Organizational Responses to Product Cycles*

Achyuta Adhvaryu† Vittorio Bassi‡ Anant Nyshadham§
Jorge Tamayo¶ Nicolas Torres‖

November 11, 2025

Abstract

We study how firms reorganize to adapt to changing production conditions over the product cycle. Using daily data from an automobile assembly plant, we show that when the firm upgrades its products, hierarchies flatten and training increases. In contrast, when production of existing products scales up, hierarchies expand without additional training. We develop and quantify a dynamic model that explains these patterns: product innovation raises productivity but temporarily increases problem complexity, making training and hierarchy short-run substitutes and long-run complements. Taken together, these results reveal an *organizational cycle*: firms compress and train during innovation, then expand hierarchies as production scales.

JEL Codes: D22, M12, M53, O3

Keywords: product upgrading, product cycles, organizational design, knowledge hierarchies, training, auto manufacturing

*We would like to thank Ricardo Alonso, Catherine Thomas, and Kate Whitefoot for insightful discussions of the paper. We also thank Daron Acemoglu, David Atkin, Augustin Bergeron, Luis Cabral, Robert Clark, Wouter Dessein, Florian Englmaier, Miguel Espinosa, Ben Friedrich, Luis Garicano, Deepak Hedge, Mitch Hoffman, Larry Katz, Amit Khandelwal, Jin Li, Rob Metcalfe, Tommaso Porzio, Andrea Prat, Esteban Rossi-Hansberg, Raffaella Sadun, Jesse Shapiro, Chris Stanton, Eric Verhoogen, Tom Wollmann, as well as participants at many seminars and conferences for useful comments. Julian Jimenez Cardenaz and Ana Maria Ramos provided exceptional research assistance. All errors are our own.

†UC San Diego, NBER, J-PAL, Good Business Lab & BREAD; email: aadhvaryu@ucsd.edu.

‡University of Southern California, NBER, J-PAL, BREAD, CEPR & IGC; email: vbassi@usc.edu.

§University of Michigan, NBER, J-PAL, Good Business Lab & BREAD; email: nyshadha@umich.edu.

¶Harvard Business School, Digital Reskilling Lab - The Digital, Data, and Design Institute at Harvard & NBER; email: jtamayo@hbs.edu.

‖University of Michigan and Good Business Lab; email: fnicolas@umich.edu.

1 Introduction

Firms must adapt to grow. When scaling up production to meet demand shocks, firms expand hierarchies by adding layers of management, reallocating knowledge across them, and redefining spans of control (Caliendo et al., 2020, 2015; Caliendo and Rossi-Hansberg, 2012). This restructuring reflects a fundamental trade-off between internal coordination and communication costs across the hierarchy on the one hand, and the efficiency gains from specialization and delegation on the other.

Yet producing more is not the only way in which firms grow. Many industries are characterized by product cycles, in which firms must periodically innovate to stay competitive, introducing new generations of products that are often more complex and technologically advanced (Aghion et al., 2005; Aghion and Howitt, 1990; Grossman and Helpman, 1991, 1994; Krugman, 1979; Vernon, 1966). Growth through innovation requires firms to acquire and proliferate new knowledge internally, as product introductions create new problems to solve. While the notion of organizational learning in the context of product upgrading is not novel (Levitt et al., 2013), we still have limited evidence on the ways in which firms adapt their internal organization to facilitate product innovation over the product cycle.[1]

In this paper, we document how these two forms of firm growth—innovation and expansion—are connected by an internal *organizational cycle*. Using high-frequency data on the hierarchical structure, training investments, and worker-specific problem-solving activities from an assembly plant of a global automaker, we document systematic adjustments in the firm's organization as it transitions between phases of product upgrading and expansion of production. During the innovation phase, the firm compresses its hierarchy and intensifies training. Then, when production of the upgraded product ramps up, the firm expands its hierarchy by adding layers and employment, without additional training.

We rationalize these patterns through a dynamic model of the firm, in which organizational design and training investments are jointly determined over the product cycle. When output is fixed in the short-run, training and hierarchy act as *substitutes*: facing new product complexity, the firm flattens its structure and invests in skill formation. As output expands in the longer-run, training and layers become *complements*: earlier training raises productivity by enabling product innovation, and the resulting efficiency gains facilitate hierarchical expansion and scale growth without further training. Quantifying the model with our data, we find that both organizational levers are central to reducing the costs of product introduction, with

[1] Related literature has emphasized other factors enabling quality upgrading, such as plant scale and input quality (Kugler and Verhoogen, 2012), competition in input markets (Amiti and Khandelwal, 2013; Fan et al., 2015), export access (Atkin et al., 2017; Verhoogen, 2008), managerial skill (Adhvaryu et al., 2023c; Bandiera et al., 2020; Bloom and Van Reenen, 2007, 2010), and knowledge spillovers (Bai et al., 2021).

short-run training flexibility playing a relatively larger role than flexibility in hierarchy.

Our data come from an Argentinian subsidiary of the automaker, and include exceptionally rich daily records on production activities and the internal organization of the plant between 2012-2019, as well as worker-specific training and problem-solving activities. This setting is ideal for our analysis for two reasons.

First, the data combine detailed operational and organizational information. We observe the number of vehicles produced and the number of defects per vehicle (DPV)—a key performance indicator of the plant. The assembly line is organized into more than 130 working groups, for which we have complete information on hierarchical composition and training activities for all employees. We also observe the wage schedule across hierarchy levels and the firm's continuous-improvement system, which records problem-solving suggestions by workers at all levels, each scored by management according to its complexity and value to the firm.[2] Working groups are stable and persistent production units that exhibit clear pyramidal structure: workers further up the hierarchy have completed more training programs, earn higher wages, and solve more complex problems. This pattern validates interpreting the organization of working groups through the lens of canonical knowledge-hierarchy models (Caliendo and Rossi-Hansberg, 2012; Garicano, 2000; Garicano and Rossi-Hansberg, 2012).

Second, during our observation period, the plant underwent multiple product cycles. Each year, upgraded model variants are introduced—what we refer to as Model changes—and the firm periodically expands planned production in discrete steps, or Volume changes. Between 2012-2019, the plant experiences seven Model changes and five Volume changes, expanding annual output by over 40% over this period. We know the exact timing of each event, which allows us to trace how the firm adjusts its internal organization over the product cycle.

We estimate event-study and regression-discontinuity-in-time specifications around Model and Volume changes to document how production and organizational outcomes respond. Although both types of changes are planned in advance, we find no evidence of anticipatory adjustments. Conversations with management indicate (and event-study pre-trends confirm) that advance restructuring is limited because production runs continuously with no downtime and because the problems arising from production changes are difficult to anticipate before implementation. The absence of anticipatory responses enables us to document empirically how the firm adjusts its hierarchy and training after Model and Volume changes.

We first show that Model changes do not immediately alter planned production or total parts but introduce a large share of new components: roughly 13–15% of the parts assembled after each upgrade are new. These components create unanticipated operational challenges.

[2] As part of the firm's continuous-improvement system, workers are incentivized (both monetarily and by way of recognition) to submit valuable suggestions of proposed solutions to new problems they encounter.

Small differences in the material or shape of a bumper or console button, for example, can substantially alter the appropriate angle or force required for assembly. Consistent with the difficulty of predicting such issues *ex ante*, defects per vehicle (DPV) spike sharply—by about two standard deviations on average—and gradually return to baseline within roughly three weeks. Even these temporary disruptions are costly: a day with DPV one standard deviation above the mean reduces output by nearly 18%, and we document that these losses are not subsequently recuperated.

We then examine how changes in the firm's internal organization enable rapid problem-solving after a Model change. Along the training margin, the firm upskills lower- and mid-level workers in problem-solving and communication skills, expanding the pool of employees capable of addressing the new problems that arise. Consequently, the aggregate knowledge level within each working group, and as a result the firm as a whole, increases.

Along the hierarchical margin, the shape of working groups shifts from distinctly pyramidal to more rectangular. The ratio of frontline workers to middle and upper-level managers (the problem-solvers' span of control in Garicano (2000)) falls by about 0.4 from a mean of 4 in the first four weeks following a Model change. Groups also become flatter, with fewer distinct knowledge layers. The firm defines 16 possible layers, with an average working group of 20 workers spanning roughly eight layers. In the first four weeks after a Model change, the number of hierarchical layers declines by about 10%. These changes, unlike the changes in aggregate knowledge, are largely temporary: once defect rates return to pre-change levels, the firm back-fills lower positions and restores the original number of layers.

Using novel data from the continuous-improvement system, we study how these adjustments affect problem-solving. Both the number and the complexity/value scores of suggestions increase following Model changes, particularly among lower- and mid-level workers who receive the additional training. These findings are consistent with an increase in the frequency of complex problems following the Model change and confirm that the new training is indeed productive: workers generate more valuable suggestions, and the effect persists over time.

We then further validate that changes in training and hierarchy both have real productivity implications. Exploiting variation in the intensity of training and hierarchical restructuring across working groups, we show that groups receiving more training and/or undergoing greater flattening of the hierarchy experience smaller spikes in DPV. These results are robust to instrumenting the extent of restructuring with worker exits around Model changes. In contrast, worker experience in the working group or the plant does not predict differences in DPV effects. In sum, adjustments in training and hierarchical structure appear to be key mechanisms enabling adaptation to new products, over and above any learning by doing.

We then contrast these findings with the firm's responses to increases in production volume, finding that the organizational response is a sustained expansion of employment and hierarchical layers, consistent with evidence from manufacturing firms in high-income countries (Caliendo et al., 2020, 2015; Ewens and Giroud, 2025; Friedrich, 2022). Importantly, this expansion is not accompanied by additional training. Because the problems that arise from higher output are routine and easily addressed by frontline workers, the pyramidal structure of working groups actually becomes *more* pronounced: their bases expand while middle- and upper-level managers are reallocated to new groups. The average stock of knowledge in working groups thus declines. Employment rises less than proportionately with output, as cars per worker increase due to economies of scale. Finally, consistent with the more hierarchical structure and lower average knowledge, the number and average score of suggestions to new problems fall sharply after Volume changes.

This comparison shows that the plant uses training very differently in response to Model and Volume changes. We probe this further using data from the in-house training programs of the plant, which show that in response to Model changes, the plant invests in improving its training programs by increasing the number of courses and teachers. Instead, we find no significant change in these outcomes after Volume changes.

We then develop a dynamic model to interpret and unify these findings, quantifying how training investments and hierarchical structure jointly help manage innovation over the product cycle. Building on Garicano and Rossi-Hansberg (2012), the firm chooses when to innovate and how to organize knowledge to minimize costs over time. Innovation permanently raises productivity but temporarily increases problem complexity, which can be managed through additional training or hierarchical reorganization.

Guided by the empirical evidence of improvements in training programs following Model changes, we extend the canonical model by allowing the firm to invest in improving the efficacy of training (by reducing training costs). More complex problems make training costlier, and deeper hierarchies further raise these costs by complicating coordination: intuitively, it is more costly to organize and deliver training in an organization with many layers. Each period, the firm chooses its training and hierarchical depth to minimize total costs—comprising training investments (with adjustment costs) and the coordination costs of maintaining a particular hierarchy—subject to production and feasibility constraints. The optimal response to phases of innovation is to temporarily flatten the hierarchy and expand training, then gradually re-expand the hierarchy and reduce training as problems become routinized. This generates short-run *substitutability* between training and layers when output is fixed, consistent with our empirical results.

We calibrate the model to match the observed dynamics around Model changes and use it to assess the relative importance of the two margins. In counterfactuals, constraining

training flexibility raises costs by 2.4% during the innovation phase, while fixing the hierarchy increases them by 1.4%, implying that both margins matter but that adaptive training plays a more important role for cost minimization. Finally, when demand rises after the innovation and learning has taken place, the firm deepens its hierarchy without additional training: higher volume creates a capacity challenge, not a complexity one. Because retraining is costly and less effective at this stage, the firm expands layers to exploit specialization and manage scale more efficiently.

The model thus clarifies the dynamic interaction between training and hierarchy. In the short run when output is fixed, the two margins act as *substitutes*: when new problems arise, the firm flattens and trains. Over time, however, training facilitates the adoption of more productive technologies, which in turn make it easier to expand output in the long run through additional hierarchical layers without further training. In this sense, short-run training and long-run hierarchy function as *complements*. Together, this defines what we term the firm's *organizational cycle*: a systematic alternation between periods of compression and expansion in the hierarchy accompanied by shifts in training investments and a resulting rise and fall in average knowledge per worker that occurs over the product cycle.

Related Literature and Contribution Our study links the literature on knowledge hierarchies and firm growth with studies of product upgrading and innovation by introducing the concept of an *organizational cycle*: a recurring pattern of adaptation in training, team structure, and problem-solving allocation over the product cycle. Studies of firm-level upgrading have emphasized the roles of plant size, input quality, export access, and managerial capability in enabling product upgrading and innovation (see Verhoogen (2023) for a recent review). Yet little is known about the internal organizational adjustments that accompany upgrading. Using granular data from a large automobile assembly plant, we confirm prior findings that volume growth coincides with an expansion of hierarchical layers and reductions in frontline problem-solving (Caliendo et al., 2020, 2015; Ewens and Giroud, 2025; Friedrich, 2022). We then provide the first evidence that product upgrades instead coincide with compression of the hierarchy, reductions in layers and spans of control, and reallocation of problem-solving to lower layers.[3]

To rationalize these findings, we extend the canonical knowledge-hierarchy framework by endogenizing both the *distribution* and *stock* of knowledge. Whereas earlier models treat the cost of knowledge acquisition as fixed or exogenous (Caliendo and Rossi-Hansberg, 2012; Garicano, 2000; Garicano and Rossi-Hansberg, 2006, 2012), we allow the firm to jointly

[3] A related literature studies organizational adaptation to changing conditions, showing that decentralization and limited specialization facilitate adaptation (Alonso et al., 2008; Dessein and Santos, 2006).

choose hierarchical structure and training capacity—an investment that lowers the cost of knowledge acquisition. This captures the core trade-off in the data: training and delegation act as short-run *substitutes* when output is fixed but become long-run *complements* as scale adjusts. The firm initially expands training to manage new-product complexity, then deepens hierarchies as problems become routinized and scale rises.[4]

We also contribute to the literature on organizational learning and "learning by doing" (Arrow, 1962; Irwin and Klenow, 1994; Thompson, 2010). The closest study to ours in this literature, Levitt et al. (2013), documents spikes in defect rates during automobile model launches and interprets learning as the accumulation of "organizational capital." We extend this view by directly measuring the mechanisms of learning: firm investments in training capacity and changes in hierarchical depth and problem-solving allocation. Whereas Levitt et al. (2013) emphasize routinization through managerial attention and whiteboard systems, we show that firms actively reconfigure span of control and invest in training to acquire and diffuse new knowledge. We find little evidence that these improvements stem from group-level experience, indicating that adaptation is largely organizational: a collective, knowledge-intensive process requiring formal changes in how knowledge is acquired, allocated, and applied to problem-solving across the firm.

Finally, the organizational responses we document mirror core principles of "lean production," widely diffused from the Toyota Production System (TPS) (Dillon, 2019; Holweg, 2007; Liker, 2020; Ohno, 2019). Our evidence that the plant expands training and flattens hierarchies aligns with the TPS focus on decentralized problem-solving and continuous improvement (*kaizen*). These mechanisms are not unique to the auto context: similar emphasis on distributed problem-solving and training underpins adaptation to technology and demand in fast food (Adhvaryu et al., 2023b, 2022c) and productivity gains from managerial and supervisory training in garment manufacturing (Adhvaryu et al., 2023a, 2022b). Our contribution is to document and quantify how two central lean production levers—training and hierarchical structure—are operationalized in the organizational cycle of an innovating firm.

Structure of the Paper Section 2 describes the setting and data. Section 3 presents the empirical strategy, and Section 4 the empirical results on the organizational responses to Model and Volume changes. The model is developed in Section 5, the calibration and counterfactuals are described in Section 6, and Section 7 extends the model to consider changes in quantity. Section 8 concludes. Additional details are in the Online Appendix.

[4] In highlighting the importance of training investments as an organizational lever over the product cycle, we also relate to the literature studying the returns to training programs within the firm (Adhvaryu et al., 2023a; Espinosa and Stanton, 2022; Sandvik et al., 2020).

2 Context, Data and Descriptives

In this section, we first describe the organization of production in our partner plant and discuss the timing and characteristics of product cycles. We then present the various data sources used in our analysis.

2.1 Context, Organization of Production, and Product Cycles

We partnered with a leading global auto manufacturer's subsidiary, whose plant is located in Argentina. The plant began operating in the 1990s and now produces more than 140,000 cars per year, employing over 3,400 workers. Around two-thirds of the production is exported to different countries in Latin America, and the rest is dedicated to the local market. We have data on the production activities of this plant for the period 2012-2019.[5]

2.1.1 Production Process

Production takes place on a production line comprised of eight sectors: Press, Welding, Painting, Frame & RX Axle, Engines, Resin, Assembly, and Quality Check. The different parts of a production unit (i.e., an automobile) are produced in parallel by different sectors: chassis and car-body components are manufactured and connected by Press and Welding, respectively, and the rest of the parts are manufactured by Frame & RX Axle, Engines, and Resin. Finally, these different parts are put together in the Assembly sector. Assembly is one of the most important but also most error-prone phases of production, with 75% of the defects per vehicle occurring at this stage of the production process. The Assembly sector is itself divided into four sub-sectors: Trim, Chassis, Final and Repair. Our study focuses on the Assembly sector, both because it is where the majority of defects occur and because we are able to track working group composition for this sector over the period of our study.

The pace of production is based on the "takt time", which is the maximum amount of time that a product (i.e., a vehicle in this case) should take to go through the assembly line.[6]

[5] We can compare the size of our plant to that of other car assembly plants studied in the literature. Our partner plant produces an average of about 2,500 cars per week or 400 cars per day. For comparison, the automaker assembly plant studied in Levitt et al. (2013) has a capacity of about 3,500-4,500 cars per week. Using information from 66 US-based automotive assembly plants for the period 1993-2007, Giardili et al. (2023) show that the average plant produced about 4,200 cars per week, with a standard deviation of 1,800. Finally, MacDuffie et al. (1996) use data from 70 assembly plants worldwide collected in 1989-90, and show that assembly plants in "newly industrialized countries" (i.e., Brazil, Korea, Mexico and Taiwan) produced about 600 units per day on average. This comparison shows that our partner plant is smaller than the average plant studied in the literature, but is by no means an outlier: for instance, our plant falls within 1 standard deviation of the mean size of the assembly plants studied by Giardili et al. (2023).

[6] For example, if the plant has a target of 100 vehicles in a day and operates for 8 hours per day, the takt time is (8 x 60 x 60)/100 = 288 seconds per car.

At a given "takt" time, the line would generally produce a fixed number of cars each day or shift. However, the assembly line operates almost entirely sequentially such that issues at one process or station along the line will lead to stoppages of the entire line, slowing the flow of completed vehicles. Therefore, the critical variation in productivity (i.e., cars completed per unit time) comes from these stoppages due to problems and any defects they cause.

2.1.2 Organization of Labor, Knowledge Hierarchies, and Training Provision

Working Groups Workers are divided into working groups within each of the four sub-sectors of the assembly line. The plant has two shifts per day (morning and afternoon) with each working group operating in one shift. There are about 130 working groups operating on the assembly line at each point in time. Each working group has about 20 workers.

Knowledge Hierarchies Employees within each working group are organized in a clear production hierarchy, characterized by different *layers*. Workers are assigned to different layers based on the training they have accumulated through in-house training programs and the tasks they perform. The different layers are therefore characterized by different skill or "knowledge" levels. Table 1 shows a diagram of the hierarchical structure of working groups.

There are three main employee categories within the hierarchy. The first are Front-line Operators (FL), who engage in production tasks on the production line. Above FL workers are Mid-line Operators (ML), who are tasked with identifying and solving production problems and closely supervising and helping FL workers whenever problems or bottlenecks occur. Finally, above ML workers are Superiors (S), who are in charge of supporting and managing the overall working group performance.

Within each main employee category (i.e., within the FL, ML and S categories), employees are then divided into different layers. FL workers are divided into three possible layers of New Front-line Operators (New FL), four layers of Regular Front-line Operators (Reg FL), and one layer of Mid Front-line Operators (Mid FL). New FL are the lowest level of the hierarchy, comprising workers who have recently started and have received only basic training. Reg FL workers are more experienced and skilled, having typically spent at least a year on the assembly line and having completed at least five training programs, and Mid FL are the highest skilled within the FL category, having more than two years of experience and having gone through 13-14 training programs. ML and S workers are subdivided into four layers each, again based on their training. So, in total, there are 16 possible employee layers that a working group can have—8 within FL workers, 4 within ML workers, and 4 within S workers. A given working group does not necessarily have an employee from each of the 16 layers.

The hierarchical structure is pyramidal in shape. The typical working group has about 20 workers, divided into about 16 FL workers, 3 ML workers and 1 S worker. If a worker lower down the hierarchy (e.g., a lower FL worker) faces an issue with the assembly process, one of the workers farther up (e.g., an upper level FL worker or an ML worker) makes a quick diagnosis of the issue and offers a solution. In case the issue exceeds the knowledge of the worker in the middle ranks, the problem is then directed further up the hierarchy, potentially all the way up to the Superior.[7] Most new employees start as FL workers with the lowest skill level (i.e., New FL Operator 1). Workers are promoted and change layers as they gain experience, complete training programs, and develop new skills, as shown in Table 1.

Training Provision The firm provides training to its employees; and through such training, employees can be promoted to higher positions in the working group and advance their professional careers. Training is provided in-house through training programs administered by trainers and is primarily focused on management, problem solving and alignment of production line activities with company targets, rather than on technical skills to operate machinery on the production line.[8]

More advanced training programs place a greater emphasis on problem solving and management. Specifically, 78% of the courses required for FL workers to be promoted to ML positions are related to identifying problems, coming up with solutions, and coordinating production to minimize waste and slow-downs. 91% of the courses required to advance from lower ML to upper ML positions are related to these topics; while 68% of the courses required to advance from ML to S positions cover these topics, with the remaining courses covering more managerial topics.[9]

Wages The wage schedule by layer over our study period is shown in Table 1.[10] These data reveal two key findings. First, wages increase monotonically within the hierarchy. This is consistent with workers farther up the hierarchy having received more training and having higher skills. Second, each of the 16 layers has a different wage, with the only exception being New FL 2 and New FL 3 workers, who are paid the same as Regular FL 1 workers. Assuming that wages reflect productivity, this set of facts confirms that training is indeed productive, and that workers in each layer have a different skill level and occupy a distinct position in the knowledge hierarchy of the firm. As promoting workers to a new layer comes

[7] Superiors are supervised by Line Superiors, who are tasked with overseeing multiple working groups.

[8] The trainers are usually employees of the plant who themselves work on the assembly line but also devote time to training others as needed.

[9] These percentages reflect the share of courses with this content, but share of training hours yields similar magnitudes.

[10] The schedule is normalized to the wage level of New FL1 workers for confidentiality reasons.

Table 1: Hierarchical Structure of Working Groups

	Employee layer	Cumulative training	Tasks/Responsibilities	Wage (normalized)
S	Superior 4	17 to 20 training programs	Support and manage overall working group performance	1.75
	Superior 3			1.68
	Superior 2			1.62
	Superior 1			1.59
ML	Mid-line Operator 4	14 to 17 training programs	Identify production problems; closely supervise front-line workers to ensure production standards are met	1.45
	Mid-line Operator 3			1.38
	Mid-line Operator 2			1.31
	Mid-line Operator 1			1.25
FL	Mid Front-line Operator	13 to 14 training programs	Engage in production tasks; support other front-line operators	1.19
	Reg. Front-line Operator 4	12 to 13 training programs	Engage in production tasks	1.14
	Reg. Front-line Operator 3	10 to 12 training programs		1.08
	Reg. Front-line Operator 2	8 to 10 training programs		1.05
	Reg. Front-line Operator 1	5 to 8 training programs		1.03
	New Front-line Operator 3	4 to 5 training programs	Engage in production tasks; learn basic skills	1.03
	New Front-line Operator 2	1 to 4 training programs		1.03
	New Front-line Operator 1	0 to 1 training programs		1.00

Note: Table 1 shows a diagram of the hierarchical structure of working groups. For each employee layer, the diagram shows: (i) the cumulative number of training programs typically received by workers in a given layer; (ii) typical tasks and responsibilities; (iii) the wage earned by workers in the layer, normalized by the wage level of New Front-line Operators 1 (for confidentiality reasons).

with real wage implications for the firm, this also rules out that promotions are simply a way to reward workers through changes in job titles (i.e., promotions are not just labeling).[11]

2.1.3 Problem-Solving in Working Groups and Continuous Improvement System

As described above, the assembly line operates sequentially. Accordingly, each working group's ability to solve problems quickly is critical. Indeed, the pyramidal hierarchical structure of the working groups and the bulk of training content together emphasize the importance of problem-solving in this context.

Given the importance of diagnosing and solving problems quickly, the firm operates a continuous-improvement system in which workers are encouraged to record proposed solutions they find to new problems they encounter. These records are reviewed and scored by upper management, with higher scores meaning a more sophisticated/complex suggestion that is more valuable to the firm. Workers are incentivized to submit valuable suggestions both monetarily (through tournament style incentives) and by way of recognition.[12] The goal is both to elicit maximal effort from workers while leveraging opportunities to proliferate

[11]The wage schedule is decided at the headquarter level and the plant does not adjust the layer-specific wages over the product cycle (wages are negotiated with the workers' union and wage changes are related to local labor market conditions (e.g., inflation), not the product cycle). In fact, over our study period we observe no change in the wage schedule (that is, any change in wages happening during this period is not differential by layer).

[12]Management only incentivizes the submission of suggestions to solve *new* problems: if a worker submits the same suggestion twice, the second one does not get counted towards the incentives.

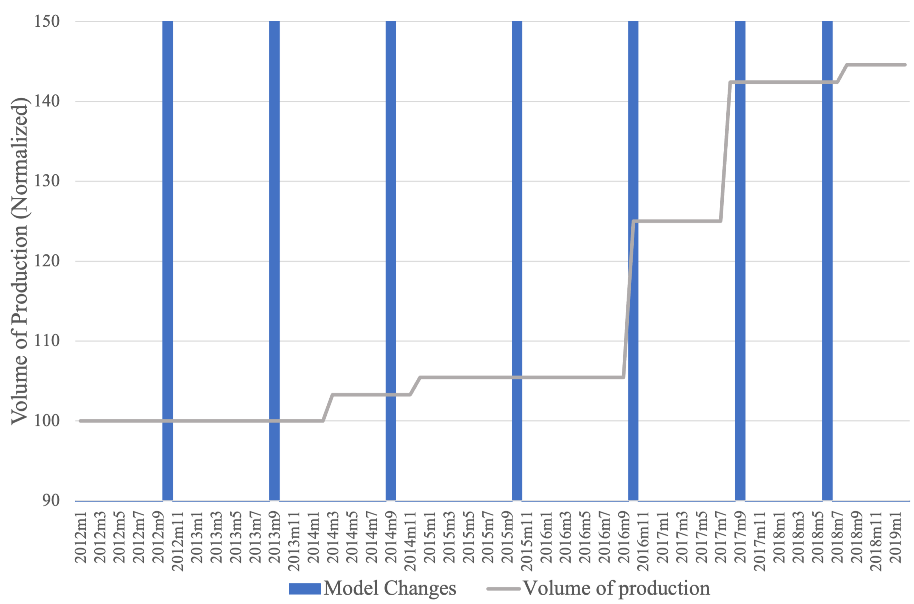
solutions which may apply more broadly across the production process. As we discuss further below, these records prove invaluable in our analysis as they allow us to study impacts on how many problems are being solved in each working group and by whom (i.e., from which knowledge layers) as well as on the value/complexity of the proposed solutions.

2.1.4 Context and Organization of Production: Discussion

Note that: (i) the pyramidal structure of working groups; (ii) the specific knowledge that higher levels obtain via required training for promotion, reflected in higher wages; (iii) the fact that productivity is increased mainly by minimizing stoppages and defects from production problems; (iv) and the firm's clear valuing of problem-solving all underscore the link between production settings of this sort and the intuition at the heart of the knowledge hierarchies models (Caliendo and Rossi-Hansberg, 2012; Garicano, 2000; Garicano and Rossi-Hansberg, 2012). We expand further on this point in the next sub-section, where we use our data to further validate the knowledge hierarchy interpretation of the organization of working groups. We then formalize this intuition in the model later in the paper.

2.1.5 Product Cycles

Figure 1: Timeline of Model and Volume Changes



Note: The figure shows the timeline of Model and Volume changes. Production volume is normalized to the January 2012 level for confidentiality. In 2016, the Model change happened in early October and the Volume change in mid-October.

Over our study period, the company manufactures two auto models, which are regularly updated through new model variants, introduced roughly once a year. Figure 1 shows the timing of the seven "Model changes" that took place over this period, each indicated by a vertical blue line. The figure also shows that Model changes are typically followed by increases in the volume of production: as shown by the gray line, over this same period, the plant gradually expanded its production volume by over 40%, through five discrete jumps in the volume of production, which we call "Volume changes".[13] The figure clearly shows that there are consistent and regular product cycles. We next describe Model and Volume changes, in turn.

Model Changes Most changes to new models tend to be made to the bumpers, lights, grille, wheels, and color options. As shown in Appendix Table A1, on average with each Model change, 13% of the car parts are modified. This share of new parts can be as high as 37% for major Model changes, and we use this heterogeneity below in our analysis. As shown in Figure 1, the number of cars the plant plans to produce does not also change at the same time.[14] Therefore, Model changes result in a significant increase in the number of *new parts* that have to be assembled (while the total number of parts that need to be assembled stays the same), with resulting changes in the types and frequency of problems that arise on the assembly line. For example, the bumper on the new model may appear to be nearly identical to the old one, but if it is of a slightly different shape or weight or made of slightly different material, the angle or force used when installing the bumper may require adjustment.

The introduction of new models is decided at the headquarters level, but the plant management is well aware of when the new product will be introduced. In this multinational company, new products (outside the headquarter country) are first produced in Thailand and then in Argentina, and therefore the Argentina plant knows in advance which new model they will start producing and when. Based on this and on the fact that the timing of Model changes is highly regular (Figure 1), it is clear that the timing of the arrival of new products can be anticipated. However, conversations with management reveal that while there is some planning beforehand, this is quite limited. Headquarters does send in new parts for 20 to 30 trial vehicles in advance for dry runs, but only a few workers are trained in advance on these vehicles and for a few repetitions each. There are two reasons why anticipatory training is limited: (i) there is a clear "training-production trade-off", as taking away workers from the production line is costly (as the factory still needs to produce the previous model and there is

[13] Production volume is normalized to the January 2012 level to protect the confidentiality of the plant. We do not have information on the reasons behind each volume increase. The fact that these are permanent changes suggests that they are not due to exchange rate fluctuations, but rather to increased demand for the plant (possibly reflecting some reallocation from other plants of the multinational).

[14] The one exception is in 2016, when a Model change occurred in early October and was followed shortly after by a Volume change in mid-October. In the analysis, we always control for the distance of each Model and Volume change from all other Model and Volume changes.

never any downtime); (ii) uncertainty also plays a role, as the factory managers do not fully know the key problems until production starts in the precise layout of the factory and with the set of workers on hand with many repetitions and under time pressure.[15] This notion is consistent with evidence from garment manufacturing in India (Adhvaryu et al., 2022a) and fast food in Latin America (Adhvaryu et al., 2022c) showing that acute production bottlenecks and resulting adjustments depend on operation sequence, worker-operation matches, and even worker skills and characteristics. As a consequence, the bulk of training and reorganization happens once the factory switches to the new model and sees firsthand where in the line and for which workers the acute production issues arise in practice.[16]

Volume Changes As shown in Figure 1 and Table A1, Volume changes result in a substantial increase in the number of cars that have to be produced, which jump up on average by 8%, but not in the complexity of what needs to be produced, since the model does not also change at the same time. Volume changes therefore entail a capacity rather than complexity challenge—the firm needs to adjust operations to be able to produce more of the same product per unit time.

2.2 Data and Descriptives

Our data for the Assembly sector comes from four main sources: (i) data on production; (ii) data on employment and training received by employees; (iii) data on the composition of each working group, in terms of the number of employees by knowledge layer; (iv) records of suggestions from the continuous-improvement system. We now describe each in turn.

2.2.1 Production Records

We have daily production data for each shift in the Assembly sector from January 2012 to February 2019. The data include the number of cars produced per day for each model and

[15] Of course managers can form some idea of this through reports of problems in Thailand, but this is imperfect, according to the factory manager.

[16] On a tour of the factory one such example was highlighted: a bumper attachment was being performed with minimal to no defects for the prior model, but when the new model was introduced, the new bumper was heavier and attached slightly differently to account for the weight. After the switch to the new model, a high proportion of bumpers were being attached imperfectly. That this process would be one that would produce defects could not easily be predicted by any of the team members as they had never tried to attach this particular part before. In this instance, an ML in the working group diagnosed that the root cause was the new weight and angle such that having to hold the bumper in place while attaching it was not as easy for this new model as it had been for the prior model. They suggested that a hydraulic lift should be used to hold the bumper at the correct height so that the worker could focus their time and attention on simply attaching it correctly. This fairly straightforward adjustment allowed the operation to be accomplished once again with minimal to no defects.

the number of defects per 100 vehicles produced (DPV). DPV is a key performance indicator that the plant uses to monitor productivity.[17]

Panel A of Table 2 presents summary statistics for these production variables at the daily level over the entire sample period.[18] The plant produces around 410 cars per day. However, there is significant variation— across days due to stoppage and defects, and dramatically from the start of our observation period to the end due to the discontinuous drops in takt time making up the Volume changes discussed above. We standardize defects per vehicle using the mean and standard deviation of the full sample, so that our DPV measure has mean 0 and standard deviation 1.[19]

Our data shows that the incidence of DPV is associated with a sizable decrease in the number of cars produced per day, thus confirming that DPV is an appropriate *inverse* measure of productivity in this context. More precisely, Appendix Table A2 displays the coefficients of a Distributed Lag Model where we regress the number of cars produced per day on an indicator variable for whether the factory experienced DPV 1 standard deviation above its mean on a given day as well as 9 daily lags.[20] Our preferred specification controlling for year and month fixed effects, a linear time trend, and the first lag of the number of cars produced shows that such a spike in DPV reduces output the same day by roughly 47 cars and by a total of about 76 cars (or 18% from the sample mean of cars of 413) over the 10 day period, with no sign of any of that lost production being recovered.[21] This confirms that high DPV on a particular day is associated with an overall permanent (i.e., unrecoverable) reduction in the total number of cars produced. The (negative) correlation we uncover between DPV and number of cars produced is consistent with the findings of Levitt et al. (2013), who document a strong correlation between defects per vehicle and the average time needed to produce a car (and therefore number of cars produced per day).[22]

[17]For the period January 2017 to February 2019, we also have data on DPV for each of the four sub-sectors comprising the Assembly line (so that for this period the DPV data is at the day-shift-subsector level).

[18]As described, production data is originally at the day-shift level. To go to the daily level, we sum the number of cars and average the defects per vehicle across the two shifts within the same day.

[19]To protect the confidentiality of the plant, we were not allowed to disclose mean and standard deviation of the DPV measure. Appendix Figure A1 shows the distribution of (standardized) DPV-day observations, confirming that there is substantial variation in defects per vehicle across days.

[20]To compute this indicator variable, we first residualize the daily DPV variable to net out a linear time trend, month and year fixed effects, and then compute the indicator variable using the residuals from this regression. We do so because the DPV variable is trending down significantly over time (see Figure A2a).

[21]Including the first lag of cars produced in this regression is important to control for other factors correlated over time that might affect the number of cars produced as well as the defect rate (e.g., worker absenteeism). As shown in Table A2, removing this control leads to an increase in the estimated loss in the number of cars.

[22]Specifically, Levitt et al. (2013) show that both defects per vehicle and average hours per car are much higher in the first few weeks after a major change in the model produced by the plant, and then show that these variables both decrease by about 70% after 9-10 weeks from the start of production of the new model.

Table 2: Descriptive Statistics on Productivity, Employment and Training

	Mean	SD
Panel A: Productivity, daily level (2012-19)		
Number of Cars	413.32	110.86
Defects Per Vehicle (DPV)	0.00	1.00
Number of observations	1,590	
Panel B: Employment and training, weekly level (2017-19)		
Number of Employees	2,488.62	460.62
Front-line Employees (FL)	2,022.59	406.519
Mid-line Employees (ML)	380.23	60.79
Superior Employees (S)	117.81	7.933
Number of Working Groups	131.68	13.32
Avg Completed Training Programs	11.56	0.74
Avg Accumulated Training Hours	54.50	7.23
Number of observations	146	
Panel C: Training courses, weekly level (2012-19)		
Number of Courses	35.65	15.51
Number of Teachers	38.95	16.81
Number of observations	370	

Note: Panel A shows descriptive statistics on productivity at the daily level, Panel B and C on employment and training averaged at the weekly level. DPV is standardized using the mean and standard deviation of DPV of the full sample.

2.2.2 Employment and Training Records

We also have employment data in the Assembly sector between January 2017 and February 2019. The data includes information on the date of hiring (and exit) from the plant, whether the worker is present at the factory on any given day, current position (i.e., layer and working group of the employee) and historical data on previous positions, as well as completion date and course type of each training program received by each worker. Panel B of Table 2 presents basic summary statistics on the number of employees in Assembly, where information on employment is averaged at the weekly level.[23]

There are on average 2,488 employees each week: 2,022 front-line employees, 380 mid-line

[23] As 10% of workers are absent each day, in all analysis of employment-related variables we prefer to use weekly averages to smooth out the noise introduced by short-term absenteeism. However, all our results are robust to conducting the analysis at the daily level, as discussed in more detail in Section 4. Note that concerns about noise are less relevant for the DPV measures, as DPV is a key daily performance indicator for the plant and is carefully recorded in the administrative data each day.

employees and 117 superiors, confirming the pyramidal structure described above. These employees are organized in about 130 working groups. The average employee has received 11.6 training programs, totaling about 55 hours of company-sponsored on-the-job training since joining the plant.[24]

Finally, we have data on the number of training courses provided by the plant each week for the entire sample period. Panel C shows that the company engages in continual on-the-job training: in the average week the plant carries out 36 separate training courses, with substantial week-to-week variation in the amount of training provided.

2.2.3 Composition of Working Groups

Our data allows us to track the exact composition of each working group in the Assembly sector for two years, between January 2017 to February 2019. That is, we know the precise layer each employee belongs to (out of the possible 16 layers described in Table 1) and their accumulated training. This enables us to identify promotions across layers within each employee category (e.g., New FL 1 to New FL 2) and across categories (e.g., ML 4 to S 1), as well as moves across working groups over this period. Our data on the production hierarchy thus has even more granularity than the data used in recent papers studying production hierarchies in Europe and the U.S. (Caliendo et al., 2020, 2015; Ewens and Giroud, 2025).[25]

Table 3 shows descriptive statistics on the structure of working groups at the weekly level (so that the unit of observation is a working group in a week).[26] The average size of each working group is about 20, which includes around 16 FL workers, 3 ML workers, and 1 S worker.[27] We note that each of the 16 layers need not be, in fact are rarely, present in a given working group at a given time. On average there are about eight layers of skills/knowledge observed across groups: five within FL, two within ML, and one within the S worker category.

Panel B reports the mean and standard deviation of the week on week change in the number of layers of each working group over the sample period. Working groups are not static: while the number of layers and ratio of FL to ML and S workers both tend to increase on average over the sample period, there is variation in the evolution of the hierarchical structure of working groups. Working groups both add and remove layers and each layer

[24] Two of the five Model changes and two of the seven Volume changes take place in the period between January 2017 to February 2019, when we have data on employment and the composition of working groups.

[25] Caliendo et al. (2015) identify five hierarchical layers in French data, and Caliendo et al. (2020) identify eight layers in Portuguese data. Ewens and Giroud (2025) use resume data to back out corporate hierarchies in U.S. firms, finding ten layers on average. In contrast, we identify 16 layers. Working with only one firm allows us to gather very granular data including on training and problem-solving, although of course comes at the expense of not being able to analyze responses across many firms, as these other papers do.

[26] In the data, a working group is defined as a set of workers with the same Superior each day.

[27] There are some periods in which Mid-line workers act as Superiors. Such cases are limited however to about 11 percent of working group-shift observations, which do not have a Superior.

can both expand and contract over our sample period. Consequently, it is precisely this organizational flexibility that we seek to analyze in the rest of the paper, exploiting our granular data on the production hierarchy of the firm and training provision.[28]

Table 3: Descriptive Statistics on Working Groups and Employee Hierarchies

	Mean	SD
Panel A: Composition, weekly level		
Number of Employees	20.25	8.23
Number of FL Employees	16.47	7.56
Number of ML Employees	2.89	1.27
Number of S Employees	0.89	0.31
FL/(ML+S)	4.56	2.23
Number of Layers	7.80	1.99
Number of FL Layers	5.02	1.65
Number of ML Layers	1.89	0.76
Number of S Layers	0.89	0.31
Number of observations	13,280	
Panel B: Change in layers, weekly level		
Change in Number of Layers	0.015	0.349
Change in Number of FL Layers	0.011	0.303
Change in Number of ML Layers	0.003	0.141
Changes in Number of S Layers	0.002	0.035
Changes in FL/(ML+S)	0.005	0.050
Number of Observations	13,060	

Note: The information presented in Panel A is averaged at the working group-shift-week level. The changes in layers in Panel B are computed as the average change in the number of layers in the working group at the weekly level.

2.2.4 Problem-Solving Records from Continuous Improvement System

As discussed above, the firm has a continuous improvement system in which workers log suggestions on how to solve new problems they encounter. These suggestions are then scored

[28] In the Appendix we show the evolution of average size of working groups (Figure A2b), average number of layers in each working group (Figure A2c) and average number of working groups (Figure A2d) over our sample period. The size of working groups grew by more than 30% during the study period. Correspondingly, the number of layers in each working group increased by 15%; and the ratio of FL workers to ML and S workers grew from less than 4 to more than 5 on average. The number of working groups also grew by 30%. Note that employee absenteeism and turnover contribute to creating some of the high frequency variation in these Figures. Over the same period (from January 2017 to February 2019), annual planned production grew by 16% (as shown in Figure 1) and training investments also grew considerably (Figures A3 and A4).

by management on up to nine possible dimensions, depending on the specific topics the suggestion relates to.[29] Higher scores mean a more complex and valuable suggestion. We have data on all suggestions logged by workers for 2018 and 2019, together with their score on each dimensions. This allows us to compute the weekly number of suggestions logged by each worker, along with the average score of each suggestion, where the score is defined as the sum of all scores it received across the different dimensions.

Table 4 shows descriptive statistics from these data at the working group level. FLs log the highest number of suggestions, followed by ML workers and finally S workers, who make the fewest suggestions (column 1). Accounting for the number of workers in each layer, however, ML workers solve more problems per worker than FL, but S workers still solve the fewest (column 2). Finally, the suggestions logged by S workers receive systematically the highest scores, followed by those of ML workers and those of FL workers (and these differences in scores are all significant).[30] These records are ideal for an analysis of how the extent and distribution of problem-solving responsibilities, as well as the complexity/value of the suggestions made, evolves with training and group compositional changes.

Table 4: Descriptive Statistics on Employee Suggestions in Working Groups

	(1) Weekly Suggestions	(2) Weekly Suggestions per Employee	(3) Average Score of Suggestions
S Employees	0.0472	0.0550	10.508
ML Employees	0.9627	0.3243	8.776
FL Employees	4.0465	0.2507	7.107

Notes: All statistics are at the working group level. The outcome in column 1 is computed as the average weekly number of suggestions logged by all employees in a given layer and working group. The outcome in column 2 is computed as the average number of weekly suggestions logged by employees in a given layer and working group, divided by the total number of employees in that layer and working group. The outcome in column 3 is computed as the average score of logged suggestions by the employees in a given layer and working group. We then average across working groups to create the statistics in the table.

2.2.5 Validating the Knowledge Hierarchy Interpretation of Working Groups

The results in Table 4 farther validate the knowledge hierarchy interpretation of our context in two ways: first, workers farther up the hierarchy solve more complex problems, consistent

[29] The nine dimensions are: cost, productivity, security, quality, environment, working conditions, initiative, creativity, effort.

[30] In Appendix Table A4, we show that, reassuringly, higher layers tend to get higher scores across all dimensions. Table A3, shows the breakdown of the share of suggestions by layer, suggestions per worker, and average score of suggestions by the 16 possible layers in our data. This broadly confirms the results in Table 4. Regarding the score of suggestions, Table A3 confirms that even within each employee category, higher layers tend to get higher scores. The pattern is less clear across the S layers, but we note that: (i) S workers submit the fewest suggestions (Table 4, column 1), so the sample size is the smallest for that group; (ii) in line with this, the differences between the S2 and S1 layer and between the S4 and S3 layer are not significant.

with them having higher knowledge; second, workers at the top of the hierarchy make fewer suggestions, consistent with them being able to solve a higher share of the problems they face relative to workers at lower ranks, so that they encounter *new* problems that have not been solved before less often. Together with the evidence discussed above that workers at higher layers of the hierarchy have accumulated more training and earn higher wages, this confirms the appropriateness of interpreting the hierarchal structure of working groups through the lenses of knowledge hierarchy models (Caliendo and Rossi-Hansberg, 2012; Garicano, 2000; Garicano and Rossi-Hansberg, 2012).

Finally, in Appendix Table A5, we show that working groups are stable production units, which further validates focusing on them as our key unit of analysis. In particular, the table shows that: (i) moves between working groups are not frequent, with only 0.4 workers moving from each working group per week; (ii) working groups are very persistent, with the average working group observed in the data for 78 weeks and very few working groups created/dissolved each week; and (iii) it is rare that an entire employee category is missing altogether from a working group: it never happens that working groups do not have at least one FL worker, and working groups do not have at least one ML and one S worker only in 4% and 11% of cases, respectively.[31]

3 Empirical Strategy

Our broad objective is to empirically document the *organizational cycle* of the firm. We do so by focusing on how the two key organizational levers of training provision and hierarchical structure evolve as a response to Model and Volume changes. To shed light on this, we use event studies and regression discontinuity-in-time methods, which leverage the high-frequency nature of our data to document how these outcomes change around the exact time of Model and Volume changes. The event studies allow us to trace out the dynamics of the effects. The discontinuity-in-time specifications allow us to provide a useful summary point estimate of the effects averaged over the time period under consideration.

Throughout the analysis, we do not treat the timing of Model and Volume changes as exogenous; rather, we think of both as part of a cycle planned in advance by the firm. Indeed, in the model in Section 5 the timing of the product cycle is endogenous and fully anticipated by the firm. However, as we will show in Section 4, the absence of anticipatory effects for outcomes related to training and hierarchical structure enables us to recover how the firm

[31] Again in line with working groups being stable, it is very uncommon for working groups to add/subtract more than one layer per week (as shown by the standard deviations in Panel B of Table 3). Also, in Section 4 we will show that Model changes do not lead to significant change in the probability of new working groups being formed or dissolved, nor in a change in the probability that workers move working group.

uses changes in these two organizational levers throughout the product cycle.

3.1 Regression Specifications

Event Studies For the event studies, we estimate the following specification:

$$Y_{it} = \theta_i + \gamma_m + \gamma_y + \sum_{k \geq -b, k \leq b, k \neq -1} \beta_k D_t^k + \delta X_t + \epsilon_{it} \tag{1}$$

where Y_{it} is the outcome for unit of observation i at time t. For the analysis of DPV, which is measured at the shift-day level, Y_{it} is the DPV of a given shift in a given day. For working-group related outcomes (e.g., average training programs completed, hierarchical structure etc.), Y_{it} is the outcome of a given working group in a given week. D_t^k is an indicator of the distance with respect to the (Model or Volume) change, where k indicates the distance to the event at time t. In our dataset we know the exact date when the event takes place. b represents the bandwidth (window of time) considered in the event study. Following Calonico et al. (2014) we find that the optimal bandwidth for Model changes when using the DPV measure as the outcome is 40 days (8 five-day working weeks) before and after the event. We use this bandwidth throughout for both Model and Volume changes.[32] Our specification controls for unit of observation (i.e., shift or working group) fixed effects (θ_i), and year and month fixed effects (γ_m and γ_y) to net out any group-specific effects and to control flexibly for time effects and seasonality. Additionally, we control for linear distance to other events (e.g., when looking at a given Model change, we control for distance to all other Model and Volume changes) and a linear time trend (X_t). Finally, standard errors are clustered by distance to the event and unit of observation (i.e., by distance to the event-shift or distance to the event-working group, depending on the outcome).[33]

Our parameters of interest are β_k. For $k \geq 0$ (i.e., for time periods corresponding to the event or later) we interpret β_k as the effect of the Model or Volume change on the outcome at time k. All β_k parameters are relative to the outcome in the time period immediately before the event $k = -1$, which is the excluded category in these regressions.

Regression Discontinuity-in-Time For the regression discontinuity-in-time (RDiT) analysis, we estimate the following specification:

$$Y_{itw} = \theta_i + \gamma_m + \gamma_y + \beta_1 I(0 \leq dis_w \leq 3)_w + \beta_2 I(4 \leq dis_w \leq 7)_w + f(dis_t) + \delta X_t + \epsilon_{itw} \tag{2}$$

[32] We use the same bandwidth for Model and Volume changes for ease of comparison of the results across the two types of events. The optimal bandwidth is similar for Volume changes (25 days or 5 weeks).

[33] Results are very similar if we include a quadratic time trend instead of a linear trend.

where Y_{itw} is the outcome of unit of observation i measured at time t and week w. For DPV, Y_{itw} are defects per vehicle in shift i, day t and week w. For working group related outcomes, Y_{itw} is simply the outcome for working group i in week w (as we always work with weekly level data for outcomes related to working groups). $I(0 \leq dis_w \leq 3)_w$ is an indicator for being four weeks (we define a week as five working days) after the event occurring at time t. $I(4 \leq dis_w \leq 7)_w$ is an indicator for being more than four weeks after the event at time t. Recall that the optimal bandwidth for the event studies was 40 days, which guides our definition of the indicator functions. We control for unit of observation fixed effects (θ_i), month and year fixed effects (γ_m and γ_y). Additionally, we control for a function of the distance to the event ($f(dis_t)$), where we choose the linear function, and for the linear distance to all other Model and Volume changes (X_t).[34] Standard errors are clustered by the level of distance to the event and unit of observation (so by distance to the event-shift for DPV and by distance to the event-working group for other working group related outcomes).

Our parameters of interest are β_1 and β_2, which capture the average effect of each event in the two windows of time. This allows us to speak to the dynamics of the Model and Volume changes, while still providing summary point estimates for the average effect of each event.

3.2 Identification

We note four points related to identification in the event studies and RDiT specifications.

First, identification relies on the assumption that, after controlling for the linear time trend and for month and year fixed effects, any discrete changes in the outcomes observed just after the event are not due to underlying (residual) trends in the outcome variables. The high-frequency nature of our data and our ability to focus on a narrow time window around the events reassures us of the validity of this assumption.

Second, the focus on a narrow time window reduces the possible concern that changes in the outcome variables after the Model or Volume changes are due to other events affecting trends in productivity or organizational structure (and we are controlling for distance to other Model and Volume changes).

Third, of course, as discussed above, Model and Volume changes are anticipated, and it is possible that the firm starts responding in advance. Whether the firm indeed makes adjustments in advance is an empirical question we can investigate using our event studies. Lack of significance of the estimated β_k in the time periods before the event (i.e., $k < -1$) would provide evidence in support of the limited role of any anticipation effects or other trends in potentially biasing our results. Consistent with limited planning in advance, we will

[34] $f(dis_t)$ and X_t are measured in days for DPV and in weeks for all other working group related outcomes.

show no evidence that our key outcomes start responding before Model and Volume changes.

Finally, while we focus on training provision and hierarchical structure, there are likely other production strategy changes happening at the same time that we cannot observe in the data (such as supply managers changing how they coordinate with suppliers). What we are able to document are the incremental organizational adjustments on the margins of training investments and hierarchical structure made by the plant over and above these other unobserved production strategy changes, and how these map to productivity effects. Given that our event studies do not show significant evidence of planning in advance for our key outcomes including end-line production outcomes like productivity (here DPV), we are reassured that any such unobserved strategy shift is unlikely to bias our results.

4 Empirical Results

In this section, we document the organizational responses to Model and Volume changes. For both, we first validate the nature of the event by studying impacts on planned number of cars to be produced and corresponding parts and share of new parts, when relevant. Then, we report impacts on defects per vehicle, to confirm that both types of changes have real implications for productivity, showing heterogeneity by share of new parts and number of new hires, respectively, to link defects to changes in complexity and capacity, respectively. Next, we present impacts on training provision (and therefore knowledge) and its distribution within working groups, as well as the hierarchical shape of working groups, to study how the organizational responses the plant puts in place to bring down defects per vehicle differ across the two types of changes. Finally, we study problem-solving activities across knowledge layers in working groups to validate the interpretation of impacts as being driven by both the level and distribution of problem-solving knowledge and responsibility. The full sets of Model changes and Volume changes results are summarized in Appendix Figures A5 and A12.

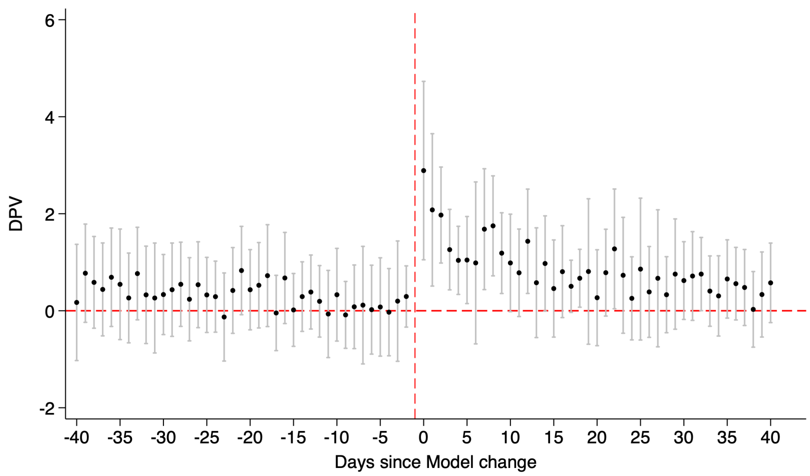
In Section 5, we will then develop a dynamic model of the firm to (i) interpret and link together these results as part of the organizational cycle of the firm; and (ii) quantify the role of each of the two organizational levers of training provision and hierarchical structure in enabling adaptation to changing production conditions over the product cycle.

4.1 Responses to Model Changes

4.1.1 Validating that Model Changes Increase Production Complexity

Appendix Table A6 reports the results of regressions like equation 2 but where the outcome variables are total number of cars planned for assembly per day, corresponding total number of parts per day, and the share of parts that are new relative to the last model variant produced just before the Model change.[35] The table shows that Model changes do not result in a significant change in the number of cars produced nor in the total number of parts used. Instead, Model changes result in a persistent increase of at least 14% in the share of parts that are new over the two months after the change.[36] This confirms that Model changes result in a substantial amount of new parts that need to be used to assemble the new model. This plausibly leads to an increase in complexity of production, at least during the period when workers are learning how to assemble the new parts.

Figure 2: Event Study of Model Changes on Defects per Vehicle



Note: Figure 2 shows the effect of Model changes on DPV in a time window running from 40 days before the event to 40 days after the event. DPV is computed as number of defects per 100 vehicles, and is standardized using the mean and standard deviation of the full sample. We control for month, year, and shift fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes and a linear time trend. Standard errors are clustered by distance to the event and shift. 95% confidence intervals are reported. Number of observations: 2 shifts x 81 days x 7 events.

4.1.2 Impacts on Productivity

Figure 2 reports the results of an event study specification following equation 1 with (standardized) DPV as the outcome. The figure shows that there is a discrete jump in defects per vehicle right after the Model change. Defects increase on average by about 2 SD right after the Model change, and start to gradually come back down the next day, reverting back to

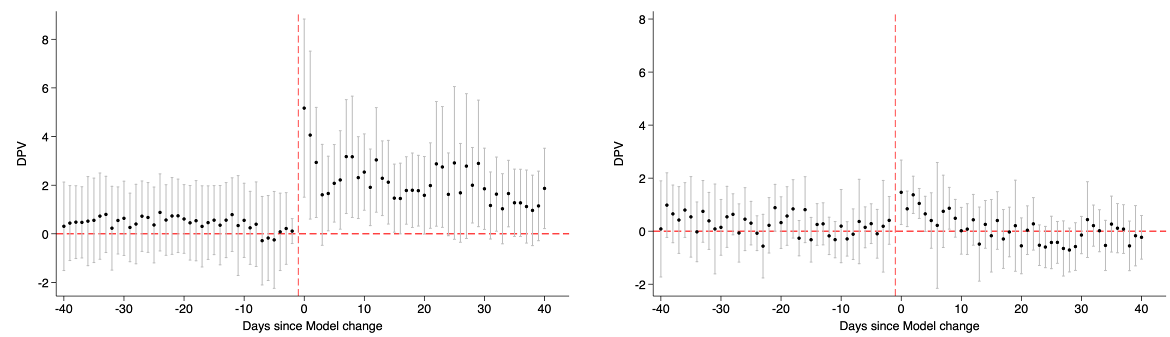
[35]The level of observation in Table A6 is a day and the sample size is 567 because there are 7 Model changes and for each event we choose a bandwidth of 40 days around the event, so 81 days in total for each event.

[36]Table A6 shows that the magnitude of the effect on the share of new parts fluctuates in the two months following the Model change. This is because of variation in product mix. For each model, the factory makes several different trim levels. Higher level trims usually see a higher share of new parts to keep up with evolving tastes of customers willing to pay for luxury features. The share of production devoted to each trim can vary based on orders.

the pre-Model change level after about three-four weeks.[37] The figure also shows that DPV does not exhibit any significant pre-trend up until the *exact day* of the Model change.

Figure 3 presents the same event studies, but with the sample split between Model changes involving high (above median) and low share of new parts. The initial spike in DPV is much larger (roughly 5 SDs in magnitude) and the subsequent decay slower for changes involving the larger shares of new parts.[38] This heterogeneity analysis lends support to the interpretation that the spike in defects following Model changes is driven by the novelty of many parts and corresponding processes.

Figure 3: Event Study of Model Changes on Defects per Vehicle by Share of New Parts



(a) High Share of New Parts

(b) Low Share of New Parts

Note: Figure 3 shows the effect of Model changes on DPV in a time window running from 40 days before the event to 40 days after the event, split between changes with high share of new parts (above the median) and changes with low share of new parts (below the median). DPV is computed as number of defects per 100 vehicles, and is standardized using the mean and standard deviation of the full sample. We control for month, year, and shift fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes and a linear time trend. Standard errors are clustered by distance to the event and shift. 95% confidence intervals are reported. Number of observations for Panel (a): 2 shifts x 81 days x 3 events. Number of observations for Panel (b): 2 shifts x 81 days x 4 events.

The results in Figure 2 and 3 show that there are real negative productivity implications of the introduction of new model variants as a result of product innovation and upgrading. At the same time, the plant is able to bring these back down fairly quickly on average. Despite the temporary nature of the rise in DPV, the costs to the firm are substantial and unrecoverable: as discussed in the previous section, days with DPV more than 1 SD above

[37] Appendix Table A7 confirms the results in Figure 2 by showing that Model changes lead to an increase in DPV of about 0.75 SD in the first four weeks after the shock, with the effect coming down to zero after that (column 1). This pattern holds, although point estimates are larger, when we restrict attention to the period 2017 to 2019 coinciding with the working group composition database (column 2).

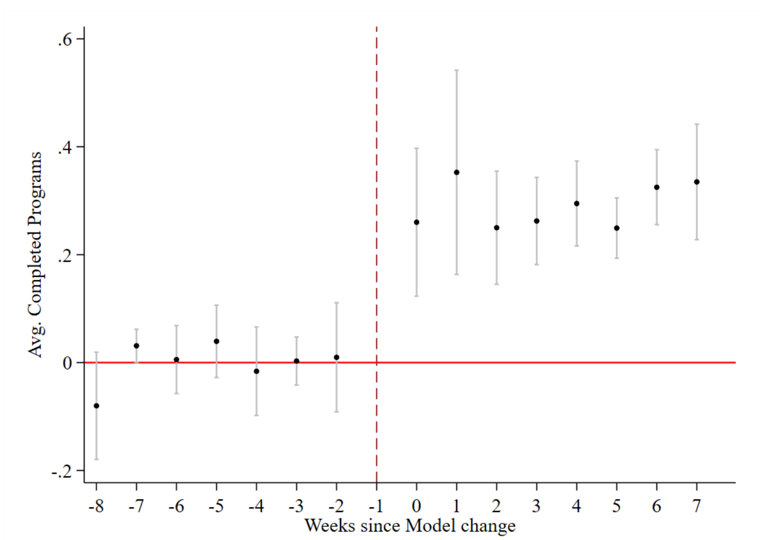
[38] Panel (a) of Figure 3 shows that DPV remains significantly elevated for more than 6 weeks after the Model change. Estimating a version of the event study with a bandwidth of 80 days reveals that the spike in DPV does eventually disappear by about 10 weeks after the Model change. These results are in line with evidence from Levitt et al. (2013), who find that it takes about 8 weeks for a similar measure of defects per vehicle to come down by about two thirds after a major change in car model, and about 10-12 weeks for most of the reduction in DPV to be realized and for DPV to stabilize. Levitt et al. (2013) also show that the speed at which DPV declines after a model change varies depending on the specific model that is introduced.

the mean are associated with a decrease in number of cars of 18% (see Appendix Table A2). Notice that in Figure 2 DPV stays elevated by 1 SD or more for more than a week following a Model change on average, thus confirming that Model changes lead to a significant reduction in productivity while the assembly line learns how to deal with the new problems that arise.

4.1.3 Training Responses

We first investigate how the cumulative number of completed courses and hours of training for the average worker in the working group changes after a Model change. Figure 4 reports the results of an event study specification following equation 1 with the average number of cumulative completed courses per worker as the dependent variable. The figure shows that in response to the Model changes, the firm increases training provision in each working group.[39] The point estimates in Appendix Table A9 show that the cumulative completed courses go up by 3% and the total hours of training by 6%, such that a working group will accrue roughly seven more completed courses or 60 more hours of training across its 20 workers.

Figure 4: Event Study of Model Changes on Cumulative Training in Working Groups



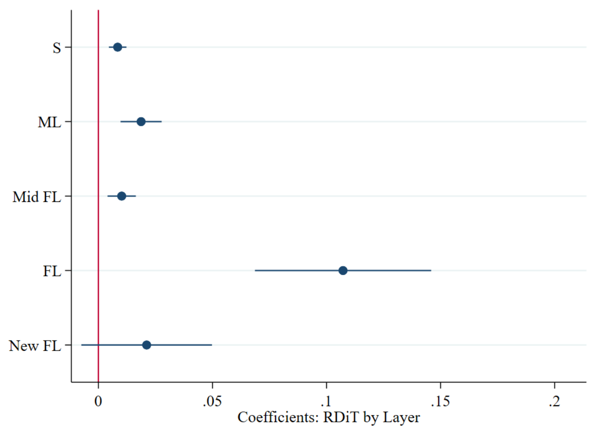
Note: Figure 4 shows the effect of Model changes on the average completed training programs within working groups in a time window running from 8 weeks before to 8 weeks after the event (where the week of the Model change is labeled as week 0 on the x-axis). We control for month, year, and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes and a linear time trend. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

These results are in line with the firm increasing the knowledge of its workers (via training) to deal with the more frequent arrival of more complex problems. Training is therefore a key organizational margin used by the firm to respond to the product cycle.

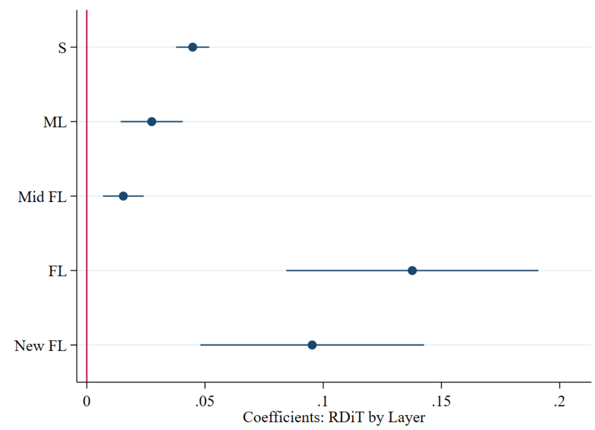
[39] As the outcome is the *cumulative* training received, it is expected that impacts are persistent.

Linking back to the discussion on identification, note that Figure 4 shows no significant evidence that the firm starts responding in advance via the training margin, thus validating the identification assumptions. This is consistent with the intuition shared by factory management that anticipatory responses are limited in practice.

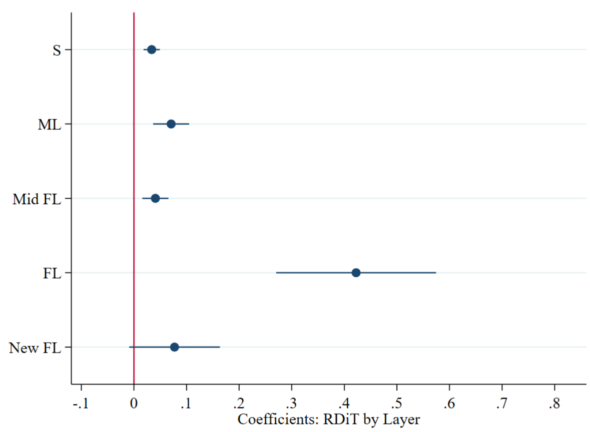
Figure 5: Impact of Model Changes on Incremental Training Provision



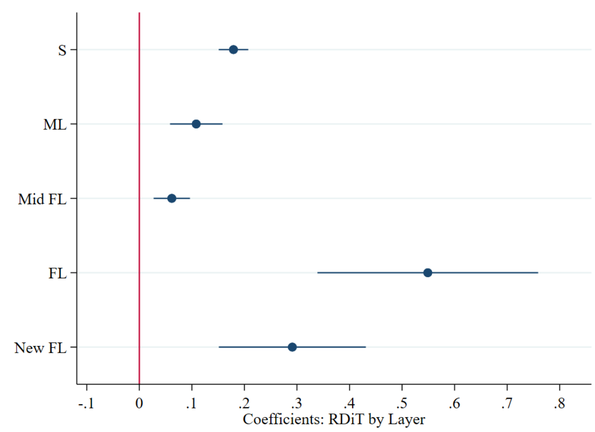
(a) Num. Courses Taken, 0-3 weeks



(b) Num. Courses Taken, 4-7 weeks



(c) Num. Employees Trained, 0-3 weeks



(d) Num. Employees Trained, 4-7 weeks

Figures 5a and 5b show the effect of Model changes on the average number of courses taken by workers by layer during 0-3 weeks and 4-7 weeks post Model change, respectively. Figures 5c and 5d show results from similar regressions but where the outcome is the number of employees trained by layer. For details on the definition of the layers see Table 1. Each coefficient is estimated from a separate regression. We control for month, year, and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

Figure 5 shows that the incremental training investments—in terms of both number of courses taken by workers and number of employees trained per week—are concentrated primarily among FL workers (in fact upper FL on the margin of promotion to ML as seen in the more disaggregated results in Appendix Figure A6, panels (a) and (c)), with magnitudes consistent with roughly 4-6 FL workers each completing the additional course they need to reach ML level skill and the remaining 1-3 courses spread across the other workers in the group. Figure A6 in the Appendix also presents this same result, but restricting attention to just courses covering problem-solving and communication content (panels (b) and (d)). The near identical pattern and similar order of magnitude in these results confirm that investments in these communication and problem-solving skills are primarily driving the training results.

4.1.4 Impacts on Hierarchical Structure of Working Groups

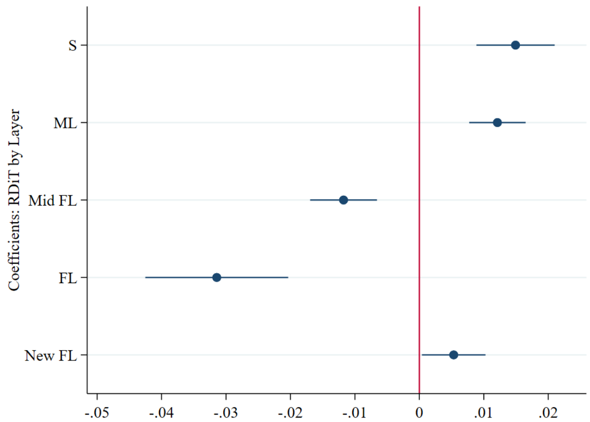
Next, we study responses in the hierarchical structure of working groups. First we verify that the number of employees in the Assembly sector, number of working groups, and size of working groups does not change as a new model is introduced (see Appendix Tables A10 and A11). This result is sensible in that the total number of cars that need to be produced has not changed.[40]

To check if Model changes lead to a reallocation of workers across hierarchical layers, in Figure 6 we study the impact of the Model changes on the hierarchical shape of the working groups. We do so by running a series of regressions like equation 2 with as dependent variables shares of workers in the working group by layer, in the four weeks (Panel (a)) and eight weeks (Panel (b)) after the Model change. This figure shows a recomposition of working groups with a reduction in upper-FL layers in exchange for an increase in the share of ML and S layer workers. Appendix Figure A9 shows a more disaggregated version of Figure 6, confirming that the recomposition is mainly driven by upper FL workers being upskilled and promoted to higher level positions. These results are consistent with Figures 5 and A6, as promotions in the plant are associated with receiving additional training. Comparing Panels (a) and (b) in Figure 6 and Figure A9, we note that the ricomposition is more pronounced in the four weeks after a Model change, with the shape of the hierarchy beginning to revert back after that.

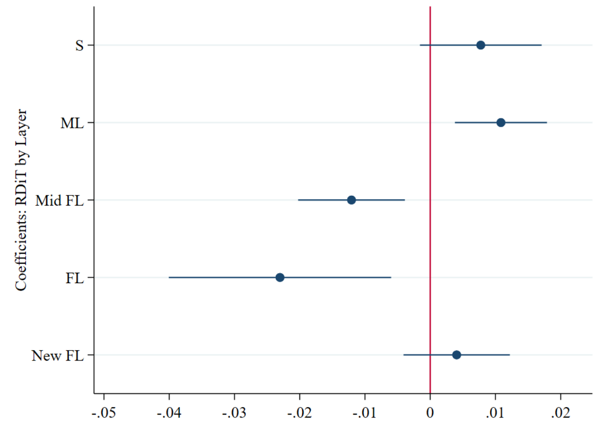
Figures 6 and A9 show that working groups are becoming flatter with less distinct knowledge layers, and more rectangular in shape, especially in the four weeks after a Model change. These results are confirmed in the corresponding regression discontinuity in time estimates reported in Table 5: column 1 shows a negative and significant effect of about 0.7 layers in the first four weeks after the Model change, corresponding to a 10% decrease from the mean. The negative effect on layers means that the plant is reducing the number of

[40]More specifically, column 1 of Table A10 shows no impact on employment, and Table A11 shows no impact on the number of working groups, group size, probability of a new group being formed, and share of workers moving between groups. The fact that these working-group related outcomes do not change further reassures us about the stability of working groups and validates focusing on them as our key unit of analysis for the study of knowledge hierarchies.

Figure 6: Impact of Model Changes on Working Group Structure and Knowledge Hierarchies



(a) Share of Workers by Layer, 0-3 weeks



(b) Share of Workers by Layer, 4-7 weeks

Note: Figures 6a and 6b show the effect of Model changes on the share of workers in the working group by layer at 0-3 weeks and 4-7 weeks post Model change, respectively. For more details on the definition of the layers see Table 1. Each coefficient is estimated from a separate regression. We control for month, year, and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are presented in the figure. Number of observations: 220 working groups x 16 weeks x 2 events.

separate knowledge levels within working groups after a Model change. Column 2 shows that the ratio of FL to ML workers also decreases by 10%, while Column 3 shows that the ratio of FL to ML and S decreases by 12%. That is, the span of control of ML and S workers (or number of FL workers from which each fields problems) is going down. Consistent with the results in Figure 6, Table 5 shows that these effects are concentrated in the four weeks after a Model change, and the hierarchy largely reverts back by eight weeks after the change.

Taken together, Figures 6 and Table 5 show evidence of a substantial temporary change in hierarchical structure, with the dynamics showing a remarkable similarity to Figure 2: as DPV comes back down, the hierarchy starts to revert back to the pre-Model change shape. We confirm this further in Appendix Figure A8a, which reports event studies on the number of layers in working groups, and shows that Model changes result in a sudden decrease by about one layer, which lasts about three weeks, after which the working groups back-fill the layer that was removed.[41] Changes in hierarchical structure are thus the second key organizational lever used by the plant to respond to Model changes.

[41] In practice, back-filling happens through some workers farther up the hierarchy leaving and being replaced by new workers at the bottom of the hierarchy. Indeed, Appendix Table A10 shows that the coefficients on both hires and separations 4-7 weeks after the Model change are large in magnitude relative to the mean (although not significant). In line with this, Panels (b) and (d) of Figure 5 show that in the 4-7 weeks after a Model change, the additional training is concentrated among Front-Line workers and New Front-Line workers (the bottom layer), which is again indicative of some back-filling starting to take place.

Table 5: Impact of Model Changes on Working Group Structure

	(1) Num. Layers	(2) FL/ML	(3) FL/(ML+S)
0 to 3 weeks	-0.666***	-0.525***	-0.430***
	(0.124)	(0.120)	(0.0867)
4 to 7 weeks	-0.285	-0.244	-0.250*
	(0.190)	(0.190)	(0.136)
Observations	7,040	7,040	7,040
Obs. Level	Group-Week	Group-Week	Group-Week
Mean	6.373	5.026	3.755

Note: Standard errors clustered by distance to event and working group. Number of observations: 220 working groups x 16 weeks x 2 events. Number of Layers is defined as the number of separate positions present in a working group. FL/ML is the ratio between Num. of Front-Line workers and Mid-Line workers. FL/(ML+S) is the ratio between Num. of Front-Line workers and Mid-Line workers plus Superiors. We use as controls month, year and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

4.1.5 Impacts on Problem-Solving in Working Groups

Finally, given which workers are being trained and the content covered in those trainings, Table 6 explicitly looks for effects on the frequency and complexity of problem-solving-related activities. First, Panels A and B study changes in the number of suggested solutions to new problems logged by workers at different levels of the hierarchy. Indeed all members (at all layers) of the working group are logging more suggestions. The impacts are large, with the number of logged suggestions approximately doubling after the Model change. This is consistent with both the frequency of new problems having increased after the Model change, and with workers being better able to solve problems because of the training received. Consistent with the training being concentrated among upper FLs with resulting promotions to ML ranks (Figures 5 and A6), we see that per employee, the impacts on reporting of suggestions are mainly driven by FLs and MLs; though in line with more complex problems arising after a Model change, each S worker is also logging more suggestions than before.

Second, Panel C shows that Model changes lead to a significant increase in the score of the suggestions at all levels of the hierarchy. This result is notable in two ways. First, this is direct evidence that the new training is productive, as workers are able to make more valuable/complex suggestions and the impacts are sustained over time. Again consistent with the upskilling of FL workers to upper FLs and MLs, the FL and ML layers are the ones that see the largest increase in scores. Second, the positive impacts on suggestion scores confirm that the introduction of new models leads to problems that are more difficult, not just different. This is in line with the new upgraded products entailing more complex operations and therefore being more sophisticated.[42]

Finally, notice that these results do not rule out knowledge depreciation over time: it is possible that the old knowledge may become (partly) obsolete when a new model is introduced, thus prompting the need for more training to solve the increasingly complex tasks brought about by Model changes. The key insight from our analysis is that the firm uses training provision as an important organizational lever to equip workers with the necessary knowledge to deal with the new complex problems that arise when new models are introduced. The extent to which this knowledge is retained as new generations of products are introduced is something that goes beyond the scope of this paper, and our data are not well suited to study these long-run dynamics.[43]

4.1.6 Changes in Training and Hierarchy both Have Productivity Implications

The results so far are notable in that they show how the firm combines both (i) in-house training and (ii) changes in hierarchical structure to adapt to the introduction of new upgraded (and more complex) products. We now validate that these two types of organizational responses both have real productivity implications for the firm. This analysis further confirms that training is productive and that changes in the hierarchy are not just a mechanical result of the implied training with no independent productivity effect. To do so, we exploit variation in the amount of training and hierarchical restructuring experienced by different working groups when new models are introduced, and study whether those working groups that receive more training or undergo a greater change in hierarchical structure indeed experience a smaller increase in DPV, our key (inverse) measure of productivity.

Productivity Implications of Training We start by focusing on how the magnitude of the DPV increase after Model changes documented in Figure 2 depends on the extent of training received by working groups.

We restrict attention to the period 2017-19, when we have information on both DPV and training/working group composition. DPV information is not available at the working group level; however, for the period 2017-2019, it is available for each of the four sectors comprising

[42] The positive impacts on the scores of suggestions rules out that the impacts on the number of suggestions are driven by management encouraging workers to report solutions soon after the Model change so that they learn the new problems: if that was the case, we would not expect an immediate increase in the score of the suggestions made. Therefore, the evidence is more consistent with suggestions increasing as a result of training than because of pressures from management.

[43] Appendix Figure A2a shows that DPV trends down over time and Table A12 shows that Model changes happening later in our sample period have a smaller impact on DPV. These results are both consistent with at least some of the knowledge generated through training being retained over time, so that there is a secular increase in productivity. However, they may also be consistent with other secular trends, such as a general improvement in managerial capabilities of the plant, so we avoid interpreting these results too strongly.

Table 6: Impact of Model Changes on Problem-Solving Activity by Layer

	(1) All	(2) FL	(3) ML	(4) S
	Panel A: Number of suggestions			
0 to 3 weeks	2.264***	1.918***	0.317***	0.0288*
	(0.331)	(0.284)	(0.0692)	(0.0169)
4 to 7 weeks	1.709***	1.437***	0.214	0.0577*
	(0.522)	(0.449)	(0.143)	(0.0310)
Mean	2.200	1.772	0.407	0.021
	Panel B: Number of suggestions per employee			
0 to 3 weeks	0.194***	0.201***	0.196***	0.0648*
	(0.0236)	(0.0261)	(0.0379)	(0.0364)
4 to 7 weeks	0.141***	0.133**	0.0831	0.126*
	(0.0488)	(0.0523)	(0.115)	(0.0646)
Mean	0.188	0.187	0.243	0.042
	Panel C: Score of suggestions			
0 to 3 weeks	1.230***	1.333***	1.067***	0.171*
	(0.217)	(0.205)	(0.218)	(0.0992)
4 to 7 weeks	0.901**	0.798**	1.450***	0.245
	(0.368)	(0.337)	(0.455)	(0.187)
Mean	7.208	6.771	8.358	10.948
Observations	3,520	3,520	3,520	3,520
Obs. Level	Group-Week	Group-Week	Group-Week	Group-Week

Note: Standard errors clustered by distance to event and working group. Number of observations: 220 shifts x 16 weeks x 1 events. We control for month, year, and shift fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. * p<0.1, ** p<0.05, *** p<0.01

the Assembly line, so that DPV observations are at the day-shift-sector level during this period. To conduct this heterogeneity analysis, we thus divide sector-shifts into those that received above vs. below median training (as measured by the number of courses taken by workers) in the three weeks following a Model change (as that is the relevant time period where we detect a spike in DPV—see Figure 2). We then estimate a heterogeneous version of our baseline discontinuity in time specification (equation 2), where we allow impacts on DPV to vary by this measure of intensity of training received. The results are in Table 7. For comparison, Column 1 reports the impact on DPV for the average sector-shift over this period (and reassuringly we note that the effects are similar to those in Figure 2).[44] Column

[44] Appendix Table A7 also shows that, reassuringly, the estimates are very similar when conducting the analysis of DPV impacts at the shift-day level or shift-day-sector level in the 2017-19 period.

2 then reports the estimates of this heterogeneous specification, and shows that the increase in DPV in weeks 0-3 following a Model change is concentrated in groups that received *below median* training: the coefficient on the interaction is negative and similar in magnitude to the main effect in weeks 0-3, indicating no statistically significant increase in DPV for groups with above median training (the estimated coefficient for groups with above median training in weeks 0-3 is 0.479 with standard error 0.323, as shown at the bottom of the table). Column 2 also shows some evidence that in groups receiving less than median training, the negative impacts on DPV last longer than four weeks, which again is not the case in groups that received above median training, as shown by the negative and significant interaction with the 4-7 weeks dummy.

We can repeat the analysis in column 2 but using the score of the suggestions logged by workers as a measure of "training". That is, we split sector-shifts into those with above and below median scores of logged suggestions in the first three weeks after a Model change. The results are in column 3, and tell a very similar story to column 2: the increase in DPV is concentrated in groups with below-average quality of suggestions. While not causal, these results again support the conclusion that the training provided not only increases the knowledge of workers (as measured by the increase in the scores of suggestions in Table 6) but also helps to attenuate the negative productivity effects of Model changes.

Productivity Implications of Hierarchical Structure To show that changes in the hierarchy and span of control also have productivity implications, over and above those implied by the additional training, we show heterogeneity of the DPV increase following Model changes by the extent of changes in the number of layers of working groups. As we have documented that Model changes lead to a decrease in layers (Table 5 and Figure A8a), we separate sector-shifts between those that experienced a decrease in layers above the median vs below the median in the three weeks following Model changes.

We then run our same discontinuity in time specifications of the impact of Model changes on DPV following equation 2 but allowing for this margin of heterogeneity. The results are in column 4 of Table 7, and they show that the increase in DPV is significantly larger in those groups that experienced a below-median change in layers. This confirms that changes in the hierarchical structure have real productivity consequences for the firm: reducing the span of control of managers allows the firm to limit the productivity losses from switching to new models by reducing communication costs within the firm in a period when new knowledge needs to be acquired quickly.

As changes in training and hierarchical structure are potentially correlated, we study their relative importance by allowing for heterogeneity along both margins in the same regression.

The results (in column 6 of Table 7) show that *both margins matter*. Finally, we can try to isolate a plausibly exogenous driver of the change in layers by exploiting heterogeneity across sector-shifts in the number of workers who left the firm in the three weeks following a Model change. The idea is that working groups affected by such separations may have been induced to reorganize by reducing the number of layers to a greater extent. This analysis indeed finds a strong first stage between separations and reorganization leading to a reduction in layers (Appendix Table A13). The results of the 2SLS estimation (in column 5 of Table 7) confirm that the DPV increase is concentrated in groups with below median reduction in number of layers.

Taken together, this analysis is consistent with the structure of the hierarchy having actual productivity consequences. These results also validate our choice to group workers into 16 layers for the analysis of hierarchical structure (see Table 1 and related discussion), as we document significant productivity consequences of hierarchical restructuring based on our definition of the hierarchy.

In terms of relative magnitudes, column 6 of Table 7 shows that changes in training have a larger effect than changes in hierarchical structure, especially in the first few weeks. We come back to this later in the paper, where we use the calibrated model to quantitatively disentangle the role of training and hierarchical structure in helping the firm minimize the cost of introducing new models.

Limited Role of Experience As teams that receive more training or undergo greater reorganization experience lower defects, this refutes the hypothesis that the productivity gains are materialized solely through learning by doing. We can speak more directly to the role of learning retained over time by contrasting the importance of training and hierarchical structure in explaining DPV with the role of experience. We do so by conducting heterogeneity analysis of the effects of Model changes on DPV with respect to the experience of workers in different working groups, measured as either tenure in the factory or tenure in their current working group. We extend the analysis in Table 7 to consider this type of heterogeneity by splitting sector-shifts in those with above vs below median experience (using the two definitions discussed above). The results are in columns 7 and 8 of Table 7, and show that there is no significant evidence that impacts on DPV are smaller in more experienced groups, as the interactions are small in magnitude and not significant.[45] This confirms that changes in training and organizational structure are a stronger predictor of reductions in defect rates. This again highlights the importance of these two organizational levers for productivity, over

[45]Consistent with these results, controlling for either measure of experience does not affect the impact of Model changes on DPV in our main specifications (see Appendix Table A7, columns 4 and 5).

and above the role of learning by doing. This also emphasizes our contribution over Levitt et al. (2013), who show direct evidence of learning by doing: we contribute precisely by uncovering the role of training and of changes in hierarchical structure in enabling learning of new problems and minimizing the productivity losses from the introduction of new products.[46]

Table 7: Heterogenous Effects of Model Changes on DPV

Var →	Baseline (1)	Training (2)	Suggestions scores (3)	Layers decrease (OLS) (4)	Layers decrease (IV) (5)	Layer decrease (OLS) (6)	Experience in firm (7)	Experience in working group (8)
0 to 3 weeks	1.047***	1.918***	2.796***	1.346***	1.640***	2.306***	0.984***	1.158**
	(0.377)	(0.501)	(1.063)	(0.428)	(0.411)	(0.588)	(0.345)	(0.478)
4 to 7 weeks	0.542	0.879*	1.789	0.775*	0.973**	1.225**	0.598	0.503
	(0.440)	(0.522)	(1.231)	(0.450)	(0.461)	(0.550)	(0.422)	(0.467)
0 to 3 weeks × High Var		-1.439***	-2.734**	-0.598**	-1.186***	-0.655**	0.147	-0.208
		(0.330)	(1.138)	(0.301)	(0.208)	(0.324)	(0.342)	(0.383)
4 to 7 weeks × High Var		-0.556**	-1.896	-0.466**	-0.862***	-0.556**	-0.138	0.0711
		(0.260)	(0.260)	(0.265)	(1.298)	(0.217)	(0.164)	(0.240)
0 to 3 weeks × High Training						-1.540***		
						(0.362)		
4 to 7 weeks × High Training						-0.664**		
						(0.278)		
Observations	1,296	1,296	1,296	1,296	1,296	1,296	1,296	1,296
Effect for High group (0-3w)		0.479	0.062	0.747	0.453	1.651	1.131	0.950
		(0.323)	(0.201)	(0.368)	(0.380)	(0.476)	(0.479)	(0.368)
Effect for High group (4-7w)		0.323	-0.107	0.309	0.111	0.670	0.460	0.574
		(0.405)	(0.293)	(0.445)	(0.442)	(0.526)	(0.497)	(0.452)
F-stat first stage					14.472			

Note: Estimates of equation 2 with DPV as dependent variable using data between 2017-2019 at the shift-sector-day level. Interactions are with dummies for being above the median ("High Var") on each of the following variables at shift-sector level: "Training", measured as the average number of courses taken in the three weeks after the Model change, "Suggestions scores", measured as the average score of suggestions in the three weeks after the Model change, "Layers Decrease", measured as the average decrease in layers in the three weeks after the Model change (so note that those groups experience a larger decrease in number of layers have a higher value of this variable), "Experience" measured as the worker's time at the firm (measured in the three weeks before the Model change), "Experience in working group" measured as the worker's time in the working group (measured in the three weeks before the Model change). In column 6, changes in layers after the Model change is instrumented with a dummy for whether the shift-sector had a number of separations above the median in the three weeks after a Model change. Standard errors clustered by distance to event and shift-sector. Number of observations: 8 shift-sector x 81 days x 2 events. * p<0.1, ** p<0.05, *** p<0.01.

4.1.7 Additional Robustness Checks

We conclude this section by presenting several robustness checks, the details of which are reported in the Appendix. Specifically: (i) using a difference-in-difference specification, we verify that the impact of Model changes on DPV, training, and number of layers is stronger

[46] In Appendix Table A12 we show that the impact on DPV is lower for Model changes happening later in our sample period. This shows that the plant might learn over time how to minimize the productivity losses from the introduction of new products. The fact that worker experience does not matter much but plant experience does is again in line with the results in Levitt et al. (2013) that learning by doing is embodied primarily in the organizational capital of the firm and is not retained by specific workers to a large extent.

in those shifts that produced relatively more of the new model (Table A14); (ii) the impacts of Model changes on DPV are even stronger if we weight defects based on their importance, as measured by how consequential the defects in different sectors of the Assembly line are in slowing down production (Table A7, column 6);[47] (iii) the impacts on number of layers reported in Table 5 are very similar if we redefine layers so that the three layers that earn the same wage (i.e., New FL 2, New FL 3 and Reg FL 1—see Table 1) are grouped into one layer (Table A15); (iv) the results of all regression discontinuity in time specifications are confirmed in the corresponding event studies (see Appendix Figures A7, A8, A10); (v) the results of the event studies are very similar if we conduct the analysis at the daily level (instead of weekly) for our key outcomes of training and number of layers (Figure A20).

4.2 Contrasting Responses to Model and Volume Changes

As shown in Figure 1, over our sample period the plant experiences several sharp and permanent increases in production volume, so that Model changes are regularly followed by Volume changes. We contrast the organizational response to Model changes documented above with that to Volume changes. The results of the analogous analysis are presented in detail in Appendix A2. In short, we find that while Volume changes lead to a quantitatively similar spike in defects per vehicle, the organizational response is very different: the plant hires more entry-level workers and adds more layers to working groups, increasing the distance—in terms of knowledge layers—between front-line workers and supervisors farther up the hierarchy. The span of control of complex problem-solvers increases, and the average skill level in the firm decreases, as no additional training is provided. Unlike the temporary flattening observed after Model changes, the hierarchical expansion following Volume changes is persistent. Consistent with the broader span of control and lower average skill, Volume changes lead to a decline in both the number and the average score of suggestions in the plant's continuous-improvement system.

The results for Volume changes are in line with the literature on knowledge hierarchies (Caliendo et al., 2020, 2015; Ewens and Giroud, 2025; Friedrich, 2022), which tends to find that as firms expand production, new layers are added to the hierarchy, because the higher volumes allow the firm to better focus the talent of highly skilled managers on solving only the more complex problems. However, the analysis of worker-specific stocks of knowledge and problem-solving activities are, to our knowledge, novel to the literature.

[47] To classify the importance of defects, we regress the total number of cars produced per day on dummies for whether each of the four sectors of the Assembly line had defects per vehicle more than 1 SD above the mean of the sector. We then use the coefficients from this regression as weights. See Table A8 for details.

4.2.1 The Role of Investment in Training Resources

The comparison of the responses to Model and Volume changes shows that training plays a key role in determining the equilibrium levels of knowledge at different levels of the hierarchy. We know from anecdotal evidence that every time the complexity of the routine tasks increases due to a Model change, the partner firm *invests* in the training programs (in terms of both content and trainers) to train workers more efficiently in problem solving and managerial skills. Therefore, we conclude this section by exploring empirically how training investments by the firm differ for Volume and Model changes, to substantiate the claim that the firm increases the productivity of training in response to Model changes.

Table 8: Impact of Model and Volume Changes on Training Investments

	(1) Number of Courses Provided	(2) Number of Trainers
	Panel A: Model Changes	
0 to 3 weeks	1.805***	1.457*
	(0.464)	(0.706)
4 to 7 weeks	1.233	1.371
	(0.848)	(1.251)
Observations	112	112
Obs. Level	Weekly	Weekly
Mean	6.255	8.055
	Panel B: Volume Changes	
0 to 3 weeks	-1.588	-2.213
	(0.943)	(1.668)
4 to 7 weeks	-0.466	-0.607
	(1.535)	(2.169)
Observations	80	80
Obs. Level	Weekly	Weekly
Mean	6.100	8.175

Note: Standard errors clustered by distance to event. Number of Courses Provided is the number of courses provided each week and Number of Trainers is the number of trainers teaching each week. We use as controls month and year fixed effects. We also control for a linear function of distance to the event (i.e., Model or Volume change) and to all other events in the data. Panel A shows the effect for Model changes and Panel B shows the effect for Volume changes. Number of observations in Panel A: 7 events x 16 weeks. Number of observations in Panel B: 5 events x 16 weeks. * p<0.1, ** p<0.05, *** p<0.01

Table 8 shows that indeed the firm increases substantially the number of *different* courses provided *in house* per week and the number of trainers per week after a new model is introduced (Panel A), especially in the first four weeks after a Model change.[48] These results

[48] Appendix Figures A11 report the corresponding event study specification for Model changes, and confirm that the effects are concentrated in the first week after the Model change.

are consistent with the anecdotal evidence provided by the partner company. It also again highlights investment in training resources as an important margin that enables the firm to adapt to Model changes and resolve defects quickly, consistent with the empirical evidence documented above. In contrast, we do not see any impact when the firm changes the scale of production (Panel B). Based on this evidence, in the theoretical model in the next section we allow the firm to make (costly) investments in the productivity of training over the product cycle.

5 Model

In this section, we develop a dynamic model of the firm to (i) provide a unifying framework for interpreting our empirical results, and (ii) quantify the role of the two organizational levers of training provision and hierarchical structure in facilitating adaptation to changing production conditions over the product cycle. We build on and extend the organizational hierarchy model of Garicano and Rossi-Hansberg (2012) to study how firms optimally adjust their internal organization in response to technological innovations over the product cycle. We develop a dynamic framework in which firms choose both the timing of technology adoption—such as the introduction of a new automobile model—and the accompanying adjustments in hierarchical structure (layers) and knowledge acquisition efforts (training investments). Adopting a new technology raises productivity but also temporarily increases problem complexity.

Relative to Garicano and Rossi-Hansberg (2012), our framework introduces a key innovation: we endogenize training costs, allowing firms to adjust training intensity optimally over time. This extension reflects the empirical evidence from the previous section showing that training investments vary across the product cycle and are a critical organizational lever used by the firm. By incorporating this flexibility, the model captures the joint dynamics of training and hierarchical depth, and characterizes the optimal combination of the two at each stage of the cycle to manage complexity and meet production requirements at minimum cost.

Our model shows that in the early phase following new product adoption, firms tend to reduce the number of layers while intensifying training investments. As workers become better at solving new problems, the need for training declines and the organizational structure gradually re-expands. These dynamics reproduce the empirical patterns for Model changes described in the previous section. The model also clarifies the underlying trade-off: training and layers act as substitutes in managing complexity in the short-run—firms can either train workers more intensively or delegate problem-solving through additional layers. The model provides conditions under which each margin is preferred, making explicit how substitution between the two channels minimizes adjustment costs.

We quantify the implications of this trade-off in counterfactual exercises in Section 6, isolating the roles of training and hierarchy in cost-effective adaptation. We then examine firm responses to exogenous output expansions following technology adoption in Section 7, showing that under certain conditions, the optimal long-run adjustment involves increasing layers without expanding training. In this way, the model links our empirical findings on Model and Volume changes into a unified *organizational cycle* of the firm.

5.1 Static Problem

We begin by formalizing the firm's static cost-minimization problem. At any point in the product cycle, which is defined by a fixed technology and a given level of problem complexity, the firm chooses the number of hierarchical layers (from a finite set) and the corresponding training investments to minimize total costs related to training, subject to meeting a fixed production requirement.

We proceed in three steps. First, we describe the production and task delegation processes, introducing the key elements of the knowledge-based hierarchy framework. Second, for a given set of organizational parameters, we characterize the firm's optimal choice of knowledge levels and explain how these choices determine net earnings per worker and total output. Finally, we formally define the firm's static optimization problem and introduce the associated cost function.

5.1.1 Hierarchical Problem-Solving Framework

We follow Garicano and Rossi-Hansberg (2012) and model a knowledge-based hierarchy where workers and managers solve production-related problems characterized by a complexity level $z \geq 0$. Problems are drawn from a distribution with decreasing density $f(z)$, so simpler problems occur more frequently. Specifically, complexity follows an exponential distribution $F(z) = 1 - e^{-\lambda z}$ (with $\lambda > 0$), yielding $f(z) = \lambda e^{-\lambda z}$.

Agents acquire knowledge to address problems up to complexity z at a cost cz, where $c > 0$ is the cost per unit of knowledge. If a worker solves a problem, the firm produces A units of output, where A represents the technology level. Unsolved problems are escalated to the next layer: the manager spends h units of time diagnosing the issue (the communication cost) and resolves it if her knowledge suffices; otherwise, the problem is passed further upward. This process continues until the problem is solved or reaches the highest managerial layer.

5.1.2 Organizational Structure and Equilibrium

Given a firm with $L+1$ layers, and fixed parameters c, λ, A, and h, agents at each layer ℓ choose their knowledge level z_L^ℓ by weighing the trade-off between problem-solving output and the costs of training and communication.[49] A higher knowledge level enables an agent to solve a greater share of problems directly, thereby increasing net earnings, but it also entails higher training costs. These choices, taken simultaneously across layers, determine the flow of problems within the hierarchy and the resulting distribution of workers across organizational levels. Formally, workers solve:

$$w_L^0 = \max_z \left\{A - e^{-\lambda z}(A - r_L^0)\right\} - cz, \tag{3}$$

$$w_L^\ell = \max_z \left\{\frac{1}{h}\left[(A - r_L^{\ell-1}) - e^{-\lambda z}(A - r_L^\ell)\right]\right\} - cz, \quad \text{for } 0 < \ell < L, \tag{4}$$

$$w_L^L = \max_z \left\{\frac{1}{h}\left[(A - r_L^{L-1}) - Ae^{-\lambda z}\right]\right\} - cz. \tag{5}$$

Here, w_L^ℓ denotes the net earnings per agent in layer ℓ, and r_L^ℓ represents the internal transfer price for unsolved problems passed to the next layer. The term $e^{-\lambda z}$ is the probability that a problem exceeds an agent's knowledge and thus must be escalated.

Solving the first-order conditions yields optimal knowledge levels:

$$z_L^0 = -\frac{1}{\lambda}\ln\left(\frac{c}{\lambda(A - r_L^0)}\right), \tag{6}$$

$$z_L^\ell = -\frac{1}{\lambda}\ln\left(\frac{ch}{\lambda(A - r_L^\ell)}\right), \quad \text{for } 0 < \ell < L, \tag{7}$$

$$z_L^L = -\frac{1}{\lambda}\ln\left(\frac{ch}{\lambda A}\right). \tag{8}$$

These expressions show that optimal knowledge at each layer depends jointly on technology A, problem complexity λ, training cost c, communication cost h, and the set of transfer prices $\{r_L^\ell\}$, which govern the internal valuation of problems passed upward in the hierarchy.[50]

These knowledge choices, taken simultaneously across layers, govern the flow of problems within the hierarchy and also determine the allocation of workers across layers, balancing

[49] The firm is aware of the workers' choices of knowledge levels. Later, we will see how it uses this information to determine the organizational parameters that lead workers to select knowledge levels that achieve the firm's desired output at minimum cost.

[50] In Sections B1.2 and B1.3 of the Appendix, we analyze these relationships in greater depth, showing that optimal knowledge increases with higher technology and lower training costs—patterns that are central in our dynamic model.

training and communication costs to deliver cost-effective problem solving.[51] Aggregating effective production across layers, the firm's total output is given by:

$$\Pi = An_L^0 F(Z_L^L), \tag{9}$$

where n_L^0 is the mass of workers in layer 0, $Z_L^L = \sum_{\ell=0}^{L} z_L^\ell$ is cumulative knowledge, and $F(Z_L^L) = 1 - e^{-\lambda Z_L^L}$ is the share of problems resolved.

In equilibrium, agents must be indifferent between roles and willing to both supply and demand problems at each layer. This requires net earnings per worker to be equalized across layers:

$$w_L^\ell = w_L^{\ell+1} \equiv \tilde{w}, \quad \text{for } 0 \leq \ell < L. \tag{10}$$

We follow Garicano and Rossi-Hansberg (2012) and introduce the following assumption.

Assumption 1. *The parameters satisfy the following conditions: $A \geq 1$, $h < 1$, and the inequality* $\dfrac{A\lambda}{c} > \dfrac{1}{h} + \ln h$.

The first condition in Assumption 1 imposes a minimum productivity requirement, ensuring that the output generated by each problem solved exceeds one. The second requires communication costs to be sufficiently low so that each problem delegated to a manager takes less time than the manager's total available time. The full set of conditions in Assumption 1 constitutes a necessary and sufficient requirement for strictly positive knowledge levels at all layers.

Under Assumption 1, there exists a unique equilibrium with transfer prices decreasing across layers. This structure implies strictly higher knowledge levels in upper layers: lower levels address frequent, simple tasks; higher levels solve rarer, more complex problems.[52]

5.1.3 Endogenous Training Costs

A key innovation of our framework is to treat training costs as endogenous. We assume that firms can invest in lowering c through programs and processes such as standardized modules, mentorship structures, or increased number of trainers/courses, but there is a penalty $\mathcal{P}$ (or a fee) for investing in reducing the training cost. We represent this penalty by:

[51] The allocation at each level is discussed in Appendix B1.1, equation 16.

[52] This set up is consistent with our empirical evidence from the previous sections that workers further up the hierarchy have accumulated more knowledge (via training) and make more complex suggestions.

$$\mathcal{P}(c, \lambda, L) = \begin{cases} \left(c - \frac{1}{\lambda} - \gamma L\right)^{-1}, & \text{if } c > \frac{1}{\lambda} + \gamma L, \\ +\infty, & \text{otherwise,} \end{cases} \tag{11}$$

Here, $1/\lambda$ reflects the minimum feasible training cost given problem complexity: this captures the fact that more complex problems inherently require more sophisticated and costly training. γL captures additional coordination and layer-specific training costs that come with organizational depth. Intuitively, as the training is layer-specific, it is more costly to organize and deliver training in an organization with many layers.

5.1.4 Firm's Static Problem

Following Caliendo and Rossi-Hansberg (2012), we model the firm's objective as the minimization of a cost function subject to a minimum production requirement. Specifically, the firm selects c so as to minimize training expenditures while ensuring that output meets the target level q:

$$\begin{aligned} \min_{c} \quad & \mathcal{P}(c, \lambda, L) \\ \text{s.t.} \quad & c \in \Gamma, \end{aligned} \tag{12}$$

where the feasible set Γ is defined as

$$\Gamma = \left\{ c \in \left(1/\lambda + \gamma L, \frac{A\lambda}{1/h + \ln h}\right] : q - A n_L^0 F(Z_L^L) \leq 0 \right\}. \tag{13}$$

The lower bound of the feasible set Γ, $1/\lambda + \gamma L$, ensures that the penalty function $\mathcal{P}(c, \lambda, L)$ is finite, reflecting the minimum cost required to generate knowledge, given the complexity of problems and the depth of the hierarchy. The upper bound, derived from Assumption 1, guarantees viability—positive knowledge and transfer prices at all layers.

Proposition 1. *Suppose that $\Gamma_{A,\lambda,L} \neq \emptyset$. Then, there exists a unique $c^* \in \Gamma_{A,\lambda,L}$ that solves the optimization problem (12).*[53]

Endogenizing training costs creates a direct link between organizational structure and knowledge acquisition. Firms jointly choose hierarchy and training intensity in response to technology upgrading over the product cycle, extending Garicano and Rossi-Hansberg (2012).

[53] This condition is equivalent to a condition on the training cost, further detailed in Appendix B1.4.

5.2 Dynamic Framework

We now extend the analysis to a dynamic setting in which firms sequentially choose their organizational structure and training investments, taking into account the adjustment costs of increasing or reducing training. This dynamic perspective is essential for understanding how organizations evolve over product cycles, which generate both technological changes and shifts in problem complexity.

Each period, the firm makes two endogenous decisions: (i) whether to change its training cost c_t, and (ii) whether to modify its number of hierarchical layers L_t. These choices are shaped by exogenous parameters such as the prevailing technology level A, problem complexity λ, and a fixed adjustment cost δ capturing the resources required to alter training investments.

The training cost c is the per-worker cost of knowledge acquisition—e.g., hours of formal instruction, mentorship, or access to learning resources. In contrast, the adjustment cost δ captures organizational frictions in changing training policies, such as the administrative costs of redesigning programs, renegotiating contracts with trainers, or reallocating staff time.[54]

The firm's decision problem can be represented by a per-period cost function $\phi_t(A, \lambda)$, which gives the minimum total cost of meeting production and other feasibility constraints, given technology A and problem complexity λ. Each period, total costs arise from two sources:

1. The *training penalty* $\mathcal{P}(c, \lambda, L)$, defined in (11), which captures the resources required to sustain a training cost c given problem complexity λ and hierarchy depth L.

2. The *adjustment cost* δ, incurred when the firm changes c from its previous period's level, reflecting the frictions of altering training policies (e.g., redesigning programs, reallocating staff time, or revising internal processes).

Changing L alone avoids δ, while changing c adds it to the new training penalty. The resulting dynamic problem is:

$$
\begin{aligned}
\phi_0(A, \lambda) &= \mathcal{P}(c_0, \lambda, L_0), \\
\phi_{t+1}(A, \lambda) &= \min_{L \in \{L_t - 1,\, L_t,\, L_t + 1\}} \min \left\{ \begin{array}{l} \mathcal{P}(c_t, \lambda, L), \\ \min\limits_{c} \left[\mathcal{P}(c, \lambda, L) + \delta\right] \quad \text{s.t. } c \in \Gamma \end{array} \right\}
\end{aligned}
\tag{14}
$$

where the feasible set Γ, (defined in (13)) depends on A, λ, L, and q. Our framework extends Garicano and Rossi-Hansberg (2012) and allows firms to adjust structure in either

[54] We model the adjustment cost as a "fixed" type of cost as it captures organizational frictions that are not worker specific.

direction—reducing it ($L_{t+1} = L_t - 1$), keeping it unchanged ($L_{t+1} = L_t$), or expanding it ($L_{t+1} = L_t + 1$).[55] Similarly, firms may either maintain their current training investment $\mathcal{P}(c_t, \lambda, L)$, or re-optimize training at cost δ. The latter involves solving $\min_c [\mathcal{P}(c, \lambda, L) + \delta]$ over $c \in \Gamma$, with δ incurred whenever c changes.

Finally, note that the dynamic system has two interdependent state variables—hierarchical depth L_t and training cost c_t—with optimization rules detailed in Appendix B1.5.[56]

5.2.1 Convergence to Steady State

We model product cycles as sequences of sub-cycles, each defined by a fixed technology A and problem complexity λ. To study firm responses to endogenous technological innovation—here, the introduction of new products that temporarily raise complexity—we first examine behavior within a single sub-cycle. Holding A and λ constant allows us to analyze the organizational structure and training investments in a stable environment.

Under mild conditions, both c_t and L_t converge to steady-state values. This steady-state analysis forms the basis for the next section, where we consider transitions across sub-cycles in which technology or complexity changes.

The next assumption guarantees that, even with fixed technology and complexity, the firm has at least one feasible organizational choice, ensuring the optimization problem is well-defined.

Assumption 2. *Let the problem complexity λ and technology level A remain constant, and suppose the firm begins with initial training cost c_0 and number of layers L_0. Then, there exists at least one $L \in \{L_0 - 1, L_0, L_0 + 1\}$ such that the feasible set Γ is non-empty.*

Lemma 1. *Under Assumption 2, the sequence $\{L_t\}$ is monotonic.*

This lemma implies that, once the firm begins adjusting its structure, it will move in a single direction—either adding layers or removing them—until reaching the steady state.

Proposition 2. *Under Assumption 2, the sequences $\{c_t\}$ and $\{L_t\}$ converge to steady-state values c_∞ and L_∞, respectively, as $t \to \infty$.*

These results reveal a fundamental tension in organizational evolution. While our framework allows for both expansion and contraction of organizational structure, Lemma 1 shows that, with unchanging levels of technology and problem complexity, firms optimally choose a consistent direction of adjustment. This directional consistency, combined with the bounded

[55] As in Garicano and Rossi-Hansberg (2012), the firm may adjust its structure by at most one layer per period, reflecting incremental rather than wholesale organizational changes.

[56] Proposition 1 ensures a unique optimal response given a nonempty feasible set.

nature of our cost function, leads to the convergence to the steady state (c_∞, L_∞) of our sub-cycle, as established in Proposition 2. The proof also implies that this convergence is reached in finite steps.

5.3 Endogenous Technological Innovation

We now augment the model to incorporate technological innovation and its impact on the firm's decisions over time, making both the timing of innovation and the transitions between product-cycle phases endogenous. In this setting, the firm chooses when to adopt new technologies by weighing the anticipated efficiency gains against the organizational restructuring and retraining costs such adoption entails.

In practice, technological innovation follows a learning curve: the introduction of a new product (e.g., a new automobile model) initially increases problem complexity as workers adapt to new processes and components. After this learning phase, agents become proficient, and complexity returns to its baseline level. We represent this temporary increase in complexity as a drop in the complexity parameter from λ to $\tilde{\lambda} < \lambda$, beginning at the innovation time τ^* and lasting until $\tau^* + T$, where T is the exogenous length of the learning phase.[57]

To capture these dynamics formally, we embed endogenous technology adoption into the firm's dynamic optimization problem, with a temporary increase in complexity during adaptation. The firm starts with technology A and complexity λ until it chooses the optimal innovation time τ^*. At that moment, it adopts an improved technology $\tilde{A} = (1+\zeta)^{\tau^*} A$ at a one-time cost $\Psi \tilde{A}$.

The resulting dynamic cost-minimization problem is:

$$\begin{aligned}\Phi(\Psi, \zeta, L_0, c_0, \boldsymbol{\lambda}, T) = \min_{\tau^*} \Bigg\{ & \sum_{t=0}^{\tau^*-1} \beta^t \phi_t(A, \lambda) + \beta^{\tau^*} \Psi \tilde{A} \\ & + \sum_{t=\tau^*}^{\tau^*+T-1} \beta^t \phi_t(\tilde{A}, \tilde{\lambda}) + \sum_{t=\tau^*+T}^{\infty} \beta^t \phi_t(\tilde{A}, \lambda) \Bigg\},\end{aligned} \tag{15}$$

where β is the discount factor, ζ is the rate of technological progress, and $\phi_t(A, \lambda)$ is the per-period cost function from Equation 14, capturing both training investments and organizational structure.

The optimal switching time τ^* balances two forces: delaying adoption avoids immediate restructuring and training costs, but also postpones productivity gains; adopting too early accelerates efficiency improvements but triggers complexity increases before they can be

[57] Our assumption that the learning phase lasts for a fixed number of periods is motivated by the evidence that the spike in DPV following Model changes is temporary (see Figure 2).

cost-effectively absorbed.

Our framework extends the knowledge-based hierarchy model of Garicano and Rossi-Hansberg (2012) by explicitly modeling the timing of innovation and the transitory complexity effects of adoption. We show that, under standard regularity conditions, there exists a unique τ^* that minimizes total discounted costs, ensuring efficient adaptation to technological change (see Appendix B1.6 for illustrations).

5.3.1 Organizational Dynamics Following an Innovation

We now examine the organizational adjustments that follow the adoption of new technology. Specifically, we identify conditions under which a firm optimally responds to a temporary increase in problem complexity by flattening its hierarchy—reducing the number of layers during the learning phase—and then re-expanding it once complexity subsides.

To characterize these conditions, we introduce the following assumption, which captures cases where heightened complexity renders the pre-innovation structure unsustainable:

Assumption 3. *Let the firm be in its steady state (c, L) with training cost c and number of layers L, such that $\Pi(A, \lambda, c, L) \geq q$, where Π is output as defined in (9). The learning-phase complexity $\tilde{\lambda}$ satisfies:*

1. *The firm cannot meet q with L layers and the minimum feasible training cost $\bar{c}_L = 1/\tilde{\lambda} + \gamma L$, i.e., $\Pi(\tilde{A}, \tilde{\lambda}, \bar{c}_L, L) < q$.*
2. *The firm can meet q by reducing to $L-1$ layers and the corresponding minimum training cost $\bar{c}_{L-1} = 1/\tilde{\lambda} + \gamma(L-1)$, i.e., $\Pi(\tilde{A}, \tilde{\lambda}, \bar{c}_{L-1}, L-1) \geq q$.*

This assumption captures situations in which the complexity increase from λ to $\tilde{\lambda}$ is large enough that the pre-innovation structure with L layers cannot produce q even if each layer operates at its maximum feasible knowledge level. However, by removing a layer and thereby concentrating more knowledge per remaining layer, the firm can still meet q. Flattening frees up resources for training and reduces communication costs, enabling the organization to cope with the temporary surge in complexity.

Lemma 2. *There exists an open interval Λ for $\tilde{\lambda}$ such that any $\tilde{\lambda} \in \Lambda$ satisfies Assumption 3.*

Lemma 2 shows that the assumption above is not a knife-edge case: there is a range of $\tilde{\lambda}$ values for which the hierarchy with L layers fails but a flatter structure succeeds. When the complexity increase is too small, the original structure still works; when it is too large, even a flatter structure cannot meet q.

With these conditions in place, we can state the main result on the firm's optimal response:

Proposition 3. *Suppose that Assumption 3 is satisfied, with $\tilde{\lambda} < \lambda$. If the firm is in its steady state (c, L) before innovation and there exists a finite solution to the optimization problem (15), then the number of organizational layers L_t exhibits the following behavior:*

- *At the time of innovation adoption $t = \tau^*$, the firm decreases its number of layers.*
- *During the learning phase $\tau^* \leq t < \tau^* + T$, the number of layers remains at this reduced level or may decrease further but does not increase.*
- *After the learning phase, at time $t = \tau^* + T$, the firm increases its number of layers.*
- *For $t > \tau^* + T$, the number of layers remains at this higher level or may increase further but does not decrease.*

Proposition 3 highlights how the adoption of new technology triggers a learning phase with higher complexity (lower λ), making the pre-existing hierarchy less effective. Flattening reduces communication costs by shortening reporting layers and, combined with intensified training, increases each agent's problem-solving capacity. In the short run, training and layers act as *substitutes*: when complexity is high, raising knowledge per layer is more efficient than adding new layers. After the learning phase, as λ recovers, re-expansion allows the firm to process a higher volume of problems, leveraging the improved technology.[58]

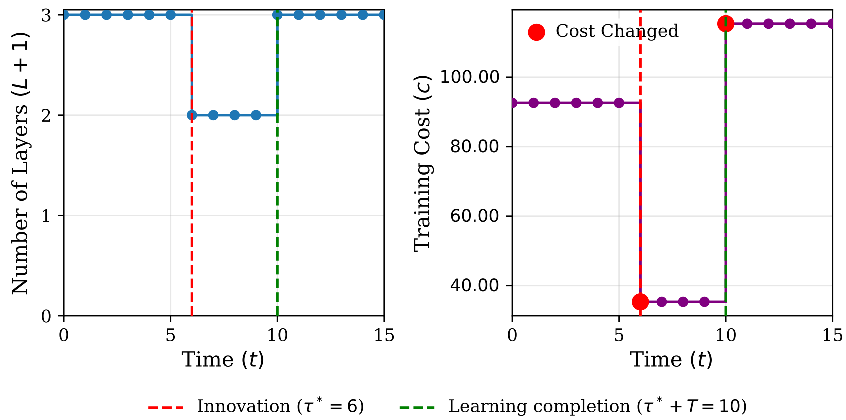
Figure 7 illustrates this adjustment.[59] Initially, the organization operates with $L = 2$ (three layers). At τ^*, adoption raises A and lowers λ. In response, the firm reduces the number of layers and invests in lowering c to maintain problem-solving efficiency, so that training provision increases. As the learning phase ends, λ returns to its initial value, prompting re-expansion and training adjustments. Higher post-learning training costs relative to pre-innovation reflect the combined effect of higher A and a re-optimized knowledge distribution across layers.

These findings underscore that organizational flexibility, and particularly the ability to adjust the hierarchy dynamically, is key to managing transitional complexity and capturing the long-run gains from innovation. The model qualitatively reproduces our empirical results on firms' responses to Model changes, both in terms of training provision and investment (Figure 4 and Table 8, Panel A) and adjustments in hierarchical structure (Table 5 and Appendix Figure A8a). Moreover, it clarifies how these patterns are jointly determined as part of the firm's dynamic optimization problem.

[58]While T is treated as exogenous for tractability, making it endogenous—e.g., as a function of training, accumulated knowledge, or hierarchy depth—would likely reinforce these predictions, as the threshold condition driving τ^* would remain.

[59]Parameter values in Figure 7 are for illustration only. The next section calibrates the model using data to assess quantitative implications.

Figure 7: Innovation with an Increase in Complexity



Notes: Parameters: $q = 65$, $A = 100$, $h = 0.43$, $\lambda = 1.5$, $\tilde{\lambda} = 0.7$, $\beta = 0.001$, $\Psi = 10^{-3}$, $\zeta = 0.01$, $\gamma = 1.4$, $\delta = 10^{-4}$, $T = 4$.

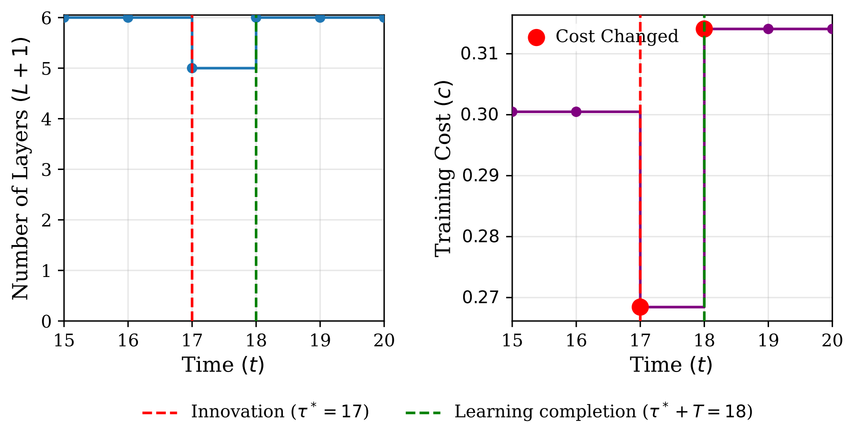
6 Model Calibration and Counterfactuals

To quantitatively assess the implications of our theoretical framework, we calibrate the model using empirical moments from the observed dynamics of hierarchical layers and training investments over the product cycle. Specifically, we match the model to average values of hierarchical structure, number of training courses, and number of instructors across three distinct phases—Initial (i.e., before the Model change), Learning (i.e., the first three weeks after a Model change, when defects spike), and Post-Learning (i.e., after the Learning phase)—constructed from weekly event-study estimates around Model changes. The calibration employs eight parameters, estimated by minimizing the normalized error between model predictions and observed data, and captures both the timing and magnitude of organizational responses to innovation. Full details of the calibration procedure, parameter values, and empirical moment construction are presented in Appendix B2.

Figure 8 shows that our calibrated model closely replicates the empirical patterns documented in the data, including temporary hierarchy flattening and intensified training during learning, followed by re-deepening and reduced training thereafter. It also reproduces the dynamic adjustment of training costs and organizational layers in response to innovation. These results establish the quantitative credibility of the model and provide a strong foundation for the counterfactual exercises that follow.[60]

[60]Each period in Figure 8 corresponds to three weeks. As discussed in more detail in Appendix B2, the average product cycle in our data lasts 51 weeks, or 17 three-week periods. The timing of innovation τ is treated as exogenous in the calibration and is assumed to occur in period 17 (51 weeks), with the learning phase lasting one period (3 weeks). Figure 8 shows dynamics of the calibrated model in periods around the innovation. Dynamics over the complete product cycle are shown in Appendix B2.

Figure 8: Firm Dynamics over the Product Cycle with Calibrated Parameters



Notes: Firm dynamics with calibrated parameters for a product cycle with a 34-period time horizon, innovation at $\tau = 17$, and a learning phase of $T = 1$.

Next, we examine the dynamics of the model under some counterfactual scenarios. Specifically, we analyze how the overall cost of the firm over the product cycle increases in two scenarios: (i) when shutting down the possibility of changing training investments (by increasing the adjustment cost δ associated with the modification of training costs (c)); and (ii) when preventing the firm from changing its organizational structure (by fixing the number of hierarchical layers (L)). This allows us to decompose the role of training investments and the hierarchy in allowing the firm to adapt to product cycles by minimizing cost, thus comparing the significance of each of these two organizational levers.

6.1 Counterfactual: Dynamics when Training Costs are Fixed

A distinctive feature of our model is that training costs are endogenously determined. When they can be adjusted at low cost, firms respond to innovation by increasing training and temporarily flattening their hierarchy. However, if adjustment costs are prohibitively high, this dynamic changes: training remains fixed, and the firm adapts instead by altering its hierarchy, often at higher overall cost.

We begin by restricting attention to cases in which the training cost at the onset of the learning phase, $c_{\tau-1}$, is feasible for at least one nearby hierarchical configuration—specifically, with $L_{\tau-1} - 1$, $L_{\tau-1}$, or $L_{\tau-1} + 1$ layers, where $L_{\tau-1}$ is the number of layers in the initial phase's steady state. This restriction is not arbitrary: it is the only setting in which we can meaningfully analyze the counterfactual where the firm, facing prohibitively high adjustment costs, keeps the same training cost from the pre-learning phase throughout the learning phase.[61]

We now examine the consequences of holding training costs fixed during the learning phase. Let (c, L) denote the steady state at the end of the pre-learning phase—that is, the training cost and number of layers to which the firm converges before the innovation. A natural possibility is that, after the innovation, this pre-learning training cost becomes insufficient to meet production requirements with either a flatter hierarchy or even with the same one, despite the technological gain. The following assumption formalizes such scenarios.

Assumption 4. *Suppose the firm enters the learning phase from a pre-learning steady state (c, L) established under technology level A, problem complexity λ, and output $\Pi(A, \lambda, c, L) \geq q$. Upon adopting improved technology $\tilde{A} > A$ at the start of the learning phase, the new problem complexity $\tilde{\lambda} < \lambda$ is such that either the first or both of the following conditions hold:*

1. *The firm cannot meet its production target q with $L - 1$ layers and the pre-learning training cost c, i.e., $\Pi(\tilde{A}, \tilde{\lambda}, c, L - 1) < q$.*

2. *The firm cannot meet its production target q with L layers and the pre-learning training cost c, i.e., $\Pi(\tilde{A}, \tilde{\lambda}, c, L) < q$.*

This assumption captures cases in which, if the firm were forced to keep its pre-learning training cost, the complexity increase would render the existing or a flatter hierarchy infeasible for meeting output targets. In such circumstances, the only way to sustain production without retraining is to deepen the hierarchy. This is not an arbitrary or irrelevant case: as the following lemma demonstrates, for sufficiently large complexity increases, scenarios of the type described in the assumption do in fact arise.

Lemma 3. *There exists a positive open interval $\Lambda \neq \emptyset$ such that any $\tilde{\lambda} \in \Lambda$ satisfies the first condition in Assumption 4. Moreover, there exists a non-empty subinterval $\Lambda' \subseteq \Lambda$ such that any $\tilde{\lambda} \in \Lambda'$ satisfies both conditions in Assumption 4.*

Lemma 3 establishes that the scenario described in Assumption 4 is not hypothetical: there is a range of complexity values for which the pre-learning training cost fails with a flatter hierarchy, and a narrower range within which it fails even with the same hierarchy.

We can now characterize the firm's optimal response when training costs remain fixed and complexity rises sufficiently during learning.

[61] If feasibility failed for all these nearby structures, keeping training fixed would be impossible from the outset, making the counterfactual impossible to run. In the restricted setting, Lemma 5 (Appendix B3.1) shows that, when adjustment costs are high, the firm optimally maintains its training cost during learning.

Proposition 4. *Under the first condition of Assumption 4, if the initial training cost satisfies $c \in \bigcup_{L' \in \{L-1,L,L+1\}} \Gamma_{L',\tilde{A},\tilde{\lambda}}$, then there exists a threshold $\delta' > 0$ such that for any adjustment cost $\delta > \delta'$, the firm maintains constant training expenditures throughout the learning phase. Furthermore, the number of organizational layers converges to a level weakly greater than the initial one. That is, if (c^*, L^*) denotes the new steady state during the learning phase, then $c^* = c$ and $L^* \geq L$.*

Proposition 4 contrasts with the low-adjustment-cost case in Proposition 3 and Figure 8, where firms respond to rising complexity by increasing training and temporarily flattening their hierarchy. When adjustment costs are high, Proposition 4 shows that firms hold training fixed and instead adapt their hierarchy—retaining or expanding it as needed. In both cases, layers and training remain substitutes in managing complexity.

If the learning phase involves a particularly large complexity increase relative to the technological gain, the firm will strictly increase its number of layers, as stated below.

Corollary 1. *Under the conditions of Assumption 4, if the initial training cost satisfies $c \in \bigcup_{L^* \in \{L-1,L,L+1\}} \Gamma_L$, then there exists a threshold $\delta > 0$ such that for any adjustment cost $\delta' > \delta$, the firm maintains constant training costs throughout the learning phase. Moreover, the number of organizational layers converges to a value strictly greater than the initial one. That is, if (c^*, L^*) denotes the new steady state during the learning phase, then $c^* = c$ and $L^* > L$.*

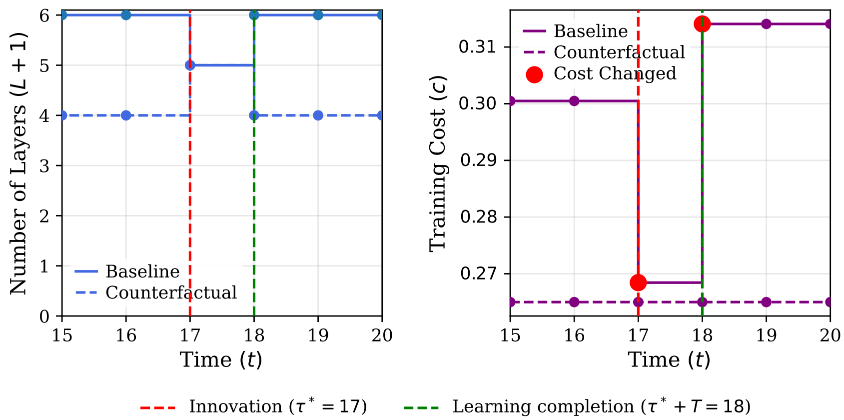
We validate these results through simulations using the calibrated parameters. In the baseline case (Figure 8), low adjustment costs allow the firm to increase training and reduce layers during the learning phase, then expand the hierarchy and reduce training thereafter. In contrast, Figure 9a shows that when adjustment costs are high and the initial training cost is feasible throughout the cycle, the firm holds training constant and adjusts only its hierarchy—deepening it during learning and returning to a shallower structure afterward.

Finally, in Figure 9b, we consider a case with a higher initial training cost. Here, even with high adjustment costs, the firm increases training during learning because the initial training level is infeasible under the higher complexity, violating Lemma 5. Anticipating this, the firm would have to adopt a suboptimal pre-learning configuration with fewer layers and higher training to keep training fixed, ultimately requiring hierarchical adjustments.

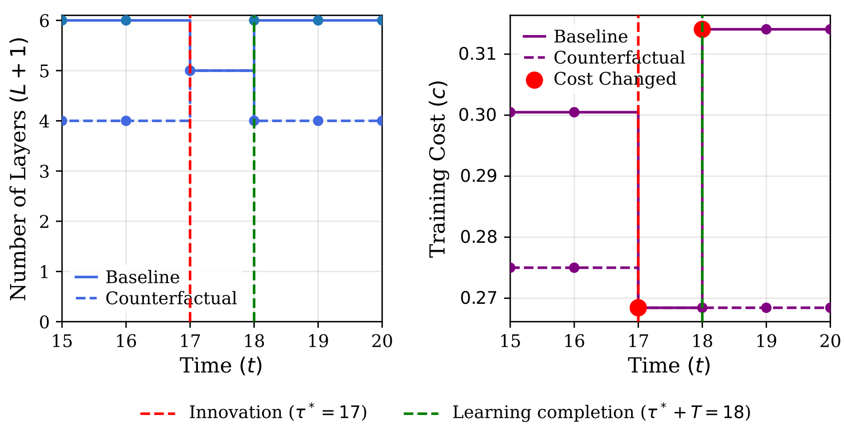
6.2 Counterfactual: Dynamics when the Hierarchy is Fixed

Figure 10 plots the dynamics of training costs when the number of organizational layers is held fixed. The overall pattern closely parallels the baseline case: training investment rises

Figure 9: Organizational Responses when Training Costs are Fixed over the Product Cycle



(a) Low initial training cost scenario (c_0=0.265).



(b) High initial training cost scenario (c_0=0.275).

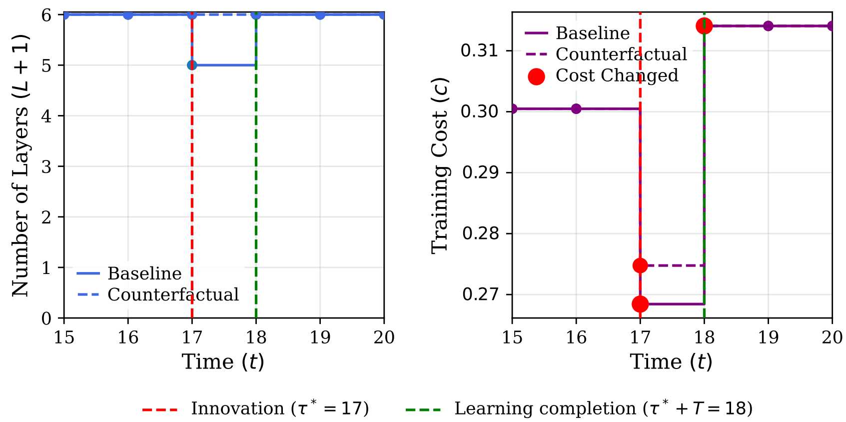
Notes: Comparison of organizational responses under different initial training costs and fixed initial layers ($L_0 = 5$). Each scenario compares baseline ($\delta = 0$) to high adjustment costs ($\delta = 10^6$).

during the learning phase and falls afterward. The key difference lies in the learning phase itself, where the firm—unable to flatten its hierarchy—reduces training costs per worker by a smaller margin. Two mechanisms account for this: first, maintaining a deeper hierarchy raises the cost of training each worker; second, with more layers in place, the amount of knowledge required at each layer to meet the production target is lower.

6.3 Cost Comparison Across Counterfactual Scenarios

To better understand the firm's organizational response during the learning phase, we compare cost increases under two counterfactual scenarios: (i) fixing training investments (illustrated

Figure 10: Organizational Responses when the Hierarchy is Fixed over the Product Cycle



Notes: Comparison of the baseline model and counterfactual with fixed organizational layers $L = 5$.

in Figure 9a), where the firm begins with a low initial training cost and is prevented from adjusting it during the cycle; and (ii) fixing the number of organizational layers (shown in Figure 10), where the hierarchy is held constant at its baseline level. Table 9 reports the percentage increase in total costs during the learning phase relative to the baseline, in which both margins are flexible.

Table 9: Cost Increase from Fixed Training and Fixed Hierarchy Counterfactuals

Scenario	**Cost Increase**
Fixed Training	2.4%
Fixed Layers	1.4%

Notes: For each of the two counterfactuals, we report the percentage increase in costs during learning phase relative to baseline.

There are two key takeaways from these results. First, both margins of flexibility—training provision and hierarchical restructuring—matter quantitatively, raising costs by 1.4–2.4% when constrained. Second, restricting training flexibility imposes a larger cost (2.4%) than restricting hierarchical adjustment (1.4%) during the learning phase. This indicates that training plays the more important role, reflecting the value of adjusting workers' knowledge in response to temporary increases in problem complexity.

Section 4 documented empirically that firms adjust both training and hierarchy in response to the introduction of new products. The counterfactuals presented here complement that evidence by quantifying the relative importance of each margin. Together, these findings constitute a central contribution of our study.

The counterfactuals also underscore the importance of allowing training costs, c, to adjust in the model. Training emerges as the main margin of adjustment, and a model with fixed training costs would fail to match the empirical evidence that the number of layers decreases temporarily following a Model change. Indeed, Figure 9 shows that with fixed training costs, the number of layers instead rises during the learning phase. In this way, we extend the theoretical literature on knowledge hierarchies (Caliendo and Rossi-Hansberg, 2012; Garicano, 2000; Garicano and Rossi-Hansberg, 2012) by highlighting the role of investment in training resources—a margin typically overlooked in existing models.

7 Model Dynamics When Scale Can Adjust

In Section 2, we documented that product cycles also involve changes in demand, as the introduction of new car models typically comes with subsequent increases in the quantity produced (Figure 1). In our model, demand enters as a minimum production constraint that the firm must meet while minimizing costs. The production target q thus serves as a fundamental scaling parameter, determining the minimum output needed to satisfy market demand (or contractual obligations). Here, we examine how discrete increases in minimum demand affect the firm's choices under calibrated parameters, and contrast these scale effects with those arising from technological change and shifts in problem complexity.

Model Dynamics with Changes in Quantity Produced To isolate the role of demand variation, we hold the technology level A and problem complexity λ fixed at the time the demand increase occurs. Specifically, consider a scenario in which after the completion of the learning phase $T + \tau$, demand rises exogenously from q to $\tilde{q} > q$ at time $\mathcal{T}$. The firm must then adjust its organizational structure and training investments to meet the higher production requirement.

While our dynamic problem treats both q and $\tilde{q}$ as exogenous, introducing a quantity shock after the learning phase is both more plausible and less costly than doing so earlier. By this stage, the firm has already adopted the new technology and completed the associated learning process, operating with a more productive configuration.[62] As a result, the production constraint is less binding, and a given output level can be achieved with a simpler, less costly structure.

In responding to a demand increase, the firm has two main adjustment margins: (i) increase worker knowledge through higher training investments,[63] which raises the capacity to solve complex problems at each layer; or (ii) expand the number of hierarchical layers, which

[62] Corollary 8 in the Appendix shows formally that output increases in A, holding other parameters fixed.

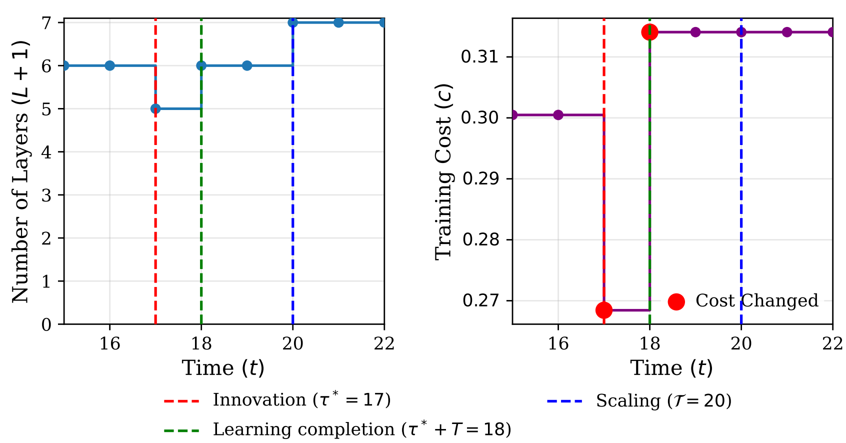
[63] Formally, Corollary 5 in the Appendix establishes that productivity Π increases as training cost c decreases, holding other factors fixed.

distributes problems more finely across the organization. Since increasing training entails a fixed adjustment cost, the firm compares this cost to the gains from higher productivity. If the cost is sufficiently high, it may be optimal to hold training constant and deepen the hierarchy instead.

Intuitively, this occurs because after adapting to the new technology, a demand increase presents a capacity rather than a complexity challenge—the nature of the problems is unchanged, but their volume rises. Without structural changes, the inflow of similar problems risks overloading existing layers, especially frontline workers. Because retraining is both costly and less effective at alleviating this bottleneck, the firm responds by deepening its hierarchy, enabling finer specialization and a greater division of labor.

Figure 11 illustrates this dynamic under the calibrated parameters. Although the adjustment cost is fixed, its bindingness depends on the context in which it is incurred. From the start through the end of the learning phase, training and layers act as *substitutes*, and the firm optimally modifies training despite the cost, as the long-run efficiency gains outweigh the investment. By contrast, when demand rises after the learning phase, the same cost becomes more restrictive: the firm maintains its existing training level and instead expands the hierarchy to meet the higher production target.

Figure 11: The Organizational Cycle of the Firm



Notes: Dynamics of the calibrated model with a learning phase followed by exogenous demand scaling over a 34-period product cycle. Adjustment cost $\delta = 0.5$; scaled demand $\tilde{q} = 0.1309$. Innovation occurs at $\tau = 17$, followed by a one-period learning phase ($T = 1$) and a demand shock at $\mathcal{T} = 20$. Panel (a) shows the evolution of organizational layers, while panel (b) displays the dynamics of training costs throughout the cycle.

The Organizational Cycle of the Firm The dynamics in Figure 11 help unify and clarify the interpretation of our empirical results on Model and Volume changes from Section 4. They reveal a nuanced relationship between training and hierarchy: in the short run, the two margins act as *substitutes*, as discussed above. However, short-run training investments facilitate the adoption of more productive technologies, which in turn make it easier to expand output in the long run through additional hierarchical layers without further training. In this sense, short-run training and long-run hierarchy function as *complements*.

By jointly considering changes in technology and demand, these results highlight the cyclical nature of organizational adjustment: periods of contraction and expansion in hierarchy combined with shifts in training costs. We refer to this as the firm's *organizational cycle*, which underpins two distinct forms of growth—quality (innovation and complexity) and quantity (scale of production). Documenting and modeling this cycle represents a central contribution of our study.

8 Conclusion

We introduce the concept of an *organizational cycle*—a recurring pattern of adaptation in hierarchy and training over the product cycle. Using detailed data from a subsidiary of a global automaker, we document systematic alternation between phases of organizational compression and expansion. Following product introductions, hierarchies flatten, spans of control narrow, and training intensifies, raising knowledge levels and shifting problem-solving to lower layers. During subsequent volume growth, hierarchies deepen and the firm hires low-skilled workers without additional training, reducing average knowledge within the organization.

To interpret and quantify these patterns, we develop a dynamic model that allows for endogenous investment in training efficacy alongside the allocation of knowledge across layers. The model rationalizes our findings: when output is fixed, training and hierarchy act as *substitutes* in managing new complexity, but over time they become *complements*, as training-induced innovation enables scale expansion through deeper hierarchies. Counterfactuals using the calibrated model highlight the quantitative importance of both organizational levers during phases of innovation, with training playing a quantitatively more important role.

This study connects the literature on knowledge hierarchies and firm growth with studies of product upgrading and innovation by providing analytical and empirical micro-foundations for an organizational cycle linking innovation and scale with training and hierarchy. In doing so, it sheds new light on how growing firms leverage their internal organization to manage complexity in the short run and scale in the long run.

References

Adhvaryu, A., Kala, N., and Nyshadham, A. (2022a). Management and shocks to worker productivity. *Journal of Political Economy*, 130(1):1–47.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2023a). Returns to on-the-job soft skills training. *Journal of Political Economy*, 131(8):2165–2208.

Adhvaryu, A., Molina, T., and Nyshadham, A. (2023b). Problem-solving skills training and continuous improvement in electronics manufacturing.

Adhvaryu, A., Murathanoglu, E., and Nyshadham, A. (2022b). On the allocation and impacts of managerial training.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2022c). An anatomy of performance monitoring. *Harvard Business School*.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2023c). Managerial quality and productivity dynamics. *The Review of Economic Studies*, 90(4):1569–1607.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The quarterly journal of economics*, 120(2):701–728.

Aghion, P. and Howitt, P. (1990). A model of growth through creative destruction.

Alonso, R., Dessein, W., and Matouschek, N. (2008). When does coordination require centralization? *American Economic Review*, 98(1):145–179.

Amiti, M. and Khandelwal, A. K. (2013). Import competition and quality upgrading. *Review of Economics and Statistics*, 95(2):476–490.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173.

Atkin, D., Khandelwal, A. K., and Osman, A. (2017). Exporting and firm performance: Evidence from a randomized experiment. *The quarterly journal of economics*, 132(2):551–615.

Bai, J., Barwick, P., Cao, S., and Li, S. (2021). Quid pro quo, knowledge spillover, and industrial quality upgrades: Evidence from the chinese auto industry. Technical report, Working Paper.

Bandiera, O., Prat, A., Hansen, S., and Sadun, R. (2020). Ceo behavior and firm performance. *Journal of Political Economy*, 128(4):1325–1369.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, pages 1351–1408.

Bloom, N. and Van Reenen, J. (2010). Why do management practices differ across firms and countries? *The Journal of Economic Perspectives*, pages 203–224.

Caliendo, L., Mion, G., Opromolla, L. D., and Rossi-Hansberg, E. (2020). Productivity and organization in portuguese firms. *Journal of Political Economy*, 128(11):4211–4257.

Caliendo, L., Monte, F., and Rossi-Hansberg, E. (2015). The anatomy of french production hierarchies. *Journal of Political Economy*, 123(4):809–852.

Caliendo, L. and Rossi-Hansberg, E. (2012). The impact of trade on organization and productivity. *The quarterly journal of economics*, 127(3):1393–1467.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Dessein, W. and Santos, T. (2006). Adaptive organizations. *Journal of Political Economy*, 114(5):956–995.

Dillon, A. P. (2019). *A study of the Toyota production system: From an Industrial Engineering Viewpoint*. Routledge.

Espinosa, M. and Stanton, C. T. (2022). Training, communications patterns, and spillovers inside organizations. Technical report, National Bureau of Economic Research.

Ewens, M. and Giroud, X. (2025). Corporate hierarchy. Technical report, National Bureau of Economic Research.

Fan, H., Li, Y. A., and Yeaple, S. R. (2015). Trade liberalization, quality, and export prices. *Review of Economics and Statistics*, 97(5):1033–1051.

Friedrich, B. U. (2022). Trade shocks, firm hierarchies, and wage inequality. *Review of Economics and Statistics*, 104(4):652–667.

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

Garicano, L. and Rossi-Hansberg, E. (2006). Organization and inequality in a knowledge economy. *The Quarterly journal of economics*, 121(4):1383–1435.

Garicano, L. and Rossi-Hansberg, E. (2012). Organizing growth. *Journal of Economic Theory*, 147(2):623–656.

Giardili, S., Ramdas, K., and Williams, J. W. (2023). Leadership and productivity: A study of us automobile assembly plants. *Management Science*, 69(3):1500–1517.

Grossman, G. M. and Helpman, E. (1991). Quality ladders and product cycles. *The Quarterly Journal of Economics*, 106(2):557–586.

Grossman, G. M. and Helpman, E. (1994). Endogenous innovation in the theory of growth. *Journal of Economic Perspectives*, 8(1):23–44.

Holweg, M. (2007). The genealogy of lean production. *Journal of operations management*, 25(2):420–437.

Irwin, D. A. and Klenow, P. J. (1994). Learning-by-doing spillovers in the semiconductor industry. *Journal of Political Economy*, 102(6):1200–1227.

Krugman, P. (1979). A model of innovation, technology transfer, and the world distribution of income. *Journal of Political economy*, 87(2):253–266.

Kugler, M. and Verhoogen, E. (2012). Prices, plant size, and product quality. *The Review of Economic Studies*, 79(1):307–339.

Levitt, S. D., List, J. A., and Syverson, C. (2013). Toward an understanding of learning by doing: Evidence from an automobile assembly plant. *Journal of Political Economy*, 121(4):643–681.

Liker, J. (2020). *The Toyota way: 14 management principles from the world's greatest manufacturer*. McGraw-Hill.

MacDuffie, J. P., Sethuraman, K., and Fisher, M. L. (1996). Product variety and manufacturing performance: evidence from the international automotive assembly plant study. *Management science*, 42(3):350–369.

Ohno, T. (2019). *Toyota production system: beyond large-scale production*. Productivity press.

Sandvik, J. J., Saouma, R. E., Seegert, N. T., and Stanton, C. T. (2020). Workplace knowledge flows. *The Quarterly Journal of Economics*, 135(3):1635–1680.

Storn, R. and Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359.

Thompson, P. (2010). Learning by doing. *Handbook of the Economics of Innovation*, 1:429–476.

Verhoogen, E. (2023). Firm-level upgrading in developing countries. *Journal of Economic Literature*, 61(4):1410–1464.

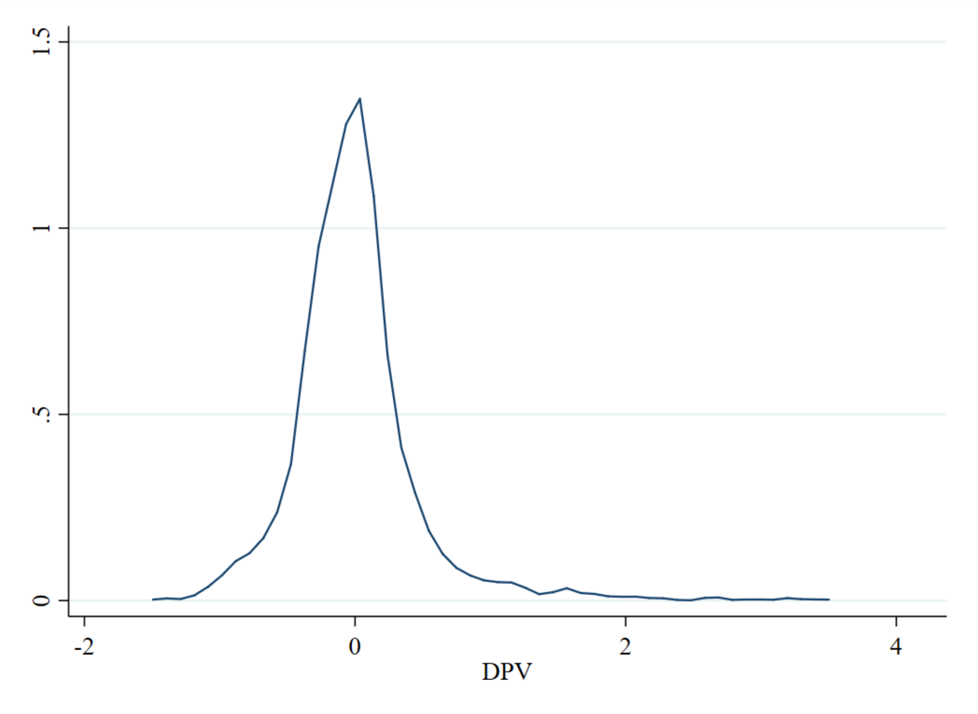
Verhoogen, E. A. (2008). Trade, quality upgrading, and wage inequality in the mexican manufacturing sector. *The Quarterly Journal of Economics*, 123(2):489–530.

Vernon, R. (1966). International investment and international trade in the product cycle. *The Quarterly Journal of Economics*, 80(2):190–207.

Online Appendix

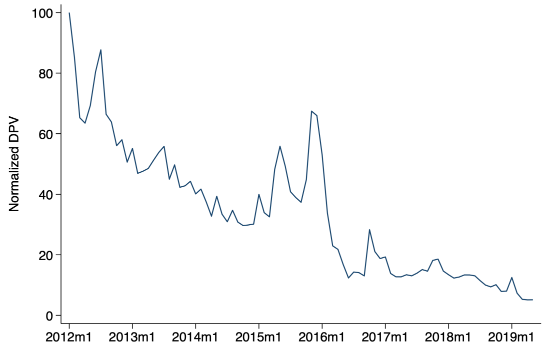
A Additional Figures and Tables

Figure A1: Dispersion in Daily Defects per Vehicle (DPV)

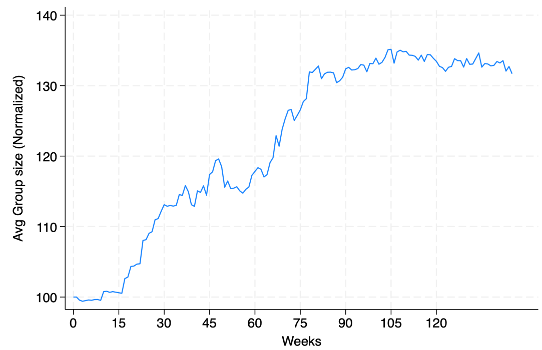


Note: Figure A1 plots the distribution of DPV-day observations, pooling across all days in the data. DPV is a standardized variable with mean 0 and standard deviation 1.

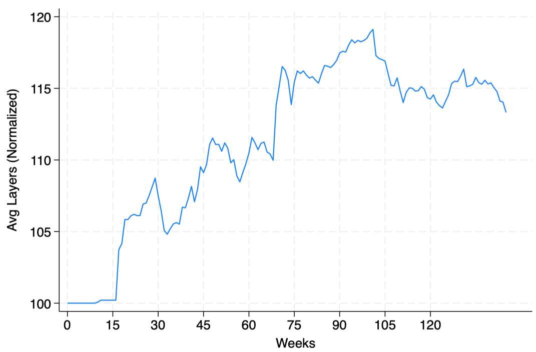
Figure A2: DPV, Average Size of Working Groups, Average Number of Layers and Number of Working Groups over Time



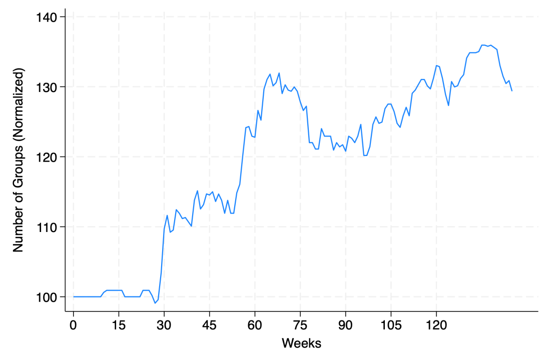
(a) DPV



(b) Average size of working groups



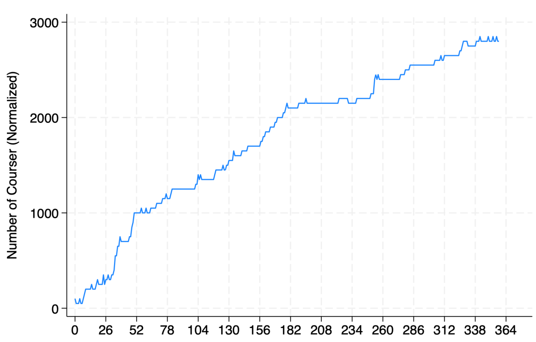
(c) Average number of layers



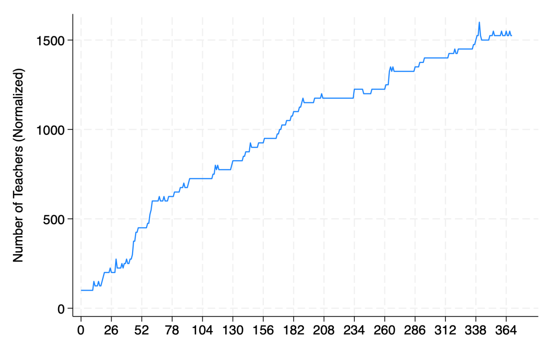
(d) Number of working groups

Note: Figure A2a plots the average monthly DPV from 2012 to 2019, normalized to 100 in January, 2012. Figure A2b plots the weekly average working group size in the period 2017-2019, normalized to 100 in the first week of 2017. Figure A2c plots the weekly average number of layers per working group in the period 2017-2019, normalized to 100 in the first week of 2017. Figure A2d plots the weekly number of working groups in the period 2017-2019, normalized to 100 in the first week of 2017.

Figure A3: Courses and Teachers over Time



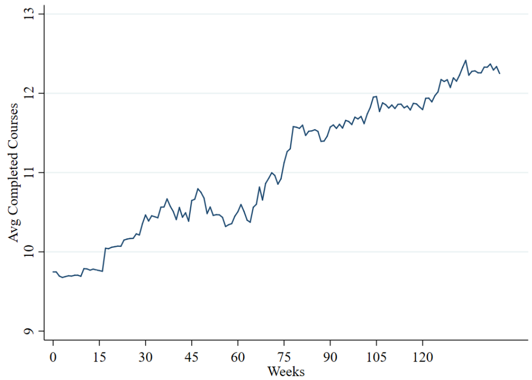
(a) Courses



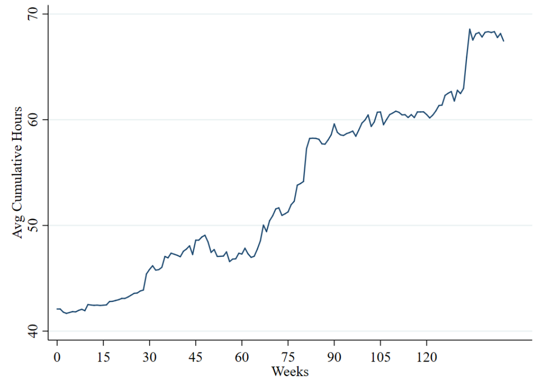
(b) Teachers

Note: Figure A3a plots the weekly number of courses provided during our period of analysis, normalized to 100 in the first week of 2012. Figure A3b plots the weekly number of teachers during our period of analysis, normalized to 100 in the first week of 2012.

Figure A4: Average Completed Programs and Average Cumulative Training Hours over Time



(a) Average Completed Programs



(b) Average Cumulative Training Hours

Note: Figure A4a plots the weekly average completed programs per working group in our period of analysis. Figure A4b plots the weekly average accumulated training hours per working group in our period of analysis.

Table A1: Descriptive Statistics on Model and Volume Changes

	Model Changes				Volume Changes			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Num of Cars	366.55	112.33	170.71	522.40	388.75	92.70	312.70	537.78
Num of Parts	1,421,760.00	834,783.50	1,838.32	2,416,469.00	1,790,505.00	505,012.30	1,425,545.00	2,577,422.00
Num of New Parts	189,047.50	174,787.90	69.96	533,349.20	0	0	0	0
Share of New Parts	0.13	0.11	0.04	0.37	0	0	0	0
Num of Models	1.44	0.74	1.00	3.10	1.26	0.52	1.00	2.19
Num of New Models	1.12	0.10	1.00	1.25	0	0	0	0
Number of Events				7				5

Note: The information presented comes from shift-day level data for the period 2012-2019. The statistics presented for Model and Volume changes are the daily average in the first month after the change happens.

Table A2: DPV and Plant Level Productivity Losses

	(1)	(2)	(3)	(4)	(5)	(6)
	Number of Cars Produced					
High DPV	-64.39***	-50.05***	-49.46***	-50.03***	-49.51***	-47.05***
	(9.989)	(9.005)	(8.958)	(9.080)	(8.978)	(8.971)
L1.High DPV	-25.86***	-14.83**	-14.04**	-14.65**	-14.29**	-10.48
	(8.655)	(7.223)	(7.058)	(7.175)	(7.214)	(6.968)
L2.High DPV	-7.807	2.425	2.654	2.269	2.902	3.799
	(8.384)	(6.909)	(6.769)	(6.881)	(6.904)	(6.619)
L3.High DPV	-20.34**	-10.44	-10.29	-10.63	-10.02	-10.52
	(8.694)	(7.140)	(6.968)	(7.070)	(7.111)	(6.867)
L4.High DPV	-4.106	4.303	4.294	4.085	4.621	5.316
	(9.064)	(7.121)	(6.988)	(7.117)	(7.073)	(6.856)
L5.High DPV	-27.09***	-18.84**	-19.23**	-19.73**	-19.01**	-19.60**
	(10.39)	(8.577)	(8.431)	(8.476)	(8.588)	(8.495)
L6.High DPV	-14.79	-6.584	-6.611	-6.918	-6.311	-4.679
	(9.279)	(7.645)	(7.510)	(7.615)	(7.600)	(7.598)
L7.High DPV	-15.79*	-6.878	-6.930	-7.191	-6.506	-5.949
	(8.546)	(6.734)	(6.724)	(6.776)	(6.771)	(6.673)
L8.High DPV	-11.37	-1.883	-1.763	-2.209	-1.432	-1.007
	(8.955)	(7.211)	(7.118)	(7.204)	(7.218)	(7.105)
L9.High DPV	-1.878	8.970	9.034	8.581	9.286	9.280
	(8.249)	(6.256)	(6.255)	(6.386)	(6.295)	(6.152)
L10.High DPV	-8.369	5.553	5.592	4.840	5.899	4.686
	(7.646)	(6.315)	(6.313)	(6.417)	(6.346)	(6.241)
Observations	1,841	1,841	1,841	1,841	1,841	1,841
Year FE	No	Yes	Yes	Yes	Yes	Yes
Month FE	No	Yes	Yes	Yes	Yes	Yes
Trend	No	No	Cubic	Quadratic	Linear	Linear
Lagged Num of cars	No	No	No	No	No	Yes
R-squared	0.140	0.571	0.579	0.577	0.572	0.576
Cumulative effect	-201.793	-88.258	-86.741	-91.582	-84.376	-76.206

Note: The table reports the estimates from a Distributed Lag Model of order 10 using different sets of controls. To define a High DPV day, we first residualize the daily DPV variable to remove a linear time trend, month FE and year FE. We then define a High DPV day as a day when the residual DPV from this regression goes over 1 SD above the mean residual daily DPV in our sample. In column 1 we estimate the model without controls. In column 2 we include year and month FE. We include year FE, month FE, and a time trend in columns 3 to 5 (non-linear and linear). In column 6, we include year FE, month FE, a linear time trend, and the lagged number of cars.

Table A3: Descriptive Statistics on Employee Suggestions by Layer

	Share of Weekly Suggestions	Weekly Suggestions per Employee	Average Score of Suggestions
S4	0.01%	0.019	10.458
S3	0.83%	0.087	8.617
S2	0.33%	0.041	11.834
S1	0.42%	0.041	12.143
ML4	0.52%	0.350	8.651
ML3	5.63%	0.371	8.929
ML2	8.87%	0.336	8.725
ML1	6.21%	0.278	8.566
Mid FL	10.50%	0.339	7.570
FL4	22.05%	0.320	7.222
FL3	17.95%	0.275	7.062
FL2	12.39%	0.276	6.675
FL1	9.19%	0.203	6.378
New FL3	3.04%	0.113	6.515
New FL2	1.60%	0.102	6.444
New FL1	0.44%	0.060	6.653

Note: Table A3 shows the distribution of weekly suggestions by layer, the number of weekly suggestions per employee by layer, and the average score of suggestions by layer, for the years 2018 and 2019.

Table A4: Average Scores of Logged Suggestions by Grading Categories and Layers

	Total	Costs	Productivity	Security	Quality	Environment	Working conditions	Initiative	Creativity	Effort
S	10.448	9.333	3.223	2.741	3.223	2.320	2.141	2.859	1.465	1.933
ML	8.813	5.780	2.675	2.604	2.947	2.303	1.817	2.566	1.462	1.839
Mid FL	7.595	6.033	2.327	2.450	2.750	2.285	1.749	2.386	1.382	1.673
FL	6.991	3.756	2.238	2.359	2.440	2.208	1.641	2.305	1.328	1.545
New FL	6.498	2.125	2.059	2.173	2.198	2.055	1.374	2.174	1.305	1.471

Notes: The table reports the average score of the logged suggestions, by category of suggestion and layer. A logged suggestion only gets scored on the categories the suggestion relates to (i.e., a suggestion pertaining to costs only gets scored on that dimension), so each logged suggestion is not necessarily scored on all categories. The total score of a suggestion is then the sum of all the scores across all categories it applies to. The main categories that suggestions relate to are the following: Initiative (24%), creativity (21%), effort (20%), security (15%) and working conditions (13%).

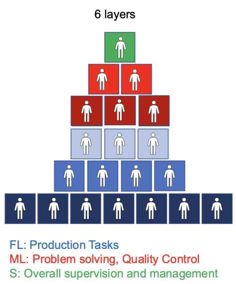
Table A5: Stability of Working Groups

	Mean	SD
Number of working groups per week	131.678	13.319
Number of workers moving between working groups per week	0.411	1.996
Distance in weeks between consecutive movements	6.553	9.019
Number of working groups dissolved per week	0.127	0.466
Number of working groups created per week	0.873	1.443
Duration of working groups (in weeks)	78.396	40.939
Number of working groups with no S per week	0.115	0.319
Number of working groups with no ML per week	0.042	0.2

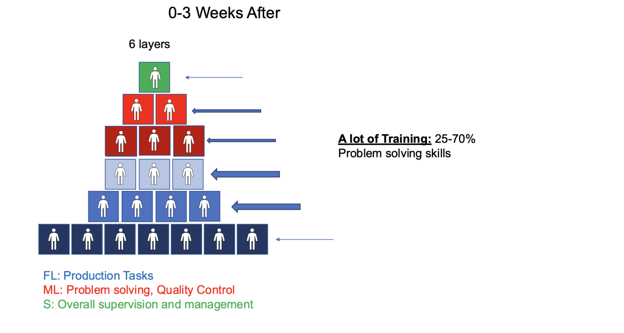
Note: The number of workers moving between working groups per week is computed as the total number of workers that leave the groups where they were for the previous week. The distance between movements is the number of weeks between two consecutive movements of workers from the same working group towards another working group.

A1 Additional Model Change Figures and Tables

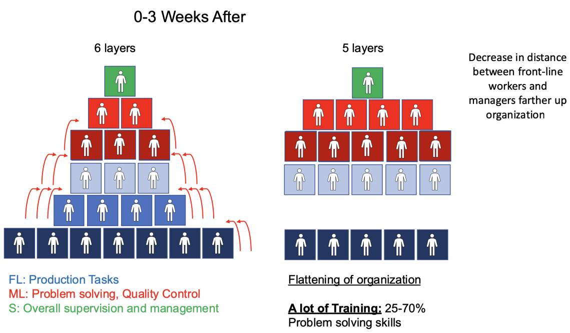
Figure A5: Schematic of Working Group Composition Movement after a Model Change



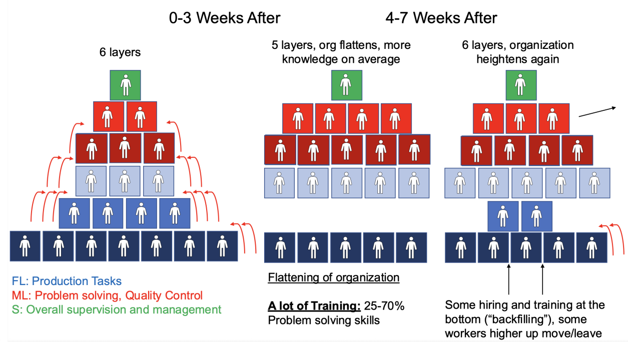
(a) Baseline



(b) Training



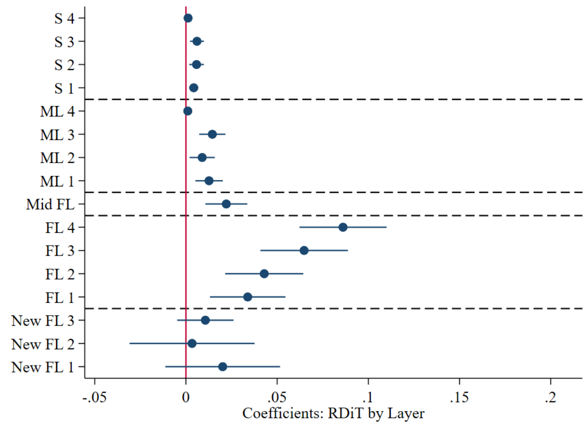
(c) Promotion at the top



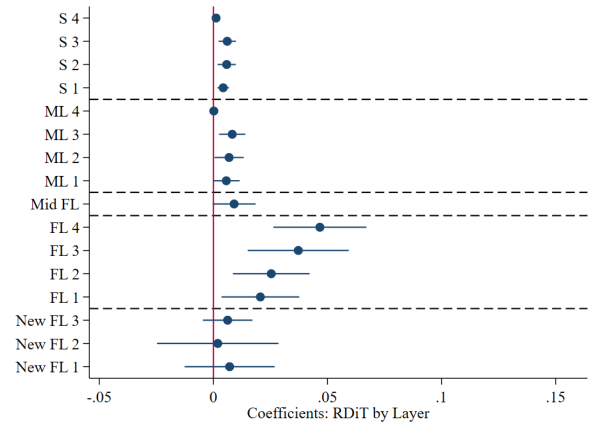
(d) Promotion at the bottom (back-filling)

Figure A5 shows an example of how the working group composition changes after a Model change, illustrating our empirical results. Panel (a) shows a schematic of a typical working group hierarchy before the model change: the working group has 20 workers and six layers, in line with the descriptives on working groups in Table 3. Panel (b) focuses on training investments along the hierarchy after model changes, summarising the results from Figure 5 and A6: the thickness of horizontal arrows represent the amount of training at each layer. Consistent with our empirical results, training investments are highest in the middle-low ranks. Panel (c) summarizes our results on the temporary restructuring of the hierarchy that follows model changes, summarizing the results in Table 5 and Figures 6, A8 and A9. Each red arrow indicates one worker moving layer. It shows that one layer is removed and the hierarchy becomes flatter and more top heavy. Panel (d) summarizes our results on the backfilling that starts to take place 4-7 weeks after model changes, thus summarizing the results from Figures 6, A8 and A9 and Table A10. It shows that the number of layers reverts back to six, and the hierarchy starts to become more pyramidal again. This happens through some hiring and training of workers at the bottom of the hierarchy, some of whom get promoted to backfill. Some workers leave higher layers (either to other working groups or leaving the assembly line), thus maintaining the total number of workers in the working group unchanged at 20.

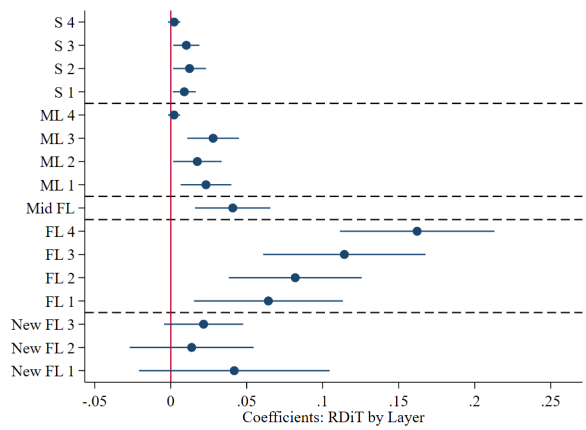
Figure A6: Impact of Model Changes on Avg Num. of Courses and Employees Trained in All Courses and Problem-Solving and Communication Specific Content (0-3 weeks)



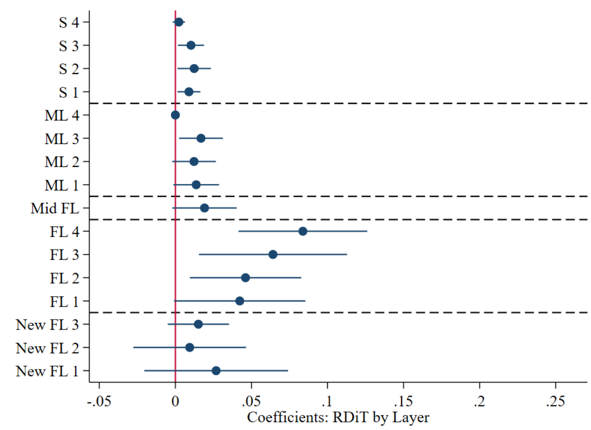
(a) Num. Courses, All courses



(b) Num. Courses, Problem-solving



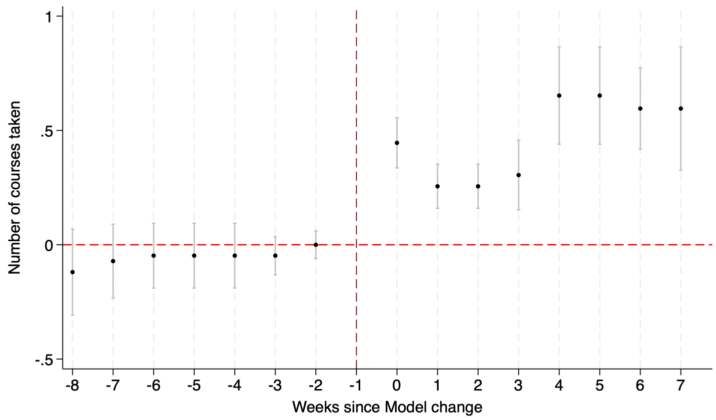
(c) Num. Employees, All courses



(d) Num. Employees, Problem-solving

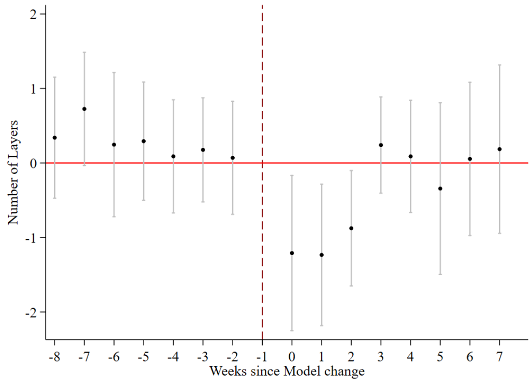
Figures A6a and A6b show the effect of Model changes on the average number of courses taken by workers by layer at 0-3 weeks post Model change in all courses and courses with problem-solving and communication content, respectively. Figures A6c and A6d show the effect of Model changes on the number of employees trained by layer at 0-3 weeks post Model change in all courses and courses with problem-solving and communication content, respectively. For more details on the definition of the layers see Table 1. Each coefficient is estimated from a separate regression. We control for month, year, and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are presented in the figure. Number of observations: 220 working groups x 16 weeks x 2 events.

Figure A7: Event Study of Model Changes on Incremental Training Provision

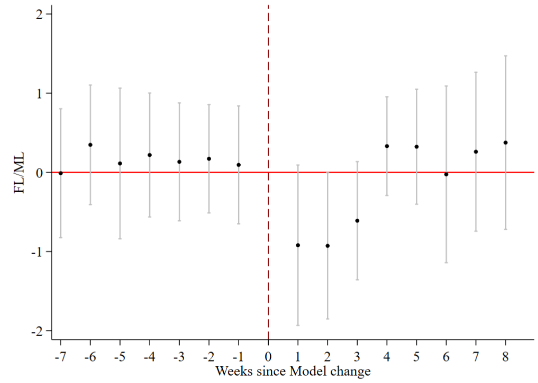


Notes: We estimate equation (1) on the number of courses taken using data between 2017-2019 at the working group-week level. Standard errors are clustered by distance to event and working group level. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

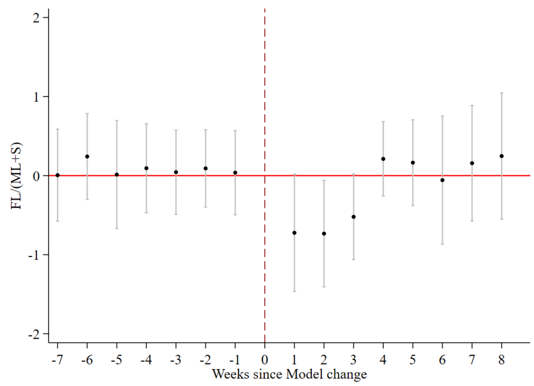
Figure A8: Event Study of Model Changes on Structure of the Hierarchy of Working Groups



(a) Number of Layers



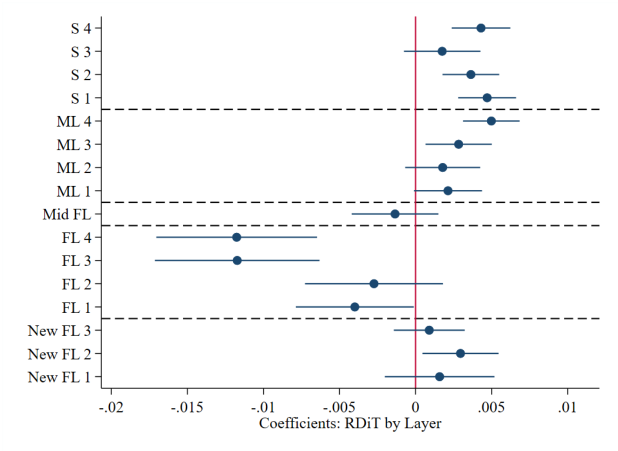
(b) FL/ML



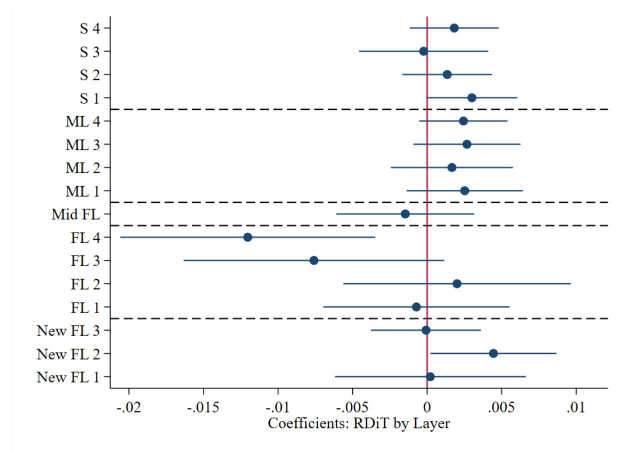
(c) FM/(ML+S)

Notes: This Figure shows the effect of Model changes on the number of layers, FL/ML and FL/(ML+S) ratios within working groups in a time window running from 8 weeks before the event to 8 weeks after the event (where the week of the Model change is labelled as week 0 on the x-axis). We control for month, year, and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes and a linear time trend. Data is for the period 2017-2019 at the working group-week level. Standard errors are clustered by distance to event and working group. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

Figure A9: Impact of Model Changes on Working Group Structure and Knowledge Hierarchies (Disaggregated)



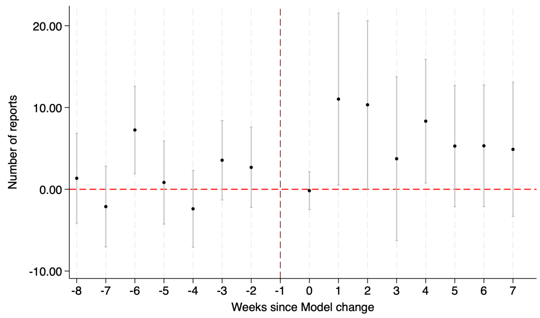
(a) Share of Workers by Layer, 0-3 weeks



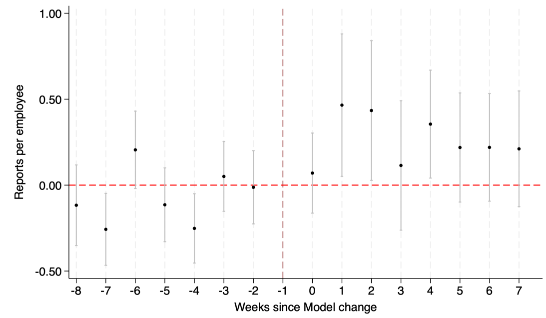
(b) Share of Workers by Layer, 4-7 weeks

Figures A9a and A9b show the effect of Model changes on the share of workers in the working group by layer at 0-3 weeks and 4-7 weeks post-shock, respectively. For more details on the definition of the layers see Table 1. Each coefficient is estimated from a separate regression. We control for month, year, and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are presented in the figure. Number of observations: 220 working groups x 16 weeks x 2 events.

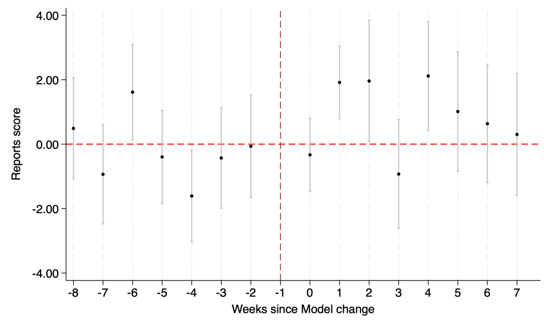
Figure A10: Event Studies of Model Changes on Problem Solving Activity



(a) Number of suggestions



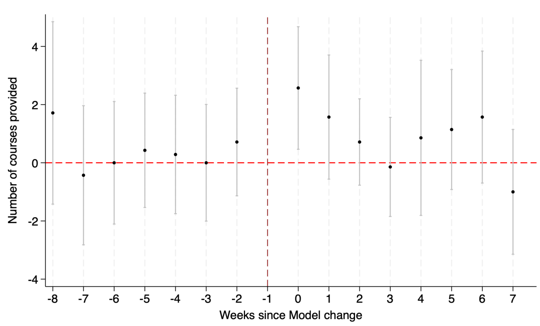
(b) Number of suggestions per employee



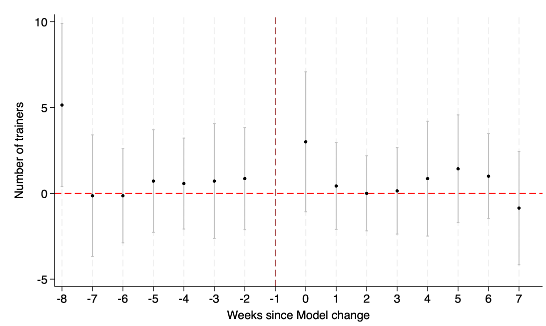
(c) Suggestions score

Notes: We estimate equation (1) on (panel (a)) the Number of suggestions, (panel (b)) Suggestions per employee, and (panel (c)) Suggestions scores for Model Changes using data from 2018 at the working group-week level. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 1 events.

Figure A11: Event Study of Model Changes on Training Investments



(a) Number of courses provided



(b) Number of trainers

Notes: Estimates of equation 1 on the Number of courses and Number of Teachers for Model Changes using data between 2012-2019 at the week level. Standard errors are clustered by distance to the event. 95% confidence intervals are reported. Number of observations: 16 weeks x 7 events.

Table A6: Impact of Model Changes on Production

	(1) Total Cars	(2) Total Parts (Millions)	(3) Share New Parts
0-3 weeks	-0.534	-0.084	30.6***
	(0.703)	(0.175)	(2.16)
4-7 weeks	-1.162	-0.258	13.5***
	(0.875)	(0.195)	(4.78)
Observations	567	567	567
Obs. Level	Day	Day	Day
Mean	105.386	1.423	0

Note: Standard errors clustered by distance to the event. Number of observations: 81 days x 7 Model changes. Total parts are expressed in millions. Share of new parts is the percentage of new parts introduced in each model change relative to those used in the previous variant of the model. Car production is reported by the plant at the daily level. We control for month and year fixed effects as well as a linear function of distance to the Model change and distance to every other Model and Volume change in the data. * p<0.1, ** p<0.05, *** p<0.01

Table A7: Impact of Model Changes on Productivity: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
				DPV		
0 to 3 weeks	0.745***	1.184**	1.047***	1.043***	1.066***	1.552**
	(0.134)	(0.526)	(0.377)	(0.365)	(0.375)	(0.664)
4 to 7 weeks	0.198	0.850	0.542	0.588	0.586	0.829
	(0.186)	(0.689)	(0.440)	(0.441)	(0.440)	(0.743)
Observations	1,134	324	1,296	1,296	1,296	1,296
Obs. Level	Shift-day	Shift-day	Shift-sector-day	Shift-sector-day	Shift-sector-day	Shift-sector-day
Years of data	2012-2019	2017-2019	2017-2019	2017-2019	2017-2019	2017-2019
Experience control	No	No	No	Firm	Working group	No
Defects weighted	No	No	No	No	No	Yes

Note: Standard errors clustered by distance to event and shift. Number of observations for column 1: 2 shifts x 81 days x 7 events. Number of observations for column 2: 2 shifts x 81 days x 2 events. Number of observations for column 3 to 6: 8 shifts-sector x 81 days x 2 events. Productivity measures are reported by the plant at the shift-day level (for the period 2012-17) and shift-sector-day level (for the period 2017-19). DPV is the number of defects per 100 vehicles, and is standardized. We control for a linear function of distance to the Model change and to all other Model and Volume changes, and month, year, shift fixed effects in all specifications. In columns 4 and 5 we control by the average experience of employees working in that shift-sector-day. Experience in the firm is measured as the worker's time at the firm. Experience in a working group is measured as the worker's time in the working group. In column 6 we weight sector-shift observations by the absolute value of the coefficients obtained in table A8. * p<0.1, ** p<0.05, *** p<0.01.

Table A8: DPV in Different Sectors of the Assembly Line and Productivity

	(1) Total Cars
High DPV in Repair	0.385
	(16.86)
High DPV in Chassis	-0.396
	(13.68)
High DPV in Final	-2.901
	(9.177)
High DPV in Trim	-32.10***
	(8.842)
Observations	1,137
R-squared	0.576

Note: We estimate the contemporaneous effect of a High DPV day in the Repair, Chassis, Final, and Trim sectors on the number of cars produced at the daily level. To define a High DPV day, we first residualize the daily DPV variable to remove a linear time trend, month fixed effects and year fixed effects. We then define a High DPV day as a day when the residual DPV from this regression goes over 1 SD above the mean residual daily DPV in our sample for that specific sector. * p<0.1, ** p<0.05, *** p<0.01.

Table A9: Impact of Model Changes on Cumulative Training in Working Groups

	(1) Avg. Completed Programs	(2) Avg. Cumulative Training Hours
0-3 weeks	0.334***	3.077***
	(0.103)	(0.543)
3-7 weeks	0.399**	3.133***
	(0.155)	(0.816)
Observations	7,040	7,040
Obs. Level	Group-Week	Group-Week
Mean	11.488	51.832

Note: Standard errors clustered by distance to the event and working group. Number of observations: 220 working groups x 16 weeks x 2 events. Avg. Completed Programs is the average number of training programs received by the employees in each working group. Avg. Cumulative Training Hours is the average number of training hours received by the employees in each working group. We use as controls month, year and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. * p<0.1, ** p<0.05, *** p<0.01

Table A10: Impact of Model Changes on Employment

	(1) Employment	(2) Hires	(3) Separations
0-3 weeks	3.894	-1.172	-0.367
	(21.73)	(4.855)	(0.758)
4-7 weeks	10.85	15.30	0.720
	(40.86)	(12.46)	(1.260)
Observations	64	64	64
Obs. Level	Shift-Week	Shift-Week	Shift-Week
Mean	1175	21	2

Note: Standard errors clustered by distance to the event and shift. Number of observations: 2 shifts x 16 weeks x 2 events. For more details on the definition of the layers see Table 1. We control for month, year, and shift fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A11: Impact of Model Changes on Number of Groups, Group Size, and Group Stability

	(1) Number of Groups	(2) Group Size	(3) Any New Group	(4) Share of Workers Moving between Groups
0-3 weeks	-0.786	0.039	-0.241	-0.0394
	(1.703)	(0.240)	(0.289)	(0.0264)
4-7 weeks	-0.789	0.105	-0.523	-0.0525
	(2.304)	(0.349)	(0.365)	(0.0383)
Observations	64	64	64	64
Obs. Level	Shift-Week	Shift-Week	Shift-Week	Shift-Week
Mean	58	20	0.4063	0.0284

Note: Standard errors clustered by distance to event and shift. Number of observations: 2 shifts x 16 weeks x 2 events. We use as controls month, year, and shift fixed effects. We control for a linear function of distance to Model change and distances to the other Model and Volume changes. Number of groups is the number of working groups in each shift-week. Group size is the average number of employees in each working group. Any new group is a dummy 1 if at least one working group is created in that shift-week. The share of workers is computed as the total number of moving works over the total number of workers in that shift-week. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A12: Heterogeneous Effects of Model Changes on DPV by Plant Experience with Model Changes

	(1)	(2)
Plant experience measure →	Order of Model Change	Dummy for Late Change
0 to 3 weeks	0.722***	0.591***
	(0.250)	(0.159)
4 to 7 weeks	-0.148	0.0392
	(0.272)	(0.208)
0 to 3 weeks × Plant experience	-0.104**	-0.674**
	(0.0491)	(0.261)
4 to 7 weeks × Plant experience	0.0398	-0.345
	(0.0485)	(0.226)
0 to 3 weeks × High share of new parts	1.061***	1.041***
	(0.174)	(0.223)
4 to 7 weeks × High share of new parts	0.686***	0.968***
	(0.131)	(0.134)
Observations	1,134	1,134

Notes: We estimate equation (2) on DPV for Model changes using data for 2012-2019 at the shift-day level, adding an interaction with a linear variable capturing the order of the Model change, with higher values indicating that the Model change happened later in our sample period (column 1) or with a dummy for whether the Model change happened in the second half of our sample period, i.e., after 2015, (column 2), and always controlling by whether the Model change has a share of new parts above the median, to control for the size of the Model change in terms of number of new parts and isolate the role of whether the Model change happened early or late in the sample period.

Table A13: Correlation between Separations and Decrease in Layers (First-stage)

	High Decrease Layers
High Separations	0.686***
	(0.180)
Observations	16
F-stat	14.472

Notes: The Table show the linear regression (first stage) between the dummy of High Decrease in Layers, which takes value 1 if the average decrease in layers in the three weeks after the Model change is higher than the median change, and 0 otherwise, against the dummy of High Separations, which takes value 1 if the average number of separations in the three weeks after the Model change is higher than the median number of separations. All the variables are measured at the shift-sector-event level, and there are 2 shifts, 4 sectors, and 2 events (i.e., Model changes). Robust standard errors reported in parenthesis. * p<0.1, ** p<0.05, *** p<0.01.

Table A14: Difference in Difference Results by Share of New Models

	(1) DPV	(2) Avg. Completed Programs	(3) Layers
0 to 3 weeks	0.849**	0.231***	-0.511***
	(0.339)	(0.0880)	(0.148)
4 to 7 weeks	0.506	0.0842	-0.249
	(0.425)	(0.133)	(0.216)
0 to 3 weeks × High share of new models	1.102*	0.202**	-0.346**
	(0.608)	(0.0915)	(0.158)
4 to 7 weeks × High share of new models	0.175	0.157	-0.204
	(0.397)	(0.119)	(0.200)
Observations	1,296	7,040	7,040
0 to 3 weeks + 0 to 3 Interaction	1.951	0.432	-0.857
	(0.721)	(0.084)	(0.143)
4 to 7 weeks + 4 to 7 Interaction	0.678	0.241	-0.453
	(0.588)	(0.123)	(0.201)

Notes: estimates of equation (2), where the 0-3 weeks and 4-7 weeks dummies are interacted with indicators for shifts where an above the median share of new models are produced. We use data between 2017–2019 at the shift-sector-week level for the defects per vehicle outcome (column 1), and at the working group-week level for the average completed training programs outcome (column 2) and the number of layers outcome (column 3). The effects on all these outcomes are significantly larger for shifts producing an above the median share of new models. * p<0.1, ** p<0.05, *** p<0.01.

Table A15: Effect of Model Changes on Alternative Definition of Layers

	(1) Num. Layers (Alt. Definition)
0 to 3 weeks	-0.573***
	(0.112)
4 to 7 weeks	-0.216
	(0.174)
Observations	7,040
Mean	5.911

Notes: Standard errors clustered by distance to event and working group. Number of observations: 220 working groups x 16 weeks x 2 events. Number of Layers is defined as the number of separate positions present in a working group with different wages. We use as controls month, year and group fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes. * p<0.1, ** p<0.05, *** p<0.01

A2 Organizational Responses to Volume Changes

We first confirm in Appendix Table A16 that Volume changes result in a sudden and sizable increase in the planned number of cars to be produced and, consequently, in the number of parts that have to be assembled. The total number of cars planned per day jumps up by about 9% and the number of parts increases by 7% after the Volume change; and these effects on production are long-lasting. Since by definition the model produced on either side of Volume changes does not change, the types of parts that have to be assembled are the same as before. That is, the share of new parts is consistently 0 through this window so we do not analyze this outcome for Volume changes.

The change in volume leads to a sudden and large increase in defects per vehicle, as shown in Figure A13. DPV shoots up by about 3 SD on average in the immediate aftermath of the change, and then comes down to pre-Volume change levels in about three weeks on average. The similarity with Figure 2 is remarkable: both Volume and Model changes lead to a similar short-term reduction in productivity.[64] Appendix Figure A14 shows that the initial spike in DPV is fairly similar, but the length of time over which DPV remains elevated is longer for Volume changes for which the factory hires more new workers. This pattern of heterogeneity supports the interpretation that the rise in DPV following Volume changes is due to the restructuring related to expanding employment and production. That is, the spike in DPV after Volume changes derives from a capacity rather than a complexity challenge.[65]

[64] Appendix Table A17 shows the corresponding discontinuity in time estimates (column 1). It also shows that this pattern holds when we restrict attention to the period 2017 to 2019 coinciding with the working group composition database (columns 2 and 3).

[65] Appendix Table A24 shows that the impact of Volume changes on DPV is significantly lower in working groups where workers are more experienced, thus suggesting that learning by doing retained over time plays

However, the organizational response of the firm to the increase in the volume of production is very different. Figure A15 reports the results of an event study specification following equation 1 with the average number of cumulative completed courses per worker as the outcome. The figure shows that in response to the Volume changes, the average level of cumulative training (and therefore knowledge) in each working group *decreases* immediately and persistently.[66]

We next ask how the decrease in the average knowledge is generated. Table A18 shows that as a response to Volume changes the firm hires about 24 additional workers per shift-week in the first four weeks after the Volume change (column 2) while separations do not increase significantly (column 3). The result is that overall shift-level employment increases by about 80 workers over the two months following the Volume change, or a 7.5% increase over a mean of about 1,100 workers per shift (column 1). This hiring is concentrated among front-line layers with a low level of skill. Column 1 of Table A19 confirms that the way this knowledge reduction is being achieved is by adding layers and increasing the relative size of the base of the pyramid and consequently the span of control of complex problem-solvers (column 2 and 3 of Table A19).[67] That is, as a result of Volume changes, the firm moves to a more pronounced pyramidal structure of working groups with a thicker base and a larger number of separate knowledge layers, resulting in an overall reduction in the average level of knowledge of working groups. This increase is persistent for at least 8 weeks after the volume change, which is consistent with the persistent increase in the volume that needs to be produced.[68]

Finally, Panels A and B of Table A20 show a reduction in the number of suggestions made by workers, especially FL workers. This is consistent with the fact that there are now more FLs in working groups and the average skill level of workers is lower. This result also

a more important role in reducing DPV for Volume changes, where more of the same needs to be produced, than for Model changes, where new tasks need to be learned. Table A17, columns 4 and 5, further show that controlling for worker experience does not alter our estimates of the effect of Volume changes on DPV.

[66] Appendix Table A21 reports the corresponding regression discontinuity in time results and shows that Volume changes lead to a persistent reduction in average completed training programs of about 7% and a reduction in average cumulative training hours of about 11% after eight weeks.

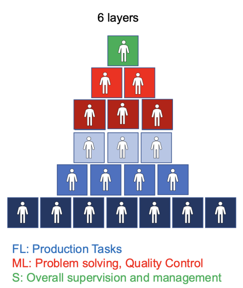
[67] Appendix Figure A18 reports the corresponding event studies on employment, showing that Volume changes lead to a persistent increase in employment over at least an eight-week period. Appendix Table A22 shows that the increase in employment is associated with an increase in the number of working groups in the Assembly sector, but not with a significant increase in the size of pre-existing working groups. In addition, Appendix Table A23 shows that despite the increase in employment, Volume changes still lead to an increase in the number of cars and number of parts per employee as well as per working group. This then justifies why we see an increase in the number of layers within working groups: the firm reorganizes production to deal with the increase in the volume of work per employee, achieving greater economies of scale. In line with this, we still find a positive and persistent effect on the number of layers when restricting the sample to pre-existing working groups only (Appendix Figure A17).

[68] Appendix Figure A16 reports the corresponding event study for the number of layers. Appendix Figure A19 reports impacts of Volume changes on the share of workers by layer, confirming that the share of FL workers in working groups increases, so that groups become heavier in the lower part of the pyramid.

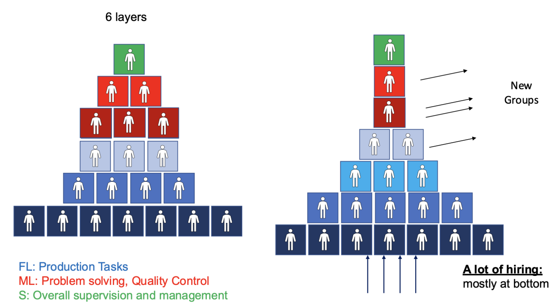
reflects a gain in economies of scale by way of an increase in the span of control of MLs and Ss and possibly some learning in the cumulative quantity produced as well (Levitt et al., 2013). In line with the stock of knowledge of working groups being now lower, we find that the average score of suggestions decreases after a Volume change (Panel C). This comparison again highlights how different and unique are organizational responses to product upgrading vis a vis responses to Volume changes.

A2.1 Volume Change Figures and Tables

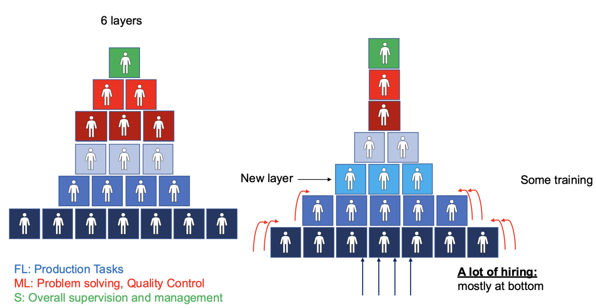
Figure A12: Schematic of Working Group Composition Movement after a Volume Change



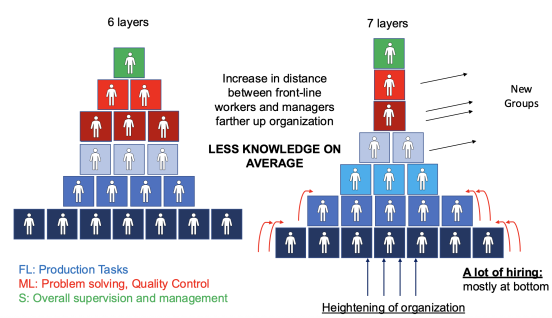
(a) Baseline



(b) Hiring and new groups formation



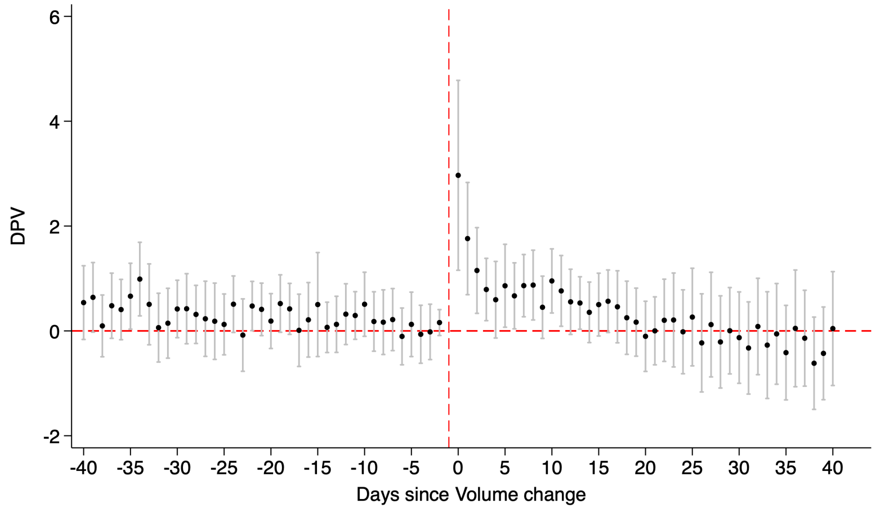
(c) Promotion at the bottom



(d) Less knowledge on average

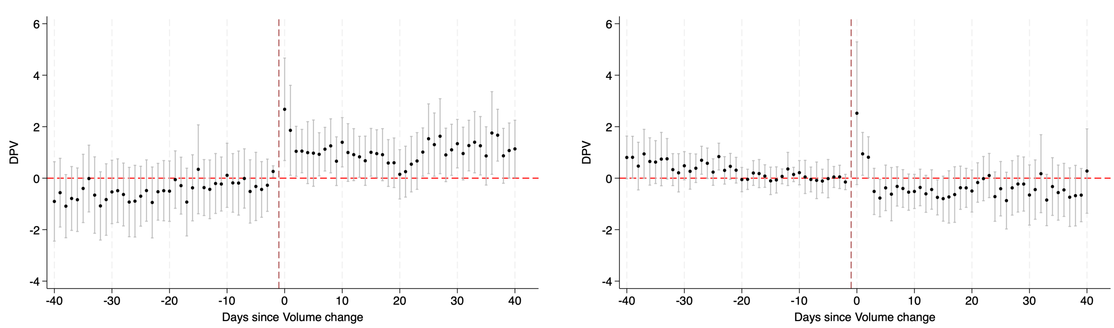
Figure A12 shows an example of how the working group composition changes after a Volume change, illustrating our empirical results. Panel (a) shows a schematic of a typical working group hierarchy before the volume change: the working group has 20 workers and six layers, in line with the descriptives on working groups in Table 3. Panel (b) focuses on hiring after model changes, summarising the results from Table A18 and Table A22: each arrow represents a worker hired or a worker moving to other working groups (as detailed in the caption). Consistent with our empirical results, we show that the plant hires heavily at the bottom, and then forms new working groups, with some higher level workers leaving old working groups to join new groups, so that total employment and the number of working groups increase but the size of working groups is unchanged. Panel (c) summarizes our results on the restructuring of the hierarchy that follows volume changes, summarizing the results in Table A19 and Figure A19. Each red arrow indicates one worker moving layer. It shows that one layer is added in the lower part of the hierarchy through some training and promotion at the bottom, and the hierarchy becomes more pyramidal and bottom heavy as a result. Panel (d) summarizes our results on knowledge levels from Figure A15. It shows that as the hierarchy is more bottom heavy, this corresponds to lower average levels of accumulated training and therefore knowledge. Note that in contrast to model changes, volume changes lead to a more persistent restructuring of the hierarchy (i.e., we find no evidence of the hierarchy reverting back up to 8 weeks after volume changes).

Figure A13: Event Study of Volume Changes on Defects per Vehicle



Note: Figure A13 shows the effect of Volume changes on DPV in a time window running from 40 days before the event to 40 days after the event. DPV is computed as number of defects per 100 vehicles, and is standardized using the mean and standard deviation of the full sample. We control for month, year, and shift fixed effects. We also control for a linear function of distance to the Volume change and to all other Volume and Model changes and a time linear-trend. Standard errors are clustered by distance to the event and shift. 95% confidence intervals are reported. Number of observations: 2 shifts x 81 days x 5 events.

Figure A14: Event Study of Volume Changes on Productivity by Hiring

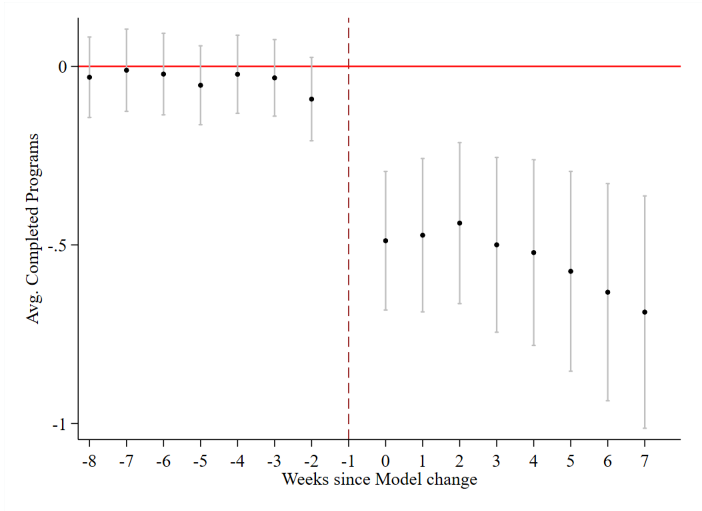


(a) High Hiring

(b) Low Hiring

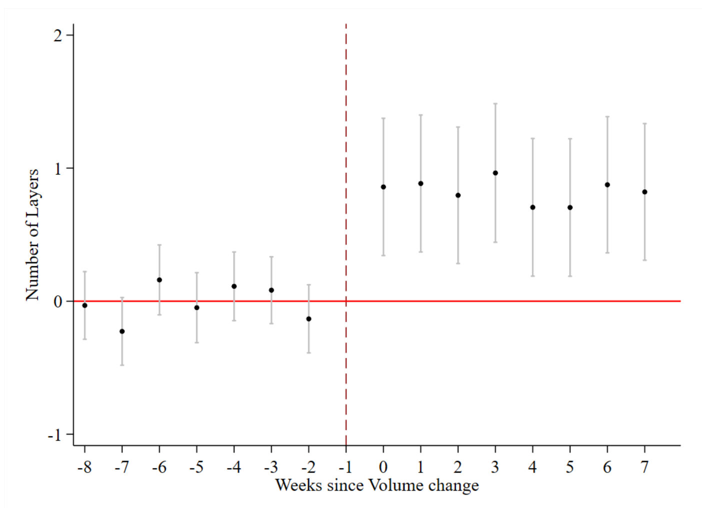
Note: Figure A14 shows the effect of Volume changes on DPV in a time window running from 40 days before the event to 40 days after the event split between changes with high number of employees hired (above the median) and changes with low number of employees hired (below the median). DPV is computed as number of defects per 100 vehicles, and is standardized using the mean and standard deviation of the full sample. We control for month, year, and shift fixed effects. We also control for a linear function of distance to the Model change and to all other Model and Volume changes and a linear time trend. Standard errors are clustered by distance to the event and shift. 95% confidence intervals are reported. Number of observations in Panel (a): 2 shifts x 81 days x 2 events. Number of observations in Panel (b): 2 shifts x 81 days x 3 events.

Figure A15: Event Study of Volume Changes on Cumulative Training in Working Groups



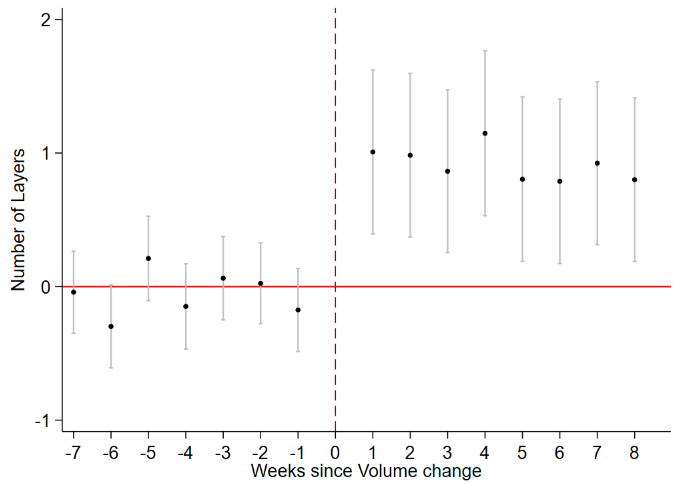
Note: Figure A15 shows the effect of Volume changes on the average completed training programs within working groups in a time window from 8 weeks before the event to 8 weeks after the event (where the week of the Model change is labelled as week 0 on the x-axis). We control for month, year, and group fixed effects. We also control for a linear function of distance to the Volume change and to all other Model and Volume changes and a linear time trend. Standard errors are clustered by the distance to the event and working group. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

Figure A16: Event Study of Volume Changes on Layers



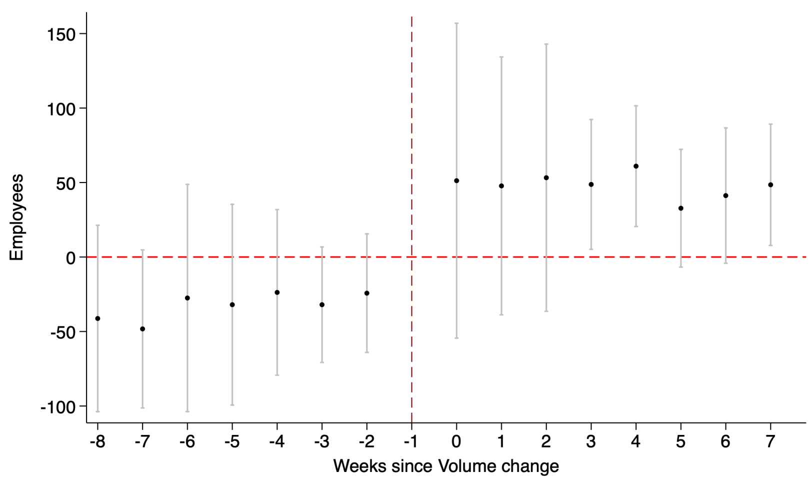
Note: Figure A16 shows the effect of Volume changes on the number of layers within working groups in a time window running from 8 weeks before the event to 8 weeks after the event (where the week of the Volume change is labelled as week 0 on the x-axis). Number of layers is defined as the number of separate positions present in a working group. We control for month, year, and group fixed effects. We also control for a linear function of distance to the Volume change and to all other Volume and Model changes and a linear time trend. Standard errors are clustered by the distance to the event and working group. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

Figure A17: Event Study of Volume Changes on Layers for Pre-Existing Groups



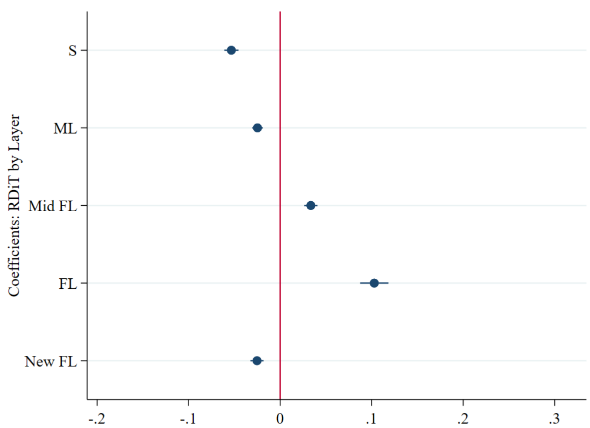
This Figure shows the effect of volume changes from 8 weeks before the event to 8 weeks after the event on the number of layers, limiting the sample to pre-existing working groups only. Number of layers is defined as the number of positions in the working group. Standard errors clustered at weekly-shift level. 95% confidence intervals are presented in the figure. We control for month, year, and group fixed effects. We also control for a linear function of distance to the Volume change and to all other Volume and Model changes and a time linear-trend. Number of observations: 167 working groups x 16 weeks x 2 events.

Figure A18: Event Study of Volume Changes on Employment

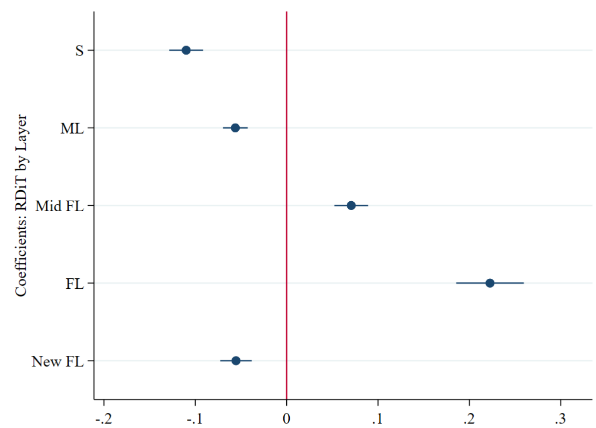


Note: Figure A18 shows the effect of volume changes on the number of employees in a time window running from 8 weeks before the event to 8 weeks after the event (where the week of the Model change is labelled as week 0 on the x-axis). We control for month, year, and shift fixed effects. We also control for a linear function of distance to the volume change and to all other Model and Volume changes and a time linear-trend and a time linear-trend. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are reported. Number of observations: 2 shifts x 16 weeks x 2 events.

Figure A19: Impact of Volume Changes on Group Structure and Knowledge Hierarchies



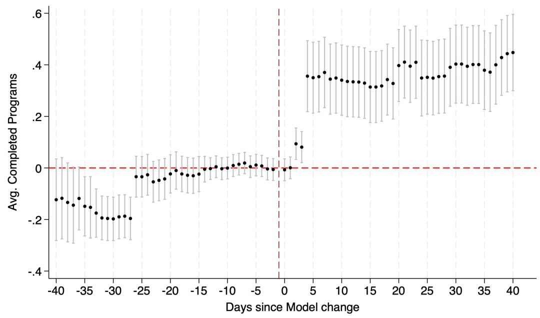
(a) Share of Workers by Layer, 0-3 weeks



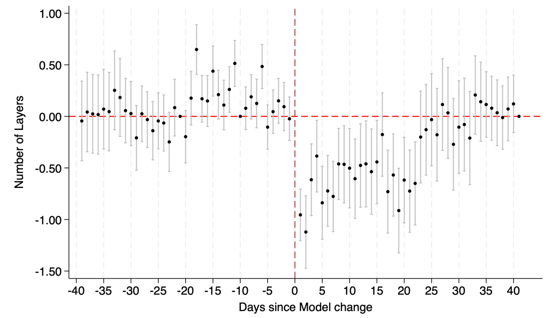
(b) Share of Workers by Layer, 4-7 weeks

Figures A19a and A19b show the effect of Volume changes on the share of workers in the working group by layer at 0-3 weeks and 4-7 weeks post-Volume change, respectively. For more details on the definition of the layers see Table 1. Each coefficient is from a separate regression. We control for month, year, and group fixed effects. We also control for a linear function of distance to the Volume change and to all other Model and Volume changes. Standard errors are clustered by distance to the event and working group. 95% confidence intervals are shown. Number of observations: 220 working groups x 16 weeks x 2 events.

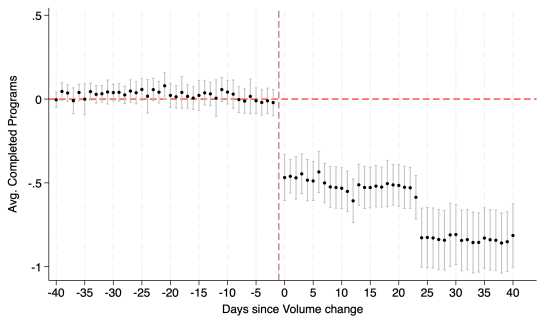
Figure A20: Event Studies of Model and Volume Changes at the Daily Level



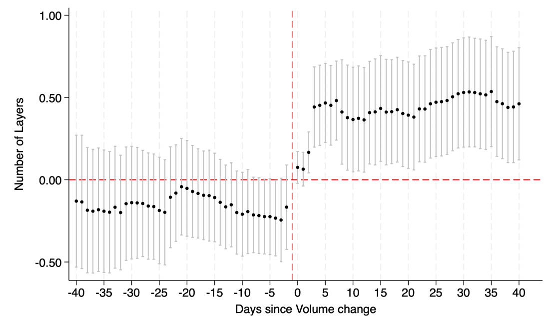
(a) Average completed courses—Model



(b) Number of layers—Model



(c) Average completed courses—Volume



(d) Number of layers—Volume

Notes: In panels (a) and (b), we estimate equation (1) on Avg Completed Training Programs and Number of Layers for Model Changes using data between 2017-2019 at working group-day level. In panels (c) and (d) we estimate equation (1) on the same two outcomes but for Volume Changes using data between 2017-2019 at working group-day level. Standard errors are clustered by distance to event-working group level. 95% confidence intervals are reported.

Table A16: Impact of Volume Changes on Production

	(1) Total Cars	(2) Total Parts
0-3 weeks	30.65***	0.120**
	(10.55)	(0.047)
4-7 weeks	57.97***	0.237***
	(18.00)	(0.079)
Observations	405	405
Obs. Level	Day	Day
Mean	356.84	1.713

Note: Standard errors clustered by distance to the event. Number of observations: 81 days x 5 events. We use as controls month and year fixed effects. Total Parts in millions. Cars production is reported by the plant at daily level. We control for a linear function of distance to the Volume change and distances to the other Volume and Model changes. * p<0.1, ** p<0.05, *** p<0.01

Table A17: Impact of Volume Changes on Productivity: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
				DPV		
0 to 3 weeks	0.696***	2.206**	1.163***	1.165***	1.184***	1.573**
	(0.167)	(0.867)	(0.401)	(0.398)	(0.400)	(0.733)
4 to 7 weeks	0.319	1.952**	0.806*	0.806*	0.855*	1.379*
	(0.201)	(0.911)	(0.444)	(0.441)	(0.439)	(0.800)
Observations	810	324	1,296	1,296	1,296	1,296
Obs. Level	Shift-day	Shift-day	Shift-sector-day	Shift-sector-day	Shift-sector-day	Shift-sector-day
Years of data	2012-2019	2017-2019	2017-2019	2017-2019	2017-2019	2017-2019
Experience control	No	No	No	Firm	Working group	No
Defects weighted	No	No	No	No	No	Yes

Note: Standard errors clustered by distance to event and shift. Number of observations for column 1: 2 shifts x 81 days x 5 events. Number of observations for column 2: 2 shifts x 81 days x 2 events. Number of observations for column 3 to 6: 8 shifts-sector x 81 days x 2 events. DPV measures are reported by the plant at the shift-day level (for the period 2012-17) and shift-sector-day level (for the period 2017-19). DPV is the number of defects per 100 vehicles, and is standardized. We control for a linear function of distance to the Volume change and to all other Model and Volume changes, and month, year, shift fixed effects in all specifications. In columns 4 and 5 we control for the average experience of employees working in that shift-sector-day. Experience in the firm is measured as the worker's time at the firm. Experience in a working group is measured as the worker's time in the working group. In column 6 we weight the observations by the absolute value of the coefficients obtained in table A8. * p<0.1, ** p<0.05, *** p<0.01.

Table A18: Impact of Volume Changes on Employment

	(1) Employment	(2) Hires	(3) Separations
0 to 3 weeks	29.46	24.08**	1.67
	(34.32)	(9.63)	(1.50)
4 to 7 weeks	82.08***	-0.58	0.25
	(28.60)	(2.48)	(0.36)
Observations	64	64	64
Obs. Level	Shift-Week	Shift-Week	Shift-Week
Mean	1094	10	1

Note: Standard errors clustered by distance to the event and shift. Number of observations: 2 shifts x 16 weeks x 2 events. We use as controls month, year, and shift fixed effects. We control for a linear function of distance to Volume changes and distances to the other Volume and Model changes. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A19: Impact of Volume Changes on Working Group Structure

	(1) Num. Layers	(2) FL/ML	(3) FL/(ML+S)
0 to 3 weeks	0.850***	0.610***	0.453***
	(0.098)	(0.094)	(0.067)
4 to 7 weeks	0.706***	0.498***	0.398***
	(0.117)	(0.116)	(0.082)
Observations	7,040	7,040	7,040
Obs. Level	Group-Week	Group-Week	Group-Week
Mean	5.284	4.158	3.094

Note: Standard errors clustered by distance to event and working group. Number of observations: 220 working groups x 16 weeks x 2 events. Number of Layers is defined as the number of separate positions present in a working group. FL/ML is the ratio between Num. of Front-Line workers and Mid-Line workers. FL/(ML+S) is the ratio between Num. of Front-Line workers and Mid-Line workers plus Superiors. We use as controls month, year and group fixed effects. We also control for a linear function of distance to the Volume change and to all other Model and Volume changes. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A20: Impact of Volume Changes on Problem-Solving Activity by Layer

	(1) All	(2) FL	(3) ML	(4) S
		Panel A: Number of suggestions		
0 to 3 weeks	-3.548***	-2.949***	-0.547***	-0.0522**
	(0.429)	(0.363)	(0.0970)	(0.0220)
4 to 7 weeks	-1.143**	-1.169**	0.00344	0.0230
	(0.567)	(0.484)	(0.157)	(0.0554)
Mean	2.476	2.005	0.445	0.026
		Panel B: Number of suggestions per employee		
0 to 3 weeks	-0.270***	-0.268***	-0.306***	-0.111**
	(0.0501)	(0.0558)	(0.0871)	(0.0469)
4 to 7 weeks	-0.0769	-0.0974*	-0.0416	0.0372
	(0.0497)	(0.0557)	(0.103)	(0.117)
Mean	0.214	0.216	0.273	0.055
		Panel C: Score of suggestions		
0 to 3 weeks	-1.068***	-1.019***	-1.470***	-0.171*
	(0.237)	(0.221)	(0.224)	(0.0915)
4 to 7 weeks	0.0857	-0.0753	0.162	-0.00388
	(0.299)	(0.283)	(0.351)	(0.154)
Mean	7.244	6.913	8.526	9.798
Observations	3,520	3,520	3,520	3,520
Obs. Level	Group-Week	Group-Week	Group-Week	Group-Week

Note: Standard errors clustered by distance to the event and working group. Number of observations: 220 working groups x 16 weeks x 1 event. We control for month, year, and shift fixed effects. We also control for a linear function of distance to the Volume change and to all other Model and Volume changes. * p<0.1, ** p<0.05, *** p<0.01

Table A21: Impact of Volume Changes on Stock of Knowledge of Working Groups

	(1) Avg. Completed Programs	(2) Avg. Cumulative Training Hours
0-3 weeks	-0.660***	-4.742***
	(0.145)	(0.793)
4-7 weeks	-0.825***	-6.734***
	(0.213)	(1.168)
Observations	7,040	7,040
Obs. Level	Group-Week	Group-Week
Mean	12.597	59.173

Note: Standard errors clustered by distance to the event and working group level. Number of observations: 220 working groups x 16 weeks x 2 events. Avg. Completed Programs is the average number of training programs received by the employees in each working group. Avg. Cumulative Training Hours is the average number of training hours received by the employees in each working group. We use as controls month, year and group fixed effects. We control for a linear function of distance to Volume changes and distances to the other Volume and Model changes. * p<0.1, ** p<0.05, *** p<0.01

Table A22: Impact of Volume Changes on Number of Groups and Group Size

	(1) Num of Groups	(2) New Group Size	(3) Old Group Size
0-3 weeks	3.765**	21.074***	1.724
	(1.737)	(0.669)	(2.805)
4-7 weeks	3.624*	21.104***	1.253
	(1.949)	(0.714)	(4.124)
Observations	64	64	64
Obs. Level	Shift-Week	Shift-Week	Shift-Week
Mean	54	0	19

Note: Standard errors clustered by distance to the event and shift. Number of observations: 2 shifts x 16 weeks x 2 events. We use as controls month, year, and shift fixed effects. We control for a linear function of distance to the Volume change and distances to the other Volume and Model changes. Number of groups is the number of working groups in each shift. Group size is the average number of employees in each working group. * p<0.1, ** p<0.05, *** p<0.01

Table A23: Impact of Volume Changes on Cars per Employee/Group and Parts per Employee/Group

	(1) Cars per Emp	(2) Cars per Group	(3) Parts per Emp	(4) Parts per Group
0-3 weeks	0.0371***	0.890***	194.4***	4,628***
	(0.00747)	(0.133)	(37.18)	(660.0)
4-7 weeks	0.0471***	1.058***	249.4***	5,583***
	(0.00942)	(0.180)	(46.37)	(881.3)
Observations	162	162	162	162
Obs. Level	Day	Day	Day	Day
Mean	0.256	5.025	1211.198	23750.390

Note: Standard errors clustered by distance to the event. Number of observations: 81 days x 2 events. Cars per Emp: number of cars produced over number of workers by day. Cars per group: number of cars produced over number of working groups per day. Parts per Emp: number of parts used in produced cars over the number of employees by day. Parts per Group: number of parts used in produced cars over the number of working groups by day. We use as controls month and year fixed effects. We control for a linear function of distance to the Volume change and distances to the other Volume and Model changes. * p<0.1, ** p<0.05, *** p<0.01

Table A24: Heterogeneous Effect of Volume Changes on DPV by Experience

	(1) Exp in firm	(2) Exp in working group
0 to 3 weeks	0.792***	1.824***
	(0.261)	(0.689)
4 to 7 weeks	0.566**	1.645**
	(0.277)	(0.696)
0 to 3 weeks × Dummy above the median	-0.347***	-1.595**
	(0.131)	(0.649)
4 to 7 weeks × Dummy above the median	0.00179	-1.485**
	(0.0783)	(0.669)
Observations	1,296	1,296
0 to 3 weeks + 0 to 3 Interaction	0.445	0.229
	(0.273)	(0.132)
4 to 7 weeks + 4 to 7 Interaction	0.568	0.160
	(0.284)	(0.177)

Notes: Estimates of equation (2) on DPV for Volume changes using data between 2017-2019 at shift-sector-day level interacted with dummies for whether the sector-shift has a level of experience above the median (High). Experience in firm is measured as the workers' average time at the firm in the three weeks before the Volume change, Experience in the working group is measured as the workers' average time in the working group in three weeks before the Volume change.

B Model

B1 Organizational Structure and Dynamic Behavior

This appendix provides the technical foundations for the analysis in Section 2, building on the framework of Garicano and Rossi-Hansberg (2012). We first characterize the firm's static equilibrium governing the firm's structure and its implications on the distribution of workers and internal problem prices. We then examine how exogenous changes in training costs and technology affect organizational design. The appendix also presents the dynamic adjustment rules for training and hierarchy, and concludes with additional results on the firm's dynamic optimization problem, including the endogenous timing of innovation and its implications.

B1.1 Equilibrium Structure

The organizational equilibrium is characterized by net earnings equalization across layers, formally stated in equation 10. Following Garicano and Rossi-Hansberg (2012), this condition guarantees that agents are indifferent across hierarchical roles, thereby supporting the endogenous allocation of positive effort and knowledge acquisition at every level.

The equilibrium also determines the number of managers at each layer. Specifically, the number of agents at layer ℓ is given by:

$$n_L^\ell = he^{-\lambda Z_L^{\ell-1}} n_L^0, \tag{16}$$

where $Z_L^{\ell-1} = \sum_{k=0}^{\ell-1} z_L^k$ denotes the cumulative knowledge up to layer $\ell - 1$. Given a unit mass of agents, the number of production workers at the bottom layer is:

$$n_L^0 = \left[1 + h \sum_{\ell=1}^{L} e^{-\lambda Z_L^{\ell-1}}\right]^{-1}. \tag{17}$$

In addition, the equilibrium condition induces a recursive structure for the internal transfer prices $\{r_L^\ell\}$, which reflect the value of problems passed upward through the hierarchy:

$$\begin{aligned} r_L^{L-1} &= \beta - hw_L^L + \alpha \log\left(\frac{\alpha}{A}\right), \\ r_L^{\ell-1} &= \beta - hw_L^\ell + \alpha \log\left(\frac{\alpha}{A - r_L^\ell}\right), \end{aligned} \tag{18}$$

where $\alpha = \frac{ch}{\lambda}$ and $\beta = A - \alpha$.

As shown in Garicano and Rossi-Hansberg (2012), this recursive formulation naturally arises from the optimal design of a hierarchy in equilibrium. The number of managers at each level is calibrated to efficiently handle the volume of unresolved problems from lower layers, ensuring a consistent matching between task complexity and solver expertise.

B1.2 Organizational Responses to Changes in Training Costs

This section examines how changes in training costs affect key elements of the firm's organizational equilibrium. We begin by analyzing the impact of higher training costs on the internal transfer prices that govern problem escalation within the hierarchy. We then explore how these cost changes influence the firm's optimal allocation of knowledge across layers and, ultimately, its total output. By tracing the effects of training costs through internal pricing, knowledge acquisition, and productivity, we reveal how organizations adapt their structure and performance in response to shifts in human capital investment costs.

Proposition 5. *Let $\{r_L^\ell\}$ be the sequence of prices for which (10) holds. Then,*

$$\frac{\partial r_L^\ell}{\partial c} < 0 \tag{19}$$

for all $0 \leq \ell < L$. Moreover, the sequence $\{\partial r_L^\ell / \partial c\}$ is increasing.

This proposition reveals that as training costs increase, the prices at which problems are sold between layers decrease. This effect is more pronounced in lower layers of the organization, indicating that the organization adapts to higher training costs by adjusting its internal pricing structure. Figure B1(a) illustrates this relationship for an economy with four layers.

Given this inverse relationship between training costs and problem prices, we can expect similar effects on net earnings per worker and knowledge levels:

Corollary 2. *The net earnings per worker at each layer at equilibrium, $\tilde{w}$, are decreasing with respect to c.*

Furthermore, we observe a similar effect on the knowledge levels across the organization:

Corollary 3. *Let $\{z_L^\ell\}$ be the sequence of knowledge levels at equilibrium. Then,*

$$\frac{\partial z_L^\ell}{\partial c} < 0 \tag{20}$$

for all $0 \leq \ell \leq L$. Moreover, the sequence $\{\partial z_L^\ell / \partial c\}$ is decreasing.

These corollaries indicate that as training costs increase, both net earnings per worker and knowledge levels decrease across all layers of the organization. The impact on knowledge levels is more pronounced in upper layers, suggesting that specialized knowledge is more sensitive to cost changes. This may lead organizations to prioritize maintaining foundational knowledge at lower levels while reducing investments in highly specialized knowledge at upper levels when faced with increased training costs.

Figure B1(b) illustrates the decrease in equilibrium net earnings per worker for different numbers of layers, while Figure B1(c) shows the reduction in knowledge levels across layers as training costs increase.

Corollary 4. *The number of problem solvers decreases as the training cost increases, or equivalently,*

$$\frac{\partial n_L^0}{\partial c} < 0.$$

Increased training costs could push organizations towards a structure that relies more heavily on higher-skilled individuals at upper levels, reducing the need for many lower-level problem solvers. Additionally, higher training costs might accelerate the adoption of technologies that can replace some functions of lower-level problem solvers, further reducing their numbers.

As a consequence of Corollary 4 we have the next result.

Corollary 5. *The net output of the firm, $\pi(c) = An_L^0 F(Z_L^L)$, is decreasing with respect to c.*

B1.3 Organizational Responses to Changes in Technology

This section explores the impact of exogenous technological advancements on the organizational structure and economic outcomes. By analyzing how changes in the level of technology A affect various aspects of the firm's equilibrium, we can gain insights into the dynamic relationship between technology, productivity, and organizational efficiency. The results presented in this section illustrate the positive effects of increased technology on net earnings per worker, knowledge levels, and the overall output of the firm.

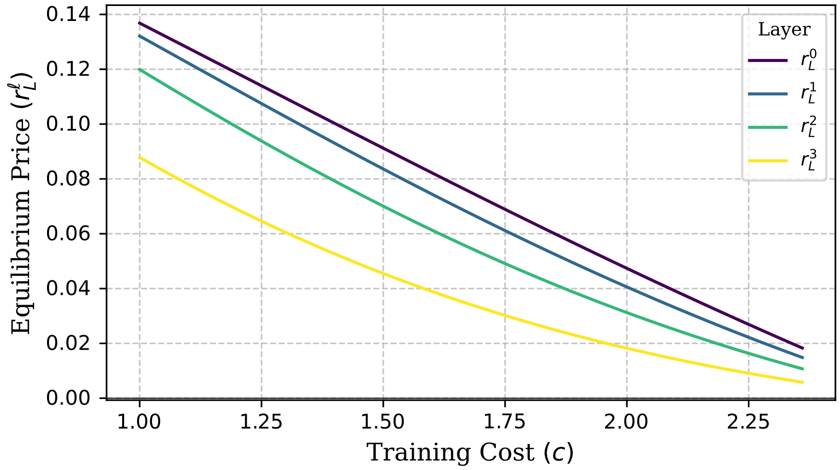
Lemma 4. *Let r_L^0 be the equilibrium price of problems sold by workers at layer 0. Then,*

$$0 < \frac{\partial r_L^0}{\partial A}.$$

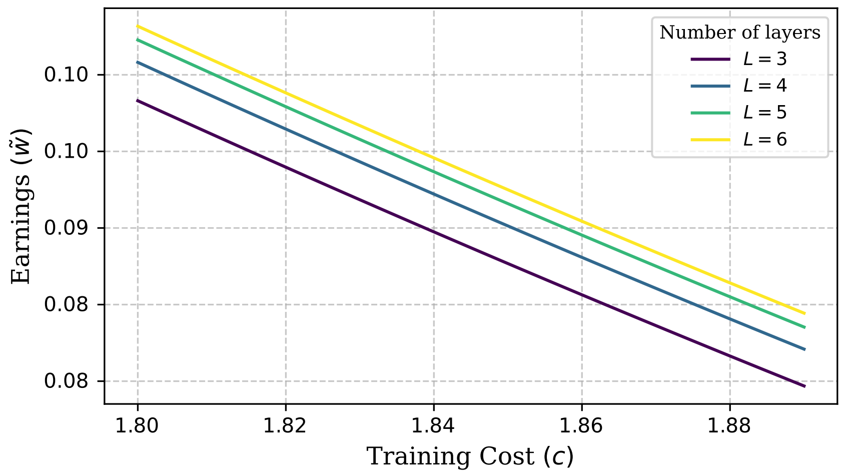
Proposition 6. *Let $\tilde{w}$ be the net earnings per worker at each layer at equilibrium. Then,*

$$0 < \frac{\partial \tilde{w}}{\partial A} < 1.$$

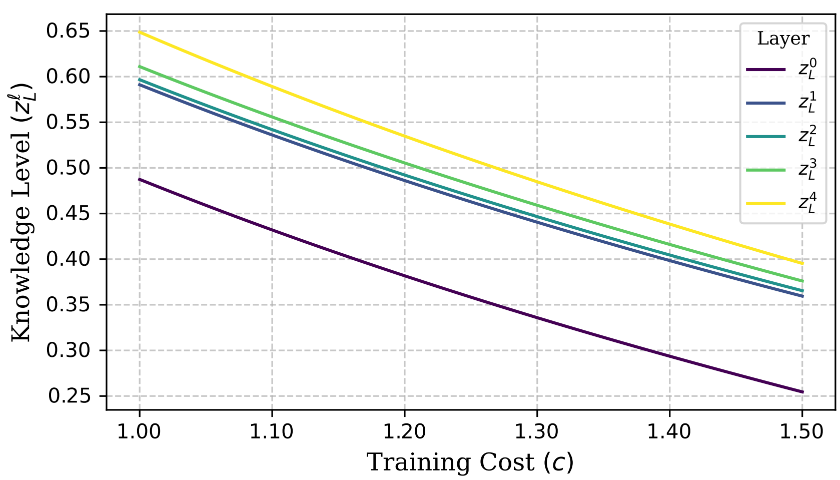
Figure B1: Effects of training costs on the organizational structure.



(a) Equilibrium prices at each layer ($L = 4$).



(b) Equilibrium net earnings per worker for different numbers of layers (L).



(c) Knowledge levels at each layer ($L = 4$).

Notes: Parameters: $\lambda = 1.6$, $A = 1.5$, $h = 0.85$.

The positive relationship between technology and the net earnings of agents at equilibrium is straightforward. An increase in technology enables agents to perform tasks more efficiently, increasing the output each agent can produce with their given ability. This higher productivity directly translates into higher net earnings at each layer within the organizational hierarchy. Figure B2a illustrates this outcome, showing an increase in equilibrium net earnings per worker as the level of technology rises for economies with four, five, six, and seven layers.

Proposition 7. *The equilibrium price of problems sold by workers at each layer is such that*

$$\frac{\partial r_L^\ell}{\partial A} < 1$$

for $0 \leq \ell \leq L-1$*, where* $\left\{r_L^\ell\right\}$ *is the sequence of prices for which* (10) *holds.*

An increase in the level of technology A enhances worker productivity and raises net earnings, as established in previous results. This productivity boost can increase the value of problems passed within the organization, affecting the equilibrium prices at which problems are transferred across layers. Proposition 7 shows that while equilibrium prices do respond to changes in technology, their rate of increase remains strictly below one. That is, for each layer, the marginal increase in internal price is bounded, ensuring that prices do not scale one-to-one with technology. Figure B2b illustrates this relationship, showing that equilibrium prices rise with A but at a rate strictly less than the growth rate of technology, in an economy with six hierarchical layers.

Corollary 6. *Let* $\left\{z_L^\ell\right\}$ *be the sequence of knowledge levels at equilibrium. Then,*

$$\frac{\partial z_L^\ell}{\partial A} > 0 \tag{21}$$

for all $0 \leq \ell \leq L$*.*

This result indicates that as technology levels rise, the ratio of output for a given knowledge level to the cost of acquiring that knowledge increases, making it more profitable for agents to invest in higher levels of expertise. This leads to an overall increase in equilibrium knowledge levels across the hierarchy. Figure B2c illustrates this, showing an increase in knowledge levels at each layer for an organization with six layers.

Corollary 7. *The number of problem solvers increases as the level of technology increases, or equivalently,*

$$\frac{\partial n_L^0}{\partial A} > 0.$$

With a rise in technology levels, equilibrium knowledge levels increase throughout the hierarchy. This increase reduces the number of problems passed up the hierarchy, as lower layers are more capable of solving them. Consequently, the proportion of front-line workers with respect to problem solvers increases, indicating a shift towards a more skilled and concentrated workforce at the lowest level.

Corollary 8. *The net output of the firm, $\pi(c) = An_L^0 F(Z_L^L)$, is increasing with respect to A.*

This result is intuitive as higher levels of technology enhance the productivity of each agent, thereby increasing the overall output of the firm.

B1.4 Static Cost Optimization Problem

The full dynamic model involves, at each point in time, choosing the optimal combination of hierarchical layers and training costs, subject to a set of constraints. The problem evaluated for each of the neighboring layers around the current one, can be decomposed into two components: computing the cost using the current training level, and minimizing cost using the optimal training level—incurring an adjustment cost if training is modified. Thus, the firm only needs to evaluate a maximum of six candidate configurations in total. The core of the static problem reduces, for each layer, to the formulation in 12, where the cost function $\mathcal{P}(c, \lambda, L)$ is minimized with respect to c, subject to the constraints defined by the feasible set Γ. As shown in Proposition 1, if the feasible set is nonempty, the problem admits a unique solution.

Using Corollary 5, a sufficient condition for the constraint $q - An_L^0 F(Z_L^L) \leq 0$ to hold within the domain of $\Gamma_{A,\lambda,L}$ is: $\pi(1/\lambda + \gamma L) > q$. [69] This insight allows us to restate Proposition 1 in a more tractable form:

Corollary 9. *Under Assumption 1, if the firm's output at minimum training cost satisfies $\Pi\left(1/\lambda + \gamma L\right) > q$, then there exists a unique $c^* \in \Gamma_{A,\lambda,L}$ solving the optimization problem (12).*

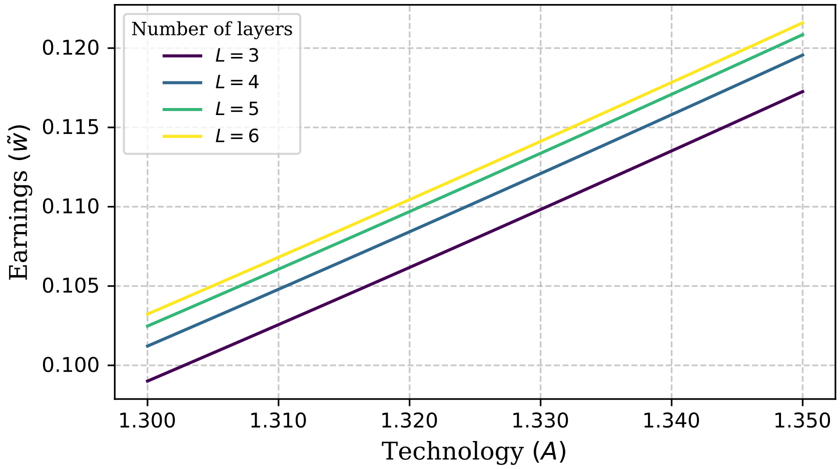
This corollary provides a sufficient condition for the existence of a solution, leveraging the results from Corollary 5.

B1.5 Dynamic Adjustment Rules

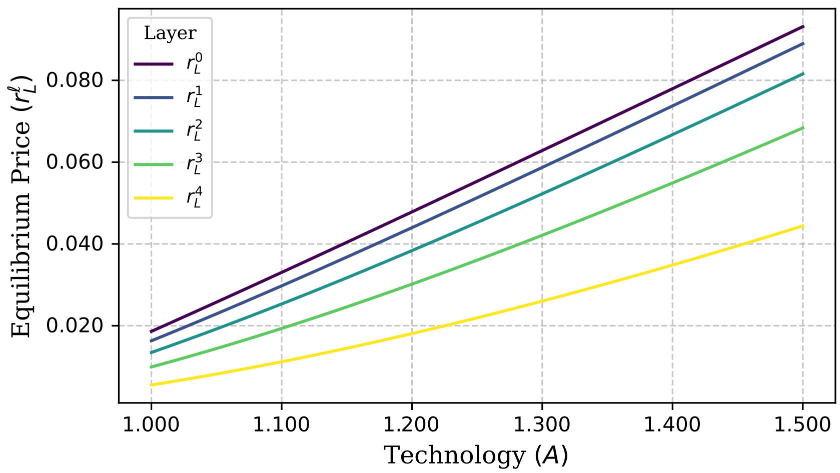
The firm's dynamic decisions are governed by two interconnected rules, both derived from the cost function defined in equation (14). These rules determine how the firm adjusts its

[69] The equality would lead to no solutions, as the highest value satisfying the production constraint will be $c = 1/\lambda + \gamma L$, for which the cost diverges. This fact will be used later in Section 5.3, where the firm is forced to decrease its number of layers.

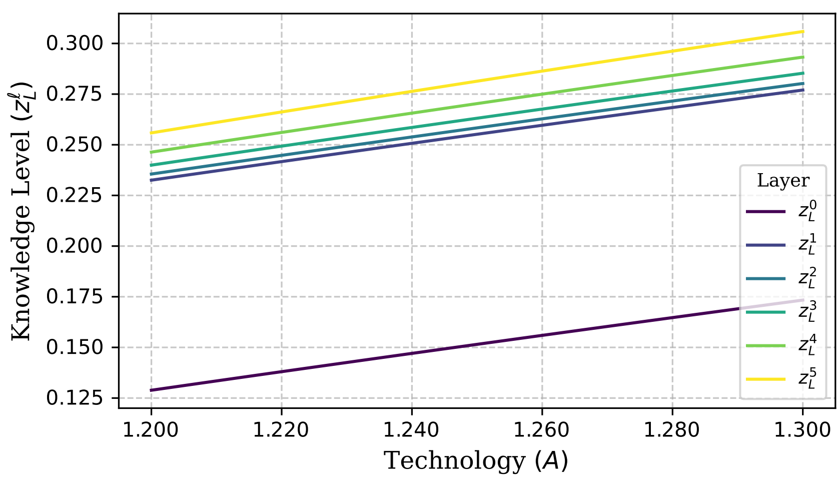
Figure B2: Effects of technology on organizational structure.



(a) Effect of technology on equilibrium net earnings per worker for different numbers of layers (L).



(b) Effect of technology on equilibrium prices at each layer ($L = 5$).



(c) Effect of technology on knowledge levels at each layer ($L = 6$).

Notes: Parameters: $\lambda = 1.6$, $c = 1.5$, $h = 0.85$.

organizational layers and training investments over time in response to changes in technology, complexity, and production requirements.

The first rule governs the optimal adjustment of the number of organizational layers. At each period t, the firm considers three possibilities: retaining its current number of layers, increasing it by one, or decreasing it by one. Formally:

$$L_t =_{L\in\{L_{t-1}-1,\, L_{t-1},\, L_{t-1}+1\}} \min \begin{Bmatrix} \mathcal{P}(c_{t-1}, \lambda, L), \\ \min\limits_{c\in\Gamma} \left[\mathcal{P}(c, \lambda, L) + \delta\right] \end{Bmatrix} \tag{22}$$

This rule captures the trade-off between maintaining the current structure and adjusting it at the cost of restructuring and retraining. Following Garicano and Rossi-Hansberg (2012), the discrete choice set reflects the limited flexibility firms face when modifying hierarchical depth.

The second rule determines whether the firm maintains its current training cost or optimally adjusts it, incurring a one-time adjustment cost δ. The evolution of training costs is described by:

$$c_t = \begin{cases} c_{t-1}, & \text{if no adjustment is optimal,} \\ c_t^*, & \text{otherwise,} \end{cases} \tag{23}$$

where

$$c_t^* =_{c\in\Gamma} \left[\mathcal{P}(c, \lambda, L_t) + \delta\right]$$

is the new optimal training cost when an adjustment is implemented.

Together, these rules govern the firm's dynamic behavior regarding the two endogenous organizational variables: the number of layers and the level of training investment. At each period t, the values of c_t and L_t depend on the prevailing technology A_t and problem complexity λ_t, which jointly shape the firm's cost function.

The dynamic cost trajectory is tracked using two key sequences:

$$\phi_t = \phi_t(A_t, \lambda_t), \quad \text{and} \quad \mathcal{P}_t = \mathcal{P}(c_t, \lambda_t, L_t),$$

which represent, respectively, the total period cost and the component of the cost associated with the firm's training and structural configuration.

These sequences capture how the firm adjusts to both endogenous and exogenous changes in its environment. Specifically, our analysis focuses on transitions across three distinct phases of the product cycle: pre-learning, learning, and post-learning, occurring at times 0, τ, and $\tau + T$, respectively. In the final phase of our analysis, we also examine how the firm responds to a change in demand—from q to $\tilde{q}$—occurring at time $\mathcal{T}$.

Accordingly, the cost function evolves as follows:

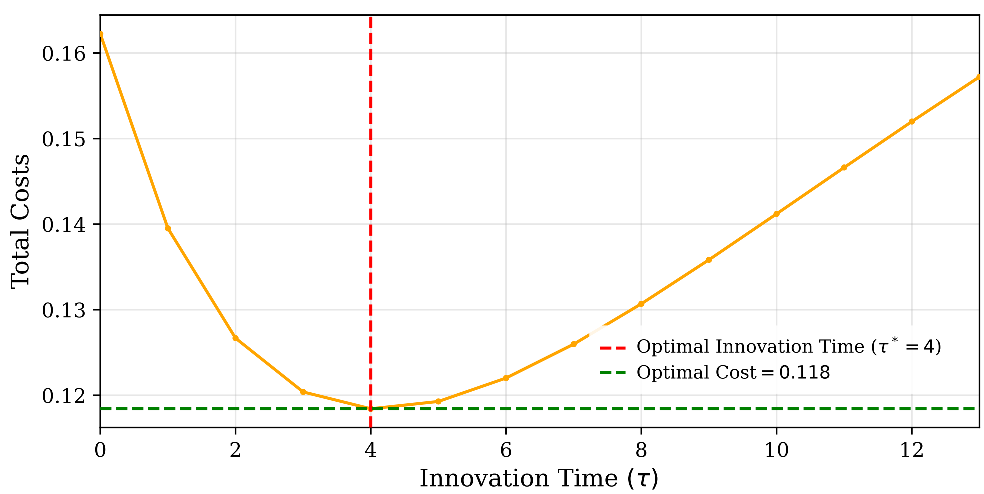
$$\phi_t = \begin{cases} \phi_t(A, \lambda, q), & \text{if } t < \tau, \\ \phi_t(\tilde{A}, \tilde{\lambda}, q), & \text{if } \tau \leq t < \tau + T, \\ \phi_t(\tilde{A}, \lambda, q), & \text{if } \tau + T \leq t < \mathcal{T}, \\ \phi_t(\tilde{A}, \lambda, \tilde{q}), & \text{if } t \geq \mathcal{T}. \end{cases} \tag{24}$$

This framework allows us to evaluate how organizations adapt their internal structures and cost strategies in response to both technological innovation and rising production requirements, while maintaining efficient knowledge transmission and effective problem resolution throughout the product cycle.

B1.6 Optimal Innovation Timing and Organizational Dynamics

The firm's choice of the optimal innovation time τ^* reflects a fundamental trade-off between the costs of early adoption and the benefits of delayed implementation. Figure B3 illustrates how total costs vary with different innovation times, suggesting the presence of a minimum that defines τ^*. The cost curve reveals that adopting too early leads to high adoption costs, while delaying too long results in forgone productivity improvements.

Figure B3: Total costs as a function of innovation time τ.



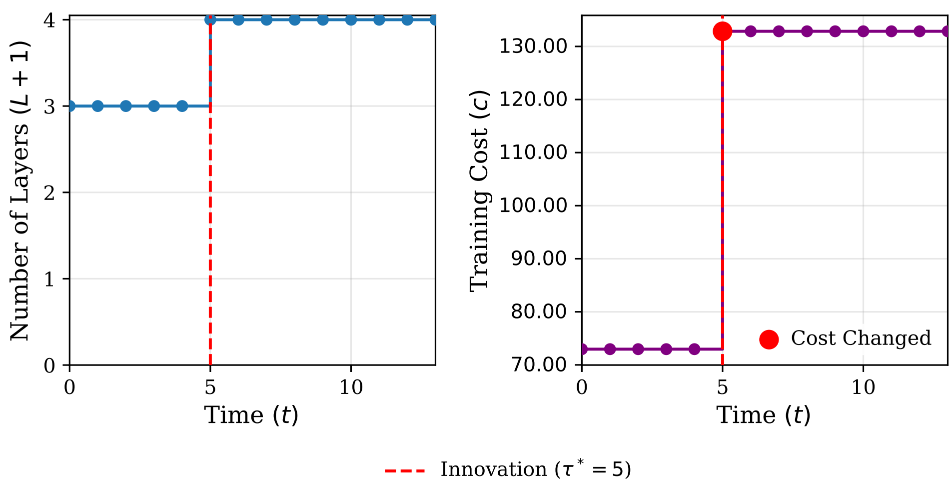
Notes: Parameters: $q = 70$, $A = 100$, $\lambda = \tilde{\lambda} = 1.5$, $h = 0.85$ $\gamma = 0.024$, $\Psi = 10^{-10}$, $\zeta = 0.05$, $\delta = 10^{-5}$.

The firm's organizational response at the optimal innovation time displays several notable patterns, as shown in Figure B4. Starting from a steady state with two layers at period

0, the firm adds a third layer at the optimal innovation time $\tau^* = 5$. This adjustment is accompanied by a sharp increase in training costs, which in our framework corresponds to a reduction in the amount of training per worker (see Corollary 5). Although the complexity of problems λ remains constant, the improved technology allows agents to maintain or even enhance productivity with lower training, leveraging automation, superior tools, or process optimizations.

Importantly, this relationship is not uniform. In our model, the interaction between training costs and hierarchical structure depends on the parameter γ in the training cost function. When γ is low, hierarchy and training tend to act as substitutes: additional layers reduce the need for highly trained workers at the base, as seen in the example. Conversely, when γ is high, deeper hierarchies may require more coordination and thus more specialized training, making hierarchy and training act as complements.

Figure B4: Evolution of organizational variables over time.



Notes: Parameters: $q = 70$, $A = 100$, $\lambda = \tilde{\lambda} = 1.5$, $h = 0.43$, $\gamma = 0.024$, $\beta = 0.92$, $\Psi = 10^{-10}$, $\zeta = 0.03$, $\delta = 10^{-5}$.

B2 Model Calibration

Several key parameters are calibrated from the existing literature. Reference values for the discount factor $\beta = 0.9998$ and the fixed innovation cost $\psi = 50$ are taken from Garicano and Rossi-Hansberg (2012).[70] The technology level $A = 0.26$, baseline complexity $\lambda = 0.9$, and

[70] In Garicano and Rossi-Hansberg (2012), the discount factor is set to 0.87, where one period corresponds to an agent's entire working life. As described in more detail below, one period in our calibration corresponds to three weeks. Converting this to a three-week period yields an equivalent value of 0.9998.

communication cost $h = 0.26$ follow Caliendo and Rossi-Hansberg (2012). These benchmark values are reported in Table B1.

The remaining parameters—q, γ, λ_1, and ζ—are calibrated by matching model-generated moments on hierarchical layers and training investments over the product cycle. Below, we describe the construction of the empirical moments used for this procedure and then present the resulting calibrated values.

B2.1 Construction of Data Moments

We target the evolution of number of hierarchical layers (our key proxy of hierarchical structure, see Table 5 and Appendix Figure A8a), as well as number of courses and teachers (our proxies for training costs, see Table 8 and Figure A11) over the product cycles. Consistent with the results shown in the empirical section, we predict the weekly value of these outcomes over the average product cycle by running event studies using equation 1 for the period from 8 weeks before the Model change to 8 weeks after (so effectively replicating Figures A8a and A11).[71] The resulting data moments comprise weekly (predicted) number of hierarchical layers, courses, and teachers (Appendix Table B3).

Based on the results in Figure 2, we set the duration of the typical learning phase (T) to 3 weeks. To facilitate analysis, the weekly data moments were aggregated into three periods—"Initial" (i.e., from 8 weeks before the Model change to the week before the change), "Learning" (i.e., the first three weeks after a Model change), and "Post-Learning" (i.e., from week 4 to week 8 after a Model change)—by averaging the weekly values within each period. The number of layers was rounded down to the nearest integer using the floor function (Table B4). So, this gives us nine data moments, that is, the average values of (i) hierarchical layers, (ii) number of teachers and (iii) number of courses for each of the three periods (Initial phase, Learning phase, and Post-learning phase).

The average length of the product cycle in our data, calculated as the average distance between two consecutive Model changes, is 51 weeks (see Figure 1). To better align the timing of innovation with a standard product cycle, period averages for the calibration were extrapolated to cover 51 weeks for the Initial phase, 3 weeks for the Learning phase, and 48 weeks (51 - 3) for the Post-learning phase.[72]

In the calibration, the time horizon is set to 34 periods, each corresponding to three weeks. In this setting, the timing of innovation, denoted by τ, is treated as exogenous and

[71] Specifically, we take the mean in the week before the Model change as our reference period, and then add the coefficients from the event study to create a predicted value of the outcome in each week.

[72] The extrapolation is done assuming outcomes remain constant in each period, so that all periods until the Model change have the same value of the outcome as the "Initial" period, and all periods after the Learning phase has concluded have the same value of the outcome as for the "Post-learning" period.

is assumed to occur in period 17 (51 weeks), with the learning phase lasting one period (3 weeks). This timing choice reflects the empirical duration of product cycles and facilitates visual alignment with the data, but does not influence the internal dynamics of the model.[73]

Finally, because we do not use cost-related moments in the calibration, and the data show adjustments in training (inferred from course and teacher registries) at each phase, we calibrate the model assuming zero training adjustment costs ($\delta = 0$). This choice simplifies the analysis without loss of generality: for a sufficiently small but positive adjustment cost, the firm would optimally choose to incur it and adjust training as observed. Thus, the same equilibrium dynamics for organizational variables would result, while the only difference would arise in total cost calculations.[74]

B2.2 Parameter Calibration and Results

Since the model outputs training costs rather than the number of courses and teachers directly, two additional parameters, θ_1 and θ_2, were introduced as weights for how the number of courses and teachers, respectively, affect the training cost. Specifically, the training cost c is given by:

$$c = \frac{1}{\theta_1 \cdot \text{Num\_Courses} + \theta_2 \cdot \text{Num\_Teachers}} \tag{25}$$

This parameterization captures the fact that as the training resources (i.e., the number of courses and teachers) available to the firm increase, the cost of training a worker c is reduced.

To calibrate the resulting eight parameters (i.e., q, γ, λ_1, ζ, θ_1 and θ_2), we target the nine empirical moments described above. Specifically, the parameters were estimated using the moment equations in Appendix B2.4. Model calibration was performed numerically by minimizing the normalized error between the targeted data moments and the model's output.[75] Table B1 summarizes the results of the calibration. The model accurately reproduces the observed layers and training costs in all phases, as shown in Table B2. Figure 8 shows that our calibrated model reproduces the dynamics for layers and training documented in the empirical section: as a response to technological innovation, the firm temporarily decreases

[73]The exogenous timing of innovation τ affects only the parameter ζ, which governs the rate of technological improvement. Since any alternative choice of τ can be accommodated by adjusting ζ such that the resulting technology level remains unchanged at the time of innovation, the model's qualitative dynamics remain unaffected.

[74]As discussed in the next section, our key counterfactuals aim to quantify the *relative* importance of hierarchical structure and training for cost minimization over the product cycle. Assuming $\delta = 0$ does not affect conclusions about the relative importance of these two organizational levers for cost minimization, but of course prevents us from calculating the total cost incurred by the firm in *levels*.

[75]To solve the optimization problem, the differential evolution algorithm (Storn and Price, 1997) was employed. This global optimization method is particularly suited for problems with multiple local minima, as it explores the parameter space efficiently within predefined bounds.

the number of layers and invests in reducing training costs, thus increasing knowledge during the learning phase.

Table B1: Model Parameters, Calibrated Values, and Sources

Parameter	Value	Range	Source
q	0.1301	[0.1292, 0.131]	Calibrated from our data
γ	0.0068	[0.0058, 0.0074]	Calibrated from our data
$\tilde{\lambda}$	0.8499	[0.8435, 0.8553]	Calibrated from our data
ζ	0.0083	[0.0073, 0.0095]	Calibrated from our data
θ_1	0.37	[0.3658, 0.3742]	Calibrated from our data
θ_2	0.0876	[0.0819, 0.0932]	Calibrated from our data
β	0.9998	—	Garicano and Rossi-Hansberg (2012)
ψ	50	—	Garicano and Rossi-Hansberg (2012)
A	0.26	—	Caliendo and Rossi-Hansberg (2012)
h	0.26	—	Caliendo and Rossi-Hansberg (2012)
λ	0.9	—	Caliendo and Rossi-Hansberg (2012)

Notes: Calibrated ranges correspond to 95% confidence intervals.

Table B2: Comparison between Empirical Data and Calibrated Model Outcomes

Moment	Data	Model
Layers (Initial)	6	6
Layers (Learning)	5	5
Layers (Post-Learning)	6	6
Training cost (Initial)	0.300	0.301
Training cost (Learning)	0.268	0.266
Training cost (Post-Learning)	0.314	0.314

Notes: Comparison between empirical data and calibrated model outcomes. Empirical training costs were inferred using equation 25 with calibrated parameter values for θ_1 and θ_2.

The accuracy of the calibration process was assessed by comparing the model's predictions to the observed data. Appendix Figure B5 (see Appendix B2.5) shows the normalized error as a function of the number of iterations. Following the elbow method, we identified the iteration at which further error reduction became marginal relative to previous improvements. This point, corresponding to iteration 41, marks the onset of a plateau in the error curve and was selected as the optimal starting point for parameter sampling. The distributions of the calibrated parameters are shown in Appendix Figure B6. Each parameter's distribution is well-defined, with clear peaks indicating the most likely values.[76]

[76] Additionally, the joint distributions of the parameter pairs are presented in Appendix Figure B7, revealing the relationships and correlations between the parameters.

B2.3 Data

Here we present the empirical data used for model calibration. Table B3 displays the weekly data, while Table B4 reports the phase-specific averages that are used in the calibration process.

Table B3: Weekly Empirical Data for Model Calibration

Weeks	Layers	Num. Courses	Num. Teachers
-8	4.2447	8.8843	12.6429
-7	4.6308	6.5766	7.5000
-6	4.1513	6.4109	7.7643
-5	4.1980	6.9421	8.8714
-4	3.9941	6.9936	8.9857
-3	4.0812	6.6956	8.8143
-2	3.9744	7.0691	8.2786
-1	3.9050	5.8571	7.0000
0	2.6960	8.4286	10.0000
1	2.6712	8.1733	9.0286
2	3.0291	7.1781	8.1286
3	4.1457	6.3210	8.2714
4	3.9938	7.2890	8.8952
5	3.5618	7.4542	9.8238
6	3.9600	7.8827	9.3952
7	4.0910	3.7429	7.3071

Notes: Weekly empirical data for model calibration. For each outcome, the values are predicted from weekly event studies around Model changes following equation 1, as described in the main text. Negative values represent the pre-introduction phase of a new car model, with an innovation event occurring at $t = 0$. The learning phase (T) is assumed to last three weeks.

Table B4: Average Empirical Data per Period

Period	Layers	Num. Courses	Num. Teachers
Initial (-8:-1)	6	6.9287	8.7321
Learning (0:2)	5	7.9267	9.0524
Post-Learning (3:7)	6	6.5380	8.7386

Notes: Average empirical data per period. Layers are rounded to the nearest integer.

B2.4 Moment Conditions for Parameter Estimation

To estimate the model parameters, we match moments computed from the observed data across the three phases: initial, learning, and post-learning. Superscripts $i = 0, 1, 2$ index the phase: 0 for the initial phase, 1 for the learning phase, and 2 for the post-learning phase. The variables L^i, c^i, and x^i denote the number of layers, training costs, and observed data vectors, respectively. Each vector $x^i = (x_0^i, x_1^i)$ corresponds to the number of courses and teachers.

The optimal number of layers and associated training costs for each phase are determined by:

$$(1) \quad L^0 = \underset{L\in\{5,6,7\}}{\arg\max}\, \pi_L^{-1}(q) - \frac{1}{\lambda_0} - \gamma L,$$

$$(2) \quad c^0 = \pi_{L^0}^{-1}(q),$$

$$(3) \quad L^1 = \underset{L\in\{L^0-1,L^0,L^0+1\}}{\arg\max}\, \pi_L^{-1}(q) - \frac{1}{\lambda_1} - \gamma L,$$

$$(4) \quad c^1 = \pi_{L^1}^{-1}(q),$$

$$(5) \quad L^2 = \underset{L\in\{L^1-1,L^1,L^1+1\}}{\arg\max}\, \pi_L^{-1}(q) - \frac{1}{\lambda_2} - \gamma L,$$

$$(6) \quad c^2 = \pi_{L^2}^{-1}(q).$$

The relationships between the observed data and the training cost parameter c are given by:

$$(7) \quad \frac{1}{\theta_1 x^0 + \theta_2 x_1^0} = c^0,$$

$$(8) \quad \frac{1}{\theta_1 x^1 + \theta_2 x_1^1} = c^1,$$

$$(9) \quad \frac{1}{\theta_1 x^2 + \theta_2 x_1^2} = c^2.$$

B2.5 Calibration Error and Parameter Distributions

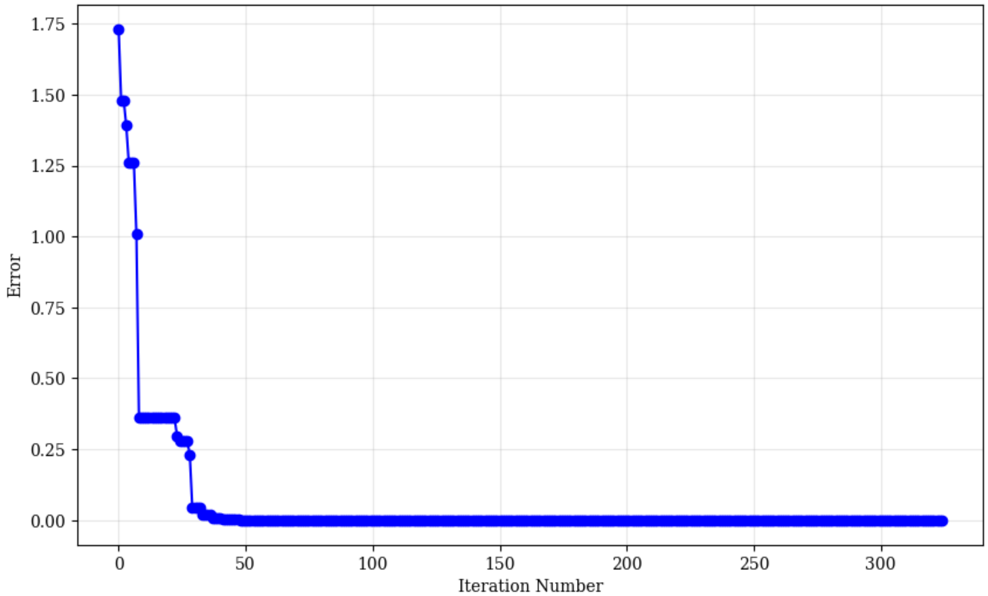
Figure B5 shows the normalized error across iterations of the calibration algorithm. The "elbow" occurs at iteration 41, where the error stabilizes at a low value. We use this iteration as the basis for sampling the calibrated parameters.

Figures B6 and B7 present the marginal and joint distributions of the calibrated parameters obtained from iteration 41 onward.

B2.6 Calibration Results over the Cycle

Figure B8 presents the complete product cycle simulated using the calibrated parameters, illustrating the dynamic evolution of organizational layers and training costs over time.

Figure B5: Normalized error as a function of the number of iterations.



Notes: The error decreases rapidly in the initial iterations and begins to plateau around iteration 41, marking the elbow point where further reductions become marginal.

B3 Counterfactuals

B3.1 Fixed training costs

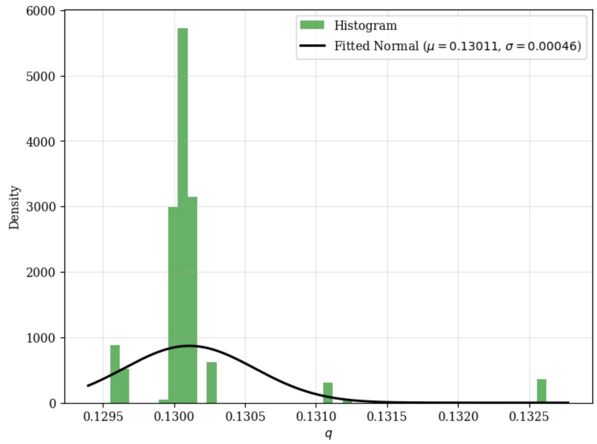
We use Lemma 5 to formally establish conditions under which training costs remain fixed throughout a single sub-cycle, during which technology and problem complexity are constant. Specifically, if the firm can meet production and all relevant constraints using its initial training cost and a hierarchical structure near its current configuration, then sufficiently high adjustment costs will lead the firm to maintain constant training expenditures.

Lemma 5. *Let A and λ be positive real numbers representing technology and problem complexity, respectively. Consider an initial state (c, L). If $c \in \bigcup_{L' \in \{L-1, L, L+1\}} \Gamma_L$, there exists a threshold $\delta' > 0$ such that for any $\delta > \delta'$, the sub-cycle determined by these conditions and adjustment cost δ results in constant training costs throughout. That is, if (c^*, L^*) represents the steady state to which the firm converges in this cycle, then $c^* = c$.*

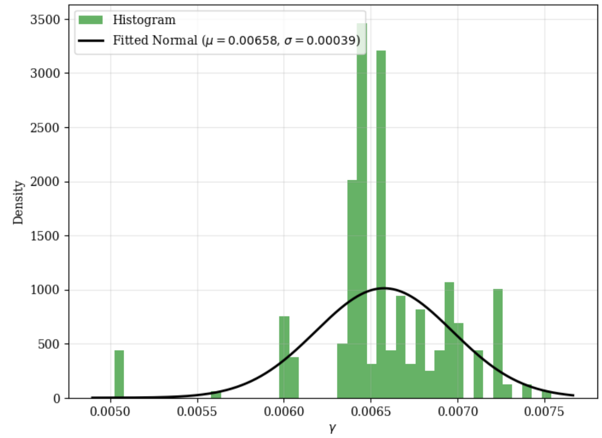
B3.2 Full simulation dynamics

Here, we present the full dynamics of the counterfactual scenarios. In both cases, we observe that the baseline scenario ($\delta = 0$) consistently converges to the calibrated steady-state values for training costs and organizational layers during the initial phase, regardless of the

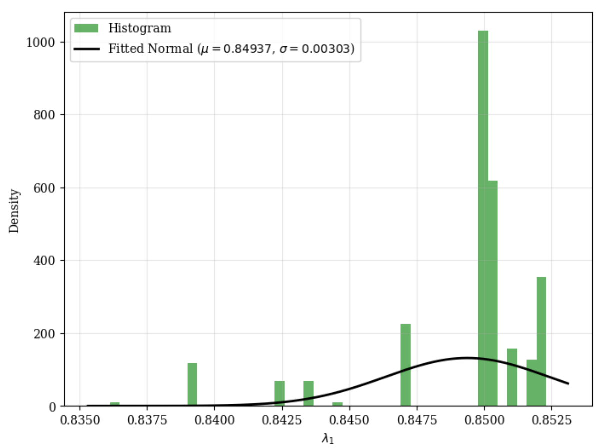
Figure B6: Distributions of the calibrated parameters starting from the 41st iteration.



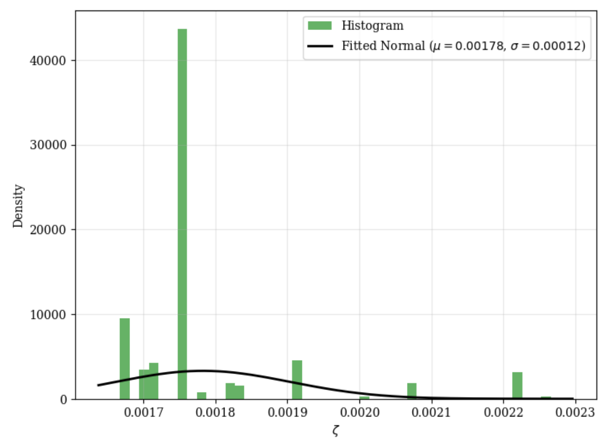
(a) Distribution of q.



(b) Distribution of γ.



(c) Distribution of λ_1.



(d) Distribution of ζ.

initial conditions. This highlights the stability of the firm's optimal response under flexible adjustment.

B4 Proofs

Proof of Proposition 5. To show the statement of this proposition, we will use some tools introduced in the proof of Proposition 2 from Garicano and Rossi-Hansberg (2012).

Namely, they introduce two kinds of profits. The first, $w_L^s(r_L^0)$, is implicitly defined by the recurrence relation (18), while the second one, $w_L^p(r_L^0)$, is defined as the profit of the workers in the first layer (deduced from (3)):

$$w_L^p(r_L^0) = A - \frac{c}{\lambda}\left(1 - \log\left(\frac{c}{\lambda\,(A - r_L^0)}\right)\right). \tag{26}$$

The value of $r \equiv r_L^0$ for which both w_L^s and w_L^p coincide is the one that equalizes all profits after using (18) to get the other elements of the sequence $\left\{r_L^\ell\right\}$. It is shown in Garicano and Rossi-Hansberg (2012) that there is a unique value of r for which this condition holds, and we will show that the derivative of this value with respect to c is negative. For that, note that the function implicitly defines the value of r

$$\phi(r) \equiv w_L^s(r) - w_L^p(r) \tag{27}$$

being equal to zero. Differentiating this function with respect to r gives

$$\frac{\partial \phi}{\partial r} = \frac{\partial w_L^s}{\partial r} - \frac{\partial w_L^p}{\partial r} = -\frac{1}{h}\left[1 + \sum_{\ell=1}^{L-1}\prod_{k=1}^{\ell}\frac{\alpha}{A - r_L^k}\right]^{-1} - \frac{c}{\lambda}\frac{1}{A-r} < 0. \tag{28}$$

On the other hand, the derivative of (27) with respect to c is composed by the derivative of w_L^p with respect to c, which is

$$\frac{\partial w_L^p}{\partial c} = \frac{1}{\lambda}\log\left(\frac{c}{\lambda\left(A-r\right)}\right) = -z_L^0 < 0,$$

while the derivative of w_L^s with respect to c is calculated using the relation $r_L^0(w_L^s(r,c),c) = r$, which is derived from the fact that r_L^0 (defined as the value from the recurrence relation (18) that defines the other r_L^ℓ) and w_L^s are inverse functions. Thus, differentiating this expression yields to

$$\frac{\partial r_L^0}{\partial w_L^s}\frac{\partial w_L^s}{\partial c} + \frac{\partial r_L^0}{\partial c} = 0,$$

from where we get that

$$\frac{\partial w_L^s}{\partial c} = -\frac{\partial r_L^0}{\partial c}\left[\frac{\partial r_L^0}{\partial w}\right]^{-1}.$$

We already know that

$$\frac{\partial r_L^0}{\partial w_L^s} = -h\left(1 + \sum_{\ell=1}^{L-1}\prod_{k=1}^{\ell}\frac{\alpha}{A - r_L^k}\right), \tag{29}$$

while the derivative of r_L^0 with respect to c can be determined using the recurrence relation (18), from where we get that

$$\frac{\partial r_L^0}{\partial c} = -hz_L^1 + \frac{\alpha}{A - r_L^1}\frac{\partial r_L^1}{\partial c},$$

$$\frac{\partial r_L^\ell}{\partial c} = -hz_L^{\ell+1} + \frac{\alpha}{A - r_L^{\ell+1}} \frac{\partial r_L^{\ell+1}}{\partial c},$$

for $\ell = 0, 1, \ldots, L-2$, and

$$\frac{\partial r_L^{L-1}}{\partial c} = -hz_L^L.$$

Thus, by using this new recurrence relation, the closed-form expression for the derivative of r_L^0 with respect to c is

$$\frac{\partial r_L^0}{\partial c} = -h\left(z_L^1 + \sum_{\ell=1}^{L-1} z_L^{\ell+1} \prod_{k=1}^{\ell} \frac{\alpha}{A - r_L^k}\right).$$

With this, we can calculate the derivative of w_L^s with respect to c, and ultimately, the derivative of ϕ with respect to c, obtaining that

$$\frac{\partial \phi}{\partial c} = \left[1 + \sum_{\ell=1}^{L-1} \prod_{k=1}^{\ell} \frac{\alpha}{A - r_L^k}\right]^{-1} \left[\left(z_L^0 - z_L^1\right) + \sum_{\ell=1}^{L-1} \left(z_L^0 - z_L^{\ell+1}\right) \prod_{k=1}^{\ell} \frac{\alpha}{A - r_L^k}\right] < 0. \tag{30}$$

Now, we can use the Implicit Function Theorem combined with (28) and (30) to get the derivative of r with respect to c, so that

$$\frac{\partial r}{\partial c} = -h\left[1 + \sum_{\ell=1}^{L-1} \prod_{k=0}^{\ell} \frac{\alpha}{A - r_L^k}\right]^{-1} \left[\left(z_L^1 - z_L^0\right) + \sum_{\ell=1}^{L-1} \left(z_L^{\ell+1} - z_L^0\right) \prod_{k=1}^{\ell} \frac{\alpha}{A - r_L^k}\right] < 0, \tag{31}$$

which is one of the results we wanted to show. We now continue with the derivative of r_L^{L-1} with respect to c, which is calculated using (18). Namely,

$$\frac{\partial r_L^{L-1}}{\partial c} = -h\frac{\partial \tilde{w}}{\partial c} + \frac{h}{\lambda} \log\left(\frac{ch}{\lambda A}\right) = h\left(z_L^0 - z_L^L\right) - \frac{\alpha}{A - r} \frac{\partial r}{\partial c},$$

where we have used the fact that $\tilde{w}(c) = w_L^p(r(c), c)$. Substituting (31) in this expression and simplifying gives

$$\frac{\partial r_L^{L-1}}{\partial c} = h\left[1 + \sum_{\ell=1}^{L-1} \prod_{k=0}^{\ell} \frac{\alpha}{A - r_L^k}\right]^{-1} \left[\sum_{\ell=1}^{L-1} \left(z_L^{\ell+1} - z_L^L\right) \prod_{k=0}^{\ell} \frac{\alpha}{A - r_L^k} + \left(z_L^0 - z_L^L\right) + \frac{\alpha}{A - r_L^0}\left(z_L^1 - z_L^0\right)\right]. \tag{32}$$

The sign of the expression inside the second squared parenthesis determines the sign of (32). Note that the first summation inside this parenthesis is negative so that it is sufficient to check that the remaining term is negative to ensure the negativity of this derivative.

Nevertheless, this is a fact, because

$$\left(z_L^0 - z_L^L\right) + \frac{\alpha}{A - r_L^0}\left(z_L^1 - z_L^0\right) < z_L^0 - z_L^L + h\left(z_L^1 - z_L^0\right) < z_L^1 - z_L^L < 0,$$

and thus $\partial r_L^{L-1}/\partial c < 0$. Finally, to estimate the derivative of the remaining r_L^m for $m = 1, 2, \ldots, L-2$ we use the recurrence relation (18), from where we get that

$$\frac{\partial r_L^m}{\partial c} = -h\left(z_L^{m+1} - z_L^0\right) - \frac{\alpha}{A - r_L^0}\frac{\partial r_L^0}{\partial c} + \frac{\alpha}{A - r_L^{m+1}}\frac{\partial r_L^{m+1}}{\partial c},$$

and rewriting this expression gives the following recurrence relation

$$\frac{\partial r_L^{m+1}}{\partial c} = e^{\lambda z_L^{m+1}}\frac{\partial r_L^m}{\partial c} + he^{\lambda z_L^{m+1}}\left(z_L^{m+1} - z_L^0\right) + he^{\lambda z_L^{m+1}}e^{-\lambda z_L^0}\frac{\partial r_L^0}{\partial c},$$

which can be applied consecutively to get the following closed-form expression for this derivative:

$$\frac{\partial r_L^{m+1}}{\partial c} = h\sum_{\ell=0}^{m}\left(z_L^{\ell+1} - z_L^0 + e^{-\lambda z_L^0}\frac{\partial r_L^0}{\partial c}\right)\prod_{k=\ell+1}^{m+1} e^{\lambda z_L^k} + \frac{\partial r_L^0}{\partial c}\prod_{k=1}^{m+1} e^{\lambda z_L^k}. \tag{33}$$

The positive term of expression (33) consists of the sum of the subtraction $z_L^{\ell+1} - z_L^0$. Nevertheless, this is compensated by other negative terms, as we will show next. First, we put (33) under the common denominator of $\partial r_L^0/\partial c$ (the dividing term in (31)), so that the positive terms are

$$h\sum_{\ell=0}^{m}\left(z_L^{\ell+1} - z_L^0\right)\left[1 + \sum_{s=1}^{L-1}\prod_{k=0}^{\ell}\frac{\alpha}{A - r_L^k}\right]\prod_{k=\ell+1}^{m+1} e^{\lambda z_L^k}.$$

Note that the term $h\sum_{s=1}^{L-1}\left(z_L^{\ell+1} - z_L^0\right)\prod_{k=0}^{\ell}\frac{\alpha}{A-r_L^k}$ is compensated by the term $-h\sum_{s=1}^{L-1}\left(z_L^{s+1} - z_L^0\right)\prod_{k=0}^{\ell}\frac{\alpha}{A-r_L^k}$ (which comes from $e^{-\lambda z_L^0}\partial r_L^0/\partial c$) for all ℓ in the summation of (33). On the other hand, the expression $h\sum_{\ell=0}^{m}\left(z_L^{\ell+1} - z_L^0\right)\prod_{k=\ell+1}^{m+1} e^{\lambda z_L^k}$ is compensated by $-h\sum_{s=1}^{L-1}\left(z_L^{s+1} - z_L^0\right)\left[\prod_{k=1}^{m+1} e^{\lambda z_L^k}\right]\left[\prod_{k=1}^{s} e^{-\lambda z_L^k}\right]$ (coming from $\prod_{k=1}^{m+1} e^{\lambda z_L^k}\partial r_L^0/\partial c$), provided that for $s \le m$ the product of products in this expression is bigger than 1.

Furthermore, expression (33) can be rewritten in terms of the previous derivative $\partial r_L^m/\partial c$ in such a way that (33) is equal to the previous derivative plus a positive term. This implies that the sequence $\left\{\partial r_L^\ell/\partial c\right\}$ is increasing, and $\partial r_L^{m+1}/\partial c < 0$ for all $m = 0, 1, \ldots, L-3$, which is the last result we wanted to show. $\square$

Proof of Corollary 2. At equilibrium, it follows that $\tilde{w} = w_L^p(r(c), c)$ (defined in (26)), where r is the value of r_L^0 for which profits are equalized. Differentiating this expression with respect to c gives

$$\frac{\partial \tilde{w}}{\partial c} = -z_L^0 + \frac{c}{\lambda} \frac{1}{A - r} \frac{\partial r}{\partial c} < 0,$$

thanks to (31). □

Proof of Corollary 3. To prove this result we differentiate (6) with respect to c to get

$$\begin{aligned}
\frac{\partial z_L^0}{\partial c} =& -\frac{1}{\lambda c}\left(1 + \frac{c}{A - r_L^0}\frac{\partial r_L^0}{\partial c}\right) \\
=& -\frac{1}{\lambda c}\left[1 + \sum_{s=1}^{L-1}\prod_{k=0}^{\ell}\frac{\alpha}{A - r_L^k}\right]^{-1}\left(1 - \frac{\lambda\alpha}{A - r_L^0}\left(z_L^1 - z_L^0\right)\right. \\
&\left. + \sum_{\ell=1}^{L-1}\left(1 - \lambda h e^{-\lambda z_L^0}\left(z_L^{\ell+1} - z_L^0\right)\right)\prod_{k=1}^{\ell}\frac{\alpha}{A - r_L^k}\right).
\end{aligned}$$

With this, note that

$$-\frac{\lambda\alpha}{A - r_L^0}\left(z_L^1 - z_L^0\right) = 1 - h e^{-\lambda z_L^0}\log\left(\frac{A - r_L^1}{h\left(A - r_L^0\right)}\right) > 1 + h\log h > 0,$$

and similarly,

$$1 - \lambda h e^{-\lambda z_L^0}\left(z_L^{\ell+1} - z_L^0\right) = 1 + h e^{-\lambda z_L^0}\log\left(\frac{h\left(A - r_L^0\right)}{A - r_L^{\ell+1}}\right) > 1 + h\log h > 0,$$

from where we can conclude that $\partial z_L^0/\partial c < 0$. Furthermore, from the fact that $\left\{r_L^\ell\right\}$ is decreasing and $\left\{\partial r_L^\ell/\partial c\right\}$ is increasing (Proposition 5), it follows that for $0 < \ell < L$,

$$\begin{aligned}
\frac{\partial z_L^\ell}{\partial c} = -\frac{1}{\lambda c}\left(1 + \frac{c}{A - r_L^\ell}\frac{\partial r_L^\ell}{\partial c}\right) &< -\frac{1}{\lambda c}\left(1 + \frac{c}{A - r_L^{\ell-1}}\frac{\partial r_L^{\ell-1}}{\partial c}\right) \\
&= \frac{\partial z_L^{\ell-1}}{\partial c} \leq -\frac{1}{\lambda c}\left(1 + \frac{c}{A - r_L^0}\frac{\partial r_L^0}{\partial c}\right) = \frac{\partial z_L^0}{\partial c} < 0,
\end{aligned}$$

and

$$\frac{\partial z_L^L}{\partial c} = -\frac{1}{\lambda c},$$

from where we conclude that the sequence $\left\{\partial z_L^\ell/\partial c\right\}$ is negative and decreasing. □

Proof of Corollary 4. This result is a direct consequence of the functional form of n_L^0 and Corollary 3. □

Proof of Corollary 5. This result is a direct consequence of Corollaries 4 and 3, and the fact that $F(z) = 1 - e^{-\lambda z}$ is an increasing function. $\square$

Proof of Lemma 4. Consider the function $\phi(r)$ defined in (27) and recall that the equilibrium price at layer zero r^* is implicitly defined by $\phi(r^*) = 0$. Consequently, by the implicit function theorem we have

$$\frac{\partial r^*}{\partial A} = -\frac{\partial \phi / \partial A}{\partial \phi / \partial r^*}. \tag{34}$$

From (28), we know that $\frac{\partial \phi}{\partial r^*} < 0$. Thus, it suffices to show that $\frac{\partial \phi}{\partial A} = \frac{\partial w_L^s}{\partial A} - \frac{\partial w_L^p}{\partial A} > 0$ to achieve $\frac{\partial r^*}{\partial A} > 0$. To see this, we first calculate $\frac{\partial w_L^p}{\partial A}$ by differentiating (26) with respect to A:

$$\frac{\partial w_L^p}{\partial A} = 1 - \frac{c}{\lambda(A - r_L^0)} = 1 - e^{-\lambda z_L^0}. \tag{35}$$

Next, we calculate $\frac{\partial w_L^s}{\partial A}$ using the relation $r_L^0(w_L^s(r, A), A) = r$. Differentiating this expression we get

$$\frac{\partial r_L^0}{\partial w_L^s}\frac{\partial w_L^s}{\partial A} + \frac{\partial r_L^0}{\partial A} = 0,$$

which solving for $\frac{\partial w_L^s}{\partial A}$ gives

$$\frac{\partial w_L^s}{\partial A} = -\frac{\partial r_L^0}{\partial A}\left(\frac{\partial r_L^0}{\partial w_L^s}\right)^{-1}. \tag{36}$$

We then calculate $\frac{\partial r_L^0}{\partial A}$ by differentiating the prices in the recurrence relation (18) and considering the first-order conditions of the original maximization problem (6). This yields

$$\frac{\partial r_L^{L-1}}{\partial A} = 1 - \frac{ch}{\lambda A} = 1 - e^{-\lambda z_L^L},$$

and for $0 \leq \ell < L - 1$,

$$\frac{\partial r_L^{\ell-1}}{\partial A} = 1 - e^{-\lambda z_L^\ell}\left(1 - \frac{\partial r_L^\ell}{\partial A}\right).$$

Then, solving the recursion for r_L^0 we get

$$\frac{\partial r_L^0}{\partial A} = 1 - e^{-\lambda z_L^1}\left(e^{-\lambda z_L^2}\left(\dots\left(e^{-\lambda z_L^{L-1}}\left(e^{-\lambda z_L^L}\right)\right)\right)\right),$$

which simplifies to

$$\frac{\partial r_L^0}{\partial A} = 1 - e^{-\lambda \sum_{\ell=1}^{L} z_L^\ell}. \tag{37}$$

Now, substituting this result and (29) into (36), we get

$$\frac{\partial w_L^s}{\partial A} = \frac{1 - e^{-\lambda \sum_{\ell=1}^{L} z_L^\ell}}{h\left(1 + \sum_{\ell=1}^{L-1} \prod_{k=1}^{\ell} \frac{\alpha}{A - r_L^k}\right)}. \tag{38}$$

Using (35) and (38), we can finally compute $\frac{\partial \phi}{\partial A}$, obtaining

$$\begin{aligned}
\frac{\partial \phi}{\partial A} = \frac{\partial w_L^s}{\partial A} - \frac{\partial w_L^p}{\partial A} &= \frac{1 - e^{-\lambda \sum_{\ell=1}^{L} z_L^\ell}}{h\left(1 + \sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=1}^{\ell} z_L^k}\right)} - 1 + e^{-\lambda z_L^0} \\
&= \frac{1 - e^{-\lambda \sum_{\ell=1}^{L} z_L^\ell} - h - h\sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=1}^{\ell} z_L^k} + h\sum_{l=0}^{L-1} e^{-\lambda \sum_{k=0}^{\ell} z_L^\ell}}{h\left(1 + \sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=1}^{\ell} z_L^k}\right)} \\
&= \frac{1 - \frac{ch}{\lambda A} e^{-\lambda \sum_{\ell=1}^{L-1} z_L^\ell} - h - h\sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=1}^{\ell} z_L^k} + he^{-\lambda \sum_{\ell=0}^{L-1} z_L^\ell} + h\sum_{\ell=0}^{L-2} e^{-\lambda \sum_{k=0}^{\ell} z_L^\ell}}{h\left(1 + \sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=1}^{\ell} z_L^k}\right)} \\
&= \frac{1 - h + h\left(e^{-\lambda \sum_{\ell=0}^{L-1} z_L^\ell} - \frac{c}{\lambda A} e^{-\lambda \sum_{\ell=1}^{L-1} z_L^\ell}\right) + h\left(\sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=0}^{\ell-1} z_L^\ell} - e^{-\lambda \sum_{k=1}^{\ell} z_L^k}\right)}{h\left(1 + \sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=1}^{\ell} z_L^k}\right)}.
\end{aligned}$$

To show $\frac{\partial \phi}{\partial A} > 0$, note that

$$S_1 e^{-\lambda \sum_{\ell=0}^{L-1} z_L^\ell} - \frac{c}{\lambda A} e^{-\lambda \sum_{\ell=1}^{L-1} z_L^\ell} = e^{-\lambda \sum_{\ell=1}^{L-1} z_L^\ell}\left(\frac{c}{\lambda(A - r_L^0)} - \frac{c}{\lambda A}\right) = \frac{c r_L^0 e^{-\lambda \sum_{\ell=1}^{L-1} z_L^\ell}}{\lambda A(A - r_0^L)} > 0,$$

$$S_2 \sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=0}^{\ell-1} z_L^k} - e^{-\lambda \sum_{k=1}^{\ell} z_L^k} = \sum_{\ell=1}^{L-1} e^{-\lambda \sum_{k=1}^{\ell-1} z_L^k}\left(e^{-\lambda z_L^0} - e^{-\lambda z_L^\ell}\right) > 0,$$

since $\{z_L^\ell\}$ is increasing. Thus, we conclude $\frac{\partial r^*}{\partial A} > 0$. Moreover, we can get an explicit expression for $\frac{\partial r^*}{\partial A}$, obtaining

$$\frac{\partial r^*}{\partial A} = \frac{1 - h + hS_1 + hS_2}{1 + h\sum_{\ell=0}^{L-1} e^{-\lambda Z_L^\ell}}.$$

□

Proof of Proposition 6. The net earnings per worker at each layer are given by w_L^p as defined in (26). Differentiating this expression with respect to A yields:

$$\frac{\partial w_L^p}{\partial A} = 1 - \frac{c}{\lambda\left(A - r_L^0\right)}\left(1 - \frac{\partial r_L^0}{\partial A}\right).$$

We know from Lemma 4 that $\gamma 1 - \frac{\partial r_0}{\partial A} < 1$, but to further analyze the sign of this derivative we consider two cases:

1. If $\gamma > 0$, we have

$$\frac{\partial w_L^p}{\partial A} = 1 - \frac{c\gamma}{\lambda\left(A - r_L^0\right)} > 1 - \frac{c}{\lambda\left(A - r_L^0\right)} > 1 - 1 = 0.$$

2. If $\gamma < 0$, then

$$\frac{\partial w_L^p}{\partial A} = 1 - \frac{c\gamma}{\lambda\left(A - r_L^0\right)} > 1 > 0.$$

Thus, in both cases, we obtain $\frac{\partial w_L^p}{\partial A} > 0$. This shows that the net earnings per worker at each layer increase with A. Next, we consider the expression for the derivative of r_L^0 with respect to A. Using the relationship from (34), we have:

$$\frac{\partial r_L^0}{\partial A} = -\frac{\frac{\partial w_L^s}{\partial A} - \frac{\partial w_L^p}{\partial A}}{\frac{\partial \phi}{\partial r}} < -\frac{\partial w_L^s/\partial A}{\partial \phi/\partial r}.$$

Substituting the known terms, we get

$$\frac{\partial r_L^0}{\partial A} = \frac{1 - e^{-\lambda \sum_{\ell=1}^{L} z_L^\ell}}{1 + h\sum_{\ell=1}^{L-1} e^{-\lambda Z_L^\ell}} < 1. \tag{39}$$

Therefore, the first case we considered ($\gamma > 0$) necessarily holds, and we conclude

$$0 < \frac{\partial w_L^p}{\partial A} < 1.$$

$\square$

Proof of Proposition 7. We analyze how the equilibrium prices r_L^ℓ respond to changes in the technology level A by studying their derivatives with respect to A.

First, consider the top layer price r_L^{L-1}. Using the recurrence relation from equation (18), we obtain:

$$\frac{\partial r_L^{L-1}}{\partial A} = 1 - h\frac{\partial w_L^L}{\partial A} - \frac{\alpha}{A}.$$

By Proposition 6, we have $0 < \frac{\partial w_L^L}{\partial A} < 1$, implying that

$$\frac{\partial r_L^{L-1}}{\partial A} < 1.$$

Next, consider a general layer $\ell < L-1$. Applying the recurrence relation again yields:

$$\frac{\partial r_L^{\ell-1}}{\partial A} = 1 - h\frac{\partial w_L^\ell}{\partial A} - \frac{\alpha}{A - r_L^\ell}\left(1 - \frac{\partial r_L^\ell}{\partial A}\right).$$

Since $0 < \frac{\partial w_L^\ell}{\partial A} < 1$ and $\frac{\partial r_L^\ell}{\partial A} < 1$ (as shown for $\ell = L-1$), the right-hand side remains strictly less than 1.

Therefore, by backward induction on ℓ, we conclude that:

$$\frac{\partial r_L^\ell}{\partial A} < 1 \quad \text{for all } 0 \leq \ell < L.$$

$\square$

Proof of Corollary 6. To prove this result we differentiate each of the equations in (6) with respect to A, obtaining

$$\frac{dz_L^0}{dA} = \frac{1 - \frac{\partial r_L^0}{\partial A}}{\lambda\left(A - r_L^0\right)} > 0, \;\; \frac{\partial z_L^L}{\partial A} = \frac{1}{\lambda A} > 0,$$

and for $0 < \ell < L$,

$$\frac{dz_L^\ell}{\partial A} = \frac{1 - \frac{\partial r_L^\ell}{\partial A}}{\lambda\left(A - r_L^\ell\right)} > 0,$$

thanks to the result achieved in proposition (7). $\square$

Proof of Corollary 7. This result is a direct consequence of the functional form of n_L^0 and Corollary 6. $\square$

Proof of Corollary 8. This result is a direct consequence of Corollaries 7 and 6, and the fact that $F(z) = 1 - e^{-\lambda z}$ is an increasing function. $\square$

Proof of Proposition 1. We start by defining the set

$$w = \left(\theta/\lambda + \theta^k L, \frac{A\lambda}{1/h + \log h}\right]$$

and the function $f : \overline{w} \to \mathbb{R}, c \mapsto q - An_L^0 F(Z_L^L)$, which is well-defined as any $c \in w$ satisfies (A1). Note then that $\Gamma \cup \{\theta/\lambda + \theta^k L\} = f^{-1}\left((-\infty, 0]\right)$, because $\Gamma \neq \emptyset$ implies that

$f(\theta/\lambda + \theta^k L) \leq 0$, thanks to Corollary 5. With this, c^* is the solution of the optimization problem (12) if and only if it satisfies

$$c^* = \sup f^{-1}\left((-\infty, 0]\right), \tag{40}$$

as we are interested in minimizing $\mathcal{P}(c, \lambda, L)$, and this is achieved by setting c to its maximum possible value. Now, given the continuity of f, $f^{-1}\left((-\infty, 0]\right)$ is closed in $\overline{w}$, and therefore compact (closed and bounded) in $\mathbb{R}$. As such, $c^* \in f^{-1}\left((-\infty, 0]\right)$, but as $\Gamma \neq \emptyset$, $c^* \in \Gamma$. $\square$

Proof of Corollary 9. The hypotheses of this result directly imply that $\Gamma \neq \emptyset$, from where we can apply Proposition 1 to deduce the desired result. $\square$

Proof of Lemma 1. The proof proceeds in two parts. First, we show that if, at any given time t, the firm optimally chooses to retain the same number of layers (i.e., $L_t = L_{t+1}$), then the sequence of layers will remain constant from that point onward. In the second part, we show that if the firm chooses to increase the number of layers at time t (i.e., $L_{t+1} = L_t + 1$), it cannot subsequently decrease the number of layers in the next period (i.e., $L_{t+2} \neq L_t$), and vice versa. Both these results imply that the sequence of layers $\{L_t\}$ is monotonic, either strictly increasing or strictly decreasing by one layer each period until it stabilizes.

To prove the first part, assume that there exists some $n \in \mathbb{N}$ such that $L_{n+1} = L_n$. If at time n the firm decides to maintain the same training costs, i.e., $c_{n+1} = c_n$, then the optimization problem at time $n+1$ is identical to the one at time n. This implies that the solutions must also be identical, leading to the equality $L_{n+2} = L_{n+1}$.

Now, let's consider the case in which the firm adjusts its training costs at time n. Here, the cost function at time n takes the value

$$\phi_n = \mathcal{P}(c_{n+1}, \lambda, L_n) + \delta = \min_{L \in \{L_n - 1, L_n, L_n + 1\}} \left\{ \min_c \left[\mathcal{P}(c, \lambda, L) + \delta\right] \text{ s.t. } c \in \Gamma_{A,\lambda,L} \right\}$$

and since $L_{n+1} = L_n$, this implies that

$$\mathcal{P}(c_{n+1}, \lambda, L_{n+1}) \leq \min_{L \in \{L_{n+1} - 1, L_{n+1}, L_{n+1} + 1\}} \left\{ \min_c \mathcal{P}(c, \lambda, L) \text{ s.t. } c \in \Gamma_{A,\lambda,L} \right\}. \tag{41}$$

Given that c_{n+1} and L_{n+1} form a feasible solution for the optimization problem at time $n+1$, inequality (41) allows us to conclude that $\phi_{n+1} = \mathcal{P}(c_{n+1}, \lambda, L_{n+1})$, from where it follows that $L_{n+2} = L_{n+1}$. By induction, we get that $L_m = L_n$ must hold for all $m \geq n$, meaning that the sequence of layers remains constant from this point onward.

Now, for the second part of the proof, assume that for some $n \in \mathbb{N}$, $L_{n+1} = L_n + 1$. If at time n the firm chooses to maintain the same training costs, i.e., $c_{n+1} = c_n$, then the cost function at time n satisfies

$$\phi_n = \mathcal{P}(c_n, \lambda, L_n + 1) \le \min \left\{ \begin{array}{c} \mathcal{P}(c_n, \lambda, L_n), \\ \min_c \left[\mathcal{P}(c, \lambda, L_n) + \delta\right] \text{ s.t. } c \in \Gamma_{A,\lambda,L_n} \end{array} \right\} \tag{42}$$

Since $c_{n+1} = c_n$ and $L_{n+1} = L_n + 1$ form a feasible solution for the problem at time $n+1$, then the cost function at this time is such that $\phi_{n+1} \le \mathcal{P}(c_n, \lambda, L_n + 1)$. Because of inequality (42), this in turn implies that L_n cannot be the optimal number of layers at time $n+1$.

Now consider the case where the firm updates the training cost at time n. In this case, we have

$$\phi_n = \mathcal{P}(c_{n+1}, \lambda, L_n + 1) + \delta \le \min_c \left[\mathcal{P}(c, \lambda, L_n) + \delta\right] \text{ s.t. } c \in \Gamma_{A,\lambda,L_n},$$

which leads us to the inequality

$$\mathcal{P}(c_{n+1}, \lambda, L_n + 1) \le \min_c \mathcal{P}(c, \lambda, L_n) \text{ s.t. } c \in \Gamma_{A,\lambda,L_n}.$$

Since $L_n + 1 = L_{n+1}$, if $c_{n+1} \in \Gamma_{A,\lambda,L_n}$, we have that

$$\phi_{n+1} \le \mathcal{P}(c_{n+1}, \lambda, L_n + 1) \le \min \left\{ \begin{array}{c} \mathcal{P}(c_{n+1}, \lambda, L_n), \\ \min_c \left[\mathcal{P}(c, \lambda, L_n) + \delta\right] \text{ s.t. } c \in \Gamma_{A,\lambda,L_n} \end{array} \right\}. \tag{43}$$

This implies that $L_{n+2} \ne L_n$. If $c_{n+1} \notin \Gamma_{A,\lambda,L_n}$, then c_{n+1} and L_n are not feasible together, leading once again to $L_{n+2} \ne L_n$. Thus, we conclude that the firm cannot decrease the number of layers in the period following an increase. An analogous argument shows that the firm cannot increase the number of layers immediately after a decrease. Therefore, the sequence $\{L_t\}$ must be monotonic, either strictly increasing or strictly decreasing until it reaches a steady state. □

Proof of Proposition 2. We will demonstrate that, under Assumption 2, the sequences of training costs and number of layers converge to finite values, denoted as c_∞ and L_∞, respectively, as $t \to \infty$. Additionally, we show that this convergence occurs within a finite number of steps. To establish this, we first prove that the sequence of layers $\{L_n\}$ is bounded and convergent, and then deduce that, from a certain point onward, the sequence of training costs $\{c_n\}$ becomes constant.

First note that, under Assumption 2, for each $n \in \mathbb{N}^+$, c_n and L_n form a feasible solution for the optimization problem governing the cost function at time n. This implies that

$c_n \in \Gamma_{A,\lambda,L_n}$, so by the definition of the set Γ_{A,λ,L_n} (13) we have

$$\theta/\lambda + \gamma L_n < c_n \leq \frac{A\lambda}{1/h + \ln h}.$$

From this, it follows that

$$L_n < \frac{A\lambda}{\gamma(1/h + \ln h)} - \frac{\theta}{\gamma\lambda}$$

Thus, the sequence $\{L_n\}$ is bounded. Since $\{L_n\}$ is also monotonic by lemma (1), it must converge to a finite value L_∞.

Since the sequence of the number of layers consists of positive integers, there exists an $m \in \mathbb{N}$ such that $L_n = L_\infty$ for all $n \geq m$. At this point, the optimization problem governing the cost function at time n can be expressed as

$$\phi_n = \min \left\{ \begin{array}{c} \mathcal{P}(c_n, \lambda, L_\infty), \\ \min_c \left[\mathcal{P}(c, \lambda, L_\infty) + \delta\right] \text{ s.t. } c \in \Gamma_{A,\lambda,L_\infty} \end{array} \right\}. \tag{44}$$

If the optimal decision at time m is to maintain the current training costs, the optimization problem for the cost function will remain unchanged from that point onward, implying that $\{c_n\}$ will become constant. On the other hand, if the firm chooses to update the training costs at time m, we would have

$$\phi_m = \mathcal{P}(c_{m+1}, \lambda, L_\infty) + \delta = \min_c \left[\mathcal{P}(c, \lambda, L_\infty) + \delta\right] \text{ s.t. } c \in \Gamma_{A,\lambda,L_\infty}. \tag{45}$$

From this and equation (44) it follows that

$$\phi_{m+1} \leq \mathcal{P}(c_{m+1}, \lambda, L_\infty) = \min \left\{ \begin{array}{c} \mathcal{P}(c_{m+1}, \lambda, L_\infty), \\ \min_c \left[\mathcal{P}(c, \lambda, L_\infty) + \delta\right] \text{ s.t. } c \in \Gamma_{A,\lambda,L_\infty} \end{array} \right\} = \phi_{m+1}.$$

Thus, at time $m+1$, the firm will optimally choose to keep the same training costs, and from this point forward, the optimization problem will always yield the same solution. Therefore, we also conclude that the sequence $\{c_n\}$ converges to a finite value c_∞. □

Proof of Lemma 2. Let $\pi(A, \lambda, c, L)$ be the output of the firm at the steady state. Then, the firm meets the production constraint, so that $\pi(A, \lambda_0, c, L) \geq q$. Now, recall that

$$\lim_{\lambda \to 0} \pi(\tilde{A}, \lambda, c, L) = \lim_{c \to \infty} \pi(\tilde{A}, \lambda, c, L) = 0, \tag{46}$$

that is, if the problems' complexity or the training cost becomes infinite, there is no production. Thus, by continuity, and due to the decreasing nature of the production function with respect to c proved in Corollary 5, there exists some λ_+ for which $\pi(\tilde{A}, \lambda_+, \theta/\lambda_+ + \gamma L, L) = q$ and taking $\lambda < \lambda_+$ will lead to $\pi(\tilde{A}, \lambda, \theta/\lambda + \gamma L, L) < q$.

Now we take the smallest $\lambda_- < \lambda_+$ such that $\pi(\tilde{A}, \lambda_-, \theta/\lambda_- + \gamma(L-1), L-1) = q$, so that any $\lambda \in \Lambda \equiv (\lambda_-, \lambda_+)$ satisfies Assumption 3. $\square$

Proof of Proposition 3. As the firm is already in its steady state, we expect that the number of layers remain constant for $t < \tau^*$. Now, at the moment of the complexity shock ($t = \tau^*$), Assumption 3 prevents the firm from maintaining its current number of layers or increasing them. If the firm were to do so, it would result in a configuration where the required training costs fall below a critical threshold, leading to an infinite cost scenario. Consequently, the firm is compelled to reduce its number of organizational layers.

Now, according to Lemma 1, the number of organizational layers must remain non-increasing during the period $\tau^* t < \tau^* + T$. This implies that the firm either maintains its reduced structure or potentially further streamlines it throughout the learning-by-doing phase. Upon the conclusion of the learning-by-doing period (at $t = \tau^* + T$), the firm is anticipated to increase its number of layers. This occurs because the complexity returns to its initial level ($\lambda_2 = \lambda_0$), and because the firm had previously established its steady-state for the initial innovation level A at (c, L).

Finally, Lemma 1 implies that the sequence of organizational layers will be non-decreasing from this point onward. $\square$

Proof of Lemma 5. Let L^* and c^* denote the optimal solutions to the problem

$$\min_{L' \in \{L-2, L-1, \ldots, L+2\}} \min_{c' \in \Gamma_{L'}} \mathcal{P}(c', \lambda, L').$$

By the hypothesis that $c \in \bigcup_{L' \in \{L-1, L, L+1\}} \Gamma_{L'}$, the optimal value of this problem is a real positive number. Let L'_0 take the smallest value in the set $\{L-1, L, L+1\}$ such that $c \in \Gamma_{L'_0}$. By choosing

$$\delta > \mathcal{P}(c, \lambda, L'_0) - \mathcal{P}(c^*, \lambda, L^*),$$

we ensure that c remains the optimal solution to problem 14 at time $t = 1$ and thereafter. This implies that, for the selected adjustment cost and any larger adjustment cost, it is optimal for the firm to maintain its training costs unchanged throughout the cycle. $\square$

Proof of Lemma 3. For an arbitrary number of layers $L' \in \mathbb{N}$, consider the production function with fixed technology $\tilde{A}$ and training cost c. This function is positive and continuous in λ throughout its domain and satisfies

$$\lim_{\lambda \to 0} \pi(\tilde{A}, \lambda, c, L') = 0.$$

Thus, at $L-1$ layers, there exists some $\lambda' > 0$ such for any $\lambda < \lambda'$ we have

$$0 < \pi(\tilde{A}, \lambda, c, L-1) < q.$$

Similarly, at L layers, there exists some $\lambda'' > 0$ such that for any $\lambda < \lambda''$,

$$0 < \pi(\tilde{A}, \lambda, c, L) < q.$$

Now, define the open intervals

$$\Lambda(0, \min\{\lambda', \lambda_0\}) \quad \text{and} \quad \tilde{\Lambda}(0, \min\{\lambda', \lambda''\}). \tag{47}$$

Both intervals are nonempty since all possible upper bounds are strictly positive. By construction, we can see that all $\lambda_1 \in \Lambda$ satisfies condition 1, while every $\lambda_1 \in \tilde{\Lambda}$ satisfies both conditions in Assumption 4. □

Proof of Proposition 4. Since $c \in \bigcup_{L' \in \{L-1, L, L+1\}} \Gamma_{L', \tilde{A}, \tilde{\lambda}}$, the firm can meet production requirements using the initial training cost c and some nearby hierarchical configuration—specifically with $L-1$, L, or $L+1$ layers. Lemma 5 then guarantees the existence of a threshold $\delta' > 0$ such that, for all $\delta > \delta'$, the firm maintains a constant training cost throughout the learning phase. That is, $c^* = c$.

Moreover, the first condition in Assumption 4 states that the training cost c is not sufficient for a flatter hierarchy with $L-1$ layers to meet the production target. As a result, the firm must either retain L layers or increase to $L+1$ at the beginning of the learning phase. By Lemma 1, it follows that $L^* \geq L$, completing the proof. □

Proof of Corollary 1. As in the previous proof, since $c \in \bigcup_{L' \in \{L-1, L, L+1\}} \Gamma_{L', \tilde{A}, \tilde{\lambda}}$, Lemma 5 ensures the existence of a threshold $\delta' > 0$ such that, for all $\delta > \delta'$, the training cost remains fixed, i.e., $c^* = c$.

Under the stronger condition of Assumption 4, neither $L-1$ nor L layers are sufficient to meet the production target with training cost c. Therefore, the firm must increase its hierarchy to $L+1$ at the onset of the learning phase. By Lemma 1, it follows that $L^* > L$. □

Figure B7: Joint distributions of the calibrated parameter pairs starting from the 41st iteration.

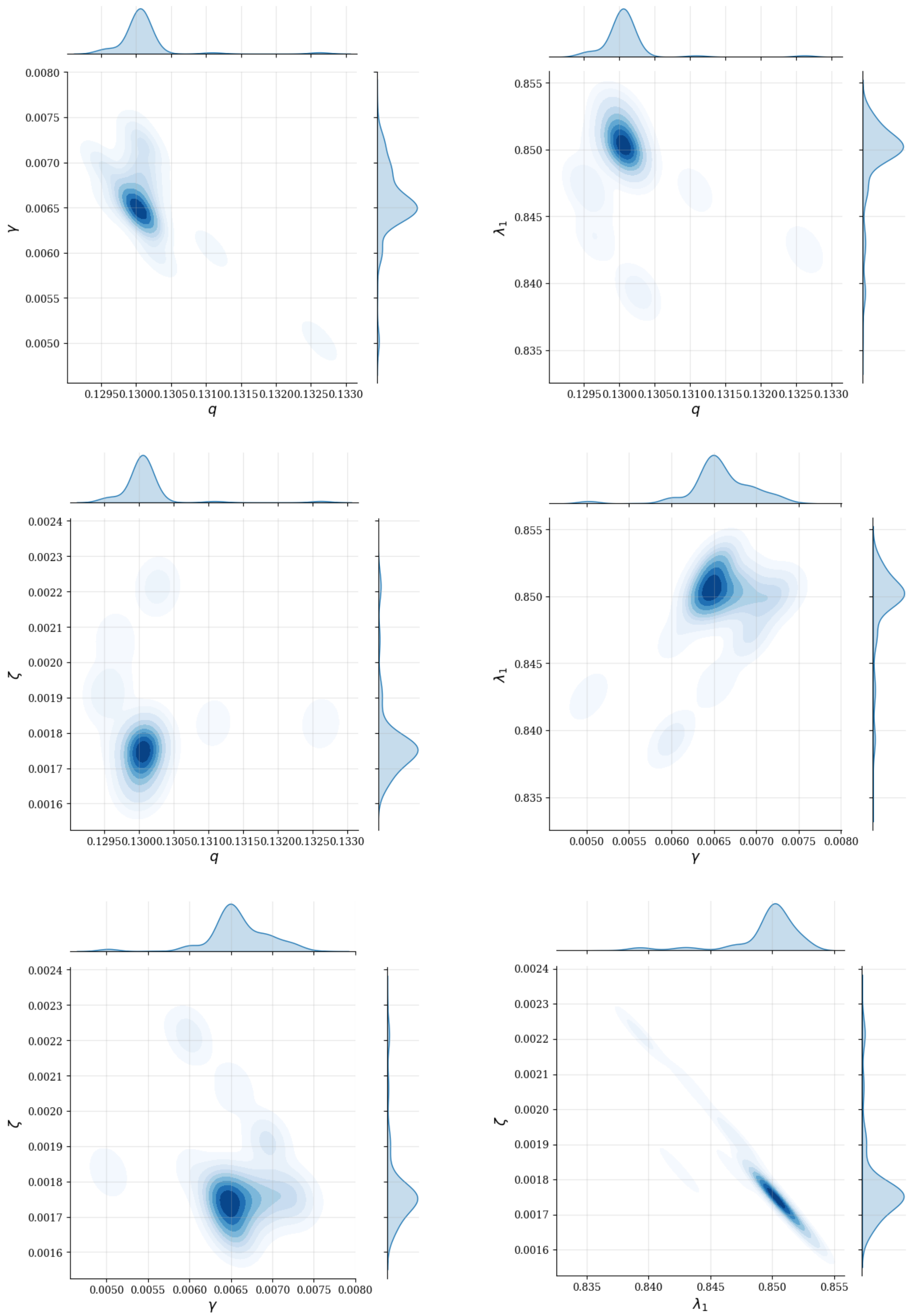


Figure B8: Calibration results for a product cycle with a 34-period time horizon, innovation at $\tau = 17$, and a learning phase of $T = 1$.

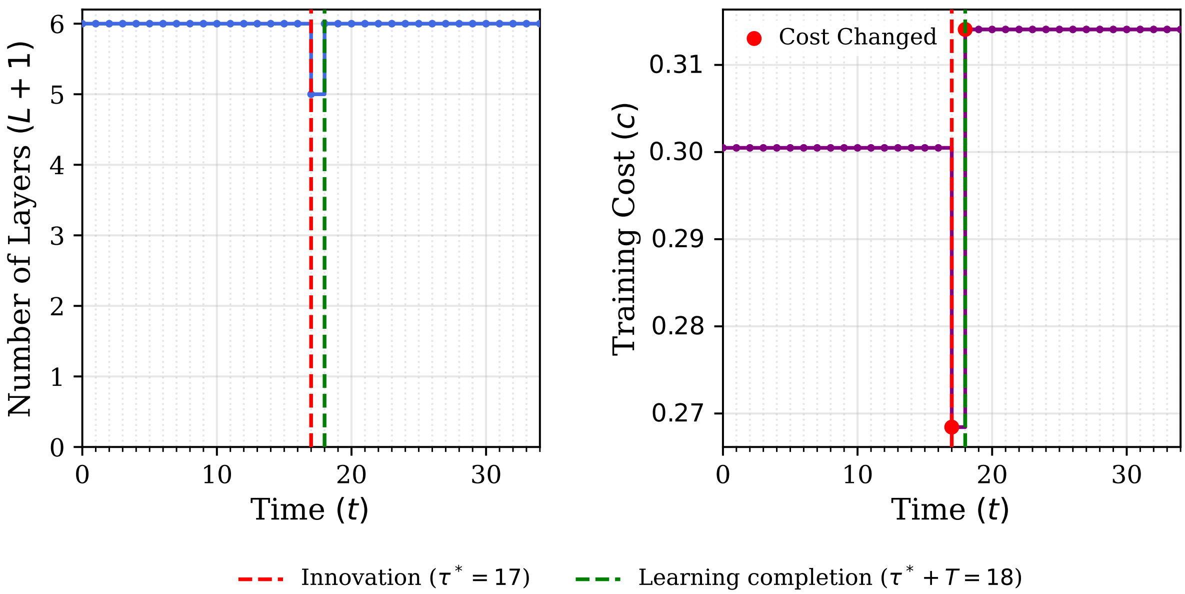
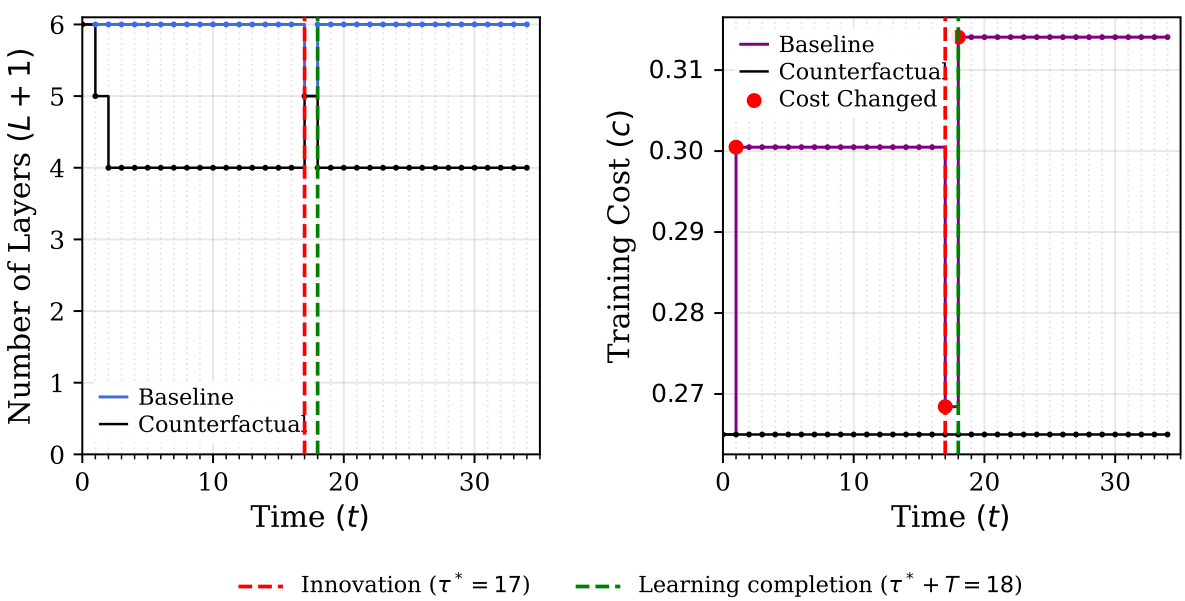
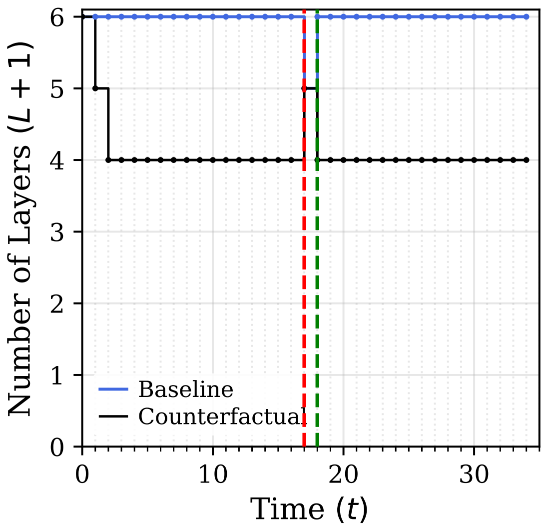


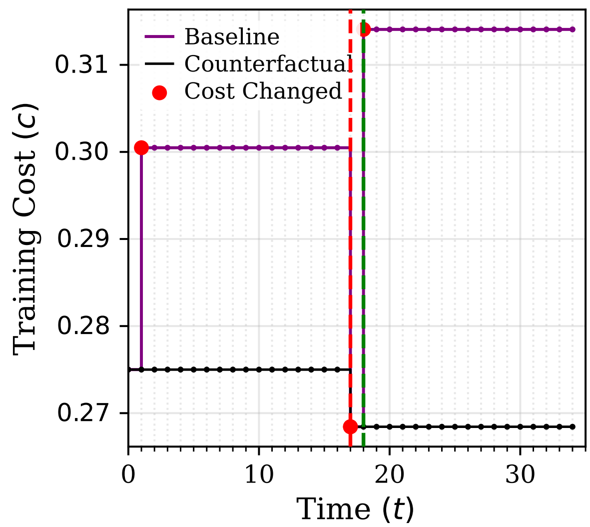
Figure B9: Organizational dynamics under a low initial training cost scenario ($c_0 = 0.265$) with fixed initial layers $L_0 = 5$.



Notes: The baseline case ($\delta = 0$) and the high adjustment cost case ($\delta = 10^6$) are shown.

Figure B10: Organizational dynamics under a high initial training cost scenario ($c_0 = 0.275$) with fixed initial layers $L_0 = 5$.

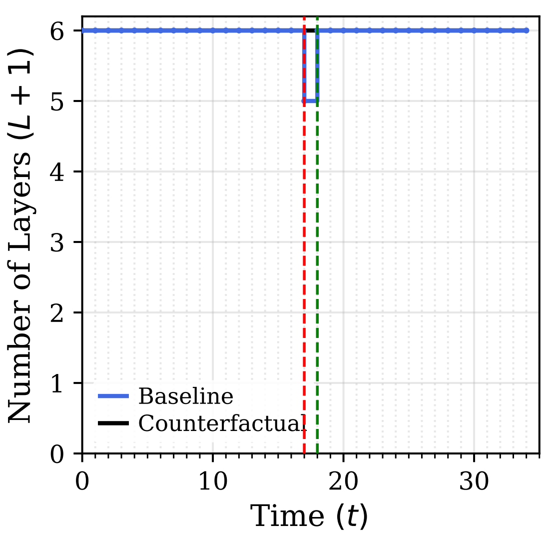


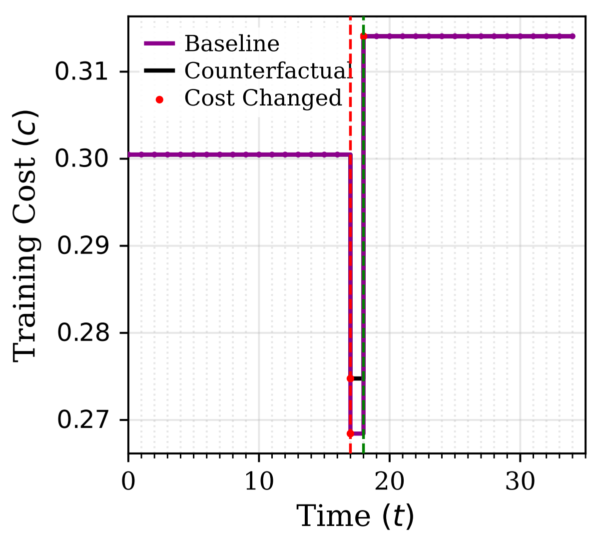


Innovation ($\tau^* = 17$) Learning completion ($\tau^* + T = 18$)

Notes: The baseline case ($\delta = 0$) and the high adjustment cost case ($\delta = 10^6$) are shown.

Figure B11: Organizational dynamics under the counterfactual scenario with fixed organizational layers ($L = 5$).





Innovation ($\tau^* = 17$) Learning completion ($\tau^* + T = 18$)