Firm Organization and Information Technology: Micro and Macro Implications*

Asier Mariscal

U. Carlos III-Madrid

December 5, 2018

The last 30 years in the US have seen: i-A rise in inequality with falling wages for the less skilled workers, ii-A notable decline in the price of Information Technology (IT) capital, and, iii-A reduction in the labor share (LS) of national income. I study a model with capital, labor and knowledge choices by firms organized as hierarchies. There are four key results. First, IT capital increases the managerial span, and a decrease of it's price reallocates knowledge away from the plant and towards managerial layers. Second, the theory is consistent with capital-labor complementarity, for each skilled and unskilled labor, as in the data; In spite of it, the calibrated model features long-run IT capital plant-labor substitution as a consequence of optimal reorganization choices. Third, organizational knowledge and IT capital are complementary choices, and increase firms' measured TFP, as in the empirical literature. Fourth, to understand recent macro trends, I perform two counterfactuals. A decrease in the demand elasticity is required for value added concentration in large firms and the decline of the aggregate labor share. On the other hand, the decline in the IT price explains the decline of plant-workers' wages and LS, and the increase of both for managers.

*I thank participants at the Annual Meetings of Society for Economic Dynamics 2018, RIDGE2018, REDg2018, and, Pol Antràs, Thomas Chaney, Antonia Díaz, Andres Erosa, Jesús Fernandez-Huertas, William Fuchs, Luis Garicano, Matthias Kredler, David Lagakos, Ellen McGrattan, Ben Moll, Ezra Oberfield, Veronica Rappoport, Devesh Raval, B. Ravikumar, Esteban Rossi-Hansberg, Raul SantaEulalia-Llopis, Anna Sanz de Galdeano and Eric Verhoogen for useful comments. I gratefully acknowledge financial help from UC3M and U.Alicante. The usual disclaimer applies.

1 Introduction

During the last 30 years there has been an intense adoption of Information and Communication Technology capital, IT, strongly impacting the labor market. In the US in 1998, Wilson (2006) reports computers alone to represent 30% of firms' investment, and BEA data shows it's price today is a third of that in 1980, Eden and Gaggl (2018). At the same time, inequality in the US is rising with the surprising feature that wages for the less skilled workers have fallen 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, IMF (2017). Many of these labor market trends are not unique to the US but shared by other European countries, see for example Autor (2010) and Karabarbounis and Neiman (2013). Similarly general is the slowing down of TFP growth during this period, OECD (2017).

This paper rationalizes these and several other micro and macro facts, within a unified framework that zooms into firms' organizational choices. In the last 30 years, the reorganization of the typical white-collar workplace has been notably visible as firms have responded to the introduction of computers, email and other internal communication technologies. Big office layouts with secretaries and clerks using typewriters, carbon paper and file cabinets have been replaced with computers and software. Simultaneously, many plant-workers have seen their autonomy, understood as independent problem solving, narrowed down by firms information systems.

An Amazon warehouse is an excellent illustration of this new organization of work. Plant-workers serving an order must deal with questions like: What is the right packaging? In which shelves are the ordered products located? How many products should I pick? Where is the order delivered? Today, answers to all these questions are provided by a computerized information system. The right box is chosen for them. Neither do workers need to find the ordered products, shelves will glide to them. If the order requires nineteen units of an item, the Jennifer headset will request the worker to pick five, then five, then five and, finally, four items. As the package leaves the premises, the address is automatically stamped on the exit belt. In this example, IT capital controlled by managers reduces the autonomy of plant-workers to an extreme degree, and, as a consequence, Amazon reduces its costs by paying plant-workers lower wages than otherwise. More generally, on Appendix Section 8.1, I document how truck drivers, low-level lawyers, or customer services have been reorganized by reducing plant-workers' autonomy, while managers fine-tune firms' information systems, suggesting these trends occur across industries.

The knowledge-based hierarchy literature is a natural framework for thinking about optimal problem solving within organizations. It was developed by Rosen (1982), and Garicano (2000). In the latter, producing requires time and knowledge. Plant-workers use their time to produce, and generate "problems", or potential production. Output in turn requires those problems to be solved. Plant-workers have knowledge that allows them to solve some of the generated problems, which realizes output. The unsolved problems are passed to managers, who solve those they know and also have the recourse of higher-level managers, until the CEO is reached. Since managerial time is fixed, using it optimally implies managers deal with unfrequent, non-routine problems. In this framework, optimal firm organization is determined by optimal employee knowledge at each layer, and the number of layers. I am closer to Caliendo and Rossi-Hansberg (2012), who study organization of firms with heterogeneous demand. Relative to them, I add optimal capital choices to understand the illustrated firm (re)organization, it's effects on workers' wages, and to draw conclusions about recent macroeconomic trends.

In this paper, firms' organizational problem requires choosing knowledge, labor and capital at each layer of the hierarchy, as well as the number of layers. Similar to the Amazon example, plant-workers use their time and (non-IT) capital to generate potential production, hence problems. Those unsolved at the plant arrive to the first managerial layer, who use IT capital to solve them, a process that potentially arrives up to the CEO. In this model, IT capital is the way that managers and the CEO leverage their fixed available time to solve problems. Bloom et al. (2014) and Garicano and Rossi-Hansberg (2006) argue that there are two types of IT when it comes to their wage inequality consequences. First, IT that decreases knowledge access costs, like computers, and second, IT that decreases communication costs. While I use a more coarse concept of IT than them, I study it's optimal adoption which allows quantification through a natural mapping to more standard production functions. Within this framework, I can study firms' optimal reorganization as IT prices decline, as they do in the data. In the model, cheaper IT capital alters the relative price of problem solving at different layers of the firm, promoting the use of managers instead of plant-level workers. Firms economize on knowledge at lower layers with numerous workers, as it becomes cheaper to solve problems by managers and IT. This is one of the main results of the paper: IT reduces wages of plant-workers and increases wages of managers.

I relate to a macro literature that studies the effects of technology on labor markets. Acemoglu and Restrepo (2018) study the race between automation and the creation of new tasks. They conclude that automation always reduces the aggregate LS and inequality increases during the transition through both channels. Their model can be viewed as a combination of directed technological change models that allow capital accumulation[1], with task-based theories of the labor market, as in Acemoglu and Autor (2011)[2]. A close mapping exists between the latter and my model: a task in that framework is a problem in mine. Hemous and Olsen (2016) also combine both of these literatures. In their paper, as the economy develops and unskilled wages increase, there is endogenous automation of unskilled labor and an increasing skill premium.

Interestingly, while these papers assume automation capital-labor substitution, I impose capital-labor complementarity at all layers of the firm, and still obtain IT capital-plant labor are gross substitutes in the long run; this is the second key result of the paper. This indirect substitution away from the unskilled due to lower IT capital prices, emerges from firms' optimal organization choices and, in particular, from less problem solving at the plant-level. It is a long run elasticity because I focus on 1980-2015 changes, consistent with an analysis across two steady states, as in Karabarbounis and Neiman (2013). Moreover, it is consistent with the declining routine wages due to adoption of Broadband Internet in Norway, Akerman et al. (2015), and, due to more IT investments induced by tax deductions in the UK, Gaggl and Wright (2017). And for the US, with the real wage decline of all percentiles below the median of the wage distribution of the largest firms in the last 30 years, Song et al. (2015).

The third key result of the paper is that firms' organizational knowledge and IT are complementary choices, as in Bresnahan et al. (2002) and Bloom et al. (2012). Put differently, IT adoption makes total firm knowledge endogenously increase as a consequence of two effects: the knowledge response for fixed output, plus the response due to the firm scale adjustment. In the former, firm reorganization due to cheaper IT reduces the cost of dealing with a larger mass of problems relative to actually solving

[1] On directed technical change models, see Acemoglu (1998), Acemoglu (2002), Acemoglu (2003), Acemoglu (2007), Kiley (1999), Caselli and Coleman (2006), Thoenig and Verdier (2003) among others.

[2] There is an ample literature studying the effects of trade on inequality an issue with which I do not deal. See, for example,Feenstra and Hanson (1999), Spitz-Oener (2006), Goos et al. (2014), Grossman and Rossi-Hansberg (2008), Autor and Dorn (2013), Rodriguez-Clare and Ramondo (2010), Acemoglu et al. (2010), and many others.

them with workers' knowledge. Put differently, for fixed output, these two factors act like substitutes, so total firm knowledge declines. On the other hand, increasing firm scale due to lower marginal costs implies that more of all factors are used, including knowledge. Overall, total firm knowledge is complementary to IT adoption. This result highlights that correct TFP measurement requires the joint study of IT, worker knowledge and firm organization, as the firm-level empirical literature finds.

I also relate to the literatures on the decline of the labor share of GDP, e.g. Karabarbounis and Neiman (2013), the increased value added concentration, Autor et al. (2017), and the increase in the aggregate markup, DeLoecker and Eeckhout (2017). The model brings production-side discipline to the understanding of these trends, which have recently received explanations based on the elasticity of demand. My baseline calibration targets micro moments, and uses two sources of variation: i-an IT price decline and ii-a demand elasticity reduction. The latter so that markups increase more for larger firms, like DeLoecker and Eeckhout (2017) find in the data. In order to disentangle the role of supply from demand in generating the aggregate trends, I conduct counterfactuals to isolate the role of each channel. I find as a consequence of IT: 1- within-firm wages decline at low layers and increase at top layers; 2-within-firm, employment share of managers increases; 3-IT capital-labor ratios increases; and, at the aggregate level, as a share of GDP: 4-LS for managers increases, 5-LS for plant-level workers decreases. On the other hand, the decline in the elasticity of demand generates: 6-the decline in the LS of value added of large firms, as in Kehrigy and Vincent (2014), and 7-the decline in the aggregate LS. Hence, the fourth key result of the paper is that the latter two results are due to increasing markups, whereas inequality is due to IT adoption.

I connect to a macro literature that studies the role of technology change on the aggregate labor share decline. Oberfield and Raval (2014) use micro data to estimate a capital-labor elasticity lower than one and conclude that technical change explains the decline in the labor share. Grossman et al. (2017) jointly explain the decline in the labor share and the productivity slowdown, while respecting capital-labor complementarity. Koh et al. (2018) argue that recent national accounting changes to measurement of intellectual property product (IPP) capital contribute substantially to the decline in the aggregate labor share. I also relate to Caselli and Manning (2018), who study the role of technology and capital on wage declines.

The empirical evidence on firm organization provides credibility to the mechanisms in my model. Caliendo et al. (2015) provide evidence based on French linked employer-employee data that is consistent with the predictions in Caliendo and Rossi-Hansberg (2012)'s theory, crucially those on wages and layers. Caliendo et al. (2016), further show that quantity and revenue-based TFP changes when firms add layers are also consistent with their firm organization theory[3]. My theory is consistent with the evidence on the substitution of routine tasks by computers in Autor et al. (2003), as in the Amazon example, and the complementarity between IT adoption, skills and firm reorganization provided by Bresnahan et al. (2002) and Caroli and Reenen (2001). My results also underscore the importance of management practices for the productivity effects of IT as found by Bloom et al. (2012).

This paper has 7 sections. On Section 2, I introduce the model. On Section 3 I solve the model and on Section 4, I provide intuitions for the effects of IT on firm organization. On Section 5, I calibrate the model using micro moments, and where both IT prices and demand elasticity change. Section 6 describes the results from the calibrated model for both micro and macro facts, including moments untargeted in the calibration, and also presents the results for two counterfactuals, each of which fixes one of the channels, while allowing the other to vary. Section 7 concludes.

2 Model

2.1 Preferences and workers

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. They maximize the straightforward generalization of the familiar CES utility,

[3] See also Antràs et al. (2006) and Antràs et al. (2008) study the implications of offshoring on wage inequality using a similar model of organizations. Bernard et al. (2018) study the effects of impact of offshoring on firms' decisions to reallocate labor from production work to technology and innovation-related occupations. Also Marin and Verdier (2012) and Marin and Verdier (2014) for the connection between offshoring and decentralization of decisions in hierarchy models. See Fuchs and Garicano (2010) for an "in-house vs outsource" analysis in the context of hierarchies and Fuchs et al. (2015) for a study on optimal contracting in hierarchies with informational problems. Altomonte et al. (2018) study detailed business group data. In their model, firms' headquarters choose whether to vertically integrate the production of intermediates, outsource them to a branch, or produce through a controlled affiliate external to the legal boundaries of the headquarter. Gumpert (2017) studies the role of communication costs on how a multinational firm is organized. Santamaria (2017) spatial sorting on inequality through two channels, spatial differences in technology and endogenous organization of production.

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

where Q is total consumption (utility), and for each variety ω, $q(\omega)$ denotes the quantity consumed, $\alpha(\omega)$ is the product-appeal shifter, and $\rho(\omega) > 1$ is the elasticity of substitution. Equation 1 is the Constant Relative Elasticity of Income and Substitution (CREIS) utility function proposed by Lashkari and Mestieri (2017). This preferences simply allow $\rho(\omega)$, the elasticity of demand for good ω, to be non-constant across ω. Still, it's implied inverse demand for ω is

$$p(\omega) \propto q(\omega)^{-1/\rho(\omega)} \alpha(\omega)^{1/\rho(\omega)} \tag{2}$$

as in a standard CES utility, and nice aggregation properties are preserved[4]. In fact, when $\rho(\omega) = \rho$, $\forall\omega$, the above collapses to CES. I use CREIS instead of CES, so that the quantitative section is, in a simple way, consistent with the evidence on markup heterogeneity in DeLoecker and Eeckhout (2017), to instead focus 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, it's demand quantity is larger. Each firm will produce one ω and have a different $\alpha(\omega)$, which is the source of exogenous firm heterogeneity in this model.

In order to consume, agents use their available unit of time in the labor market, and obtain a wage. They can work in production or in the education sector, and in equilibrium they will earn w in both, regardless of the occupation they sort into. Agents that acquire knowledge to produce, bear a training cost of learning their occupation which firms exactly compensate for, leaving them indifferent across sector and occupations with a net wage of w. More precisely, the cost of acquiring knowledge z is assumed to be wcz. It is proportional to w because learning requires teachers in the schooling sector who are paid w. In the next subsection, I introduce the benefits of knowledge, which arise from firms' production choices.

[4]More generally, CREIS also allows non-constant income elasticities across ω, a feature I do not use in this paper.

2.2 Firms

Entrepreneurs obtain a draw from CDF $G(\alpha)$, which determines the level of demand for their product. Organizing production implies choosing optimally the number of managerial layers and the amount of each of three factors at each layer, namely, capital, k_l, labor, n_l, and knowledge, z_l. Labor and capital is used both in production at the plant and by managers solving problems.

The number of layers satisfies $1 \leq L \leq 3$, a range chosen following the empirical description of firm organization in Caliendo et al. (2015). Within an organization, agents can be broadly grouped into workers and managers. The first are dedicated to production and the second to problem solving. Production workers are combined with non-IT capital at layer 1 and the input bundle is $y_1 \equiv \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{1/\beta_1}$. Per unit of input bundle, one problem arises which if successfully dealt with, yields A produced quantity. Accordingly, total potential output is Ay_1. The higher layers, $l > 1$, are specialized in problem solving and it's managers use IT capital to deal with and solve them.

Each production possibility at layer 1 is associated to a problem drawn from $F(x) = 1 - exp(-\lambda x)$. For a problem z drawn from that density to become realized output, the worker must have a knowledge set that includes such problem. Workers can deal with y_1 problems. 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 immediately above layer. Assume for exposition that such managers exist. In layer 2 managers deal with unsolved problems from $l = 1$, and the required managerial input bundle $y_2 \equiv \left(k_2^{\beta_2} + n_2^{\beta_2}\right)^{1/\beta_2}$ is given by the following restriction,

$$\left(k_2^{\beta_2} + n_2^{\beta_2}\right)^{1/\beta_2} = \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{1/\beta_1} (1 - F(z_1)) \tag{3}$$

The left-hand side (LHS) of Equation 3 is the managerial input at $l = 2$, whereas the right-hand side (RHS) is the amount of problems unsolved at layer 1, which have been passed onto layer 2. To deal with such a volume of problems in layer 2 the firm optimally chooses the managers, their knowledge and IT capital. Assume managers learn interval $[z_1, z_1 + z_2]$, the amount of problems passed onto layer 3 are $y_1(1 - F(z_2))$, i.e. the unsolved problems at all previous layers; those above z_2. More generally, at layer $l > 1$, managers learn $[Z_{l-1}, Z_l]$ with $Z_l = \sum_{k=1}^{l} z_k$, and satisfy $\left(k_l^{\beta_l} + n_l^{\beta_l}\right)^{\frac{1}{\beta_l}} = y_1(1 - F(Z_{l-1}))$ with $\beta_l \equiv \frac{\sigma_l - 1}{\sigma_l}$. The process continues until it reaches the top layer, the CEO. As the process of adding layers and knowledge continues, more of the potential production is realized, such that for a firm with total knowledge Z_L, realized production is $AF(Z_L)y_1$. 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. Section 3 describes the problem and solution of an arbitrary firm with it's full set of constraints.

3 Model Solution

The owner of the firm solves the organization problem in two steps. First, conditional on L, he chooses production factors, capital, labor and knowledge at each layer. Second, he compares profits across organizational modes, i.e. across the different L=1,..,4. On a first step, the problem of an entrepreneur is to solve, for fixed L:

$$\max_{q,\{z_l,n_l,k_l\}_{l=1}^{L}} p(q)q - \sum_{l=1}^{L} (n_l w(cz_l + 1) + p_l k_l) \tag{4}$$

subject to:

$$\begin{cases} p(q) = \left(\frac{q}{\alpha R}\right)^{-1/\rho(\alpha)} P^{\frac{\rho(\alpha)-1}{\rho(\alpha)}} \frac{M}{m(\alpha)} \\ q = A[1-\exp(-\lambda Z_L)] \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{\frac{1}{\beta_1}} & : \phi \\ \left(k_l^{\beta_l} + n_l^{\beta_l}\right)^{\frac{1}{\beta_l}} = \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{\frac{1}{\beta_1}} \exp(-\lambda Z_{l-1}), L > l > 1 & : \psi_l \\ 1 + Bk_L^{\beta_L} = \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{\frac{1}{\beta_1}} \exp(-\lambda Z_{L-1}) & : \psi_L \\ n_L = 1 \\ z_l > 0, \forall l \geq 1 & : \theta_l \end{cases} \tag{5}$$

where $m(\alpha) \equiv \frac{\rho(\alpha)}{\rho(\alpha)-1}$, and $M \equiv \int \frac{p(\alpha)q(\alpha)}{R} m(\alpha) dG(\alpha)$. In this problem, profits are the objective function, where the first term is the revenue and the second is total cost. The latter is composed of the wage bill and capital spending, where p_l are prices of capital at each layer. The first restriction is the demand curve of the firm for it's α variety. The second is the production function. The rest are the managerial input constraints.

Unlike managerial labor in previous layers, the restriction at the CEO level assumes CEO time is fixed at one unit, which implies there is no within-layer capital-labor substitution. I allow the CEO to use capital, k_L, to relieve his time constraint, although it's use is subject to decreasing returns to scale, $0 < \beta_L < 1$; Section 4.7 discusses the modeling of capital and labor at the top, which is chosen in order to match empirical moments. Previewing here the discussion in that section is useful. Fixing CEO time will make wages grow with firm size conditional on L, whereas allowing capital at the top allows CEOs to increase their span and the average cost increases less steeply with firm size. IT capital increases the managerial span of control by allowing managers and the CEO to leverage their fixed amount of time. Corporations' information systems that relieve the CEO from time consuming activities, allow him to focus on non-routine problems, and manage what is truly exceptional problems.

To capture the idea that managers control the information systems, I assume production workers use production (non-IT) capital, k_1, with price p_1, while managers at layers $l > 1$ use potentially different amounts of IT capital, k_l. In an effort to further discipline the quantitative results, I assume IT capital at any layer has price p_2. With the same intention, I impose $\beta_l = \beta_2$ for $L > l > 1$.

In order to understand the solution, it is instructive to solve the cost minimization problem first. The interested reader can find detailed derivations in the Appendix. By obtaining the first order conditions (FOCs) and combining them one can gain relevant intuitions on the trade-offs faced in organizing production. For example, except at the top layer, the capital-labor trade-off is solved at every layer $l < L$:

$$\frac{k_l}{n_l} = \left(\frac{p_l}{w(cz_l + 1)} \right)^{-\sigma_l} \tag{6}$$

This is the familiar condition arising in two-factor CES production functions with elasticity parameter $\sigma_l \equiv \frac{1}{1-\beta_l}$ with a crucial difference: in this model, knowledge is chosen optimally and hence gross wages are not exogenous to the firm. The top layer is different because CEO time is fixed:

$$1 + Bk_L^{\beta_L} = \left(k_1^{\beta_1} + n_1^{\beta_1} \right)^{\frac{1}{\beta_1}} \exp(-\lambda Z_{l-1}) \tag{7}$$

Literally one can interpret the volume of problems the CEO can tackle to be determined by his time plus the IT capital he uses. In this layer, since labor is not variable at the top, k_L is obtained a "residual", i.e. it does not come from a FOC but from the time constraint of the CEO.

Since a defining element in this theory is knowledge, it is useful to gain intuition of its optimal choice. From FOCs, we obtain that the knowledge trade-off across layers 1 and 2 is:

$$wcn_1 - \psi_2 \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{\frac{1}{\beta_1}} \lambda \exp(-\lambda z_1) = wcn_2 \tag{8}$$

where ψ_l is the multiplier associated with layer l "time" or inputs, and can be interpreted as the cost per extra unit of managerial product at l. Mathematically,

$$\psi_l = \begin{cases} \dfrac{w(cz_l + 1)}{\left(k_l^{\beta_l} + n_l^{\beta_l}\right)^{\frac{1-\beta_l}{\beta_l}} n_l^{\frac{-1}{\sigma_l}}} \\ \dfrac{p_l}{\left(k_l^{\beta_l} + n_l^{\beta_l}\right)^{\frac{1-\beta_l}{\beta_l}} k_l^{\frac{-1}{\sigma_l}}} \end{cases} \tag{9}$$

Equation 10 shows the trade-off in knowledge across layers at the optimum, as a slight rewriting makes apparent. Note that since the multiplier is associated to both capital and labor, there are two channels through which this trade-off operates. In particular, the same equation where $\psi_2 > 0$ is substituted on the "capital channel" becomes,

$$\underbrace{wcn_1}_{\text{MC of } z_1} - \underbrace{\overbrace{\frac{p_2}{\left(k_2^{\beta_2} + n_2^{\beta_2}\right)^{\frac{1-\beta_2}{\beta_2}} k_2^{\frac{-1}{\sigma_2}}}}^{\text{Cost per problem}} \overbrace{\left(k_1^{\frac{\sigma_1-1}{\sigma_1}} + n_1^{\frac{\sigma_1-1}{\sigma_1}}\right)^{\frac{\sigma_1}{\sigma_1-1}} \lambda \exp(-\lambda z_1)}^{\text{Change in Mass of Problems}}}_{\text{Savings in Managerial Inputs}} = \underbrace{wcn_2}_{\text{MC of } z_2} \tag{10}$$

This equation is the first key result of the paper. On the (LHS), an extra unit of knowledge in layer 1 has marginal cost of wcn_1 but allows to spend less on communication costs to layer 2; Such cost savings can be expressed as the cost per problem times the reduction in the amount of problems received in layer 2. At the optimum, the marginal cost of that z_1 knowledge net of the communication costs are equal to the (RHS), ie the marginal cost of knowledge in layer 2. Moreover, this equation illuminates an important insight: A reduction in IT prices p_2, makes layer 2 input relatively cheaper, and tends to increase knowledge at layer 2 and decrease it at layer 1. This mechanism generates inequality within the firm as a consequence of IT and provides the essence of the second key result: the indirect IT capital-plant-labor elasticity substitution, on Section 4.6. It is indirect because I impose capital-labor complementarity at all layers, but organizational choices involve problem solving reallocation across layers that make IT capital and plant-labor gross substitutes.

It is useful to further operate on the FOCs and restrictions of the cost minimization problem to obtain the conditional factor demands,

$$\begin{cases} n_l = \dfrac{\exp(-\lambda Z_{l-1})}{A[1-\exp(-\lambda Z_L)]}\left(\dfrac{(w(cz_l+1))}{P_l}\right)^{-\sigma_l} q, & \forall l < L \\ k_l = \dfrac{\exp(-\lambda Z_{l-1})}{A[1-\exp(-\lambda Z_L)]}\left(\dfrac{p_l}{P_l}\right)^{-\sigma_l} q, & \forall l < L \\ \ldots \\ 1 + Bk_L^{\beta_L} = \dfrac{\exp(-\lambda Z_{L-1})}{A[1-\exp(-\lambda Z_L)]} q \end{cases} \tag{11}$$

where $P_l = \left(p_l^{1-\sigma_l} + (w(cz_l+1))^{1-\sigma_l}\right)^{1/(1-\sigma_l)}$, $Z_0 \equiv 0$.

Conditional factor demands in Equation 11 are not unfamiliar since they inherit the CES structure to a large extent. Knowledge, which will be optimally chosen in this theory, is the differing element. For each $l < L$, it appears in three places: wages, the denominator and as a shifter in the numerator, inside the exponential term. The reason for it's appearance in wages is clear and the intuition for the latter two is also simple. Knowledge acts as a Hicks-neutral TFP term in that more of it delivers more production, similar to the role of A. On the other hand, the numerator contains an exponential term which 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 $k \geq l$ are needed to attain production q. The top layer L, is slightly different, as explained before.

The maximization problem of the firm can be split in two parts: optimization given L, and the choice of L. The first is the problem I have dealt with in this section, and whose solution I characterize on Propositions 1 and 2, in the next Section 4.1. On Section 4.2, I also explain the choice of L.

4 Understanding the Mechanisms

4.1 Firm Choices Given L

Before stating Propositions 1 and 2 that characterize the solution, let me first provide context for them. When minimizing cost, I allow the firm to choose zero knowledge, but solutions of that sort will never occur. In Proposition 5 in the Appendix, 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. A detailed description of the parameter Assumptions, and their intuition, is provided on Section 8.

Proposition 1 formalizes the key characteristics of factor demands, for an arbitrary L. It states that factor demands given L are increasing in scale. Intuitively, given L, as the firm expands it's scale it requires more of all factors with their relative use determined by the optimality conditions. As output increases, so does knowledge and the relative price of labor to capital, which induces the firm to increase the capital-labor ratio at every layer.

Proposition 1 *Given the number of layers L,*

1. *the knowledge of all employees, z_l $\forall l$,*
2. *the number of employees at $l < L$,*
3. *capital at $l < L$, and*
4. *capital-labor ratios at $l < L$,*

all increase with q. Moreover, k_L is weakly increasing in q.

As a consequence of this result, the following properties of the Marginal and Average costs given L, denoted MC_L and AC_L respectively, can be proven:

Proposition 2 *The properties of Marginal and Average costs given L are:*

- ☐ *Marginal Cost is:*
  - ∘ *positive and increasing in q with $\lim_{q \to \infty} MC_L = \infty$.*
- ☐ *Average Cost satisfies*

- $\lim_{q\to 0} AC_L = \lim_{q\to\infty} AC_L = \infty$.
- *AC_L decreases at $q \to 0$, and increases at $q \to \infty$.*
- *AC_L has a U-shape with a a unique minimum at q^*, the Minimum Efficient Scale.*

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 (CRS), and ii) a model without capital at the CEO level (NOk_L). Constraints at the top layer for each case are given by:

$$\begin{cases} \left(k_L^{\beta_L} + n_L^{\beta_L}\right)^{\frac{1}{\beta_L}} = \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{\frac{1}{\beta_1}} \exp(-\lambda Z_{L-1}) & (CRS) \\ 1 = \left(k_1^{\beta_1} + n_1^{\beta_1}\right)^{\frac{1}{\beta_1}} \exp(-\lambda Z_{L-1}) & (NOk_L) \end{cases} \tag{12}$$

In the CRS model, the associated cost function is constant returns to scale and hence optimal knowledge is independent from scale. Under NOk_L on the other hand, CEO time is fixed but $k_L = 0$, and increasing production given L necessarily involves increasing knowledge at all layers. Finally, in the benchmark model, the firm can adjust k_L which reduces the marginal cost for any z_l. Moreover, as long as k_L is not "too advantageous" relative to z_L, both margins are used to increase production, k_L does not increase too much relative to z_L, and the marginal cost increases. This is guaranteed under Assumption 3. If the Assumption was not satisfied due to for example too large increasing returns to capital, $\beta_L > 1$ in the benchmark model, then as scale increased the firm could find optimal to reduce knowledge at all layers and increase capital with potentially lower marginal costs. Finally, MC_L converging to infinity as output increases is a direct consequence of $\frac{\partial MC_L}{\partial q}$.

The behavior of the average cost can be understood by noting that zero knowledge at all layers implies a positive "fixed" cost but no production, and hence AC_L goes to infinity. As knowledge increases optimally with scale, the "fixed" part of the cost can be spread over more units, initially reducing AC_L until the Minimum Efficient Scale (MES), where the AC_L becomes increasing with output due to the increasing marginal cost. On the next section, I illustrate these properties and compare the functions across L.

To connect cost minimization and profit maximization, I conclude this section illustrating how firm idiosyncratic demand, α, affects choice variables and costs conditional on $L = 2$. On Figure 1, both quantities and revenues increase, since despite the increasing marginal cost, and hence price, consumers derive larger utility from higher α varieties. To satisfy the larger demand, more factors are used, including knowledge, and wages at all layers increase; and as wages increase, the firm towards capital increasing it's capital-labor ratio. The results on wages are present in the data, Caliendo et al. (2015), and proven in Caliendo and Rossi-Hansberg (2012); I extend them to a case with optimal capital.

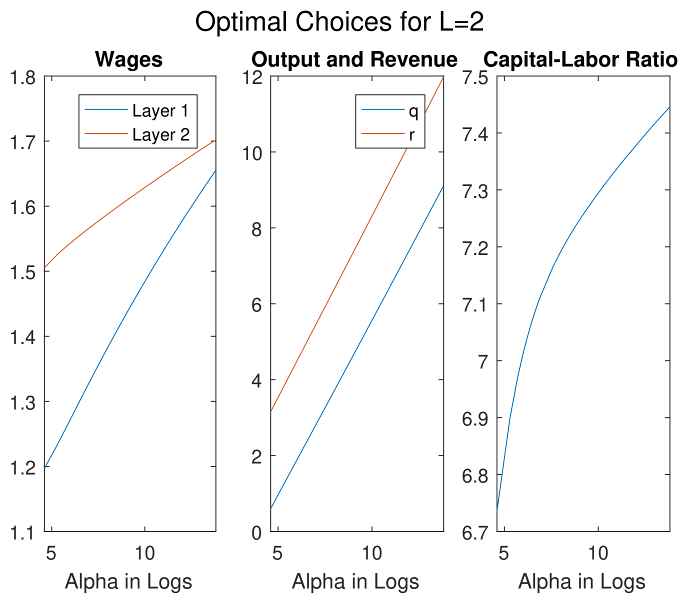

Figure 1: Optimal choices in the α cross-section conditional on $L = 2$. Vertical axis in logs.

Figure 2 shows that marginal costs increase in α and that the average cost is U-shaped as a consequence.

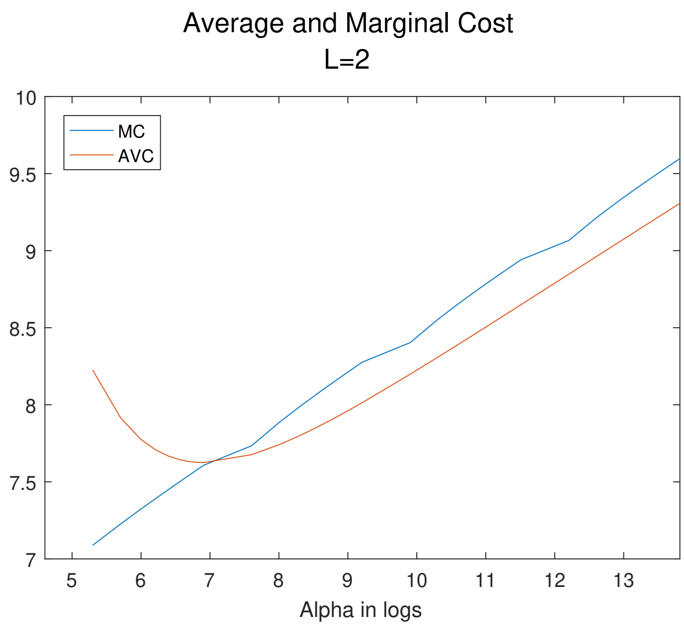

Figure 2: Average and marginal costs conditional on $L = 2$.

4.2 Firm Organization Choice

The previous Section provided intuition for the solution for fixed L, and I now turn to illustrating the optimal choice of L. In the model, entrepreneurs choose among 3 organizational modes; I exclude self-employment for now. I choose the range for L following the detailed empirical mapping of occupations to layers in business hierarchies in Caliendo et al. (2015); more on this in Section 5.1.

Firms choose $2 \leq L \leq 4$, and they do so by maximizing profits across L. Since profits are intimately related to costs, in this Section I graphically illustrate cost for $L = 2, 3$ with optimal $q(\alpha)$. Figure 3 shows average cost for given L have a U-shape, as proven above. Also note that more layers imply a larger fixed cost, so at a low production scale, $AC_3 > AC_2$. As α increases, and output is larger, the more complex organization features a lower average cost. A more complex organization also features a lower marginal cost for a given q. Intuitively, at the point of indifference between L and L+1, the more complex organization is able to economize on knowledge of the lower layers by having a new layer with managers. So the trade-off when choosing between L and L+1 is: a higher fixed cost with L+1, in exchange for a lower marginal cost. As a consequence of this trade-off, a firm with a larger α maximizes profits by choosing a more complex organization.

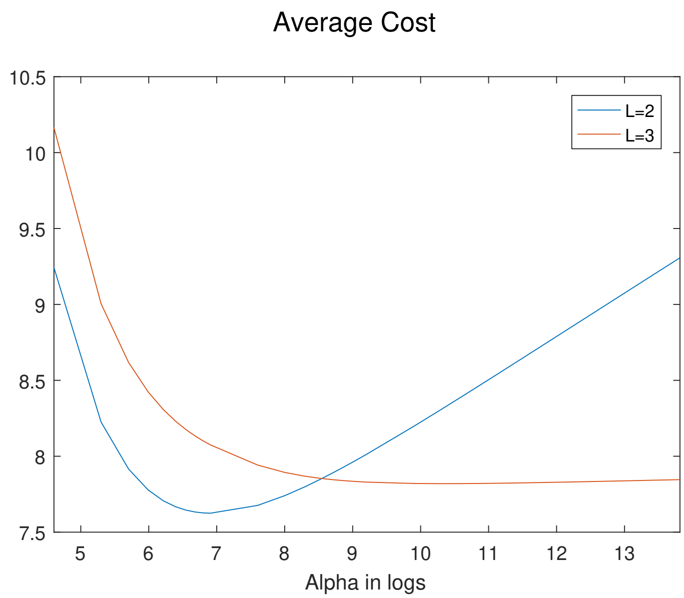

Figure 3: Average costs in the α cross-section conditional on $L = 2, 3$.

4.3 Firm Reorganization with Declining IT Prices

In this Section, I describe the effects of declining IT prices on firm reorganization using Propositions 3 and 4, and quantitative results for the median firm with optimal $L = 3$. Several key theoretical results on the effects of cheaper IT are summarized on Proposition 3. It describes firms' optimal response in terms of knowledge and wages as well as marginal cost, holding q constant, i.e. ignoring scale effects. As suggested by the intuition of Equation 10, the first result shows that when IT prices decline, the firm optimally reallocates problem solving away from the lowest layer. The second result shows that total knowledge of the firm falls, which implies the fraction of problems solved by the firm declines[5]. The third result shows for the simpler case $L = 2$: the decline in IT prices makes IT capital increase and CEO knowledge as well. Finally,

[5] Z_L falling also implies a lower firm-level unweighted average wage, $\sum_{l=1}^{L} \frac{dw_l}{dp_2} > 0$. That z_1 and Z_L decline have an interesting parallel in Caselli and Manning (2018). On their analysis based on first principles, they find that, with better technology the wage of at least one worker-type increases and that the average wage increases[6]. One important caveat is that their analysis is general equilibrium while, at this stage, I am dealing with partial equilibrium. It is ongoing work to extend this Proposition to a GE analysis.

marginal costs fall when IT becomes cheaper.

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

☐ *Knowledge is such that:*

1. *For any L, $\frac{dz_1}{dp_2} > 0$.*
2. *For any L, $\sum_{l=1}^{L} \frac{dz_l}{dp_2} = \frac{dZ_L}{dp_2} > 0$.*
3. *For $L = 2$, $\frac{dk_L}{dp_2} < 0$, and $\frac{dz_2}{dp_2} < 0$.*

☐ *Cost is such that:*

4. *For any L, $\frac{dMC_L}{dp_2} > 0$.*

The behavior of Z_L as IT prices decline is relevant for the connection between productivity and IT. Recall that 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 it's respective input quantity index. Ignoring knowledge as a production factor, a naive measure of the Solow Residual in this model would be $TFP \equiv AF(Z_L)$. From this perspective, the behavior of TFP is associated to that of $F(Z_L)$ so I will spend the rest of this section detailing it's behavior in response to cheaper IT. First, to deepen the intuition, focus on the simpler case L=2, and note that using the CEO level constraint, the production function can be rewritten as

$$q = A \underbrace{[1 - \exp(-\lambda Z_2)]}_{(1)} \underbrace{\exp(\lambda z_1)}_{(2)} \underbrace{(1 + Bk_L^{\beta_L})}_{(3)} \tag{13}$$

Each term plays a conceptually different role and I will highlight it in the next paragraphs. Term (1) is the fraction of problems solved, $F(Z_L)$. Grouping terms (2) and (3) delivers the mass of problems, which in turn can be decomposed into (2) the inverse of the unsolved fraction of problems, $\frac{1}{1-F(Z_L)}$ and (3) the CEO-level inputs, $1+Bk_L^{\beta_L}$. Operating on this equation highlights the reorganization effects of IT. Taking logs in Equation 13, and differentiating with respect to p_2 we obtain:

$$0 = \underbrace{\frac{\lambda \sum_{l=1}^{2} \frac{dz_l}{dp_2} \exp(-\lambda Z_2)}{[1-\exp(-\lambda Z_2)]}}_{\substack{\text{Fraction of problems solved (1)} \\ \text{Sign ?}}} + \underbrace{\lambda \frac{dz_1}{dp_2}}_{\substack{\text{(Inverse of) unsolved problems at plant (2)} \\ >0}} + \underbrace{\frac{\beta_L B k_L^{\beta_L}}{(1+Bk_L^{\beta_L})} \frac{dk_L}{dp_2}}_{\substack{\text{CEO capital (3)} \\ <0}} \tag{14}$$

Equation 14 shows production reorganization as a consequence of IT price declines and underscores the channel they affect in the production function, Equation 13. The first term shows how the fraction of problems solved changes, and is a function of how Z_L changes. The second term tells us how the inverse of plant workers' unsolved fraction of problems changes. We know this term is positive because knowledge of plant workers falls as a consequence of IT, which implies that the proportion of problems that get to the CEO increases. The third term is the effect of the CEO-level capital, and is negative since IT capital increases when it's price falls. The opposing effects of terms (2) and (3), render the sign of the first term, i.e. total knowledge, unclear. However, once the IT capital-plant knowledge optimal trade-off is taken into account the indeterminacy is resolved[7]: ,

$$0 = \underbrace{\frac{\lambda \sum_{l=1}^{2} \frac{dz_l}{dp_2} \exp(-\lambda Z_2)}{[1-\exp(-\lambda Z_2)]}}_{\substack{\text{Fraction of problems solved (1)} \\ <0}} + \underbrace{\left[\frac{e^{-\lambda Z_2} \frac{1}{\phi} \frac{wc\overline{C_1}}{A\lambda^2}}{\left(1+e^{-\lambda Z_2} \frac{1}{\phi} \frac{wc\overline{C_1}}{A\lambda^2}\right)}\right] \frac{G_k}{G} \frac{dk_L}{dp_2}}_{\substack{\text{Total number of problems (2+3)} \\ >0}} \tag{15}$$

In Equation 7, the term in brackets is positive, and since $\frac{dk_L}{dp_2} < 0$, the sum of terms 2 and 3 in Equation 14 is negative: the firm tackles a larger number of problems. If the first term in Equation 7 was constant, as a consequence of these two terms (2+3), output would increase. Since when cost minimizing scale is kept fixed, reorganizing production implies the fraction of problems, and total knowledge, must decrease. Cheaper IT capital allows the firm to handle more problems (2+3), but a smaller share is solved (1). Put differently, lower IT prices make the firm reorganize away from total worker knowledge and towards IT capital.

[7]The optimal trade-off showing that more IT capital reduces knowledge at the plant level is simply the result of comparative statics and is mathematically expressed as $\frac{dz_1}{dp_2} = \frac{-1}{\left(1+e^{-\lambda Z_2} \frac{1}{\phi} \frac{wc\overline{C_1}}{A\lambda^2}\right)} \frac{G_k}{G\lambda} \frac{dk_L}{dp_2}$, which used into Equation 14 delivers Equation .

Summarizing all of the above intuitions for how production changes as a consequence of cheaper IT: i-a smaller fraction of problems is solved, i.e. (1) falls, ii- as a whole, the firm attempts to solve more problems, i.e. (2+3) increase, and iii-the CEO deals with more problems thanks to more IT capital, i.e. $\left(1 + Bk_L^{\beta_L}\right)$ increases; and iv-the CEO solves a larger interval of problems, i.e. $\frac{dz_2}{dp_2} > 0$. On the next subsection I extend these results to the case when q is allowed to adjust, and I focus on the effects on measured TFP.

4.4 Productivity, Firm Reorganization and IT

The results in Proposition 3 fix q, and highlight how reorganization occurs. Mapping the model to data requires allowing q to adjust as p_2 declines. Proposition 4 does just that, and focuses on the new results.

Proposition 4 *Firm responses to a decline in the IT price, when q adjusts, have the same characteristics as in Proposition 3, except, for any L,*

- ☐ $\frac{dq}{dp_2} < 0$.

- ☐ $\sum_{l=1}^{L} \frac{dz_l}{dp_2} = \frac{dZ_L}{dp_2} < 0$.

Proposition 4 shows that firms increase output when the price of IT declines, which occurs because marginal cost falls. As a consequence of the q increase, Z_L also increases because, as shown in Proposition 1, knowledge increases with output.

Propositions 3 and 4 imply opposite results for Z_L, and combining both is illustrative. Denote by γ_x, the growth rate of x when p_{IT} falls, $q \equiv \bar{q}$, output before p_{IT} and $q \equiv q^*$ optimal output after p_{IT} changes,

$$\gamma_{TFP}\big|_{q=q^*} = \underbrace{\gamma_{TFP}\big|_{q=\bar{q}}}_{\substack{\text{Solow Paradox} \\ <0}} - \rho \underbrace{\Lambda(z_L)\gamma_{MC_L}\big|_{q=q^*}}_{\substack{\text{Scale Expansion} \\ <0}} > 0 \tag{16}$$

where $\Lambda(Z_L) > 0$, as defined in the Appendix. Equation 16 shows that TFP growth, when q adjusts as in the data, can be decomposed as the difference between two effects: 1-TFP growth, when q is fixed, minus 2-a term proportional to MC_L growth. The first term can be labeled as a "Solow Paradox" effect[8]: since $\frac{dZ_L}{dp_2}$ is positive by Proposition 3, cheaper IT makes TFP growth lower; in the words of Robert Solow, "You can see the computer age everywhere but in the productivity statistics". 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 $\forall l$ increase with q, as shown on Proposition 1. The overall effect of both terms is positive: measured TFP grows as a consequence of IT capital deepening.

This is the third key result of the paper: a precise mechanism for the complementarity between organizational knowledge and IT in generating productivity effects. 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. These predictions are consistent with the findings in the empirical literature. Bloom et al. (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 European companies. The former also have higher productivity effects of IT capital than the latter, which is suggested to be due to larger organizational capital of US parents being transplanted to their European affiliates. In fact better "people management" practices account for most of the differential output elasticity IT capital across firm types. Bresnahan et al. (2002) use detailed firm-level data, and show evidence of complementarity among IT and workplace organization 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 firm's IT capital-labor ratios, which in turn raises the relative demand of skilled to unskilled workers, as in the model in this paper. They suggest that similar results should hold for other measures of organizational capital. Similarly, Brynjolfsson and Hitt (2003) show that TFP, measured using standard growth accounting, is positively correlated with computer investments in a panel of large US firms. Using five different instrumental variables for computer investments confirms the mentioned OLS results.

The mathematical discussion above highlights that disregarding firm organization, and worker knowledge as a production factor, have important implications for estimating the TFP effects of IT. My results also emphasize the need of a richer view of how labor is dealt with when estimating productivity, and, in fact, Fox and Smeets (2011)

[8]Evidence on the existence of the 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 US, and conclude that IT usage has little impact on productivity.

find that doing so reduces the TFP dispersion in the firm cross-section.

4.5 Quantitative Effects of Declining IT Prices

I turn now to the quantitative responses on the within-firm, across-layer response to declining IT capital prices for the median firm with optimal $L = 3$. In order to evaluate the effects I use the calibration described on Section 5.2. Regarding the evolution of IT capital prices, Eden and Gaggl (2018) have recently documented them using US, BEA data. Their definition of IT capital is on the Appendix.

Figure 4 shows the response of wages at the three layers. All figures in this section report changes relative to the baseline year, with relatively high IT capital prices. As suggested by Equation 10, a decline in IT prices triggers an optimal reallocation of knowledge across layers: managerial problem solving becomes cheaper, hence managers learn a larger set of problems, some of which used to be solved by the plant-level workers. Firms find it cheaper to allocate knowledge to few managers instead of many workers, because now the former can leverage their time with IT capital.

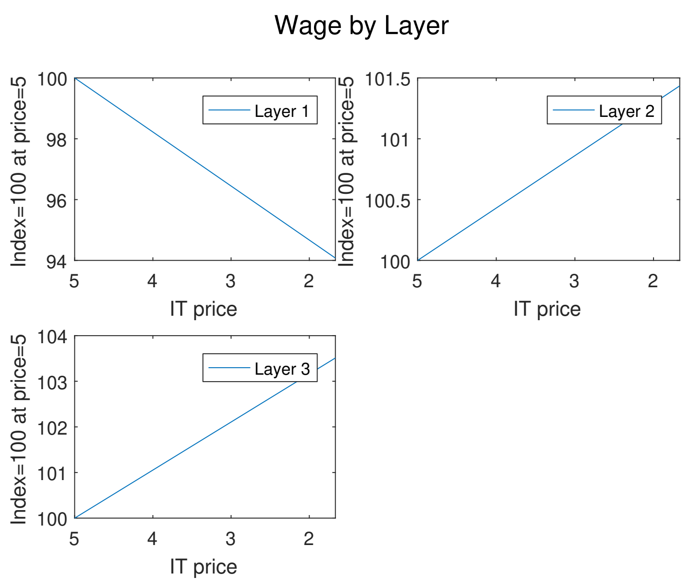

Figure 4: Wage response to IT capital price decline. Relative to baseline price.

On Figure 5 intermediate managers increase their employment share whereas blue-collar workers decrease it. There are two effects, when thinking about this result using the conditional factor demands, Equation 11. First, both n_1 and n_2 increase as the firm optimally increases it's scale. Second, since knowledge at lower layers has gone down, more inputs at higher layers are used. Overall, the increase in managerial labor is larger than that of plant workers and hence the firm-level employment shares move in opposite directions. With the described employment and wage responses, naturally the wage bill share of plant-level workers in total labor payments falls whereas that of managers increases.

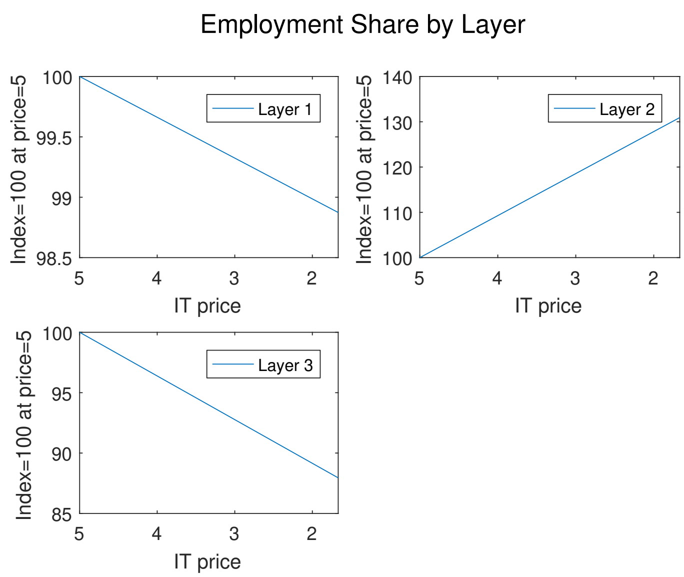

Figure 5: Employment share response to IT capital price decline. Relative to baseline price.

Capital-labor ratios respond as shown in Figure 6. Since wages increase at layer 2 and decrease at layer 1, optimal substitution makes the capital-labor ratios to move in opposite directions in each of those layers. As the firm expands, the firm responds allocating more capital to the CEO as well.

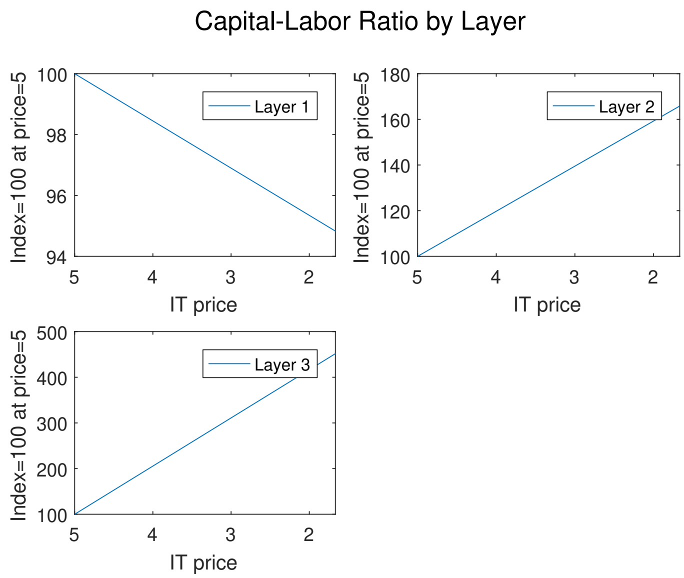

Figure 6: Capital-labor ratio response to IT capital price decline. Relative to baseline price.

Finally, in preparation for the labor share effects of IT, I study how the firm-level labor share of production costs (LSC) responds to falling IT capital prices

$$\frac{dLSC}{dp_2} = \frac{1}{TC}\left[\left(\underbrace{\sum_{l=1}^{L}\frac{dw_l}{dp_2}n_l}_{(1)} + \underbrace{\sum_{l=1}^{L}\frac{dn_l}{dp_2}w_l}_{(2)}\right)(1-LSC) - LSC\left(\underbrace{\sum_{l>1}^{L}k_l}_{(3)} + \underbrace{\sum_{l=1}^{L}\frac{dk_l}{dp_2}p_l}_{(4)}\right)\right] \tag{17}$$

where $LSC \equiv \sum_{l=1}^{L}\frac{w_l n_l}{\sum_{j=1}^{L}(w_j n_j + p_j k_j)}$, and $TC = \sum_{l=1}^{L}(w_l n_l + p_l k_l)$.

Equation 17 highlights the conceptually different channels with (1) being unique to this model from a more standard CES production function: optimal knowledge. While all terms are influenced by optimal knowledge at the optimal solution, only item number one is unique to the theory in this paper. Note also that (4) sincludes the reaction of capital at the top layer, another margin that will also play an important role for the LSC behavior. From this equation, we gain several pieces of intuition. First, as we have seen above, knowledge and wages respond optimally to capital prices and move in opposite directions at $l = 1$ versus $l = 2, 3$. However, as long as there are more workers

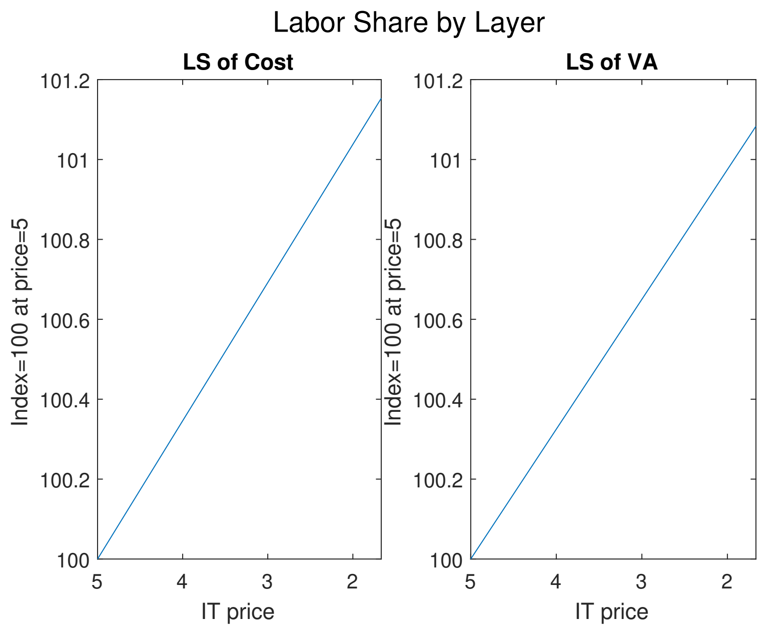

Figure 7: Labor share response to IT capital price decline. Relative to baseline price.

at the lower layer than at higher levels, term (1) will tend to be positive. Hence, the response of z contributes to a decline in the LSC when p_2 falls.

Terms (2) and (4) capture the response of conditional factor demands to changing prices. Overall, in the calibration, they are negative due to the increased firm production. Finally, (3) is positive. Overall, each of the two summands (1+2) and (3+4) feature conflicting effects and are influenced by the level of the LSC. When the LSC is large, (3+4) play a larger role, and when LSC is low (1+2) matters more. The degree and nature of heterogeneity in LSC across α is studied on Section 4.7.

$\frac{dLog(LSC)}{dLog(p_2)}$	**Total Change=**	**(1)**	**(2)**	**(3)**	**(4)**
0.50	100	1.71	-2.92	-3.67	104.87

Table 1: Decomposition of the change in the labor share of cost to 1% lower IT capital price. Each column reports values as percentage of the total change and add up to 100. Reported values are for the median α firm for $L = 2$.

It is instructive to decompose the change in the labor share of cost into each of it's constituting elements as reported in Equation 17. For simplicity I focus on the median α firm for $L = 2$. Table 1 reports the elasticity of the LSC to the IT price and the contribution of each channel to the change in the labor share of cost as a result of decline in IT capital prices of 1%, all in percentage terms. As reported on the first column, the LSC decreases 0.5% as IT prices fall 1%. In the remaining columns, negative values on the table contribute to an increase of LSC and positive values have the opposite interpretation and they all add up to the total change, normalized to a 100. Notice how (1) and (4) contribute substantially to a decline in the LSC dominating (2) + (3) which contribute to an increased LSC.

4.5.1 The Heterogeneous Effect of IT Prices across Firm Organizations

Table 2 compares the response of firms with organizations $L = 2$ and $L = 4$. I use the calibrated parameters and focus on the median firm with each organization. In both cases, wages at the lowest layer fall, though substantially more in the $L = 4$ firm. The large decline in wages at the bottom of the wage distribution in large firms is consistent with the results reported in Song et al. (2015). On the other hand, wages increase for both the CEO in $L = 2$, and the last two layers in $L = 4$ firms. The large firm expands quantities and revenues by more than the L=2 firm, as in the documented increased sales concentration at larger firms in Autor et al. (2017). Through purely a decline in IT prices, both the LSC and the LS of value added (LSVA) at large firms increase, the latter being a prediction at odds with the results in Kehrigy and Vincent (2014), who document a substantial reduction in the LSVA in the largest firms. As I discuss later on, this fact allows me to identify changes in the elasticity of demand in Section 5.2 when calibrating the model.

Organization:		
	L=2	**L=4**
Variable		
Wage, layer 1	-3.4	-6.3
Wage, layer 2	2.5	-2.4
Wage, layer 3	.	3.5
Wage, layer 4	.	3.7
Output	5.8	10.6
Revenue	3.7	6.8
LS Cost	-0.2	1.2
LS of VA	-0.4	1.2

Table 2: Effects of declining IT capital price across firms with different optimal organizations. Responses of calibrated model for IT price changes from 1980 to 2015. Values are changes in percentage for the median α firm for each organization.

4.6 The IT capital-Plant Labor Elasticity of Substitution

It is interesting to obtain the IT capital-plant labor elasticity of substitution in this model and connect it to other measures in the related literature. To do so, I first provide a quick overview of the elasticities used in the the closest macro work. 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' 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),

$$\sigma_{ij}^{A} \equiv -C(p,y)\frac{\frac{dlog x_i}{dlog p_j}}{x_i x_j} \tag{18}$$

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 (MES),

$$MES_{ij} = \sigma_{ij}^{M} \equiv -\frac{dlog\left(\frac{x_j}{x_i}\right)}{dlog\left(\frac{p_j}{p_i}\right)} \tag{19}$$

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 MES is the multiple input equivalent. The MES fixes output, but all inputs are allowed to adjust. Importantly, in it's simplest form, requires that only the j-th price, in the ratio p_j/p_i, varies[9]. This implies the MES is naturally not symmetric, $\sigma_{ij} \neq \sigma_{ji}$.

Before turning to my model, I briefly highlight some features of the above elasticities. In the case of two factors, the AES and MES coincide. In particular, for a two factor CES production function, $AES = MES = \sigma$, where σ is the CES parameter. On the other hand, for multiple inputs, the AES and MES are, in general, different and have properties as described above.

In the macro literature, it is common to define a convenient production function and impose a substitution pattern between capital and unskilled labor, or estimate the parameters governing it. For example, in Krusell et al. (2000), the production function is

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

$$y = Ak_s \left(\mu u^{\frac{\rho-1}{\rho}} + (1-\mu) \left(\lambda k_e^{\frac{\sigma-1}{\sigma}} + (1-\lambda) s^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1} \frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}$$

where u and s and unskilled and skilled labor, k_s and k_e are structure and equipment capital, and the rest are parameters. Their definition of the equipment-unskilled labor elasticity is simply $\rho = 1.67$, a constant number. More generally, nested CES do not produce constant AES and MES, Blackorby and Russell (1989).

More recent relevant work includes Acemoglu and Restrepo (2018) and Hemous and Olsen (2016). Both papers provide careful discussions of the capital-labor substitution in final output production, which is a CES aggregation of tasks or intermediates, in the former and the latter, respectively. Both models are similar in that they impose substitution of automation capital and unskilled workers, and push a set of questions about economic growth and the properties of it's dynamics[10]. Similarly, Eden and Gaggl (2018) study routine and non-routine labor shares and their relation to IT in the US. They provide a detailed analysis of the values of several elasticities including the IT capital-routine labor, which in their baseline model are substitutes.

Unlike these papers, I impose complementarity at each level of the hierarchy and, as I show next, still obtain IT capital-plant labor substitution due to optimal organizational choices. The specific mechanisms in the model imply that there is not one perfect mapping to the elasticities in the literature. This is so because, usually factor prices in the denominator are exogenous to the firm, but in my model changes to p_2 affect w_1 through firms' FOCs that determine knowledge. For my model, I focus on an elasticity defined as in Robinson (1933),

$$\sigma_{n_1,k_l} \equiv - \frac{dlog \left(\frac{k_l}{n_1} \right)}{dlog \left(\frac{p_2}{w_1} \right)} \tag{20}$$

for any layer l, for fixed output, when all other inputs are allowed to adjust. This definition of the elasticity of substitution is intuitive since it tells us how much k_l IT capital to plant-labor ratio adjusts to changes in their relative prices[11]. I will next

[10]The two models differ in several respects and I review their mechanisms more carefully in the Appendix subsection 8.5.

[11]Note that comparing the elasticity across different organizations will always be imperfect. More complex firms, have more layers and, hence, more IT capital. Focusing on one layer reduces this "extensive-margin" concern. In any case, the values for the elasticity are very similar when using total

report quantitative results for Equation 20, when only parameter p_2 changes (but $w_l, \forall l$, optimally respond). Note that absent the endogenous response of all w_l, this is the Morishima elasticity as described above. A useful property shared by the MES and my elasticity is that they both quantitatively inform of the direction of relative factor spending shares. In fact,

$$\frac{dlog\left(\frac{w_1 n_1}{p_2 k_l}\right)}{dlog\left(\frac{p_2}{w_1}\right)} = \sigma_{n_1,k_l} - 1 \tag{21}$$

According to Equation 21, if $\sigma_{n1,k_2} > 1$, that is, with sufficient IT capital-plant labor substitution, the relative spending in plant-labor to IT capital, declines when p_2 falls. This is what I find. Note also that the overall IT capital-plant labor elasticity satisfies:

$$\sigma_{n_1,IT} \equiv -\frac{d\left(\frac{\sum\limits_{l>1}^{L} k_l}{n_1}\right)}{d\left(\frac{p_2}{w_1}\right)} \frac{\frac{p_2}{w_1}}{\frac{\sum\limits_{l>1}^{L} k_l}{n_1}} = \sum_{l>1}^{L} \sigma_{k_l,n_1} s_l^k$$

where $s_l^k \equiv \frac{k_l}{\sum\limits_{j>1}^{L} k_j}$. Since the layer 2 share of total IT capital is largest, the layer 2 elasticity is quantitatively more important than that of layers $l > 2$, in determining the quantitative value of $\sigma_{n_1,IT}$. This is why I next describe, first, σ_{n_1,k_2} and, I relegate σ_{n_1,k_L} to the Appendix Section 8.2.

Figure 8 shows Equation 20 for $l = 2$ and the calibrated parameter values[12]. Like Karabarbounis and Neiman (2013), Raval (2017), and Oberfield and Raval (2014), the calibration uses long-run capital-labor elasticities, in my case at each layer of the hierarchy. Consistent with this calibration, and like Karabarbounis and Neiman (2013), I focus on the effects of long-run IT price changes, which capture movements from an initial to a final steady state for the period 1980-2015. On Figure 8, σ_{n_1,k_2} is larger than one for the output levels in the firm cross-section, implying $k_2 - n_1$ substitution. This implies the relative spending of n_1 to IT has fallen during the period 1980-2015, as in

IT capital instead, i.e. when reporting $-\frac{dlog\left(\frac{\sum_{l=2}^{L} k_l}{n_1}\right)}{dlog\left(\frac{p_2}{w_1}\right)}$.

[12]More specifically, using my calibration, I compute $-\left[\frac{\Delta\left(\frac{k_2}{n_1}\right)}{\left(\frac{k_2}{n_1}\right)_0}\right] \Big/ \left[\frac{\Delta\left(\frac{p_2}{w_1}\right)}{\left(\frac{p_2}{w_1}\right)_0}\right]$, where $\Delta\left(\frac{y}{x}\right) \equiv \left(\frac{y}{x}\right)_F - \left(\frac{y}{x}\right)_0$ the change between final and initial period.

the data, something which is not targeted in my calibration on Section 5.2.

IT capital at $l = 2$ and plant labor are gross substitutes, despite capital-labor complementarity at every layer. This is the second key result of the paper, as mentioned in the introduction. There are substantial differences in the elasticity across organizations: in the simpler $L = 2$ firms, layer 2 capital and plant-labor are very substitutable. To understand why, note that the numerator is just the difference in growth rate of k_2 and n_1, whereas the denominator has an equivalent interpretation as the percent change in relative prices. As I show next, the latter turns out to be similar across organizations, so I next focus on the numerator. For $L = 2$, k_2 is the only IT capital the firm uses so it is very elastic. On the other hand, the most complex firms with $L = 4$ have multiple layers whose capital becomes cheaper when p_2 declines. Moreover, for $L = 4$, capital-labor ratios at each layer below $L = 4$ increase when p_2 falls but, unlike for firms with $L = 2$, both k_2 and n_2 increase. This is because cheaper IT capital makes the firm shift problem solving towards managerial layers and more of all inputs at $l = 2$ are used, which dampens the percent response of k_2. For these reasons, the percent response of layer 2 capital is less in L=4 relative to L=2 firms. On the other hand, due to reorganization effects, n_1 responds positively to p_2 declining in all organizations but, percent-wise, much more in L=4 firms. Recall cheaper IT prices makes the firm reorganize by: i- lowering $F(Z_L)$ and ii-increasing the number of problems, i.e. the production input bundle. The latter implies increasing n_1, in particular. This reorganization effect is naturally larger for $L = 4$ organizations because they use more IT, and also contributes to a lower elasticity in these firms.

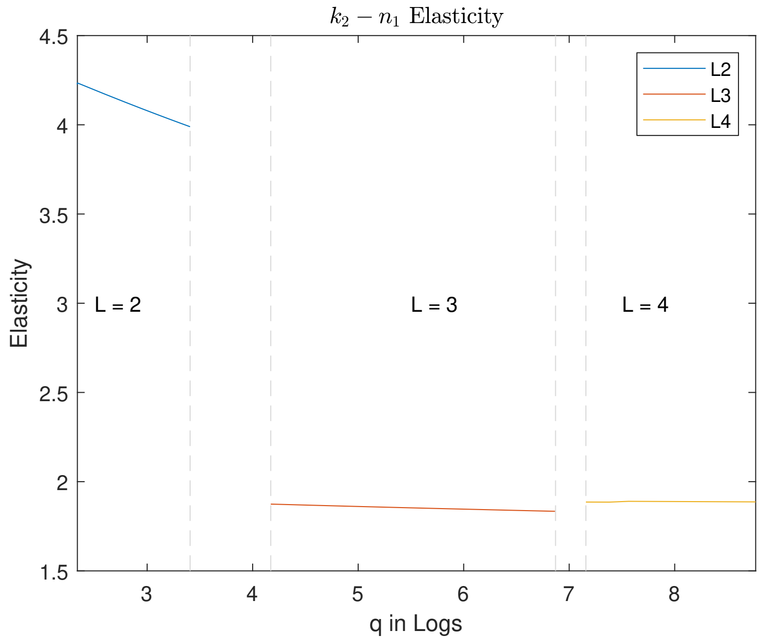

Figure 8: Layer 2 Capital-Plant Labor Elasticity.

A complementary way to understand the elasticities in Figure 8 involves using a decomposition that focuses only on optimal knowledge decisions. Mathematically, in my model, the $k_2 - n_1$ elasticity for $L = 2$, when p_2 changes is approximately[13]:

$$- \frac{dlog\left(\frac{k_2}{n_1}\right)}{dlog\left(\frac{p_2}{w_1}\right)}\Bigg|_{L=2} \approx \frac{W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} + O_1^L(\lambda, \beta_L)\frac{\partial z_1}{\partial \log(p_2)} + O_L(\lambda, \beta_L)\frac{\partial z_2}{\partial \log(p_2)}}{1 - \varepsilon_{w_1, p_2}} \tag{22}$$

whereas for $L = 3, 4$:

$$- \frac{dlog\left(\frac{k_2}{n_1}\right)}{dlog\left(\frac{p_2}{w_1}\right)}\Bigg|_{L=3,4} = \frac{W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} + O_1^L(\lambda, 1)\frac{\partial z_1}{\partial \log(p_2)} + W_2(\sigma_2)\frac{\partial z_2}{\partial \log(p_2)}}{1 - \varepsilon_{w_1, p_2}} \tag{23}$$

where

[13] The elasticity expression for $L = 2$ is an approximation. Instead of using the actual conditional 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 $k_2 = \left[\frac{\exp(-\lambda z_1)}{AB[1-\exp(-\lambda Z_L)]}q\right]^{1/\beta_L}$.

$$
\begin{aligned}
W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} &\equiv -\sigma_1\left[\frac{\partial w_1}{\partial \log(p_2)} - \frac{\partial \log(P_1)}{\partial \log(p_2)}\right] < 0 \\
W_2(\sigma_2)\frac{\partial z_2}{\partial \log(p_2)} &\equiv \sigma_2\left[1 - \frac{\partial \log(P_2)}{\partial \log(p_2)}\right] \\
O_1^L(\lambda, \beta_L) &\equiv \left(1 + \frac{1-\beta_L}{(\exp(\lambda Z_2) - 1)}\right)\frac{\lambda}{\beta_L} > 0 \\
O_L(\lambda, \beta_L) &\equiv \frac{(1-\beta_L)}{(\exp(\lambda Z_2) - 1)}\frac{\lambda}{\beta_L} > 0 \\
\varepsilon_{w_1,p_2} &\equiv \frac{\partial \log(w_1)}{\partial \log(p_2)} > 0
\end{aligned}
$$

Note several properties of Equations 22 and 23. There are two types of terms: i-related to within layer substitution, denoted by W, and ii-related to organizational choices, denoted by O. The first term, $W_1(\sigma_1)$, is common to all organizations. It is driven by the within-layer 1 capital-labor elasticity of substitution, as well as the response of z_1 to p_2. It is always negative because it captures the effect of increased n_1 as w_1 falls. The second term, the function $O_1^L(\lambda, \beta_L)$, captures the organizational effects of z_1. It captures two channels. The first is the role of z_1 on $F(Z_L)$. It is a TFP-type effect, which decreases conditional capital and labor demands in all layers, as shown on Equation 11. The second is the role of z_1 in determining which problems get to the next layer. Across firms with different L, the only difference is that for L=2 both effects are related to the CEO capital. Only for these firms $O_1^L(\lambda, \beta_L)$ depends on β_L, because the parameter governs the strength of the marginal response of IT to price changes. On the other hand, for firms with $L > 2$, $O_1^{L>2}(\lambda, \beta_L) = O_1^{L=2}(\lambda, 1)$, because layer 2 capital does not depend on β_L. Across Equations 22 and 23 the third term is different, but always captures the role of z_2. More specifically, this third term is: i-for $L = 2$, $O_L(\lambda, \beta_L)$, and, similar to O_1^L captures the organizational effects above, so it depends on λ and β_L as well; instead ii-for $L > 2$, function $W_2(\sigma_2)$, depends on σ_2, the within-layer 2 capital-labor substitution. Finally, The denominator is common to all organizations and is one minus the elasticity of w_1 to p_2.

Equations 22 and 23 are better understood by looking at the denominator and the numerator separately. Focus first on the denominator, $1 - \varepsilon_{w_1,p_2}$. Table 3 shows ε_{w_1,p_2} is small and rather similar across organizations (and also across firms within L). So the denominator of Equations 22 and 23 is close to one for all organizations, and hence most of the heterogeneity in σ_{n_1,k_2} comes from the numerator, to which I turn next.

	Organization		
	L=2	L=3	L=4
ε_{w_1,p_2}	0.03	0.08	0.09

Table 3: Percent plant-wage response to a percent change in the IT price, ε_{w_1,p_2}, for the median firm across organizations.

The contribution of each term in the numerator of Equations 22 and 23 is shown in Figure 9, with the sum of all being one hundred. Several lessons can be drawn from this decomposition. First, $O_1^L(.)$ always plays the largest quantitative role across L. The IT capital-plant labor elasticity is fundamentally determined by optimal knowledge in organizations, in particular, by how the range of problems/tasks have changed with IT. The Amazon example suggests they have decreased in the data, and the evidence from wages on the next section confirms this. Second, $W_2(.)$ is non-negligible for more complex firms at around 20%, and is related to the within-layer 2 reallocation of inputs. Third, the within-layer 1 term is negative and small for all organizations. Fourth, that $L = 2$ firms have a much larger elasticity compared to $L = 3, 4$ is quantitatively explained by the fact that $O_1^{L=2}\lambda, \beta_L) > O_1^{L=3,4}(\lambda, \beta_L)$; In words, only for $L = 2$ firms, $O_1^{L=2}\lambda, \beta_L)$ captures the increase in CEO capital due to a lower z_1, and it plays a quantitatively large role in the elasticity.

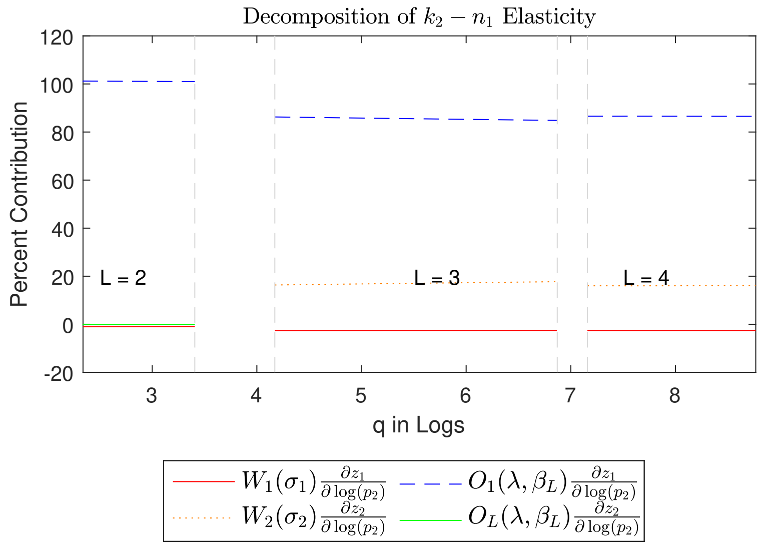

Figure 9: Decomposition of Layer 2 Capital-Plant Labor Elasticity as shown in Equations 22 and 23.

In summary, quantitatively, the organizational effects of the decline in knowledge at layer 1 as IT prices fall is the most important organizational term driving the elasticity. Fortunately, the plant-wage evidence reviewed in the next section suggests it is relevant also in the data.

4.6.1 The Evidence on the Organizational Channel of IT

There is ample evidence that is consistent with the mechanisms in this model. The influential Autor et al. (2003) is a solid departure point for thinking about the impact of IT on firm 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 US, they show that, within occupations, computerization is associated with reduced labor routine manual and routine cognitive tasks and increased labor of nonroutine cognitive tasks. Their empirical mechanism is tightly connected to the theory in this paper. In the model, plant-workers' tasks are more routine compared to those of managers, whereas problem solving by managers involves dealing with exceptions, i.e. the less frequent, less routine tasks (problems). Firm reorganization due to lower IT prices involves a reallocation problem solving across layers along the lines of Autor et al. (2003): more IT capital reduces the range of the more routine tasks done by plant-workers and increases the amount of problem solving by managers, the non-routine tasks.

In other related early work, Caroli and Reenen (2001) for the UK showed that organizational change is related to a lower unskilled manual wage-bill share. Bresnahan et al. (2002) provide evidence for the US that IT defined as i) Percent of workers using general purpose computing; ii) Percent of workers using email; computerization of work; and iii) computing power are all related to firms using a lower unskilled employment share, and a larger managerial employment share.

More recently, Akerman et al. (2015) study the introduction of broadband Internet in Norway, which under appropriate controls can be regarded as a quasi-experiment in availability. Using worker data, they show declines (increases) in hourly wages of unskilled (skilled), just like the model predicts. Moreover, their results are consistent with the task interpretation of labor market outcomes from Autor et al. (2003) and Acemoglu and Autor (2011): they also find declines (increases) in hourly wages of routine (abstract). Akerman et al. (2015) also study firms and, in particular, following a Levinson-Petrin production function estimation approach, find a decline in the output elasticity of unskilled labor due to the availability of broadband. Moreover, their results are robust to using firms' DSL adoption instrumented with availability.

A policy experiment in the UK is studied in Gaggl and Wright (2017). They focus on a 100% first year tax allowance on IT investments to small firms, those with less than 50 employees, and study differences in firms' outcomes around the threshold in a regression discontinuity design. From the perspective of the model in this paper, small firms are the least likely to be influenced by reductions on IT prices, so finding model-consistent results in these set of firms would be a strong confirmation. Their first result is that, in fact, IT investment, defined as spending in hardware and software, actually increase due to the policy. More importantly for this paper, and consistent with Akerman et al. (2015), they find reductions in weekly earnings in routine cognitive workers, and increases for non-routine workers' earnings. The experiment is particularly clean given that it compares routine cognitive workers in firms with slightly less than 50 employees with firms with slightly more than 50 workers. Moreover, the paper finds that firms introduce "Advanced Management Techniques" and change "Organizational Structure", which is also in line with the model.

From a firm organization perspective, Bloom et al. (2014) study the effects of several technologies on autonomy of workers and managers. Autonomy is defined in the same way as Bresnahan et al. (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 head-quarters; 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. With these definitions, autonomy in the data can be mapped to knowledge-based hierarchy models as the length of the knowledge interval, or how many problems are solved by each employee type. They find that plant-worker/managers' autonomy falls with intranet and plant-managers' autonomy increases with ERP, again consistent with the model in this paper.

Finally, my results are in line with Song et al. (2015), who use the US Social Security employer-employee data, and find that, from 1980-2015, in firms with 10,000+ employees median wages have declined. This is a surprising fact begging for an explanation and the model in this paper suggests one: large firms use IT intensively and as a consequence when IT prices fall, they reallocate problems solving away from the plant-level to managerial layers. Think of an airport counter. In 1980, the airline may have needed a relatively well trained clerk, who takes notes of names of passengers, checks passports and confirms payments, deals with travel agencies, etc. Today most airlines use self-check-in machines and plant-workers only needs to do the very repetitive and simple task of showing passengers how to operate a touch-screen. The design of these software and other IT managerial tools by CEOs must be the result of careful cost-saving considerations, of which deskilling the numerous workers in plant-level occupations seems a natural outcome.

4.7 The Relation Between Firm Scale and the Micro Labor Share

Oberfield and Raval (2014) emphasize that understanding the micro labor share in the cross-section of firms is crucial to understand the macro labor labor share. In this subsection I discuss the mechanisms that determine firms' LSC in my model and also compare it to two other natural benchmarks, an exercise that allows to perform model selection. In this model a firms' labor share response to α is,

$$\frac{dLSC}{d\alpha} = \frac{1}{TC}\left[\sum_{l=1}^{L}\left(\underbrace{\frac{dw_l}{d\alpha}n_l}_{(1)} + \underbrace{\frac{dn_l}{d\alpha}w_l}_{(2)}\right)(1-LSC) - \underbrace{LSC\sum_{l=1}^{L}\frac{dk_l}{d\alpha}p_l}_{(3)}\right] \tag{24}$$

The three terms are positive. Again, (1) is specific to the theory in this paper as it captures the response of knowledge to scale: as α increases, the firm responds by increasing knowledge so it can produce more. Hence, labor becomes more expensive and firms switch away from labor and towards capital. These mechanisms highlight a unique feature of this model: as a consequence of the knowledge margin, $\frac{dLSC}{d\alpha}$ is driven by wages and scale.

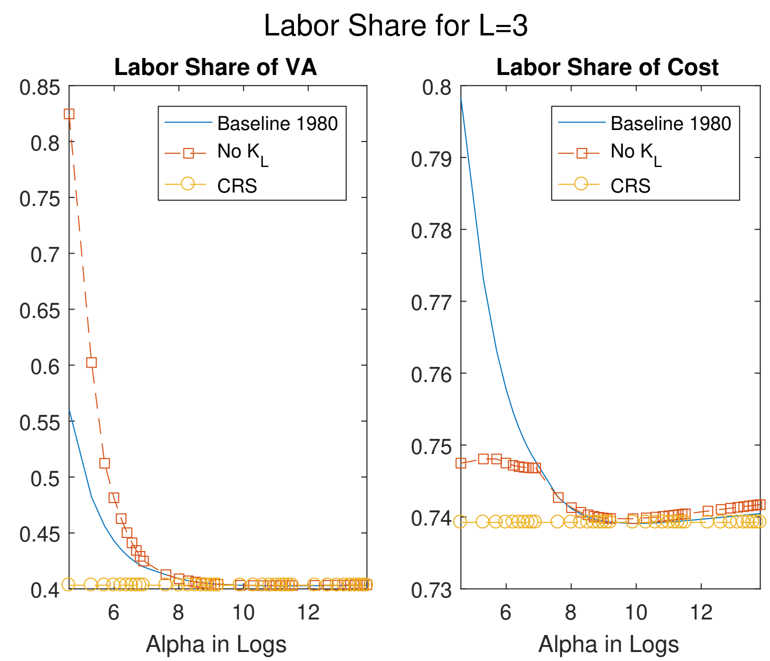

Figure 10: Labor share of cost and the α cross-section. The figure compares the benchmark model to two alternatives: i-a model with constant returns to scale at all layers (CRS), and ii-a model without capital at the CEO level (NOk_L). Same calibration except at top layer.

It is useful to compare the LSC in the benchmark model and the two alternatives: i-a model with constant returns to scale at all layers (CRS), and ii-a model without capital at the CEO level (NOk_L). The LSC for each model is shown in Figure 10, where the three models share the calibration except at the top layer constraint. Let's focus first on model (CRS). This model features constant returns to scale, so z_l are chosen irrespective of α, or the production scale. This implies a constant LSC across α, and also that wages do not change with firm size, features that are at odds with the empirical facts.

On the other hand, under model (NOk_L), firms cannot respond to a larger α by increasing capital at the top, hence they miss a part of margin (3) in Equation 24, i.e. $\frac{dk_L}{d\alpha}p_2$. Since (3) contributes to a negative overall sign, the net $\frac{dLScost}{d\alpha}$ still falls as α decreases but much less so. I conclude that the span of control of the CEO is a key identifying parameter that allows model selection.

4.8 Equilibrium

The equilibrium in this economy is a set of prices (w, p_l, P) such that:

- Consumers maximize utility subject to their budget constraint, as described in Section 2.1.
- Firms maximize profits, as described in Section 2.2.
- Labor, capital and goods markets clear.

For now in the quantitative Section 5 I will focus on Partial Equilibrium results.

5 Quantitative Analysis

5.1 Empirical Mapping of Occupations to Hierarchy Layers

The mapping of occupations to layers, follows Caliendo et al. (2015) and allows an empirical mapping from the theory in this paper to the data, and to focus 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 4.

Category	Occupational Description
2.	Firm owners receiving a wage, which includes the CEO or firm directors;
3.	senior staff or top management positions, which includes chief financial officers, heads of human resources, and logistics and purchasing managers;
4.	employees at the supervisor level, which includes quality control technicians, technical, accounting, and sales supervisors;
5.	qualified and nonqualified clerical employees, secretaries, human resources or accounting employees, telephone operators, and sales employees;
6.	blue-collar qualified and non-qualified workers, welders, assemblers, machine operators, and maintenance workers.

Table 4: Occupations and Hierarchical Layers.

Through out 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.

5.2 Calibration

In Table 5 there are two sets of parameters according to whether they are are calibrated externally or internally. Several parameter values are normalizations follow Caliendo and Rossi-Hansberg (2012), namely (A, R, P). At this stage, in partial equilibrium, their role is similar and related to average firm scale. The long-run capital-labor elasticities come from Raval (2017), and are aligned with economists' prior that both skilled and unskilled labor are complementary to capital though the former more so than the latter, Autor (2015). I borrow the change in IT capital prices from Eden and Gaggl (2018) who use BEA data.

There is a set of parameters that are chosen to match data moments. The relative complexity of the model makes it difficult to identify each parameter separately, but I will provide intuition for how each parameter is informative of each simulated moment. ρ, the demand elasticity shifts up firms' LS of VA, by altering their markups. I choose ρ to target values at the top of the size distribution and match the values reported in Kehrigy and Vincent (2014) for what they call "Highly-productive plants"; the model equivalent of the latter are $L = 4$ firms. In the calibration, I assume in 1980 markups were constant in the cross-section. For 2015, I allow the value of ρ to gradually decrease with α starting at the same 1980's value for the smallest firm and lowering it as α increases[14]. As shown on the Table below, the largest firm has $\rho = 2.2$ whereas the smallest firm $\rho = 2.8$.

Increases in w, the wage, makes the LS increase, I choose it to match a cross-sectional average of 0.5. As shown in Section 4.7, β_L generates dispersion in the LS of cost, so I choose it to match the slope of a regression of the factor cost ratio $FC \equiv \frac{CapitalCost}{WageBill}$ on value added, which I obtain from Raval (2017). Capital prices also play a role in the degree of LSC dispersion delivered by the model in the firm cross-section. Higher capital prices makes substitution towards capital less profitable and, hence, there is less dispersion. I use the $p75/p25$ values from the Factor Cost Ratio in Raval (2017). c and λ play similar roles as they both govern the importance of knowledge. c is the cost per unit of knowledge and reducing it makes knowledge less relevant which favors simpler organizations. λ is a shifter of the density of problems, with larger values implying more mass at low z values, reducing the incentives to add layers. Finally, the α distribution is assumed to be log-normal with parameters $(\mu_\alpha, std_\alpha))$. These four parameters are chosen to match the US aggregate employment shares by firm size-class obtained from the BLS as well as the employment of the smallest $L = 2$ firm.

[14] See Figure 18 in the Appendix for a visual illustration of the markups I use for sthe calibration.

Parameter	Value	Description	Source/Target
		Calibrated Externally	
A	5	TFP	Normalization
$p_l = p_2$, $l > 1$	[5, 5/3]	Capital price at l=2,3,4. Change.	Eden and Gaggl, (2017)
σ_1	0.87	Capital-labor elasticity at plant	Raval, (2011)
σ_l, $l > 1$	0.45	Capital-labor elasticity at l, $L > l > 1$.	Raval, (2011)
P	1	Price Index	Normalization
R	8.5	Aggregate income	Normalization
		Calibrated Internally	
ρ	[2.8, 2.2]	Demand elasticity	Average LS of VA at L=4
w	2	Wage	LS of cost, average
β_L	0.3	Capital exponent at layer L	Slope(log(FC),log(VA))
p_1	0.01	Plant-level capital price	p75/p25 FC distribution
c	0.5	Training cost	Employment at smallest $L = 2$
λ	4	Mean of $F(.)$	Employment share by firm size (5 bins)
μ_α	6.5	Mean of α	Employment share by firm size (5 bins)
$std(\alpha)$	2.2	Standard deviation of α	Employment share by firm size (5 bins)

Table 5: Full set of calibrated parameters in the baseline.

Table 6 shows the targeted moment data and simulated values, and their source[15]. The simulated and real data moments are close in this preliminary calibration. There are several margins with room for improvement. First, the average LS of VA at top firms in 1980 needs to be slightly higher and the value in 2015 lower. The correlation between FC and value added is relatively close though the model falls a bit short on generating FC dispersion measured as $p75/p25$. The simulated employment share distribution has a large explained variance.

[15] * indicates the moment is under construction.

Moment	Model	Data	Data Source
Firm-level moments			
Average LS of VA at L=4	[0.42, 0.38]	[0.6, 0.3]	Kehrigy and Vincent, (2017)
LS of cost, Average	0.67	$*$	Pending
Slope(log(FC),log(VA))	0.16	[0.05-0.1]	Raval, (2017)
p75/p25 FC distribution	1.44	2.1	Raval, (2017)
Employment share by firm size (5 bins)	R^2= 0.83		BLS

Table 6: Targeted Moments: Data and simulated. Data on levels are from 1980 and data on changes use the 1980-2015 period; Alternatively I use the most proximate years available in the cited papers.

6 Aggregate and Firm-level Changes from 1980 to Today

This section describes the results for the calibrated economy. It views the changes from 1980 to today as both decreasing IT capital prices and decreasing elasticity of demand, the latter more so at the larger firms. Regarding markups, I design the experiment in this way to be consistent with the evidence in DeLoecker and Eeckhout (2017), who show that in the US: i-larger firms have larger markups and ii-the increase in the average markup from 1980 to today comes entirely from the top half of the markup distribution. Figure 18 in the Appendix shows the calibrated markups, both before and after.

Figure 11 shows how heterogeneous firms optimally choose L. Firms with larger α sell more, so they find it optimal to add more layers and their fixed cost, in return for a lower marginal cost. Lower IT capital prices makes it cheaper to increase layers, through lower managerial input price indexes.

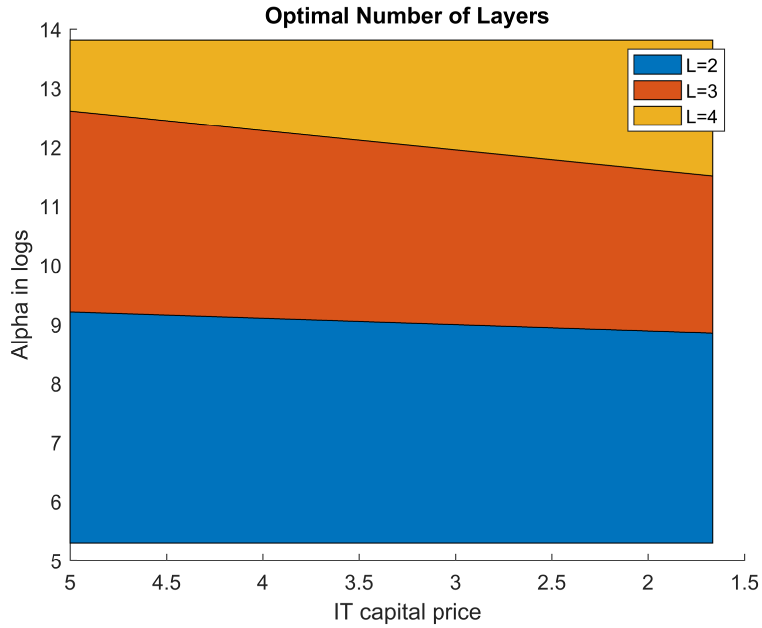

Figure 11: Optimal number of layers as function of (α, p_2).

Figure 12 shows the distribution of the LS of VA and LSC as a function of α. In the cross-section of firms, as value added increases, the LS decreases. This is the result of the mechanisms described in Section 4.7 and the L choice, where the latter causes minor discrete jumps. The figure shows the distribution in 1980 and today. As explained in Section 4.4, for $L = 4$ the LSC tends to increase as p_2 falls, but the opposite holds for L=2 firms. On the right-hand figure, because mark-ups are increasing with α the LS of VA shifts downward gradually with α.

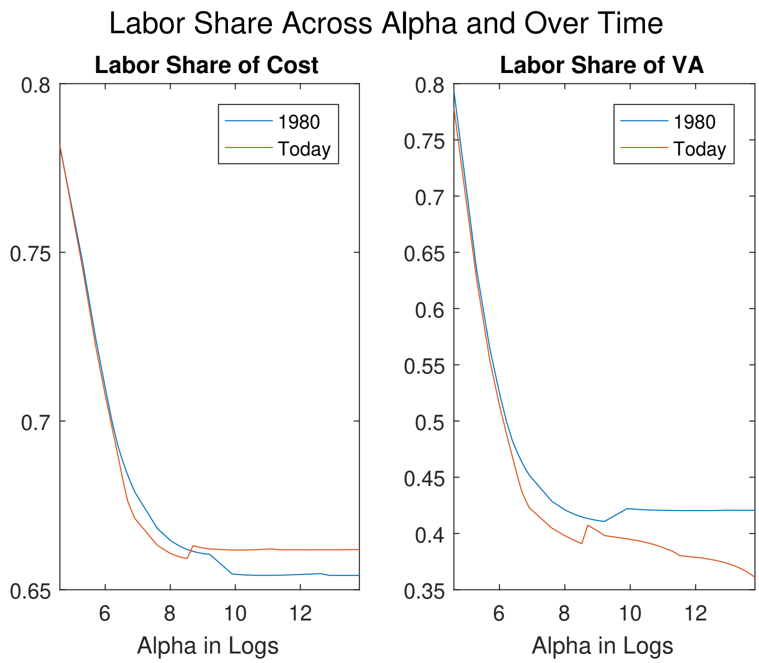

Figure 12: Simulated labor share as function of α.

Table 7 shows untargeted moments, both at the firm level and aggregate. The model delivers simulated moments that are close to the data on a variety of outcomes. Regarding aggregate moments, revenue concentration increases and the aggregate LS declines. The wage bill of managers increases and that of plant-workers falls. This is a consequence of the within-firm changes due to IT: wages of managers increase and so does their employment share, whereas the opposite happens to plant-workers' wages. The data for the US change in the wage bills by occupation type is from Eden and Gaggl (2018), who assign occupations to routine or non-routine labor; in the model, their classification corresponds to non-managers and managers, respectively. As a consequence of all the described movements the aggregate LS falls.

On the lower part of Table 7, I report firm-level moments. The capital-labor ratio in the cross-section increases with firm value added, a feature observed in the data. On the other hand, changes in wages in large firms during the last 30 years are in accordance with the results in Song et al. (2015). In particular, wages fall at the bottom and increase at the top. The data shows larger wage increases at the CEO level than the model delivers, and it might be due to mechanisms that the model lacks. For example, changes to CEO compensation packages, an issue the theory does deal with.

Moment	Model	Data	Data Source
A.Aggregate moments			
Revenue Concentration, CR4 % change	45	[12, 40]	Autor et al., (2017)
Slope(log(CR4),log(Aggregate LS))	-0.26	[0.12, -0.40]	Autor et al.,(2017)
Routine Aggregate Labor Share, % change	-1.14	-42.86	Eden and Gaggl, (2017)
Non-Routine Aggregate Labor Share, % change	21.13	23.33	Eden and Gaggl, (2017)
Aggregate Labor Share	-9.14	-13.11	Karabarbounis and Neiman, (2012)
B. Firm-level moments			
Slope(log(Capital-Labor Ratio),log(VA))	0.18	*	Pending
p50 Wage, % change in top 1% firms	-6.3	-7	Bloom et al., (2017)
p75 Wage, % change in top 1% firms	10.96	64	Bloom et al., (2017)
p99 Wage, % change in top 1% firms	7.44	137	Bloom et al., (2017)

Table 7: Untargeted Moments: Data and simulated. Data on levels are from 1980 and data on changes use the 1980-2015 period; Alternatively I use the most proximate years available in the cited papers. Brackets indicate ranges of estimates in the data moments.

6.1 Counterfactuals

6.2 The Labor Share Decline: Disentangling the Roles of IT and the Elasticity of Demand

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 wage inequality and macroeconomic aggregates. The model provides production-side discipline to explanations for macroeconomic trends and, in particular, allows to determine what is the role demand elasticity changes, a channel emphasized by Autor et al. (2017) and DeLoecker and Eeckhout (2017).

Moment	IT Price	Mark-ups
A. Targeted moments		
Average LS of VA at L=4	110	99
LS of cost, Average	100	100
Slope(log(FC),log(VA))	100	100
p75/p25 FC distribution	100	100
B. Untargeted moments: Aggregate		
Revenue Concentration, CR4 % change	2	101
Slope(log(CR4),log(Aggregate LS))	-760	118
Routine Aggregate Labor Share, % change	123	-44
Non-Routine Aggregate Labor Share, % change	123	-44
Aggregate Labor Share	-17	119
C. Untargeted moments: Firm-level		
Slope(log(Capital-Labor Ratio),log(VA))	100	100
p50 Wage, % change in top 1% firms	100	0
p75 Wage, % change in top 1% firms	41	62
p99 Wage, % change in top 1% firms	66	36

Table 8: Baseline vs Counterfactual Simulated Moments: values in table reported as % of baseline. Column title refers to an experiment where only the label changes as in the baseline, with the rest fixed at 1980 values. Numbers are rounded to closest integer.

Table 8 shows the results of the counterfactual experiments. Values on the table are in percentage terms, relative to the baseline values. Declining IT prices makes the LS of VA for large firms increase, whereas increasing markups delivers results very close to the baseline, i.e., the LS declines. Recall the IT channel makes the LS of VA increase for $L = 4$ firms, whereas larger markups, by increasing value added for a given α decrease the LS of VA. Because the micro LS does not move in the right direction when only IT prices change, and VA concentrates on those firms, this mechanism cannot deliver the decline in the aggregate LS. The IT experiment does generate VA concentration as in the data, but much smaller than the baseline, which together with the aforementioned increased in the LS for large firms generates an increase in the aggregate LS. As a consequence, the IT experiment cannot generate the correlation between LS and VA concentration either. On the other hand, these results suggest that markups are responsible for the decline in the micro LS of VA in the baseline experiment. In fact, because IT increases the micro LS of VA, the markup channel in isolation generates an aggregate LS decline that is too large, relative to the data. Moreover, the markup experiment generates too much value added concentration, which coupled with too much micro LS of VA decline, results in a decline in the aggregate LS that is 20% larger than the baseline.

Turning to inequality, declining IT prices explain wage declines of plant-workers and the wage bills of routine and non-routine workers. Recall the IT price decline reallocates problem solving away from the lowest layer, thereby decreasing their wages, with the opposite changes in managerial layers. As a consequence of cheaper IT capital, the relative demand for managers increases and, on the aggregate, their labor share increases, with the opposite happening for plant-workers. None of these changes can be explained by markups, which shows the importance of IT in generating the observed inequality trends. When markups increase, firms expand[16], which implies firms' demand more knowledge at all layers: plant-workers' wages increase very slightly, unlike on the IT experiment, and their employment grows as firms expand. Managers also have their 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 wage bill share fall, whereas plant-workers increase their wage bill share.

In summary, these results show that inequality is the result of the observed IT price decline. Also, I side with DeLoecker and Eeckhout (2017) in emphasizing increasing markups as relevant for value added concentration in large firms, and the decline in the aggregate LS. Moreover, as in Autor et al. (2017) and Oberfield and Raval (2014), here both the LS in the cross-section of firms, as well as value added reallocation are important for the decline in the aggregate labor share. One lesson from my analysis is that discriminating across alternative explanations for secular macroeconomic trends is possible because of the discipline that the firm organization model imposes on firms' cost structure, something which I discuss in more detail next.

[16]I turn to the heterogenous response of revenues in the cross-section of firms in Section 6.2.1.

6.2.1 How does the firm cross-section matter for macro outcomes?

Understanding the role of demand in the decline of the aggregate labor share requires disentangling both supply and demand factors for each of the counterfactual experiments. Doing so is tightly linked to how organizational choices affect the key outcomes in the cross-section of firms. Figure 13 shows the percent response of revenues, wage bill, marginal costs and LS of value added in the cross-section. While marginal costs fall for all firms when IT prices fall, the effect is stronger for the larger firms which have more layers and use IT more intensely. However, this channel contributes to a relatively small share of the total revenue response in the baseline. This is because, with the calibrated elasticities, even a 10% change in marginal costs does not amount to much revenue increases. As a consequence of the limited production increase, the wage bill also responds little. Finally, for the larger firms, due to capital-labor complementarity and the limited role of CEO IT capital, the falling IT prices makes the LS increase. It is worth emphasizing that while IT does not generate the degree of concentration we observe in the data, it is still crucial in generating the shape of marginal costs which, I will show next, crucially affect how increases in markups are transformed into revenue increases.

For the demand experiment, Figure 13 shows that the response of revenues and the wage bill is very close to that of the baseline, although always below. In percentage terms, the revenue response to the markup experiment is larger for larger firms. This agrees with the results on Table 8, where the lower demand elasticity explains most of the increase in value added concentration in large firms. This result is in contrast to the increased demand elasticity that a pure Melitz (2003) model would require for value added concentration. Why this difference? A simple mathematical argument clarifies this point. I can approximate the response of revenues, r, to changes in the demand elasticity[17] with,

$$\varepsilon_{r\rho} \equiv \frac{dr}{d\rho}\frac{\rho}{r} = (mMC)^{-2\rho}\left[1-(\rho-1)\ln\left(\frac{MC}{m}\right)\right] \tag{25}$$

Note that when the marginal cost is sufficiently large the expression is negative. So in a pure Melitz (2003) model, where small firms have large marginal costs, an increase in demand elasticity reallocates revenues away from small and towards large firms. In this

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

production organization theory, the behavior of the marginal cost in the cross-section of firms is more nuanced: it grows with q given L and jumps down when the firm increases L. Equation 25 highlights the importance of understanding firms' cost structure for inference on how demand might have changed in this period. In other words, the theory of production organization with IT permits identification of the direction in which demand elasticity needs to move to generate revenue concentration and other macroeconomic outcomes, including the decline in the labor share. Importantly, the discipline in the discussion is achieved because the behavior of the MC in this paper is supported by the empirical findings in Caliendo et al. (2016).

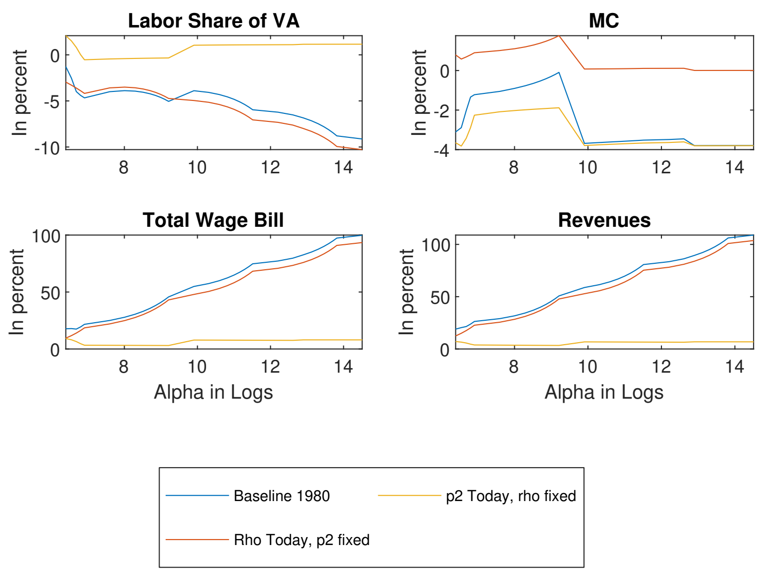

Figure 13: Counterfactual responses of revenues, wage bill, marginal costs and LS of value added in the α cross-section. Values in percentage terms.

Equation 25 is also helpful in that it shows that the total revenue response to ρ in Figure 13 is the result of a composition effect driven by two elements: i-the percent response of revenues to a 1% change in ρ, and ii-the size of the calibrated ρ changes. I discuss both elements in the next paragraphs. Figure 14 presents the first element, the elasticity of revenues to ρ, for the calibrated model. Note how the response is negative meaning that firms expand, with smaller firms relatively less, which generates concentration on the larger firms. Also note that, given the limited variation in the cross sectional response, the size of the calibrated ρ changes must explain an important part of Figure 13.

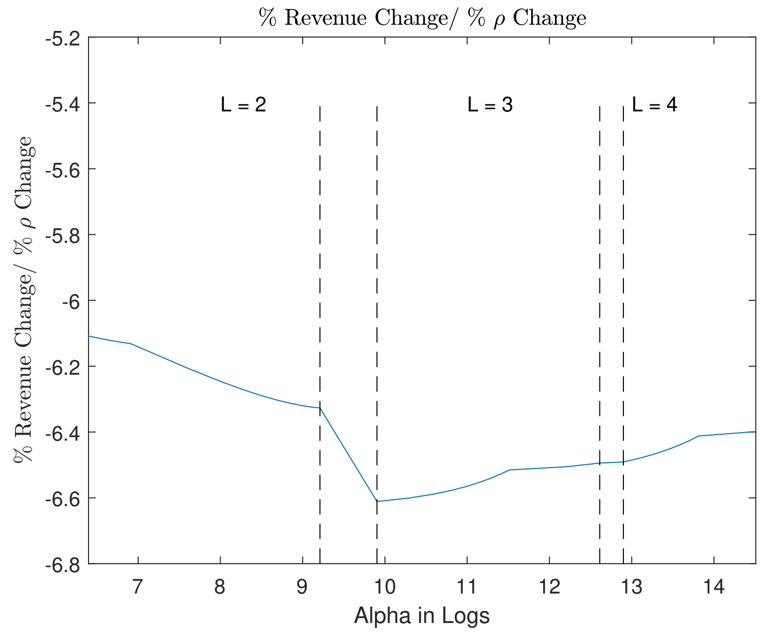

Figure 14: Percent revenue response to a percent ρ change.

I now turn to the role of the aforementioned second element that drives Figure 14, the size of the calibrated ρ changes. I evaluate Equation 25 for two cases: 1-Fix MC at the smallest α and change ρ, and 2-fix ρ at the smallest α and change MC. In Figure 15, I plot each of these channels as a percent of the total implied by Equation 25.

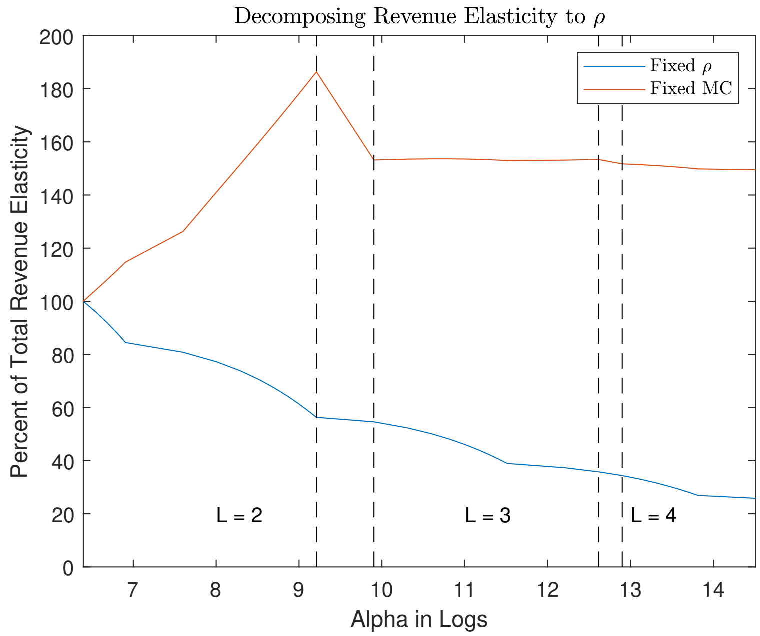

Figure 15: Decomposing the percent revenue response to a percent ρ change. Values in percentage terms relative to the revenue elasticity according to Equation 25.

By construction, at the lowest α both channels explain 100%. Note that both channels contribute in the same direction since the percent explained is always positive. In the case when ρ is fixed, the marginal cost in the cross-section changes, and the ability to explain $\frac{dr}{d\rho}\frac{\rho}{r}$ decreases with α. This shows that the importance of ρ to explain the cross-section of $\frac{dr}{d\rho}\frac{\rho}{r}$ increases with size. In fact, when MC is fixed, the ability of ρ to explain $\frac{dr}{d\rho}\frac{\rho}{r}$ increases very fast for low α and then remains at a high level, in fact delivering larger values than the baseline. This also highlights that, in the baseline, the increase in MC for low α is large and contributes towards less revenue response relative to the fixed ρ experiment. Moreover, the discrete jumps down of the MC as L increases, implies the baseline revenue elasticity increases discretely at those points, reducing the importance of ρ.

Summarizing, as argued above, a key element that affects the percent response of revenues to the calibrated demand elasticity changes is the magnitude of changes in ρ. In the calibration of the baseline economy, the decline in ρ is larger in percentage terms for larger firms, as shown in Figure 18 in the Appendix. The size-related markup increase I impose following DeLoecker and Eeckhout (2017), is important to obtain the increased concentration, since as we have seen the variation in the cross-section for $\frac{dr}{d\rho}\frac{\rho}{r}$

is not too large.

7 Conclusion

In the last 30 years, the reorganization of the typical white-collar workplace has been notably visible as firms have responded to the introduction of computers, email and other internal communication technologies. Big office layouts with secretaries and clerks using typewriters, carbon paper and file cabinets have been replaced with computers and software. Simultaneously, many plant-workers have seen their autonomy, or range of tasks, narrowed down by managerially controlled information systems, as an Amazon warehouse illustrates.

This paper opens the black-box of firm (re)organization with IT. It proposes a theory that formalizes and quantifies the role of firm organization for both micro and macro outcomes. The first key result is that cheaper IT induces firm reorganization that lowers knowledge and wages for plant-workers whereas increases them for managers. While I assume capital-labor complementarity at all layers, as in the data, I quantify the IT capital and plant-labor elasticity to be larger than one in the long-run. This is the second key result of the paper. Such elasticity in this model is a function of organizational choices, in particular, the problem solving reallocation away from the plant and towards the managerial layers. Interestingly, the empirical evidence from Akerman et al. (2015) and Gaggl and Wright (2017) is consistent and closely connected with the reorganization channel: IT reduces wages in routine occupations and increases that of non-routine. This mechanism complements those in Acemoglu and Restrepo (2018) and Hemous and Olsen (2016), who study the effects of automation in endogenously growing economies, where unskilled labor and automation are substitutes.

The third key result of the paper is that firms' organizational knowledge and IT are complementary choices, as in e.g. Bresnahan et al. (2002) and Bloom et al. (2012). Ignoring worker knowledge, leads to a lower TFP which is positively correlated with IT adoption. Hence, correct TFP measurement requires the joint study of IT, worker knowledge and firm organization. These lessons have opened a research avenue in which productivity is estimated using the structure of the model together with linked employer-employee data. This project will lead to a better understanding of the IT effects across managerial and non-managerial layers.

The model is able to replicate key facts related to firms' technology choices. For example, the positive correlation between capital-labor ratios and value added, and the positive firm-size wage premium. As a consequence, it provides production side discipline to look into macro questions, as Oberfield and Raval (2014) emphasize. My baseline calibration for the 1980-2015 period uses two sources of variation: i-an IT price decline and ii-a demand elasticity reduction. I conduct counterfactuals to isolate the role of each channel. I find IT to be responsible for the: 1- within-firm wage declines at low layers and increases at top layers; 2-within-firm, employment share of managers increases; 3-IT capital-labor ratios increases; and, at the aggregate level, as a share of GDP: 4-LS for managers increases, 5-LS for plant-level workers decreases. On the other hand, the decline in the elasticity of demand generates: 6-the decline in the LS of value added of large firms, and 7-the decline in the aggregate LS. Hence, the fourth key result of the paper is that the latter two results are due to increasing markups, whereas inequality is due to IT adoption.

References

Daron Acemoglu. Why do new technologies complement skills? directed technical change and wage inequality. *Quarterly Journal of Economics*, 113, No.4:1055–1089, 1998. 4

Daron Acemoglu. Directed technical change. *Review of Economic Studies*, 113, No.4: 1055–1089, 2002. 4

Daron Acemoglu. Labor- and capital-augmenting technical change. *Journal of European Economic Association*, 1, No.1:1–37, 2003. 4

Daron Acemoglu. Equilibrium bias of technology. *Econometrica*, 75, No.5:1371–1410, 2007. 4

Daron Acemoglu and David Autor. Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 2011. 4, 36

Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares and employment. *American Economic Review*, 2018. 4, 29, 54

Daron Acemoglu, Gino Gancia, and Fabrizio Zilibotti. Offshoring, innovation and wages. *Mimeo*, 2010. 4

Daron Acemoglu, David Autor, David Dorn, Gordon H. Hanson, and Brendan Price. Return of the solow paradox? it, productivity, and employment in us manufacturing. *American Economic Review P&P*, 104, 2014. 21

Anders Akerman, Ingvil Gaarder, and Magne Mogstad. The skill complementarity of broadband internet. *The Quarterly Journal of Economics*, page 17811824, 2015. 4, 36, 54

C. Altomonte, G.I.P. Ottaviano, and A. Rungi. Business groups as knowledge-based hierarchies. *WP LSE*, 2018. 6

Pol Antràs, Luis Garicano, and Esteban Rossi-Hansberg. Offshoring in a knowledge economy. *Quarterly Journal of Economics*, 121:31–77, 2006. 6

Pol Antràs, Luis Garicano, and Esteban Rossi-Hansberg. Organizing offshoring: Middle managers and communication costs. *In The Organization of Firms in a Global Economy, Elhanan Helpman, Dalia Marin, and Thierry Verdier, ed. Cambridge, MA: Harvard University Press.*, 2008. 6

D. Autor and D. Dorn. The growth of low-skill service jobs and the polarization of the u.s. labor market. *American Economic Review*, 103, No.5:1553–97, 2013. 4

David Autor. Market structure, organizational structure and wage structure. *MIT Lecture Notes 14.662*, 2015. 40

David Autor, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. The fall of the labor share and the rise of superstar firms. *NBER Working Paper No. 23396*, 2017. 5, 26, 46, 48

David H. Autor. 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, 2010. 2

David H. Autor. Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, 344(6186):843851, 2014. 2

David H. Autor, Frank Levy, and Richard J. Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4):1279–1333, 2003. 6, 35, 36

Andrew B. Bernard, Teresa C. Fort, and Frederic Warzynski Valerie Smeets. Heterogeneous globalization: Offshoring and reorganization. *Aarhus-Dartmouth U. Mimeo*, 2018. 6

Charles Blackorby and R. Robert Russell. The morishima elasticity of substitution; symmetry, constancy, separability, and its relationship to the hicks and allen elasticities. *The Review of Economic Studies*, 48, 1981. 28

Charles Blackorby and R. Robert Russell. Will the real elasticity of substitution please stand up? (a comparison of the allen/uzawa and morishima elasticities). *The American Economic Review*, 79, 1989. 28, 29

N. Bloom, R. Sadun, and J. Van Reenen. Americans do it better: Us multinationals and the productivity miracle. *American Economic Review*, 102:167201, 2012. 4, 6, 21, 54

Nicholas Bloom, Luis Garicano, Raffaella Sadun, and John Van Reenen. The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60(12):2859–2885, 2014. 3, 36

BLS. Competition drives the trucking industry. *Monthly Labor Review*, April, 1998. 64

T. F. Bresnahan, E. Brynjolfsson, and L. M. Hitt. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117:339376, 2002. 4, 6, 21, 36, 37, 54

E. Brynjolfsson and L. M. Hitt. Computing productivity: Firm-level evidence. *Review of Economics and Statistics*, 85:793808, 2003. 21

Erik Brynjolfsson and Eric McAfee. *Race Agains the Machine*. 2011. 64, 65

Lorenzo Caliendo and Esteban Rossi-Hansberg. The impact of trade on organization and productivity. *Quarterly Journal of Economics*, 127:3:1393–1467, 2012. 3, 6, 15, 40, 68

Lorenzo Caliendo, Ferdinando Monte, and Esteban Rossi-Hansberg. The anatomy of french production hierarchies. *Journal of Political Economy*, 123:4:809–852, 2015. 5, 8, 15, 16, 39

Lorenzo Caliendo, Giordano Mion, Luca David Opromolla, and Esteban Rossi-Hansberg. Productivity and organization in portuguese firms. *CEPR DP*, 2016. 6, 50

E. Caroli and J. Van Reenen. Skill biased organizational change? evidence from a panel of british and french establishments. *Quarterly Journal of Economics*, 2001. 6, 36

Francesco Caselli and Wilbur John Coleman. The world technology frontier. *American Economic Review*, 96, No.3:499–522, 2006. 4

Francesco Caselli and Alan Manning. Robot arithmetic: new technology and wages. *American Economic Review: Insights.*, 2018. 5, 17

Jan DeLoecker and Jan Eeckhout. The rise of market power and the macroeconomic implications. *NBER WP*, Aug 2017. 5, 7, 43, 46, 48, 52

M Eden and P Gaggl. On the welfare implications of automation. *Review of Economic Dynamics*, 2018. 2, 22, 29, 40, 45, 70

Robert Feenstra and Gordon Hanson. The impact of outsourcing and high-technology capital on wages: estimates for the united states. *Quarterly Journal of Economics*, 114, No.3:907–940, 1999. 4

Jeremy T. Fox and Valrie Smeets. Does input quality drive measured differences in firm productivity? *International Economic Review*, 52(4):961–989, 2011. 21

William Fuchs and Luis Garicano. Matching problems with expertise in firms and markets. *Journal of the European Economic Association*, 8, No.2-3:354–364, 2010. 6

William Fuchs, Luis Garicano, and Luis Rayo. Optimal contracting and the organization of knowledge. *Review of Economic Studies*, 82, No.2:632–658, 2015. 6

Paul Gaggl and Greg C. Wright. A short-run view of what computers do: Evidence from a uk tax incentive. *American Economic Journal: Applied Economics*, 2017. 4, 36, 54

Luis Garicano. Hierarchies and the organization of knowledge in production. *Journal of Political Economy*, 108(5):874–904, 2000. 3

Luis Garicano and Esteban Rossi-Hansberg. Organization and inequality in a knowledge economy. *The Quarterly Journal of Economics*, 121(4):1383–1435, 2006. 3

M. Goos, A. Manning, and A. Salomons. Explaining job polarization: Routine-biased technological change and offshoring. *The American Economic Review*, 104 (8):2509–2526, 2014. 4

G. M. Grossman, E. Helpman, E. Oberfield, and T. Sampson. The productivity slowdown and the declining labor share: A neoclassical exploration. *CEPR Discussion Paper 12342*, 2017. 5

Gene Grossman and Esteban Rossi-Hansberg. Trading tasks: a simple theory of offshoring. *American Economic Review*, 98, No.5:1978–1997, 2008. 4

Anna Gumpert. The organization of knowledge in multinational firms. *Journal of the European Economic Association*, 2017. 6

Tim Harford. 50 things that made the modern economy: Robots. *BBC News*, 8 March, 2017. 65

David Hemous and Morten Olsen. The rise of the machines: Automation, horizontal innovation and income inequality. *WP U. Zurich*, 2016. 4, 29, 54

John R. Hicks. *Theory of Wages*. 1932. 27

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

L. Karabarbounis and B. Neiman. The global decline of the labor share. *The Quarterly Journal of Economics*, 129(1):61–103, oct 2013. 2, 4, 5, 30

Matthias Kehrigy and Nicolas Vincent. Growing productivity without growing wages: The micro-level anatomy of the aggregate labor share decline. *Working Paper*, 2014. 5, 26, 40

Michael T. Kiley. The supply of skilled labor and skill-biased technological progress. *The Economic Journal*, 109, No.458:708–724, 1999. 4

Will Knight. Inside amazons warehouse, human-robot symbiosis. *MIT Technology Review*, July 7, 2015. 65

Dongya Koh, Ral Santaeullia-Llopis, and Yu Zheng. Labor share decline and intellectual property products capital. *BGSE WP927*, 2018. 5, 70

Per Krusell, Lee E. Ohanian, Jos-Vctor Ros-Rull, and Giovanni L. Violante. Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica*, 68 (5):1029–1053, 2000. 28

Danial Lashkari and Mart Mestieri. Gains from trade with heterogeneous income and price elasticities. *Harvard WP*, 2017. 7

D. Marin and T. Verdier. Globalization and the empowerment of talent. *Journal of International Economics*, 86, No.2:209–223, 2012. 6

D. Marin and T. Verdier. Corporate hierarchies and international trade: Theory and evidence. *Journal of International Economics*, 94, No.2:295–310, 2014. 6

Marc J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, Vol. 71, No. 6, 2003. 49

Ezra Oberfield and Devesh Raval. Micro data and macro technology. *NBER WP20452, Econometrica R&R*, 2014. 5, 30, 37, 48, 55

OECD. *Special Issue from the First OECD Global Forum on Productivity*. 2017. 2

Rachel Premack. Truck driver salaries have fallen by as much as 50 percent. *Business Insider*, 26 Sep, 2018. 64

Devesh Raval. The micro elasticity of substitution and non-neutral technology. *Revise and Resubmit, RAND Journal of Economics*, 2017. 30, 40, 41

Joan V. Robinson. *The economics of imperfect competition*. 1933. 28, 29

Andres Rodriguez-Clare and Natalia Ramondo. Growth, size and openness: A quantitative approach. *Mimeo*, January, 2010. 4

S. Rosen. Authority, control, and the distribution of earnings. *The Bell Journal of Economics*, 13:2, 1982. 3

Clara Santamaria. Small teams in big cities: Inequality, city size, and the organization of production. *Princeton Thesis*, 2017. 6

Jae Song, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. Firming up inequality. *NBER WP21199 QJE R&R*, 2015. 4, 26, 37, 45

Alexandra Spitz-Oener. The impact of outsourcing and high-technology capital on wages: estimates for the united states. *Journal of Labor Economics*, 24, No.2:235–270, 2006. 4

Mathias Thoenig and Thierry Verdier. A theory of defensive skill-biased innovation and globalization. *American Economic Review*, 93, No.3:709–728, 2003. 4

Daniel J. Wilson. Investment behavior of u.s. firms over heterogeneous capital goods: A snapshot. *Federal Reserve Bank of San Francisco Working Paper 2004-21*, March, 2006. 2

8 Appendix

8.1 Firm Organization, Information Technology and Wages: Case Studies

A feature of the hierarchy literature is that a firm is an efficient problem-solving organization, where lower layers deal with the most routine problems, called tasks in Autor parlance. In this way, the limited time of CEOs, and other managers, is leveraged by practicing a "management by exception", whereby their input is used only if nobody else can solve the problem. My contribution to the theory is to model the adoption of information technology capital, which is a complementary way to leverage managerial time, since this type of capital is, perhaps 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 done by the lower layers. Reorganizing in this way is cost minimizing, since it lowers compensation to the numerous workers at the plant-level, as they now require less skill. In this section I illustrate the ubiquity of this organizational mechanism.

The organization of an Amazon warehouse is an excellent illustration of the model. At the center of the warehouse is a storage space containing square shelves packed with the products. In previous generations of its order fulfillment center, workers would have roamed these shelves searching for the products needed to fulfill each new order. Now the shelves themselves glide quickly across the floor carried atop robots about the size and shape of footstools. These robots either rearrange the shelves in neatly packed rows, or bring them over to human workers, who stack them with new products or retrieve goods for packaging. Amazon's robotic shelves make stacking and picking more efficient by automatically bringing empty shelves over to packers or the right products over to pickers. Workers pack products into boxes for shipping with help from Amazon's central computer systems. Items retrieved from storage shelves are automatically identified and sorted into batches destined for a single customer. The computer knows the dimensions of each product and will automatically allocate the right box, and even the right amount of packing tape. This organization of work reduces the degree of autonomy of plant-workers to an extreme degree, as yet another gadget nicely illustrates: the Jennifer Unit, a headset, tells plant-workers what to do down to the smallest detail; if seventeen equal products have to be picked from the shelves, the system will request five, then five, then five and, finally, two items. It seems safe to conclude that, removing decision-making from workers' occupations through IT capital, allows Amazon to reduce costs by paying lower wages than otherwise. Such a conclusion seems not only reasonable but also agrees with worldwide news of Amazon plant-workers requesting wage increases.

Trucking jobs have seen similar trends. As early as 1998, a BLS analysis of the trends in the sector highlighted that many new technologies had been adopted, BLS (1998). Examples listed were electronic data interchange, new vehicle location detection systems, or voice and data communication services. The leader provider of such technology in the transportation sector is Qualcomm, who 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 it's "OmniTRACS" system. It involves an "in cab" communication device, the Automatic Vehicle Location (AVL) unit, which allows the driver to communicate with their 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 pre-formatted 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, like, 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 tasks is extreme and similarly low wages are observed 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 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, such as 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 were language skills are necessary.

Skilled jobs are not immune to this deskilling trend. On a 2011 New York Times article titled "Armies of Expensive Lawyers, Replaced by Cheaper Software", John Markoff highlighted 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 prohibitively large, as the required legal staff hours pile. The article reports how, 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, Blackstone Discovery of Palo Alto, Calif., helped analyze 1.5 million documents for less than \$100,000, Brynjolfsson and McAfee (2011). The type of software used goes well beyond finding documents with relevant terms. They 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.

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[18].

8.2 Layer L Capital-Plant Labor Elasticity

While the role of the $k_L - n_1$ elasticity in determining the IT capital-plant labor elasticity is not quantitatively as relevant as the role of $k_2 - n_1$, it is still interesting to understand it's determinants. Why? Because it can be written as a function of: i-the response of each z_l, something that can be mapped to data, or ii-intuitive terms, which help understanding the model's mechanisms. For both (i) and (ii), I use the same reasoning as for σ_{n_1,k_2}, and decompose only the numerator because the denominator is close to one. Both decompositions are made such that the sum of all elements add to one hundred. The first decomposition of the $k_L - n_1$ elasticity uses the same terms to those defined above, except for the last:

[18] The Amazon example is based on a Tim Harford Podcast on Robots, Harford (2017), and on a MIT Technology Review article, Knight (2015).

$$
-\,dlog\left(\frac{k_L}{n_1}\right) \propto W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} + \cdots
$$

$$
O_1^L(\lambda,\beta_L)\frac{\partial z_1}{\partial \log(p_2)} + O_L(\lambda,\beta_L)\frac{\partial z_L}{\partial \log(p_2)} + O_1^L(\lambda,\beta_L)\left(\sum_{l>1}^{L-1}\frac{\partial z_l}{\partial \log(p_2)}\right) \tag{26}
$$

where all functions are as defined in the main text. The first term is the within-layer 1 effect. The rest are all organizational terms, each corresponding to a layer. The last term, the only new one, involves a summation of all intermediate layers so it is zero for $L = 2$ firms. Decomposing the elasticity in these terms is done in Figure 17. It shows a similar message to that for the $k_2 - n_1$ elasticity: the organizational term associated to layer 1 knowledge is the fundamental contributor to the substitution. In $L > 2$ firms, the intermediate layers also play a substantial role, against $k_2 - n_1$ substitution. This is because, other things equal, more knowledge in those layers is related to: i-fewer problems arriving to the top layer, and ii-a larger n_1 demand.

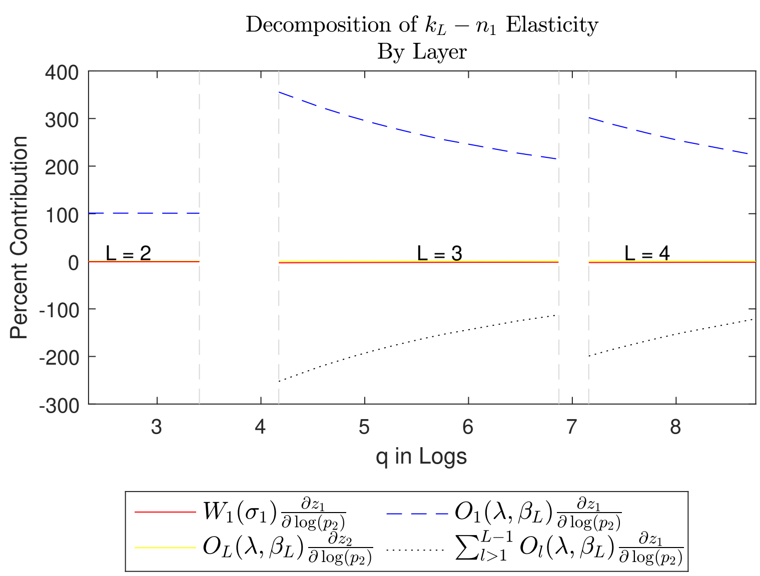

Figure 16: Decomposition of the Layer L Capital-Plant Labor Elasticity numerator as in Equation 26.

The alternative decomposition of the numerator of the $k_L - n_1$ elasticity, uses the production function to write:

$$- dlog\left(\frac{k_L}{n_1}\right) \propto W_1(\sigma_1)\frac{\partial z_1}{\partial \log(p_2)} - \frac{1}{F(Z_L)}\frac{dF(Z_L)}{d\log(p_2)} - \frac{1}{\beta_L}\frac{G_k}{G}\frac{dk_L}{d\log(p_2)} \tag{27}$$

where recall $G \equiv 1 + Bk_L^{\beta_L}$ and $G_k = \frac{\partial G}{\partial k_L}$. In Equation 27, only the second and third terms are new with respect to the layer-by-layer decomposition above. The second term captures the percent response of the fraction of problems solved, and, by Proposition 3, it is negative. The term is connected to the Solow Paradox, and, quantitatively turns out to be small in terms of its contribution to the $k_L - n_1$ elasticity, as shown on Figure 17. The third term captures the percent response of the number of problems attempted by the CEO. It is positive for $L = 2$, by Proposition 3, and quantitatively positive, for all L; It is large, outweighs the other two terms, and essentially determines the $k_L - n_1$ elasticity. Put differently, conceptually, $k_L - n_1$ are substitutes because the CEO leverages his time using IT capital, and, through automatization, he tackles more problems.

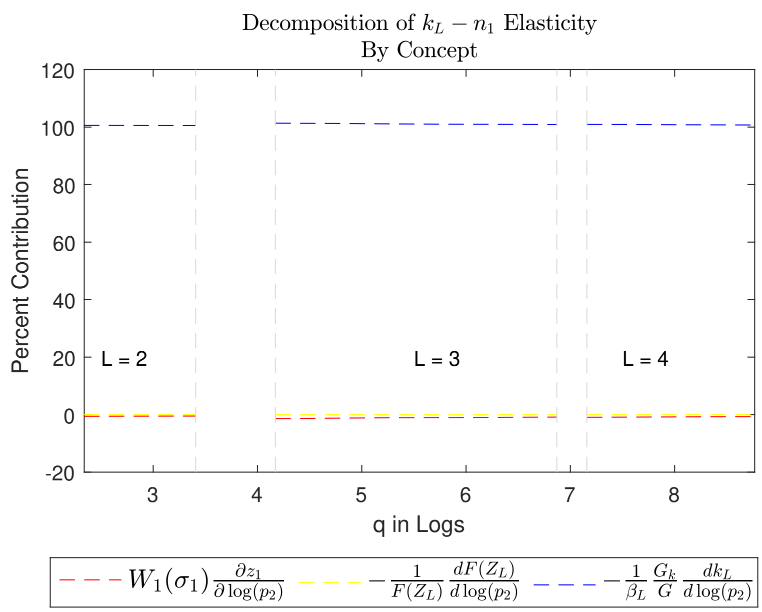

Figure 17: Decomposition of the Layer L Capital-Plant Labor Elasticity numerator as in Equation 27.

8.3 Technical Appendix

8.3.1 Assumptions

Assumption 1 *Parameters* (λ, c, σ_l) *are such that* $\frac{\lambda}{c} > \sigma_l$, $\forall l$.

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

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

B) $\forall l > 1$*:*

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

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

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

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

Before proving any results I provide intuition of the role of Assumptions. Under the conditions on parameters in Assumption 1, I show that $k_L > 0$ and that $z_l > 0$, $\forall 1 < l \leq L$. For simplicity, I focus on the simple case L=2.

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 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 communication cost has to be low. Back to my case, I can also rule out solutions where $z_1 = 0$ but $z_L > 0$, because no firm would want to hire workers without knowledge, as it could produce the same at a lower cost without them, ie as a self-employed firm.

Moreover, solutions where $z_1 > 0$ but $z_L = 0$ are ruled out by a parameter restriction where the learning cost of the CEO is lower than the effect of adding layer 1 when $z_1 = 0$, ie,

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

The first element in the left side of the inequality is the 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. This second term captures how fast variable costs of knowledge change. 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 at such z values. The term on the right of the inequality is the

CEO learning cost of knowledge adjusted for the mass of problems at 0. Intuitively, for high λ, the variable cost effect of an extra layer is low, but so is for the CEO knowledge; an extra layer however has a fixed cost, hence, zero CEO knowledge is not optimal.

Assumption $c/\lambda - \frac{1}{\sigma_1} > 0$ is enough to guarantee that z_1 increases with scale. Finally, to obtain the natural result that k_L, and z_L are increasing with scale two parametric conditions are enough. The first involves not too strong decreasing returns to scale on k_L, ie $\beta_L > 1/2$. The second requires that productivity B not be too large, ie

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

The intuition in this discussion can be analogously extended to more complex organizations.

8.3.2 Proofs

[FULL TECHNICAL APPENDIX AND PROOFS AVAILABLE UPON REQUEST]

Lemma 1 *Given the number of layers L, under Assumption 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_0 and finally z_L. Moreover, $k_L > 0$ always holds.*

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

8.4 Calibration and Data

I follow the the definition of IT in Eden and Gaggl (2018). It includes the following asset categories from the BEA detailed fixed asset accounts. Within non-residential 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). In their paper, they report a decline of roughly 2/3-ths in the price of IT capital relative to that of final goods. Note this definition differs from that in Koh et al. (2018) for IPP. For IPP, prices behave similar to other assets, whereas for the IT capital aggregate used by Eden and Gaggl (2018) and this paper, there is a decline in the relative price of IT relative to the non-IT capital. In fact the latter relative decline is just as large as the decline in the relative price of IT capital to the GDP deflator. 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.

Figure 18 illustrates the calibration of markups.

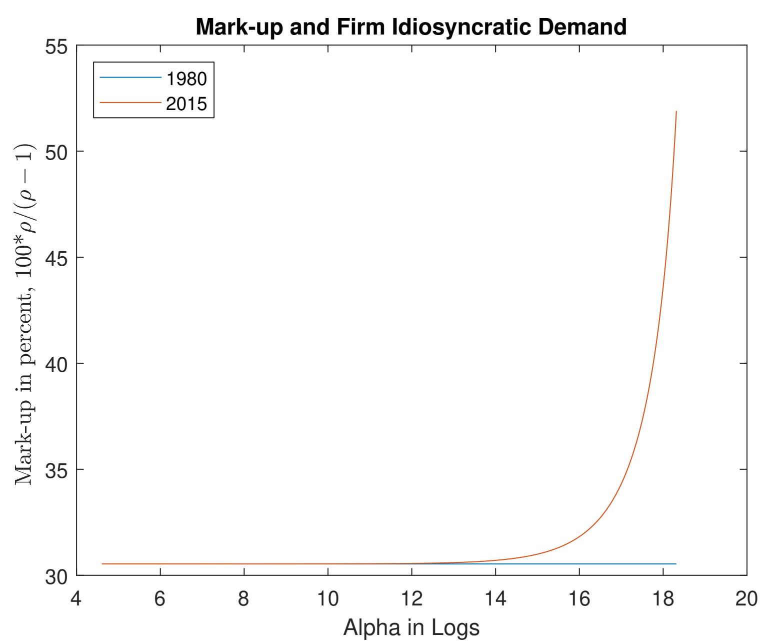

Figure 18: Calibrated markups as function of α.

8.5 Recent Automation Models

IN PROGRESS