} in the main text is closely related to the ME, except unlike in the latter, changes to one input price (IT), affects all other prices (wages).

C. Quantitative Appendix

C.1. Quantitative Implementation

For each parameter vector guess, I find the equilibrium using a procedure similar to the inner-loop in the algorithm in Burstein and Vogel (2017): (1) Guess aggregates (P, M, R), (2) solve firms' profit maximization given L, which contains six, nine and

[35] More generally, the ME can also be defined for price changes in noncoordinate directions using directional derivatives, as Blackorby and Russell (1981) show, which is helpful in terms of having a mapping between the ME for any other production function and my results.

twelve choice variables for $L = 2, 3, 4$ respectively, and choose the L that maximizes profits, and (3) obtain the implied aggregates using the results in step (2) and compare them to their values in step (1); if they are closer than a standard tolerance measure in percentage terms across the subsequent iterations, equilibrium has been found, and otherwise, use the values obtained at the last iteration as initial guess and repeat the sequence.

While the code is written in MATLAB, I use the solver KNITRO, which is reliable and an order of magnitude faster than Matlab's Optimization Toolbox. To improve the performance of the solver, I supply analytical first derivatives, which are coded automatically using MATLAB's Symbolic Toolbox. Furthermore the code is executed in parallel.

C.2. Calibration and Data Map

C.2.1 IT and non-IT Prices To measure the current cost values of different types of assets, Eden and Gaggl (2018) aggregate BEA's industry-level estimates from the BEA detailed fixed asset accounts. Broadly, there are three types of capital: residential assets, consumer durables and nonresidential assets. Within both the nonresidential and consumer durables categories, there are IT and non-IT assets.

Their definition of IT is the following. Within nonresidential assets, software (classification codes starting with RD2 and RD4) and equipment related to computers (codes starting with EP and EN). Within consumer durables, PCs and peripherals (1RGPC); software and accessories (1RGCS); calculators, typewriters, other information equipment (1RGCA); telephone and fax machines (1OD50). Non-IT are the complementary categories. The aggregation in Eden and Gaggl (2018) produces capital measures that are similar to those in the Groningen Growth and Development Centre, the EU KLEMS Growth and Productivity Accounts, and the Conference Board's Total Economy Database. TO calibrated changes in p_l for $l = 1, 2$, I use rental rates from their Figure 3B. For p_1, prices are flat. For p_2, to avoid measurement error, I use the average of the 10 years before 1980 as initial price and year 2013 (their last observation) as the final price; this delivers a conservative estimate of -2/3 for the price change relative to the initial price. Note this definition of IT differs from that in Koh, Santaeulàlia-Llopis, and Zheng (2018) for intellectual property products (IPP). For IPP, prices are similar to the non-IT prices, that is, constant.

C.2.2 Mapping of Occupations to Hierarchy Layers The mapping of occupations to layers follows Caliendo, Monte, and Rossi-Hansberg (2015) and allows an empirical mapping from the theory in this paper to the data, focusing on $2 \leq L \leq 4$. They separate workers according to their hierarchical level in the organization, that is, on the basis of the number of layers of subordinates that employees have below them. In their French manufacturing data, the occupational classification is named PCS-ESE and includes five occupational categories as presented in Table XII. Throughout the paper, they merge classes 5 and 6, since the distribution of wages of workers in these two classes is remarkably similar, indicating similar levels of knowledge.

TABLE XII: MAP FROM OCCUPATIONAL DATA TO THEORY

Task-based Category	Knowledge-based Category	U.S. Occupational Category Description (used in task-based approach)	French Occupational Category Description (used in knowledge-based approach)
Nonroutine cognitive	Layer 3	Managerial, professional, and technical occupations	3. Senior staff or top management positions, which includes chief financial officers, heads of human resources, and logistics and purchasing managers
Nonroutine manual	Layer 2	Service occupations	4. Employees at the supervisor level, which includes quality control technicians, technical, accounting, and sales supervisors.
Routine cognitive	Layer 1	Sales, clerical, administrative occupations	5. Qualified and nonqualified clerical employees, secretaries, human resources or accounting employees, telephone operators, and sales employees.
Routine manual	Layer 1	Production, craft, repair, operative occupations	6. Blue-collar qualified and nonqualified workers, welders, assemblers, machine operators, and maintenance workers.

Notes. For the knowledge-based labor categories I follow the layer to occupations mapping in Caliendo, Monte and Rossi-Hansberg (2015) for French data. For the task-based labor categories, I borrow the information reported by Gaggl and Wright (2017), who in turn use the classification from Acemoglu and Autor (2011). The descriptions for the task-based and the knowledge-based labor categories are remarkably close.

C.2.3 Mapping of Evidence in Gaggl and Wright (2017) to this Theory

GW study a policy experiment in the UK in which small firms (those with fewer than 50 employees) were granted a 100% first year tax allowance on IT investments. They identify the effect of the policy on firm and worker outcomes using a regression discontinuity (RD) design around the policy threshold. For the worker results, they use the UK Annual Survey of Household Earnings (ASHE), an annual representative 1% sample of workers from National Insurance records for the working population, which contains information on earnings and hours worked as well as occupation of UK workers. The survey also includes the number of employees associated with each worker's firm, which allows to implement the RD design.

To map their worker-level results, which use a task-based labor classification, to the labor types in this knowledge-based theory, I use occupational classification descriptions. As shown on Table XII in Appendix Section C.2.2, their routine vs nonroutine labor classification, which they borrow from Acemoglu and Autor (2011), is remarkably close to that in Caliendo, Monte, and Rossi-Hansberg (2015) for layers. I map routine labor (using the reported employment weights) to layer 1, nonroutine manual to layer 2, nonroutine cognitive to layer 3. Bundling the two routine labor types into layer 1 is the result of the Caliendo, Monte, and Rossi-Hansberg (2015) mapping for France but my reported results are not a consequence of this aggregation. Both routine labor categories independently deliver results consistent with the model. Whenever firm shares are needed, I use values for treated firms from Table 1 of descriptive statistics. One caveat is that GW do not report results for owner's/CEO, so I assume them in layer 3 together with the top management positions, and do not report results for CEOs specifically.

Wages by layer come from GW, Table 4. The next two moments in Table II are ratios of layer to firm-level variables. I construct them with estimates (in levels) from Table 4, by first converting the layer-level treatment effects to percentages to which I subtract the firm total. For a variable x_l, I compute $\frac{\tau_{x_l}}{\overline{x_l}}$, where τ_{x_l} is the estimated treatment effect from the RD, reported in levels, and $\overline{x_l}$ is the average value for the variable for treated firms, from Table 1. For the share of layer l workers hours in total firm hours, I calculate $\frac{\tau_{x_l}}{\overline{x_l}} - \frac{\tau_x}{\overline{x}}$, where $\frac{\tau_x}{\overline{x}} \equiv \sum_{l=1}^{L} \frac{\tau_{x_l}}{\overline{x_l}} \frac{\overline{x_l}}{\sum_{l=1}^{L} \overline{x_l}}$, and $x_l = t_l v_l$, with t_l and v_l respectively denoting hours per worker and number of workers at layer l. With these definitions, total hours in a layer satisfies $n_l = t_l v_l$ both in the model and the data. I proceed analogously for the third row in Table II, the wage bill share of layer l in the firm wage bill, defining instead $x_l = w_l n_l$ with $\frac{\tau_x}{x}$ adjusted accordingly.

Firm-level RD effects for investment by capital type are obtained using the Quarterly Capital Expenditure Survey (QCES), from the UK Office of National Statistics, which provides IT and other types of capital spending. I construct the change in the IT capital-labor ratio as the difference in the growth rate of IT capital (investment) and total worker hours. For the former, I use software and hardware policy results (Table 2) and aggregate them using their shares in investment (Table 1) for treated firms. For total worker hours, I construct $\frac{\tau_x}{\bar{x}}$, using the aforementioned methodology and results for $x_l = t_l n_l$. Finally, for the response of total knowledge, I compute $\sum_{l=1}^{3} \tau_{w_l}$ using GW, Table 4, which ignores the CEO layer since they do not report results for this layer specifically, and hence the change in Z_L underestimates the true change.

C.3. Decomposition of IT Capital-to-Production-Labor Elasticity.

A quantitative decomposition, focusing only on optimal knowledge decisions, highlights the specific organizational channels that determine the results for ε_{k_2,n_1} in Figure IV. Mathematically, I can approximate[36] ε_{k_2,n_1} for $L = 2$:

$$\varepsilon_{k_2,n_1}\big|_{L=2} \approx \frac{W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} + O_1^L(\lambda, \beta_L)\frac{\partial z_1}{\partial \log(p_2)} + O_L(\lambda, \beta_L)\frac{\partial z_2}{\partial \log(p_2)}}{1 - \varepsilon_{w_1,p_2}} \tag{93}$$

whereas for $L = 3, 4$:

$$\varepsilon_{k_2,n_1}\big|_{L=3,4} = \frac{W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} + O_1^L(\lambda, 1)\frac{\partial z_1}{\partial \log(p_2)} + W_2(\sigma_2)\frac{\partial z_2}{\partial \log(p_2)}}{1 - \varepsilon_{w_1,p_2}} \tag{94}$$

[36]The elasticity expression for $L = 2$ is an approximation. Instead of using the actual factor demand given by $k_2 = \left[\frac{\exp(-\lambda z_1)}{AB[1-\exp(-\lambda Z_L)]} q - \frac{1}{B}\right]^{1/\beta_L}$, I use expression $k_2 = \left[\frac{\exp(-\lambda z_1)}{AB[1-\exp(-\lambda Z_L)]} q\right]^{1/\beta_L}$. In my calibration, the omitted term $1/B$ is close to zero at 0.2.

where

$$W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} \equiv -\sigma_1\left[\frac{\partial w_1}{\partial \log(p_2)} - \frac{\partial \log(P_1)}{\partial \log(p_2)}\right] < 0$$

$$W_2(\sigma_2)\frac{\partial z_2}{\partial \log(p_2)} \equiv \sigma_2\left[1 - \frac{\partial \log(P_2)}{\partial \log(p_2)}\right]$$

$$O_1^L(\lambda, \beta_L) \equiv \left(1 + \frac{1-\beta_L}{(\exp(\lambda Z_2) - 1)}\right)\frac{\lambda}{\beta_L} > 0$$

$$O_L(\lambda, \beta_L) \equiv \frac{(1-\beta_L)}{(\exp(\lambda Z_2) - 1)}\frac{\lambda}{\beta_L} > 0$$

The intuition of Equations 93 and 94 are best understood by looking at the denominator and the numerator separately. The denominator is $1 - \varepsilon_{w_1,p_2}$, and is close to one for all L by Table III in Section III.C in the main text. Hence, the heterogeneity in ε_{k_2,n_1} comes almost exclusively from the numerator of those equations, to which I turn next.

There are two types of terms in the numerator of Equations 93 and 94: (i) related to within layer l substitution, denoted by W_l, and (ii) related to organizational choices in layer l, denoted by O_l. In those equations, the first term, $W_1(\sigma_1)$, is common to all organizations, as it is the within-layer 1 capital-labor substitution. It is always negative because, absent knowledge reorganization ($\frac{\partial z_1}{\partial \log(p_2)}$), it is the standard factor substitution effect: increased n_1 as w_1 falls with intensity governed by σ_1. The second term captures the knowledge reorganization effects of z_1 and $O_1^L(\lambda, \beta_L)$, which depends on λ, and the problem CDF, $F(.)$.

Across Equations 93 and 94 the third term is different, but is always connected to changes in z_2. More specifically, this third term is (i) for $L = 2$, $O_L(\lambda, \beta_L)$, and, analogously to O_1^L, captures cross-layer reorganization effects and it depends on λ and β_L; instead (ii) for $L > 2$, the function $W_2(\sigma_2)$ captures within-layer substitution and depends on σ_2, the within-layer 2 capital-labor elasticity.[37]

Table XIII shows the percent contribution of each term in the numerator of Equations 93 and 94, across organizations for the median α. The decomposition shows that $O_1^L(.)\frac{\partial z_1}{\partial \log(p_2)}$ always plays the largest quantitative role across all organizations. The IT capital-to-production-labor elasticity is quantitatively determined by the endogenous response of z_1 to p_2 together with λ, not by the standard parameters governing substitution, σ_l. Hence, this indirect substitution mechanism would not be a relevant

[37] Contrary to W_1 it cannot be signed a priori because in layer 2, both $p2$ and w_2 change, and W_2 depends on both.

implication of the theory if empirically we found the opposite (i.e., if $\frac{\partial z_1}{\partial \log(p_2)} < 0$). As I showed when reviewing empirical literature on the IT effects in Section III.B, there is ample evidence supporting that production-worker wages decline with IT adoption, suggesting the indirect IT capital-to-production-labor substitution is at work.

TABLE XIII: DECOMPOSITION OF THE NUMERATOR OF EQUATIONS 93-94.

	Organization		
	$L = 2$	$L = 3$	$L = 4$
$W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)}$	-1	-2	-2
$O_1^L(\lambda, \beta_L)\frac{\partial z_1}{\partial \log(p_2)}$	101	97	97
$O_L(\lambda, \beta_L)\frac{\partial z_2}{\partial \log(p_2)}$	0	–	–
$W_2(\sigma_2)\frac{\partial z_2}{\partial \log(p_2)}$	–	4	5

Notes. The table decomposes the numerator in Equations 93 and 94 into each of the components. Results for the median α given L using the calibration in Section IV with only IT price changing between 1980 and 2015. Values in percentages relative to total, rounded to the nearest integer.