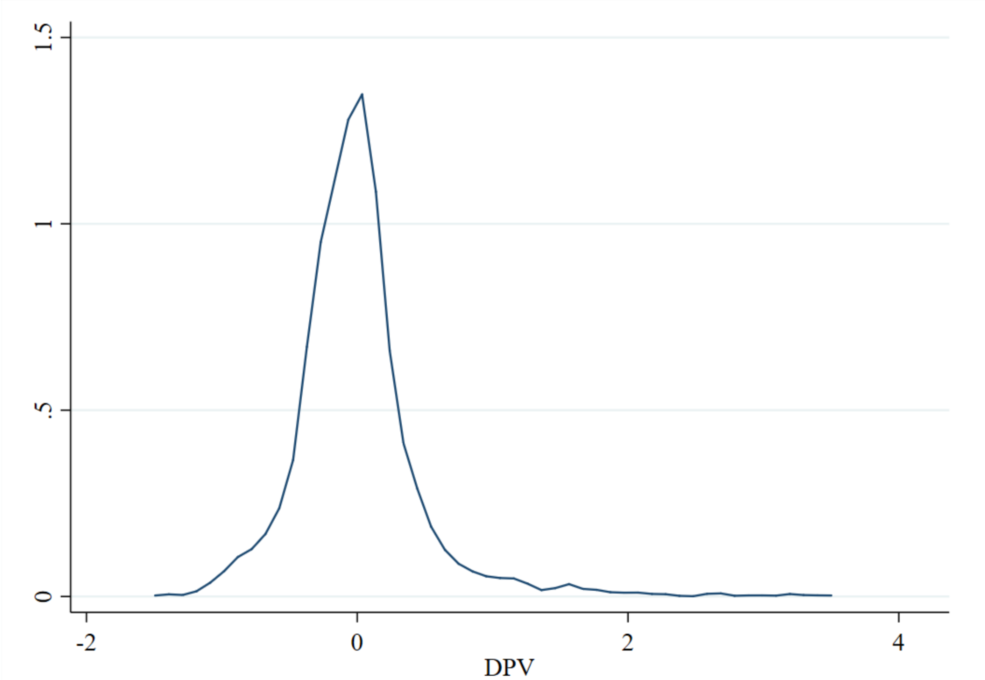
Online Appendix

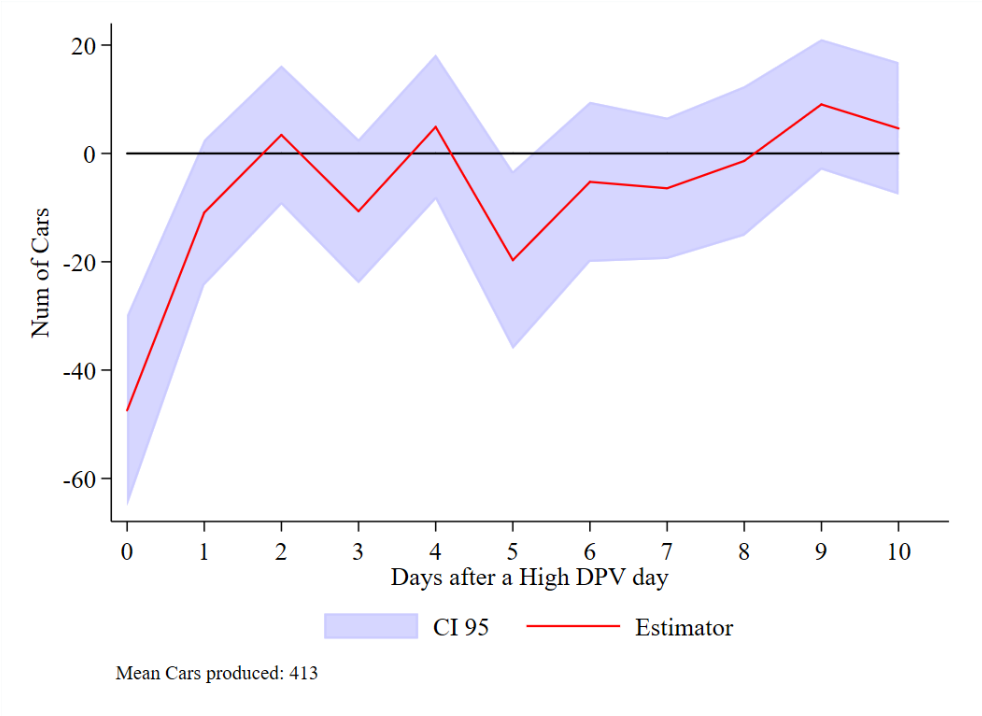
A Tables and Figures

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



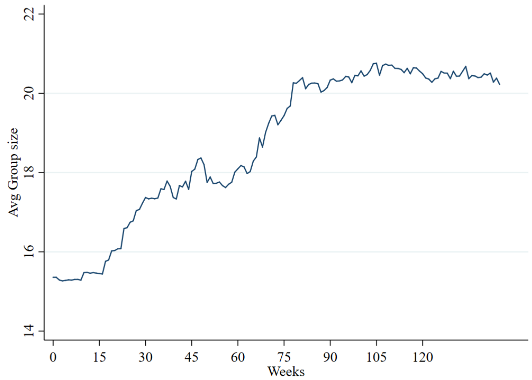
Note: Figure A1 plots the distribution of DPV-day observations, pooling across all days in the data. DPV is a standardized variable with mean 0 and standard deviation 1.

Figure A2: DPV and Plant-level Productivity Losses

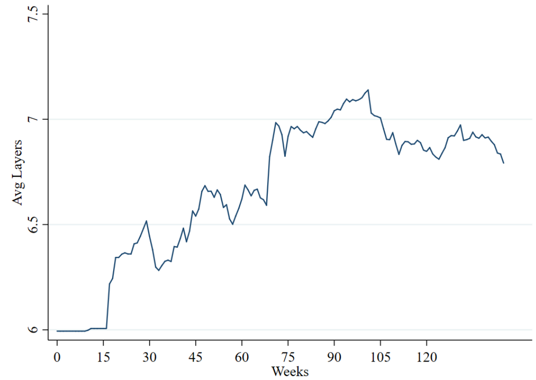


Note: Figure A2 plots the coefficients of a Distributed Lag Model of order 10. We define a High DPV day as a day when DPV goes over 1 SD above the mean in our sample. We control for a quadratic trend, year and month fixed effects, and first lag of number of cars produced. The cumulative effect of the High DPV occurrence over the 10 days is -79.799 cars (with robust standard error 10.157). Confidence Intervals are computed using robust standard errors.

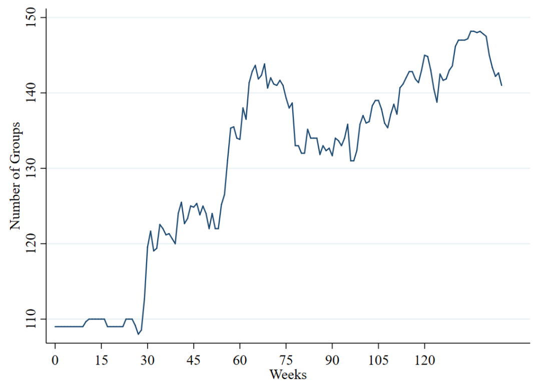
Figure A3: Average Size of Working Groups, Average Number of Layers and Number of Working Groups over Time



(a) Average size



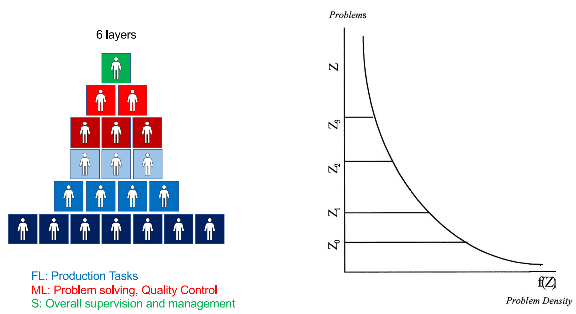
(b) Average number of layers



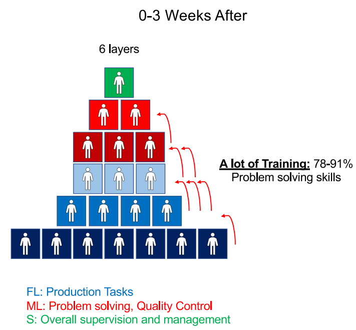
(c) Number of working groups

Note: Figure A3a plots the weekly average working group size in our period of analysis. Figure A3b plots the weekly average number of layers per working group in our period of analysis. Figure A3c plots the weekly number of working groups in our period of analysis.

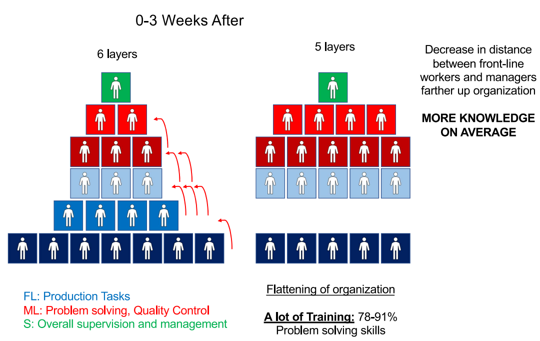
Figure A4: Schematic of Working Group Composition Movement after a Model Change



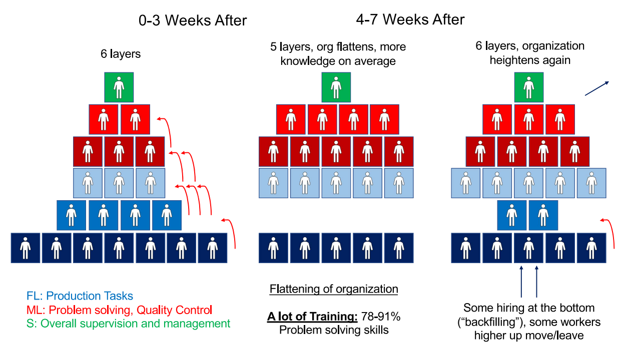
(a) Baseline



(b) Training



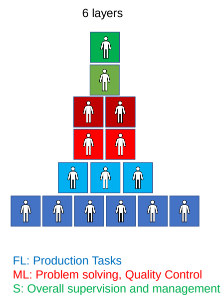
(c) Promotion at the top



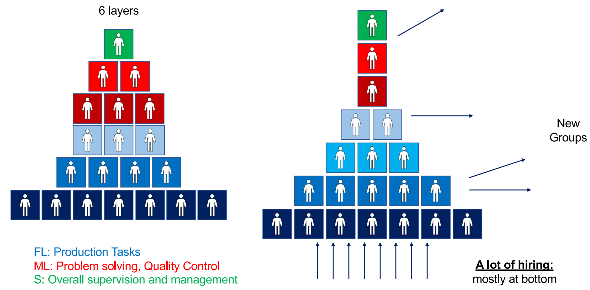
(d) Promotion at the bottom (back-filling)

Figure A4 shows an example of how the working group composition changes after a Model change, illustrating our empirical results.

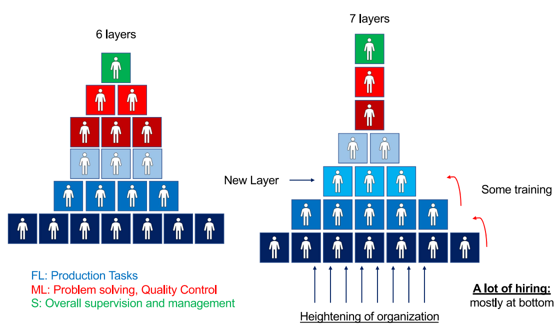
Figure A5: Schematic of Working Group Composition Movement after a Volume Change



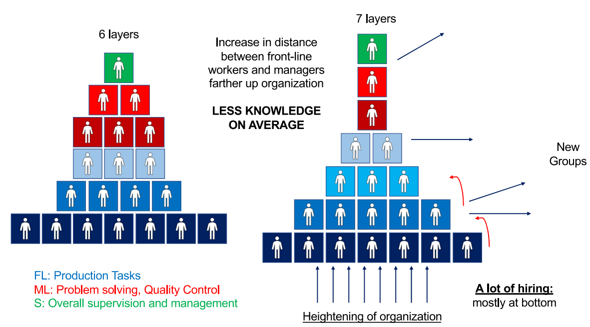
(a) Baseline



(b) Hiring and new groups formation



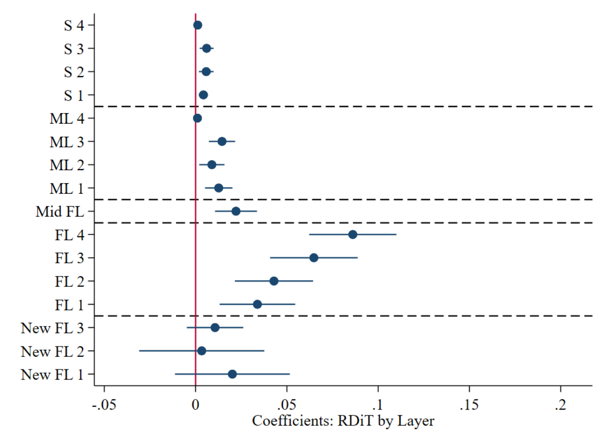
(c) Promotion at the bottom



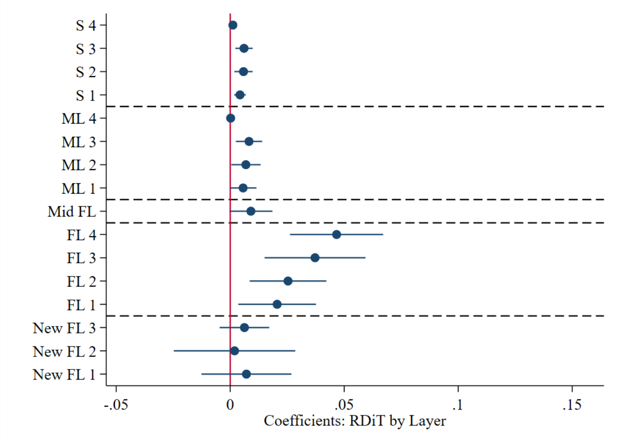
(d) Less knowledge on average

Figure A5 shows an example of how the working group composition changes after a Volume change, illustrating our empirical results.

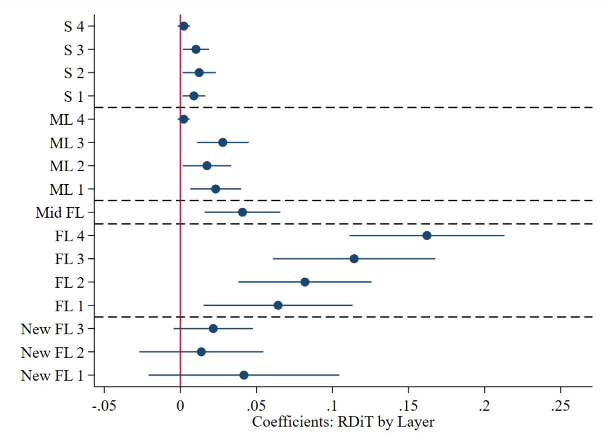
Figure A6: Impact of Model Changes on Avg Num. of Courses and Employees Trained in All Courses and Problem-Solving and Communication Specific Content (0-3 weeks)



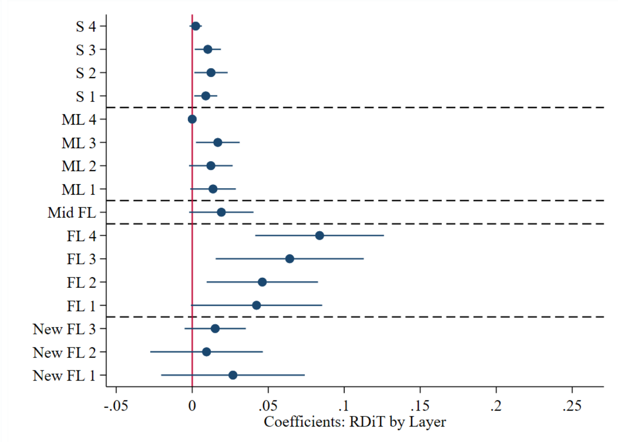
(a) Num. Courses, All courses



(b) Num. Courses, Problem-solving



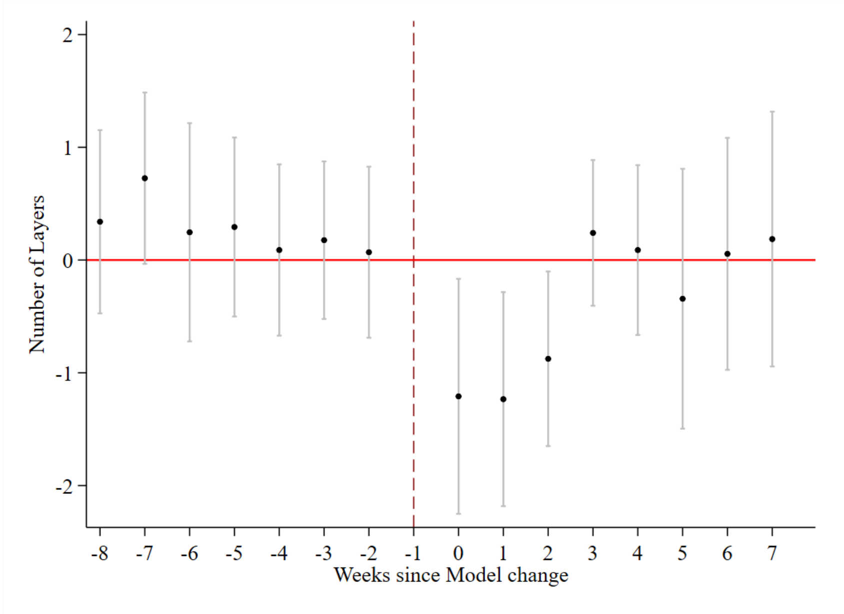
(c) Num. Employees, All courses



(d) Num. Employees, Problem-solving

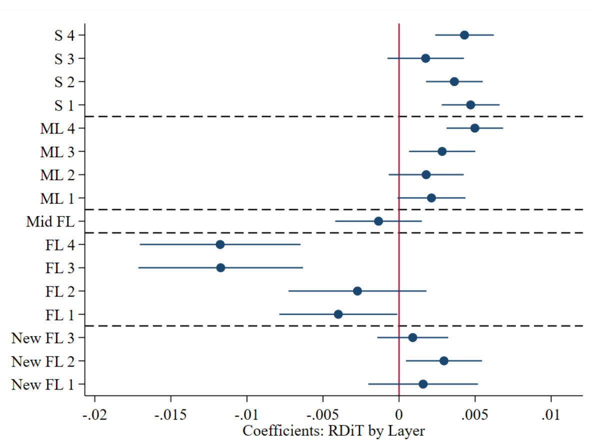
Figures A6a and A6b show the effect of Model changes on the average number of courses taken by workers by layer at 0-3 weeks post Model change in all courses and courses with problem-solving and communication content, respectively. Figures A6c and A6d show the effect of Model changes on the number of 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 Model change 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 Number of Layers

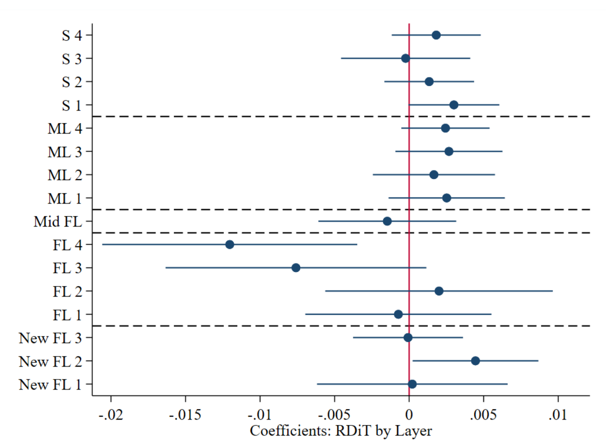


Note: Figure A7 shows the effect of model changes on the number of layers within working groups in a time window running from 8 weeks before the event to 8 weeks after the event (where the week of the Model change is labelled as week 0 on the x-axis). Number of layers is defined as the number of separate 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 Model change and to all other Model and Volume changes and a linear time trend. Standard errors are clustered at week-working group level. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

Figure A8: Impact of Model Changes on Working Group Structure and Knowledge Hierarchies (Disaggregated)



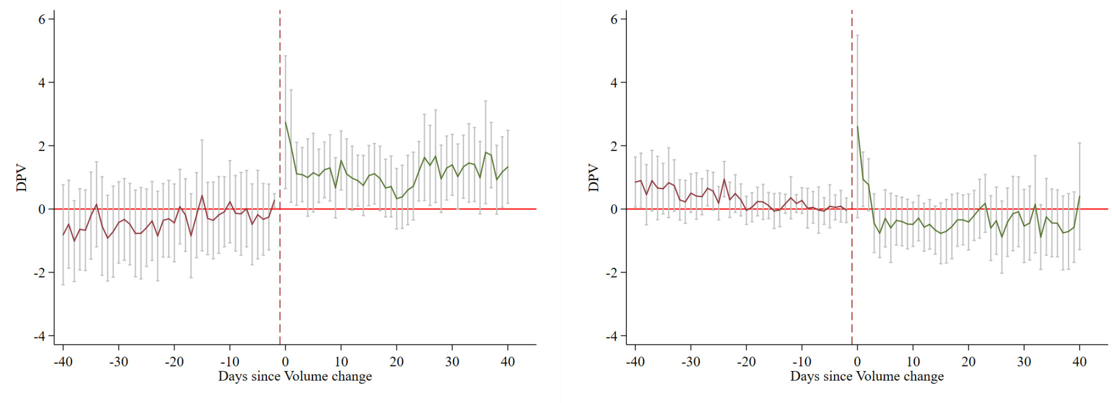
(a) Share of Workers by Layer, 0-3 weeks



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

Figures A8a and A8b 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 Model change and working group. 95% confidence intervals are presented in the figure. Number of observations: 220 working groups x 16 weeks x 2 events.

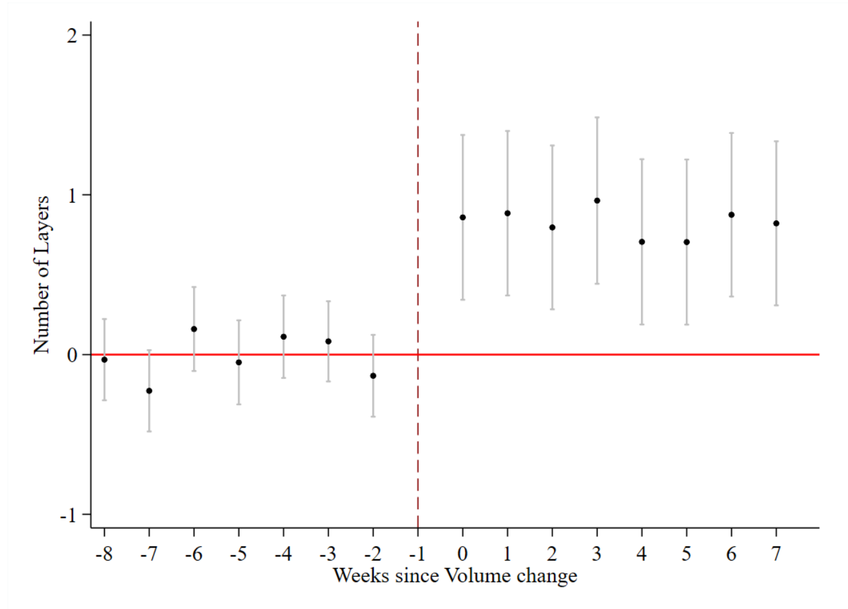
Figure A9: Event Study of Volume Changes on Productivity by Hiring



(a) High Hiring (b) Low Hiring

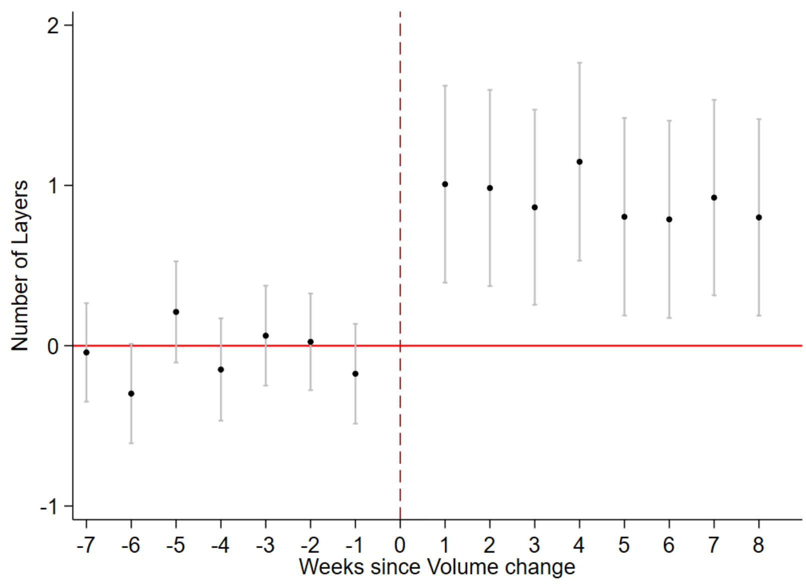
Note: Figure A9 shows the effect of Volume changes on DPV in a time window running from 40 days before the event to 40 days after the event split between changes with high number of employees hired (above the median) and changes with low number of employees hired (below the median). DPV is computed as 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-shift level. 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 A10: Event Study of Volume Changes on Layers



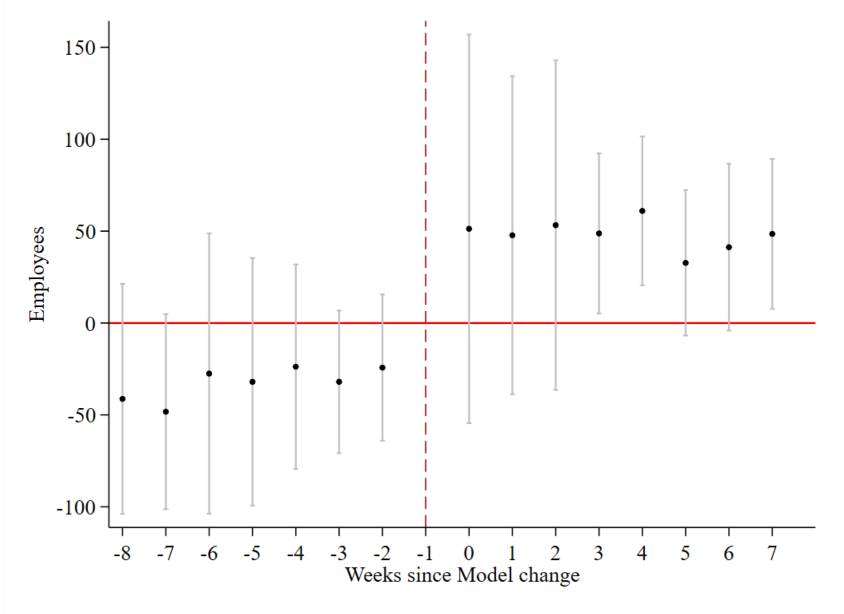
Note: Figure A10 shows the effect of Volume changes on the number of layers within working groups in a time window running from 8 weeks before the event to 8 weeks after the event (where the week of the Volume change is labelled as week 0 on the x-axis). Number of layers is defined as the number of separate 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 at week-working group level. 95% confidence intervals are reported. Number of observations: 220 working groups x 16 weeks x 2 events.

Figure A11: Event Study of Volume Changes on Layers for Pre-Existing Groups



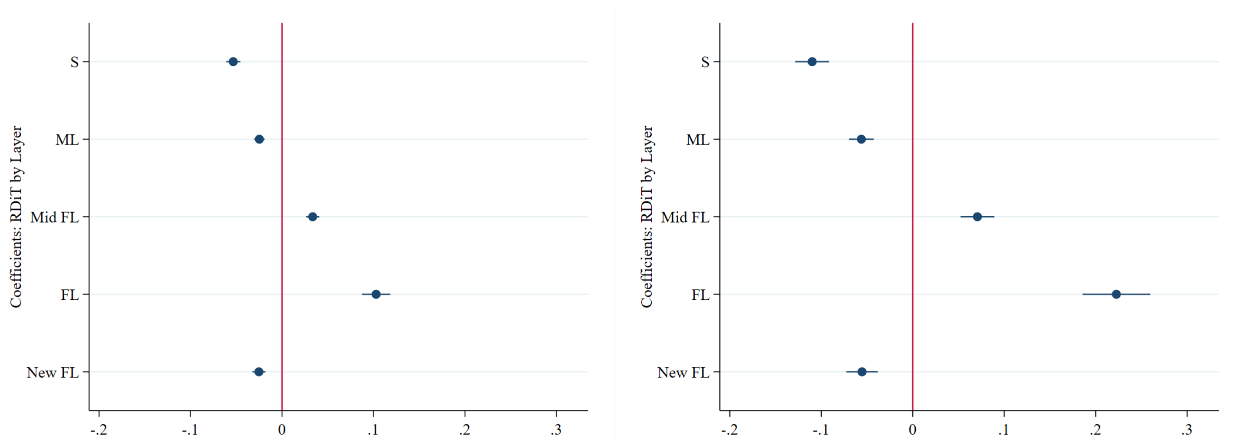
This Figure shows the effect of volume changes from 8 weeks before the event to 8 weeks after the event on the number of layers, limiting the sample to pre-existing working groups only. Number of layers is defined as the number of positions in the working group. Standard errors clustered at weekly-shift 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 A12: Event Study of Volume Changes on Employment



Note: Figure A12 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 at week-working group level. 95% confidence intervals are reported. Number of observations: 2 shifts x 16 weeks x 2 events.

Figure A13: Impact of Volume Changes on Working Group Structure and Knowledge Hierarchies

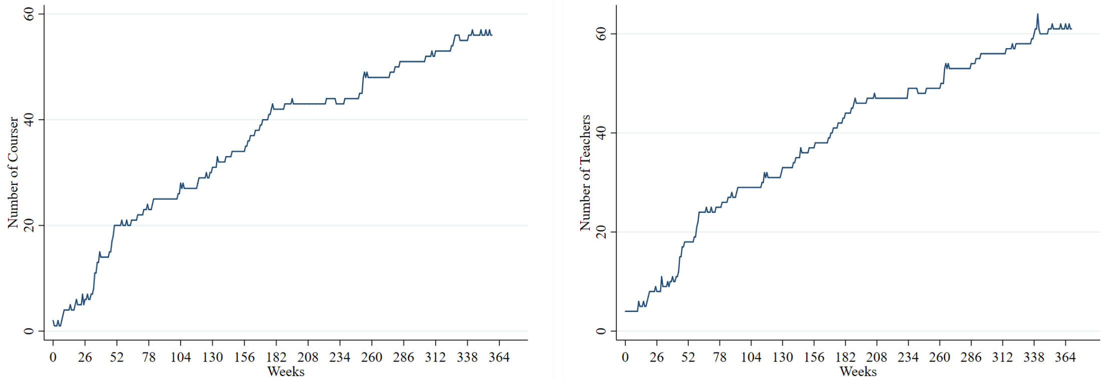


(a) Share of Workers by Layer, 0-3 weeks

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

Figures A13a and A13b 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-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 Volume change and to all other Model and Volume changes. Standard errors are clustered by distance to Model change and working group. 95% confidence intervals are presented in the figure. Number of observations: 220 working groups x 16 weeks x 2 events.

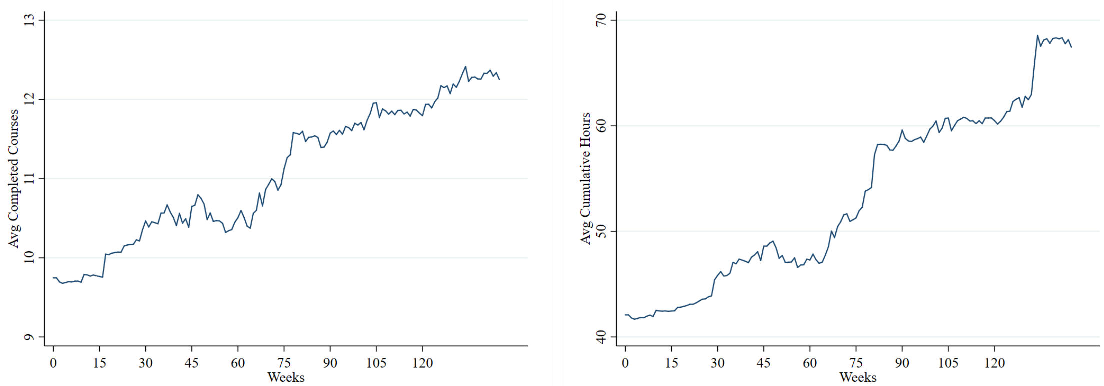
Figure A14: Courses and Teachers over Time



(a) Courses (b) Teachers

Note: Figure A14a plots the weekly number of courses provided during our period of analysis. Figure A14b plots the weekly number of teachers during our period of analysis.

Figure A15: Average Completed Programs and Average Cumulative Training Hours over Time



(a) Average Completed Programs (b) Average Cumulative Training Hours

Note: Figure A15a plots the weekly average completed programs per working group in our period of analysis. Figure A15b plots the weekly average accumulated training hours per working group in our period of analysis.

Table A1: Distribution of Reporting of Problem-Solving Activities By Knowledge Layer

	Share of total reports	Reports per employee
S4	0.002	0.001
S3	0.006	0.003
S2	0.004	0.002
S1	0.003	0.002
ML4	0.006	0.018
ML3	0.050	0.019
ML2	0.078	0.019
ML1	0.062	0.016
Mid FL	0.106	0.018
FL4	0.201	0.017
FL3	0.172	0.014
FL2	0.131	0.015
FL1	0.102	0.012
New FL3	0.043	0.008
New FL2	0.024	0.007
New FL1	0.011	0.007

Note: Table A1 shows the distribution of total reports by layer and the rate of reports per employee by layer for the years 2018 and 2019.

Table A2: 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 information of production from 2012 to 2019. The Model and Volume changes information is the average of daily information in the month after the change happens.

Table A3: 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 Model change. 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 A4: Impact of Model Changes on Productivity

	(1) DPV
0-3 weeks	0.745***
	(0.134)
4-7 weeks	0.198
	(0.186)
Observations	1,134
Obs. Level	Shift-Day
Mean	0.000

Note: Standard errors clustered by distance to event and shift. Number of observations: 2 shifts x 81 days x 7 events. Productivity measures are reported by the plant at the shift-day level. DPV is the number of defects per 100 vehicles, and is standardized. 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 A5: Impact of Model Changes on Productivity during 2017 to 2019

	(1) DPV
0-3 weeks	1.184**
	(0.526)
4-7 weeks	0.850
	(0.689)
Observations	324
Obs. Level	Shift-Day
Mean	0.000

Note: Standard errors clustered by distance to event and shift. Number of observations: 2 shifts x 81 days x 2 events. Productivity measures are reported by the plant at the shift-day level. DPV is the number of defects per 100 vehicles, and is standardized. 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 A6: Impact of Model Changes on Stock of Knowledge of 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 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 A7: 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 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 A8: Impact of Model Changes on Number of Groups and Group Size

	(1) Num of Groups	(2) Group Size
0-3 weeks	-0.786	0.039
	(1.703)	(0.240)
4-7 weeks	-0.789	0.105
	(2.304)	(0.349)
Observations	64	64
Obs. Level	Shift-Week	Shift-Week
Mean	58	20

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. Group size is the average number of employees in each working group. * p<0.1, ** p<0.05, *** p<0.01

Table A9: 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 Volume change. 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 A10: Impact of Volume Changes on Productivity

	(1) DPV
0-3 weeks	0.696***
	(0.167)
4-7 weeks	0.319
	(0.201)
Observations	810
Obs. Level	Shift-Day
Mean	0.000

Note: Standard errors clustered by distance to event and shift. Number of observations: 2 shifts x 81 days x 5 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. * p<0.1, ** p<0.05, *** p<0.01

Table A11: Impact of Volume Changes on Productivity during 2017 to 2019

	(1) DPV
0-3 weeks	2.206**
	(0.867)
4-7 weeks	1.952**
	(0.911)
Observations	324
Obs. Level	Shift-Day
Mean	0.000

Note: Standard errors clustered by distance to event and shift. Number of observations: 2 shifts x 81 days 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. * p<0.1, ** p<0.05, *** p<0.01

Table A12: 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 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 A13: 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 Volume change 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 A14: 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 Volume change. 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

B Model

We use the model of Caliendo and Rossi-Hansberg (2012) to understand how product cycles affect the organization of the firm (i.e., optimal production structure of the firm like the number of layers, number of production workers and the knowledge they acquire at each layer). As we mentioned before, every time a new model is introduced, the share of new parts in the car increases, increasing the complexity of the problems solved by the workers. Anecdotal evidence shared by the partner firm and the empirical evidence presented in Section 4, suggest an increase in the stock of knowledge and a reduction in the number of layers every time the company faced a "model change." In this section, we explore under what conditions this anecdotal evidence can be rationalized by the model and is optimal for the firm. We contrast these results with the impacts of a positive volume change that increases quantity produced, for which we show the organizational response is a monotonic, permanent increase in both employment and management layers, consistent with prior evidence from manufacturing in high-income countries (Caliendo et al., 2020, 2015).

Layers Problem: Suppose that a firm pays a wage w to each of its workers and wishes to produce q units of output. The firm chooses the optimal number of layers L in order to minimize the cost of producing q units of output while paying a wage w to each worker and manager. The firm solves

$$C\left(q;w\right) \equiv \min_{L\geq 0}\left\{C_L\left(q;w\right)\right\}, \tag{7}$$

where $C(q;w)$ denotes the minimum variable cost of producing q units of output and $C_L(q;w)$, the minimum cost of producing q units of output with an organization with $L+1$ layers, it is defined by (10) below.

Workers and Knowledge Problem: Suppose that a firm has chosen an organization with $L+1$ layers. The amount of workers the firm hires at the lowest layer ($l=0$) is denoted by n_L^0, and the knowledge they acquire is denoted by z_L^0. At an intermediate layer l ($0<l<L$), it hires n_L^l managers, each one with knowledge z_L^l. Since there is only one entrepreneur in the firm, then $n_L^L=1$. z_L^L denotes the entrepreneur's knowledge.
If the firm hires n_L^0 workers at the lowest layer, each of which possess knowledge z_L^0, then each of these workers is capable to solve a fraction $F(z_L^0)$ of the problems that the firm faces. The fraction of unsolved problems $1-F(z_L^0)$ is left for the next layer, $l=1$. Note that managers at layer $l=1$ spend a fraction h of their unit of time listening to the workers' problems, which implies that each manager can deal with at most $\frac{1}{h}$ problems. It follows that n_L^1 must be proportional to the amount of unsolved problems they can deal with, i.e.,

$$n_L^1 = hn_L^0 \left(1 - F\left(z_L^0\right)\right). \tag{8}$$

Note that as the cost of communication h increases, n_L^1 increases. Similarly, the amount of managers at layer l $(l > 1)$ must be proportional to the amount of unsolved problems at that point,

$$n_L^{l+1} = n_L^l \left(1 - F\left(z_L^l\right)\right) \text{ for all } 0 < l < L. \tag{9}$$

Given a sequence of knowledge $\{z_L^l\}_{l=0}^L$, equation (9) gives us an *evolution law* for the population within the firm. Note that, since $n_L^L = 1$, given a sequence of knowledge $\{z_L^l\}_{l=0}^L$, the evolution law completely determines the values of n_L^l for $0 \leq l < L$. It follows that the firm only has to find the optimal knowledge sequence $\{z_L^l\}_{l=0}^L$ that allows it to produce q units of output. That is, the firm solves

$$\begin{aligned} C_L(q; w) \equiv & \min_{\left\{n_L^l, z_L^l\right\}_{l=0}^L \geq 0} \sum_{l=0}^{L} \overbrace{n_L^l}^{\text{labor}} \underbrace{w\left(cz_L^l + 1\right)}_{\text{wages}}, \\ \text{s.t. } & A \cdot F\left(Z_L^L\right) n_L^0 \geq q, \\ & n_L^l = n_L^0 h e^{-\lambda Z_L^{l-1}}, \end{aligned} \tag{10}$$

where $F(z) = 1 - e^{-\lambda z}$, for $0 < l < L$, and $n_L^L = 1$. Here, $Z_L^L \equiv \sum_{l=0}^{L} z_L^l$ represents the cumulative knowledge of the firm. Note that in (10), the firm is minimizing the cost of the labor plus the cost of educating the workers.

B1 Stock of Knowledge

Proposition 1. *If a firm wants to increase its cumulative workforce knowledge by adding a new layer, such that $Z_{L+1}^L - Z_L^{L-1} = \varepsilon$, for some $\varepsilon > 0$, then $z_L^L > z_{L+1}^{L+1}$ and $z_L^l > z_{L+1}^l$ for $0 \leq l < L$.*

System (13) provides explicit formulas to determine the knowledge at every layer, and Proposition 1 uses this information to depict how a firm redistributes its total knowledge when it changes layers. More specifically, when $\varepsilon \to 0$ this transference of knowledge results into a more efficient organizational structure, since the firm would not have to invest in increasing its cumulative knowledge directly. Instead, by disclaiming less information per layer it will be able to afford additional layers, and even reduce its average costs for levels of production q large enough. Furthermore, the proof seen in the Appendix B4 shows how the knowledge of the workers at higher layers is the most affected by moving from L to $L+1$ layers, while the entrepreneur is the one that gives up the least amount of knowledge with the transition.

B2 Model Changes

Using this general framework, we study how the organization of the firm changes endogenously in response to product cycles. As we mentioned before, every time a new model is introduced, the complexity of the problems increases as the share of new parts in the car increases, which we model as a reduction in λ. A lower λ, decreases the production level q_L at where the firm should move from L to $L+1$ layers (Caliendo and Rossi-Hansberg, 2012). However, if the firm is uninterested in changing its number of layers, it must invest in modifying other parameters in order to balance λ's impact. From anecdotal evidence, investing in reducing the training cost c seems like a plausible option, so we asked: *What are the admissible ratios $\Delta c : \Delta\lambda$ that prevent the firm from increasing layers?*[28]

To do so, define the production level where the average cost of the firm working with L or $L+1$ layers intersect as a function of λ and c (*i.e.* $q_L := q_L(\lambda, c)$). Therefore, the unitary vector $\vec{v}_- := -\alpha\hat{\lambda} - \sqrt{1-\alpha^2}\,\hat{c}$ (or $\vec{v}_+ := -\alpha\hat{\lambda} + \sqrt{1-\alpha^2}\,\hat{c}$) for $\alpha \in [-1,1]$ encodes the directional derivative of q_L at $\vec{p} := (\lambda_0, c_0)$ as

$$D_{\vec{v_-}} q_L(\vec{p}) = -\alpha \left(\frac{\partial q_L}{\partial \lambda}\right)\bigg|_{\vec{p}} - \sqrt{1-\alpha^2}\left(\frac{\partial q_L}{\partial c}\right)\bigg|_{\vec{p}}.$$

We support the usage of $\vec{v}_-$ over $\vec{v}_+$ because empirical tests suggest that $\partial q_L/\partial\lambda > 0$ and $\partial q_L/\partial c < 0$. Since model changes imply a drop in λ, the firm should restrict itself to $\alpha \geq 0$. For this case,

$$D_{\vec{v_-}} q_L(\vec{p}) \geq 0 \quad \text{if} \quad 0 \leq \alpha^2 \leq \left(\frac{\partial q_L}{\partial c}\right)^2 \left[\left(\frac{\partial q_L}{\partial \lambda}\right)^2 + \left(\frac{\partial q_L}{\partial c}\right)^2\right]^{-1} =: \kappa_\alpha, \quad \kappa_\alpha \leq 1.$$

Hence, the admissible directions on the third quadrant of the λc-plane in which q_L increases belong to the interval $\mathcal{D} := [0, \sqrt{\kappa_\alpha}]$. As a consequence, the firm can choose any $\alpha \in \mathcal{D}$ and establish a $\Delta c : \Delta\lambda$ ratio of $\sqrt{1-\alpha^2} : \alpha$ aiming to maintain or even reduce its optimal number of layers.

The appropriate sign of $\partial q_L/\partial\lambda$ and $\partial q_L/\partial c$ might vary depending on the operating point $\vec{p}$. To verify which one is the case, we also provide explicit formulas for these partial derivatives in the proof of Lemma 1, in Appendix B4.

[28]We consider Δ as the absolute change of a variable, that is $\Delta\lambda := |\lambda_1 - \lambda_0|$ and $\Delta c := |c_1 - c_0|$.

Lemma 1. *For any $L > 1$, if there exists a unique z_L^L that satisfies system (13), then $\partial q_L/\partial\lambda$ and $\partial q_L/\partial c$ can be explicitly and uniquely determined.*

In particular, if q_L presents the usual behavior at $\vec{p}$,[29] then $D_{\vec{v}_-}q_L(\vec{p})$ increases for $\alpha \to 0$, while $D_{\vec{v}_-}q_L(\vec{p}) \to 0$ for $\alpha \to \kappa_\alpha$, and thus, the firm might be tempted to select very small values of α. However, the new operating point $\vec{p_1} = (\lambda - \Delta\lambda, c_0 - \Delta c)$ must have positive coordinates and since Δc is inversely correlated to α, the firm has to be aware of not choosing an α small enough for $\Delta c > c_0$. We formalize this analysis in the following proposition.

Proposition 2. *Suppose $(\partial q_L/\partial\lambda)|_{\vec{p}} > 0$ and $(\partial q_L/\partial c)|_{\vec{p}} < 0$, for $\vec{p} = (\lambda_0, c_0)$, and $\lambda_1 < \lambda_0$. If the firm can invest in decreasing the training cost freely and wants to maintain a production level $q \in [q_{L-1}, q_L]$, then it has to reduce c_0 by at least*

$$\Delta c = \frac{(\lambda_0 - \lambda_1)\sqrt{1-\alpha^2}}{\alpha}, \quad \text{where } \alpha^2 = \left(\frac{\partial q_L}{\partial c}\right)^2 \left[\left(\frac{\partial q_L}{\partial \lambda}\right)^2 + \left(\frac{\partial q_L}{\partial c}\right)^2\right]^{-1}\Bigg|_{\vec{p}} \text{ and } \alpha \geq 0,$$

to avoid increasing its number of layers. Moreover, for any $c' < c_0 - \Delta c$, there exists levels of production q at which the firm opts for an organization with fewer layers.

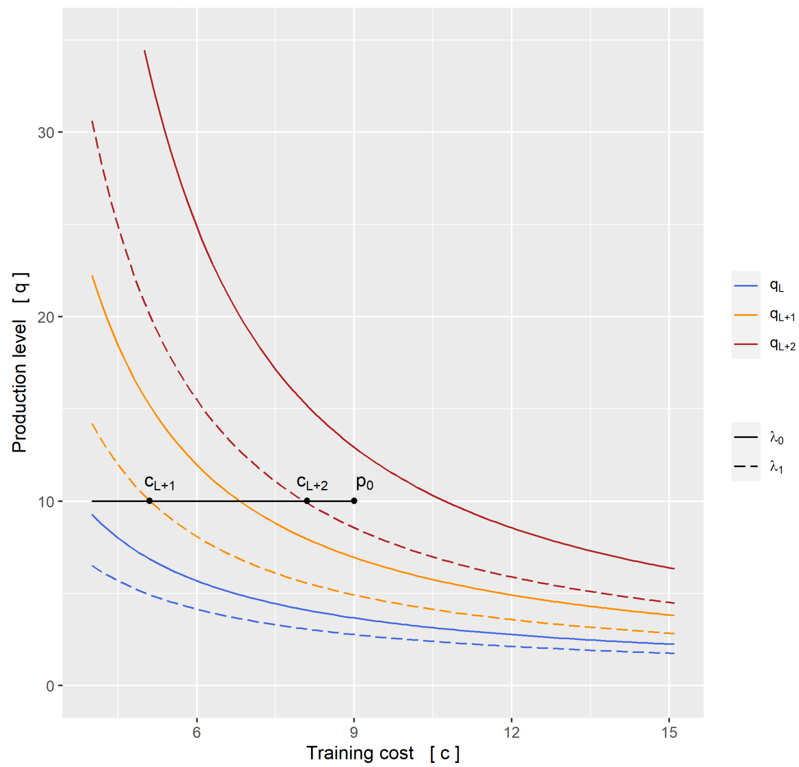
Essentially, the partial derivatives $(\partial q_L/\partial\lambda)|_{\vec{p}} > 0$ and $(\partial q_L/\partial c)|_{\vec{p}} < 0$ quantify the opposing effects on the intersecting point q_L of reducing λ or c. The former constitutes the negative impact of the firm having to face more complex problems, while the latter is the favorable scenario where it can train its workers more efficiently. In that sense, Proposition 2 provides the minimum investment in training costs required to balance the effect of problem complexity.

In addition, there is a simple graphical method, to determine the training costs c that would result in a reduction of layers despite a model change. It consists on the firm calculating the intersecting levels of production q_L for some training costs $c > 0$. Then, for a fixed production q, it identifies the active number of layers L at the original training cost c_0, which will indicate how much cost reduction is needed to reach the q_L or q_{L-1} zone.

For example, in Figure B1, we suppose that the firm maintains a production level $q = 10$, and that it has an original operating point $p_0 = (2, 9)$, with 5 active layers. Then, the model changes from $\lambda_0 = 2$ to $\lambda_1 = 1.5$, and so, if the firm preserves the original training cost $c_0 = 9$ then it has to level up to 6 layers. Conversely, if the firm wants to keep operating with 5 layers, it must reduce its training cost to some point in the interval $(c_4, c_5] := (5.1, 8]$.

[29] $\partial q_L/\partial\lambda > 0$ and $\partial q_L/\partial c < 0$.

Figure B1: Level Curves and Model Changes



Note: The parameters used in the simulation are $L = 3$, $\lambda_0 = 2$, $\lambda_1 = 1.5$, $h = 0.9$, $w = 1$, and $A = 1$.

B3 Endogenous Training cost

Here, we developed a variation for model (7) based on introducing a penalty $\mathcal{P}$ (or a fee) for investing in reducing the training cost, which results in model (11):

$$C\left(q;w\right) = \min_{\{L,\,c\}\geq 0} C_L\left(q;w\right) + \mathcal{P}(c,\lambda,L), \qquad \text{where } C_L\left(q;w\right) \text{ solves (10).} \tag{11}$$

In (11), for each c the firm has to establish the optimal distribution of knowledge $\{z_L^l\}_{l=0}^L$, and then it selects the optimal training cost c. In addition, the fee $\mathcal{P}$ is designed to reinforce the effect of training over the one of problem complexity, and reflect that reducing training costs is inexpensive for high values of c but gradually becomes costly. Moreover, since the fee is introduced to counter model changes, we consider penalties $\mathcal{P}$ that are constant for any level of production q. That is, we introduce the following assumption.

Assumption 1.

1. *$\mathcal{P}(c,\lambda,L)$ is decreasing with respect to c.*

2. $\mathcal{P}(c, \lambda, L)$ is independent of the production levels q.

Intuitively, choosing the optimal training cost for model (11) is a trade-off between lower costs $C_L(q;w)$ and the price to pay to achieve them represented by $\mathcal{P}$. Ideally, the penalty should be calibrated to be sensitive to problem complexity changes so it can induce sufficient drops in the training costs. Nonetheless, the firm must be wary of extremely costly or volatile penalties, as their effects can overshadow those of the total cost function $C_L(q;w)$, which is the main object of study.

In this paper, we propose the penalty $\mathcal{P}(c, \lambda, L) = (c - (\vartheta\lambda + \vartheta^k L))^{-1}$, where $\vartheta \in \mathbb{R}^+$, $\vartheta > 1$ and $k \in (0,1)$. Using this penalty, we define the auxiliary function

$$\Psi(q, \lambda) = \min_{c\geq 0} \left(C_L\left(q;w\right) + \mathcal{P}(c, \lambda, L)\right) - \min_{c\geq 0} \left(C_{L+1}\left(q;w\right) + \mathcal{P}(c, \lambda, L+1)\right). \qquad (12)$$

Suppose that $q_L(\boldsymbol{\lambda})$ is the value of q for which the firm should move from L to $L+1$ layers, for $\boldsymbol{\lambda}$ fixed. Proposition 3 proposes sufficient conditions to guarantee that $q_L(\lambda) < q_L(\lambda_1)$ for $\lambda - \lambda_1 > 0$ sufficiently small.

Proposition 3. *If $\partial\Psi(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda})/\partial\lambda > 0$, then there is a neighborhood of $\boldsymbol{\lambda}$ where we can parameterize $q_L \equiv q_L(\lambda)$, and this parameterization satisfies $\partial q_L/\partial\lambda < 0$.*

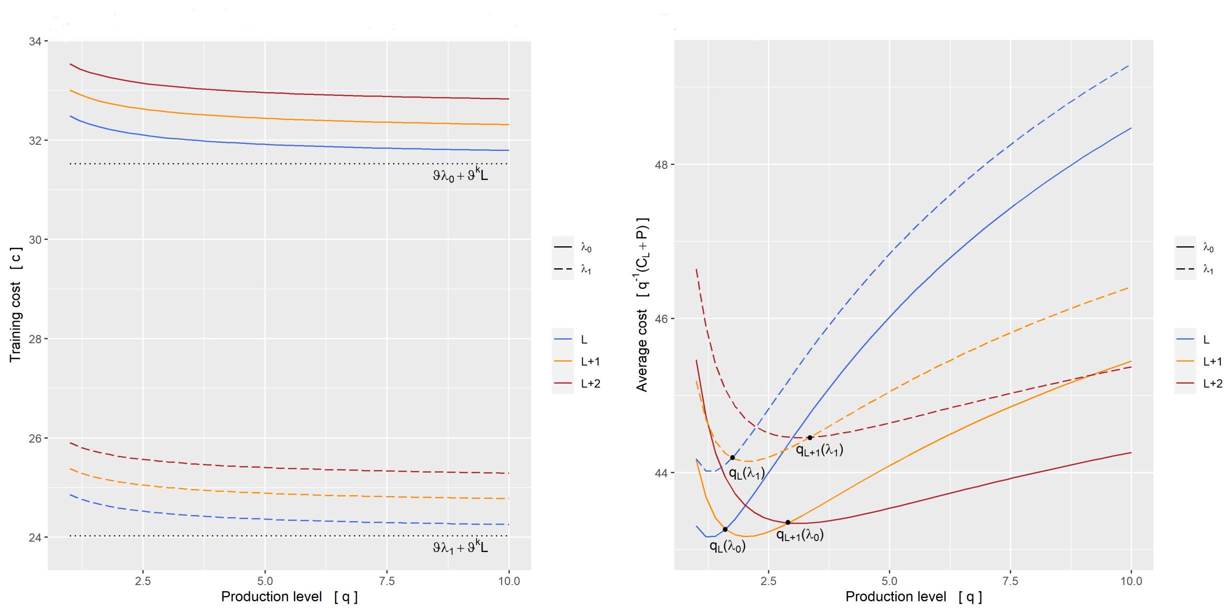
Figures B2a and B2b show the effect of this penalty on the optimal training cost c, and the resulting average cost modelled according to (11). More specifically, in Figure B2a, we see that for every level of production q the optimal training cost drops for a model change, which is a consequence of the $\vartheta\lambda$ component. Also, the training curves tend asymptotically to $\vartheta\lambda + \vartheta^k L$. This separates the optimal training costs for firms with different number of layers, primarily due to the $\vartheta^k L$ component. Notice that k acts as a weight between the problem complexity and layers components of the penalty. That said, the fact that $k \in (0,1)$ is intended to imply that the penalty is more sensitive to λ than to L. Graphically, this is seen in Figure B2a with the cost curves between a firm with different layers being closer that those between a firm facing two different problem complexities.

Besides, with this penalty we guarantee that $c_{L+1} > c_L$ to break $\partial C_L/\partial q$ and $\partial C_{L+1}/\partial q$ apart.[30] These changes seem to lead us to the desired endogenous response as seen in Figure B2b where, despite the model change, the firm in general operates with the same amount of layers, and the particular production levels where it does not is because it has the option to

[30] A graphical intuition about the changes in q_L is that it will move to the right (*left*) if and only if $\partial(C_L + \mathcal{P})/\partial q < \partial(C_{L+1} + \mathcal{P})/\partial q$ ($>$). Furthermore, from the envelope theorem together with the first order conditions for (7) we know that $\partial C_L/\partial q = whc_L e^{\lambda z_L^L}/\lambda A$ for any $L > 1$, and since $\mathcal{P}$ does not depend on q, then $\partial(C_L + \mathcal{P})/\partial q = \partial C_L/\partial q$.

reduce them. More specifically, the production levels where the firm considered can reduce its layers are $q \in [q_L(\lambda_0), q_L(\lambda_1)] \approx [1.6, 1.7]$ or $q \in [q_{L+1}(\lambda_0), q_{L+1}(\lambda_1)] \approx [2.9, 3.3]$.[31]

Figure B2: Endogeneous Training Cost



(a) Training cost curves.

(b) Endogenous average cost.

Note: The parameters used in the simulation are: $L = 3$, $\lambda_0 = 2$, $\lambda_1 = 1.5$, $h = 0.9$, $w = 1$, and $A = 1$.

[31]The firm modelled has the parameters: $h = 0.9$, $w = 1$, and $A = 0.5$.

B4 Proofs

Proof of Proposition 1. From Caliendo and Rossi-Hansberg (2012) it follows that (10) is equivalent to solving the system

$$\begin{cases} z_L^L = \frac{1}{\lambda}\ln\left(\frac{A}{Ae^{\lambda Z_L^{L-1}}-hq}\right), \\ z_L^0 = \frac{h}{\lambda}e^{\lambda z_L^L} - \frac{1}{\lambda} - \frac{1}{c}, \\ z_L^1 = \frac{1}{h\lambda}e^{\lambda z_L^0} - \frac{1}{\lambda} - \frac{1}{c}, \\ z_L^l = \frac{1}{\lambda}e^{\lambda z_L^{l-1}} - \frac{1}{\lambda} - \frac{1}{c} \text{ for } 1 < l < L. \end{cases} \tag{13}$$

Also, it follows that

$$z_L^L = \frac{1}{\lambda}\ln\left(\frac{A}{Ae^{\lambda Z_L^{L-1}} - hq}\right) \quad \text{and} \quad z_{L+1}^{L+1} = \frac{1}{\lambda}\ln\left(\frac{A}{Ae^{\lambda Z_{L+1}^{L}} - hq}\right).$$

Therefore, if we define $\varepsilon_L := z_L^L - z_{L+1}^{L+1}$ we have that

$$\varepsilon_L = z_L^L - z_{L+1}^{L+1} = \frac{1}{\lambda}\ln\left(\frac{Ae^{\lambda Z_{L+1}^{L}} - hq}{Ae^{\lambda Z_L^{L-1}} - hq}\right) > 0,$$

since $Z_{L+1}^L - Z_L^{L-1} = \varepsilon > 0$. Now, if $\varepsilon_l := z_L^l - z_{L+1}^l$ for $0 \leq l < L$, we obtain that

$$\begin{aligned} \varepsilon_0 &= \frac{h}{\lambda}\left(e^{\lambda z_L^L} - e^{\lambda z_{L+1}^{L+1}}\right) = \frac{he^{\lambda z_L^L}}{\lambda}\left(1 - e^{-\lambda \varepsilon_L}\right) > 0, \\ \varepsilon_1 &= \frac{1}{h\lambda}\left(e^{\lambda z_L^0} - e^{\lambda z_{L+1}^0}\right) = \frac{e^{\lambda z_L^0}}{h\lambda}\left(1 - e^{-\lambda \varepsilon_0}\right) > 0 \quad \text{and} \\ \varepsilon_l &= \frac{1}{\lambda}\left(e^{\lambda z_L^{l-1}} - e^{\lambda z_{L+1}^{l-1}}\right) = \frac{e^{\lambda z_L^{l-1}}}{\lambda}\left(1 - e^{-\lambda \varepsilon_{l-1}}\right) > 0 \quad \text{for } 1 < l < L, \end{aligned}$$

which concludes the proof. □

Proof of Lemma 1. From Caliendo and Rossi-Hansberg (2012) we deduced that the functional form of model (7) is given by:

$$C_L(q,w) = \begin{cases} \frac{wc}{\lambda}\left(\frac{hq}{A}e^{\lambda z_L^L} + (1 - e^{\lambda z_L^{L-1}}) + \lambda z_L^L + \frac{\lambda}{c}\right) & \text{for } L > 1, \\ \frac{wc}{\lambda}\left(\frac{hq}{A}e^{\lambda z_1^1} + \left(1 - \frac{e^{\lambda z_1^0}}{h}\right) + \lambda z_1^1 + \frac{\lambda}{c}\right) & \text{for } L = 1, \\ w\left(\frac{c}{\lambda}\ln\left(\frac{A}{A-q}\right) + 1\right) & \text{for } L = 0. \end{cases} \tag{14}$$

Therefore if q_L is the production level at which the average cost of a firm operating with L

and $L+1$ layers intersect for any $L>1$, then q_L satisfies the equation:

$$q_L = \frac{A}{h}\left(\frac{e^{\lambda z_L^{L-1}} - e^{\lambda z_{L+1}^{L}} + \lambda\left(z_{L+1}^{L+1} - z_L^L\right)}{e^{\lambda z_L^L} - e^{\lambda z_{L+1}^{L+1}}}\right), \tag{15}$$

where z_L^{L-1}, z_L^L, z_{L+1}^L, and z_{L+1}^{L+1}, satisfy the system (13). Notice that the optimal knowledge per layer depends both on λ and c. In particular, the entrepreneur's knowledge also depends on the distribution of knowledge of all the subordinate layers. Hence, the strategy to establish the formula for $\partial q_L/\partial\lambda$ (and *mutatis mutandis* for $\partial q_L/\partial c$) is to first, write $\lambda(\partial z_L^l/\partial\lambda)$ as $K_1^\lambda(z_L^l) + K_2^\lambda(z_L^l)(\partial z_L^L/\partial\lambda)$ for every $l<L$, and subsequently, $\partial z_L^L/\partial\lambda$ as $K_1^\lambda(z_L^L) + K_2^\lambda(z_L^L)(\partial q/\partial\lambda)$. Notice that this is valid for both L and $L+1$. Then, deriving both sides of (15) by λ and replacing $\partial z_L^L/\partial\lambda$ in $\lambda\partial z_L^{L-1}/\partial\lambda$ and $\partial z_{L+1}^{L+1}/\partial\lambda$ in $\lambda\partial z_{L+1}^L/\partial\lambda$ leads to an equation where we can solve $\partial q_L/\partial\lambda$.

(i) To determine $\partial q_L/\partial\lambda$:

Step 1: From system (13) we establish that

$$\lambda\frac{\partial z_L^l}{\partial\lambda} = \begin{cases} \frac{1}{\lambda}\left(1 - he^{\lambda z_L^L}\right) + he^{\lambda z_L^L} z_L^L + h\lambda e^{\lambda z_L^L}\frac{\partial z_L^L}{\partial\lambda}, & \text{for } l=0,\\ \frac{1}{\lambda} + \frac{e^{\lambda z_L^0} z_L^0}{h} + \frac{e^{\lambda(z_L^0+z_L^L)}}{\lambda}\left(\lambda z_L^L - 1\right) + \lambda e^{\lambda(z_L^0+z_L^L)}\frac{\partial z_L^L}{\partial\lambda} & \text{for } l=1,\\ \frac{1}{\lambda} + \sum_{k=1}^{l-1} e^{\lambda Z_L^{[k,l-1]}} z_L^k + \frac{e^{\lambda Z_L^{l-1}} z_L^0}{h} + \frac{e^{\lambda(Z_L^{l-1}+z_L^L)}}{\lambda}\left(\lambda z_L^L - 1\right) + \lambda e^{\lambda(Z_L^{l-1}+z_L^L)}\frac{\partial z_L^L}{\partial\lambda} & \text{for } l<L, \end{cases}$$

from which we obtain the expressions for $K_1^\lambda(z_L^l)$ and $K_2^\lambda(z_L^l)$ for any $l<L$.

Step 2: Since $\lambda z_L^L = \ln(A) - \ln(Ae^{\lambda Z_L^{L-1}} - hq)$ it follows that[32]

$$z_L^L + \lambda\frac{\partial z_L^L}{\partial\lambda} = -e^{\lambda z_L^L}\left[e^{\lambda Z_L^{L-1}}\left(Z_L^{L-1} + \sum_{l=0}^{L-1} K_1^\lambda(z_L^l) + K_2^\lambda(z_L^l)\frac{\partial z_L^L}{\partial\lambda}\right) - \frac{h}{A}\frac{\partial q}{\partial\lambda}\right]. \tag{16}$$

Hence, solving for $\partial z_L^L/\partial\lambda$ leads to

$$K_1^\lambda(z_L^L) = \frac{-1}{K_3^\lambda(z_L^L)}\left[z_L^L + e^{\lambda Z_L^L}\left(Z_L^{L-1} + \sum_{l=0}^{L-1} K_1^\lambda(z_L^l)\right)\right] \quad \text{and} \quad K_2^\lambda(z_L^L) = \frac{he^{\lambda z_L^L}}{AK_3^\lambda(z_L^L)}, \tag{17}$$

where $K_3^\lambda(z_L^L) = \left(\lambda + e^{\lambda Z_L^L}\sum_{l=0}^{L-1} K_2^\lambda(z_L^l)\right)$.

[32]As an abuse of notation we omit the λ super index in $K_1^\lambda(z_L^l)$ and $K_2^\lambda(z_L^l)$, $0\le l<L$.

Step 3: Once we derive both sides of (15) by λ we obtain that

$$\frac{\partial q_L}{\partial \lambda} K_3(q_L) = K_1(q_L) \left(e^{\lambda z_L^{L-1}} \frac{\partial(\lambda z_L^{L-1})}{\partial \lambda} - e^{\lambda z_{L+1}^L} \frac{\partial(\lambda z_{L+1}^L)}{\partial \lambda} + z_{L+1}^{L+1} - z_L^L + \lambda \left(\frac{\partial z_{L+1}^{L+1}}{\partial \lambda} - \frac{\partial z_L^L}{\partial \lambda} \right) \right)$$
$$+ K_2(q_L) \left(e^{\lambda z_L^L} \frac{\partial(\lambda z_L^L)}{\partial \lambda} - e^{\lambda z_{L+1}^{L+1}} \frac{\partial(\lambda z_{L+1}^{L+1})}{\partial \lambda} \right),$$

where $K_1(q_L) = e^{\lambda z_L^L} - e^{\lambda z_{L+1}^{L+1}}$, $K_2(q_L) = e^{\lambda z_{L+1}^L} - e^{\lambda z_L^{L-1}} + \lambda(z_L^L - z_{L+1}^{L+1})$, and $K_3(q_L) = A^{-1}h(K_1(q_L))^2$. Substituting $\partial(\lambda z_L^{L-1})/\partial\lambda$ and $\partial(\lambda z_{L+1}^L)/\partial\lambda$ according to *Step 1* and $\partial z_L^L/\partial\lambda$ and $\partial z_{L+1}^{L+1}/\partial\lambda$ according to *Step 2*, yields that $\partial q_L/\partial\lambda$ solves the equation:[33]

$$\begin{aligned}
\frac{\partial q_L}{\partial \lambda} K_3(q_L) = {} & K_1(q_L)\, e^{\lambda z_L^{L-1}} \left(z_L^{L-1} + K_1(z_L^{L-1}) + K_2(z_L^{L-1}) K_1(z_L^L) \right) \\
& - K_1(q_L)\, e^{\lambda z_{L+1}^L} \left(z_{L+1}^L + K_1(z_{L+1}^L) + K_2(z_{L+1}^L) K_1(z_{L+1}^{L+1}) \right) \\
& + K_1(q_L) \left(z_{L+1}^{L+1} - z_L^L + \lambda \left(K_1(z_{L+1}^{L+1}) - K_1(z_L^L) \right) \right) \\
& + K_2(q_L) \left(e^{\lambda z_L^L} \left(z_L^L + \lambda K_1(z_L^L) \right) - e^{\lambda z_{L+1}^{L+1}} \left(z_{L+1}^{L+1} + \lambda K_1(z_{L+1}^{L+1}) \right) \right) \\
& + \frac{\partial q_L}{\partial \lambda} K_1(q_L) \left(e^{\lambda z_L^{L-1}} K_2(z_L^{L-1}) K_2(z_L^L) - e^{\lambda z_{L+1}^L} K_2(z_{L+1}^L) K_2(z_{L+1}^{L+1}) \right) \\
& + \frac{\partial q_L}{\partial \lambda} \lambda K_1(q_L) \left(K_2(z_{L+1}^{L+1}) - K_2(z_L^L) \right) \\
& + \frac{\partial q_L}{\partial \lambda} \lambda K_2(q_L) \left(e^{\lambda z_L^L} K_2(z_L^L) - e^{\lambda z_{L+1}^{L+1}} K_2(z_{L+1}^{L+1}) \right).
\end{aligned}$$

(ii) To determine $\partial q_L/\partial c$:

Step 1: From system (13) we establish that

$$\lambda \frac{\partial z_L^l}{\partial c} = \begin{cases} \frac{1}{c^2} + h e^{\lambda z_L^L} \frac{\partial z_L^L}{\partial c}, & \text{for } l = 0, \\ \frac{1}{c^2} \left(1 + \frac{e^{\lambda Z_L^{l-1}}}{h} + \sum_{k=1}^{l-1} e^{\lambda Z_L^{[k,l-1]}} \right) + e^{\lambda (Z_L^{l-1} + z_L^L)} \frac{\partial z_L^L}{\partial c} & \text{for } 0 < l < L, \end{cases}$$

from which we obtain the expressions for $K_1^c(z_L^l)$ and $K_2^c(z_L^l)$ for any $l < L$.

Step 2: Since $\lambda z_L^L = \ln(A) - \ln(A e^{\lambda Z_L^{L-1}} - hq)$ it follows that

$$\lambda \frac{\partial z_L^L}{\partial c} = -e^{\lambda z_L^L} \left[\lambda e^{\lambda Z_L^{L-1}} \left(\sum_{l=0}^{L-1} K_1^c(z_L^l) + K_2^c(z_L^l) \frac{\partial z_L^L}{\partial c} \right) - \frac{h}{A} \frac{\partial q_L}{\partial c} \right].$$

[33] As an abuse of notation we omit the λ super index in $K_1^\lambda(z_i^j)$ and $K_2^\lambda(z_i^j)$, for $i \in \{L, L+1\}$ and $j \in \{L-1, L, L+1\}$.

Hence, solving for $\partial z_L^L/\partial c$ leads to

$$K_1^c(z_L^L) = -\frac{e^{\lambda Z_L^L}}{K_3^c(z_L^L)} \sum_{l=0}^{L-1} K_1^c(z_L^l) \quad \text{and} \quad K_2^c(z_L^L) = \frac{he^{\lambda z_L^L}}{\lambda A K_3^c(z_L^L)},$$

where $K_3^c(z_L^L) = \left(1 + e^{\lambda Z_L^L} \sum_{l=0}^{L-1} K_2^c(z_L^l)\right)$.

Step 3: Once we derive both sides of (15) by c we obtain that

$$\frac{\partial q_L}{\partial c}\frac{K_3(q_L)}{\lambda} = K_1(q_L)\left(e^{\lambda z_L^{L-1}}\frac{\partial z_L^{L-1}}{\partial c} - e^{\lambda z_{L+1}^L}\frac{\partial z_{L+1}^L}{\partial c} + \frac{\partial z_{L+1}^{L+1}}{\partial c} - \frac{\partial z_L^L}{\partial c}\right) + K_2(q_L)\left(e^{\lambda z_L^L}\frac{\partial z_L^L}{\partial c} - e^{\lambda z_{L+1}^{L+1}}\frac{\partial z_{L+1}^{L+1}}{\partial c}\right), \tag{18}$$

where $K_1(q_L) = e^{\lambda z_L^L} - e^{\lambda z_{L+1}^{L+1}}$, $K_2(q_L) = e^{\lambda z_{L+1}^L} - e^{\lambda z_L^{L-1}} + \lambda(z_L^L - z_{L+1}^{L+1})$, and $K_3(q_L) = A^{-1}h(K_1(q_L))^2$. Substituting $\partial z_L^{L-1})/\partial c$ and $\partial z_{L+1}^L/\partial c$ according to *Step 1* and $\partial z_L^L)/\partial c$ and $\partial z_{L+1}^{L+1}/\partial c$ according to *Step 2*, yields that $\partial q_L/\partial c$ solves the equation:[34]

$$\begin{aligned}\frac{\partial q_L}{\partial c}K_3(q_L) = {} & K_1(q_L)\left(e^{\lambda z_L^{L-1}}\left(K_1(z_L^{L-1}) + K_2(z_L^{L-1})K_1(z_L^L)\right)\right) \\ & - K_1(q_L)e^{\lambda z_{L+1}^L}\left(K_1(z_{L+1}^L) + K_2(z_{L+1}^L)K_1(z_{L+1}^{L+1})\right) \\ & + K_1(q_L)\left(K_1(z_{L+1}^{L+1}) - K_1(z_L^L)\right) + K_2(q_L)\left(e^{\lambda z_L^L}K_1(z_L^L) - e^{\lambda z_{L+1}^{L+1}}K_1(z_{L+1}^{L+1})\right) \\ & + \frac{\partial q_L}{\partial c}K_1(q_L)\left(K_2(z_L^L)\left(e^{\lambda z_L^{L-1}}K_2(z_L^{L-1}) - 1\right) + K_2(z_{L+1}^{L+1})\left(1 - e^{\lambda z_{L+1}^L}K_2(z_{L+1}^L)\right)\right) \\ & + \frac{\partial q_L}{\partial c}K_2(q_L)\left(e^{\lambda z_L^L}K_2(z_L^L) - e^{\lambda z_{L+1}^{L+1}}K_2(z_{L+1}^{L+1})\right).\end{aligned}$$

□

Proof of Proposition 2. Let q be a fixed level of production for a firm operating with L layers (*i.e.* $q \in [q_{L-1}(\lambda_0, c_0), q_L(\lambda_0, c_0)]$). Without loss of generality, consider $q_L(\lambda, c)$ and its directional derivative[35]

$$D_{\vec{v}}q_L(\vec{p}) = -\alpha\left(\frac{\partial q_L}{\partial \lambda}\right)\bigg|_{\vec{p}} - \sqrt{1-\alpha^2}\frac{\partial q_L}{\partial c}, \quad \text{for} \quad \vec{v} = -\alpha\hat{\lambda} - \sqrt{1-\alpha^2}\hat{c},\ \alpha \in [0,1].$$

[34] As an abuse of notation we omit the c super index in $K_1^c(z_i^j)$ and $K_2^c(z_i^j)$, for $i \in \{L, L+1\}$ and $j \in \{L-1, L, L+1\}$.

[35] $\hat{\lambda}$ and $\hat{c}$ denote unitary vectors.

Moreover, since $(\partial q_L/\partial\lambda)|_{\vec{p}} > 0$ and $(\partial q_L/\partial c)|_{\vec{p}} < 0$, for

$$\alpha^2 = \left(\frac{\partial q_L}{\partial c}\right)^2 \left[\left(\frac{\partial q_L}{\partial \lambda}\right)^2 + \left(\frac{\partial q_L}{\partial c}\right)^2\right]^{-1}\Bigg|_{\vec{p}}, \tag{19}$$

we have that $D_{\vec{v}} q_L(\vec{p}) = 0$. That is, moving in the direction $\vec{v} = -\alpha\hat{\lambda} - \sqrt{1-\alpha^2}\hat{c}$ does not change the production level at which the firm transitions from L to $L+1$ layers, as long as α satisfies (19). In particular, the vector

$$\vec{w} = \Delta\lambda\ \hat{\lambda} + \frac{\Delta\lambda\sqrt{1-\alpha^2}}{\alpha}\hat{c} = \Delta\lambda\hat{\lambda} - \Delta c\hat{c}, \quad \text{where} \quad \Delta\lambda := \lambda_1 - \lambda_0,$$

has the same direction as $\vec{v}$. Consequently $D_{\vec{w}} q_L(\vec{p}) = D_{\vec{v}} q_L(\vec{p}) = 0$ for α, and if $q < q_L(\lambda_0, c_0)$ then $q < q_L(\lambda_1, c_0 - \Delta c)$, which means that the firm remains with at most L layers. Concluding that $q > q_{L-1}(\lambda_1, c_0 - \Delta c)$ is analogous.

Now, if $c_0 - c' > \Delta c$, then the unitary vector $\vec{v}_{c'} = -\alpha'\ \hat{\lambda} - \sqrt{1-(\alpha')^2}\ \hat{c}$ associated is such that $\alpha' < \alpha$ and thus, $D_{\vec{v}_{c'}} q_L(\vec{p}) > D_{\vec{v}} q_L(\vec{p}) = 0$, since $(\partial q_L/\partial\lambda)|_{\vec{p}} > 0$ and $(\partial q_L/\partial c)|_{\vec{p}} < 0$. This means that by moving in the direction $\vec{v}_{c'}$, the intersection between L and $L+1$ layers occurs at a higher production level. As a consequence, for every $q \in [q_L(\lambda_0, c_0), q_L(\lambda_1, c')]$, the firm would have initially operated with $L+1$ layers and then drop to L layers. □

Proof of Proposition 3. Model 11 is minimizable with respect to c because

$$\lim_{c\to\infty} C_L(q; w) = \lim_{c\to(\vartheta\lambda+\vartheta^k L)^+} \mathcal{P}(c, \lambda, L) = \infty.$$

Therefore, for any pair (q, λ), there exists $c_L(q, \lambda)$ that minimizes Model 11. Figure B2a suggests that this minimum is unique. Nevertheless, if it is not unique, it suffices to take the connected component containing one of those minimums to fully parameterize $c_L \equiv c_L(q, \lambda)$.[36] With this, we can consider the function

$$\Phi_L(q, \lambda) = C_L(q, \lambda, c_L(q, \lambda)) + \mathcal{P}(c_L(\lambda, q), \lambda, L) \tag{20}$$

to be the minimum cost for a firm with L layers, production level q and complexity level $\boldsymbol{\lambda}$. $q_L(\boldsymbol{\lambda})$ then satisfies the equation $\Psi(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = \Phi_L(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda}) - \Phi_{L+1}(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$. Moreover, as $q_L(\boldsymbol{\lambda})$ is the value for which the firm moves from L to $L+1$ layers, then $\frac{\partial\Psi}{\partial q}(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \geq 0$.

[36] This connected component always exists because of the Implicit Function Theorem, and to the fact that $\frac{\partial^2 C_L}{\partial c^2} + \frac{\partial^2 \mathcal{P}}{\partial c^2} > 0$ when evaluated at a minimum.

First, we assume that $\frac{\partial \Psi}{\partial q}(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda}) > 0$. By continuity, there exists a neighborhood $\mathcal{D}_1 \subset \mathbb{R}^2$ of $(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda})$ for which $\frac{\partial \Psi}{\partial q}(\vec{p}) > 0$ if $\vec{p} \in \mathcal{D}_1$. Therefore, by the Implicit Curve Theorem, we can parameterize q_L as a function of λ for $\lambda \in \text{proj}_2(\mathcal{D}_2)$, with $\mathcal{D}_2 \subset \mathcal{D}_1$. Moreover, the slope of this parameterization is given by

$$\frac{\partial q_L}{\partial \lambda} = -\frac{\partial \Psi}{\partial \lambda} \Big/ \frac{\partial \Psi}{\partial q}.$$

From $\frac{\partial \Psi}{\partial \lambda}(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda}) > 0$, we can conclude that there is a neighborhood $\mathcal{D}_3 \subset \mathbb{R}^2$ such that $\frac{\partial \Psi}{\partial \lambda}(\vec{p}) > 0$ if $\vec{p} \in \mathcal{D}_3$. This implies that for $\lambda \in \text{proj}_2(\mathcal{D})$, $D := \mathcal{D}_2 \cap \mathcal{D}_3$, the slope of the parameterization $q_L \equiv q_L(\lambda)$ is negative.

On the other hand, if $\frac{\partial \Psi}{\partial q}(q_L(\boldsymbol{\lambda}), \boldsymbol{\lambda}) = 0$, the order of the first derivative that does not vanish must be odd, which implies we repeat the previous argument to the left and right of $\boldsymbol{\lambda}$ to parameterize $q_L \equiv q_L(\lambda)$ and show that the slope of this parameterization is again negative. $\square$