

NBER WORKING PAPER SERIES

TOWARD AN UNDERSTANDING OF LEARNING BY DOING:
EVIDENCE FROM AN AUTOMOBILE ASSEMBLY PLANT

Steven D. Levitt
John A. List
Chad Syverson

Working Paper 18017
<http://www.nber.org/papers/w18017>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2012

We received generous help from numerous individuals at the auto manufacturer from which our data are obtained, but we would especially like to thank Rich D., Shannon, and Robert M. Marco O. was an invaluable resource in helping us understand details of the production process and the associated data infrastructure. We also thank Francis Kramarz and seminar participants at Chicago, Columbia, Minnesota, and Toronto Rotman for helpful comments. We are grateful to David Greis, Kris Hult, Maria Ibanez, Carter Mundell, and Laura Rivera for their extensive research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2012 by Steven D. Levitt, John A. List, and Chad Syverson. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant

Steven D. Levitt, John A. List, and Chad Syverson

NBER Working Paper No. 18017

April 2012

JEL No. D2,L2,L6,O3

ABSTRACT

Productivity improvements within establishments (e.g., factories, mines, or retail stores) are an important source of aggregate productivity growth. Past research has documented that learning by doing—productivity improvements that occur in concert with production increases—is one source of such improvements. Yet little is known about the specific mechanisms through which such learning occurs. We address this question using extremely detailed data from an assembly plant of a major auto producer. Beyond showing that there is rapid learning by doing at the plant, we are able to pinpoint the processes by which these improvements have occurred.

Steven D. Levitt
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
slevitt@midway.uchicago.edu

Chad Syverson
University of Chicago
Booth School of Business
5807 S. Woodlawn Ave.
Chicago, IL 60637
and NBER
chad.syverson@chicagobooth.edu

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Learning by doing has occupied a central place within economics ever since Arrow (1962) used the concept as a workhorse in his theory of endogenous growth. Arrow conceptualized learning by doing within the actual activity of production, with cumulative gross investment as the catalyst for experience. Nearly two decades later, the role of experience in shaping and driving productivity growth was central in Lucas' (1988) explanations of increasing returns to human capital. Indeed, as Lucas (1988, p. 27) stresses "on-the-job-training or learning-by-doing appear to be at least as important as schooling in the formation of human capital." Yang and Borland (1991) furthered this line of thought by theoretically linking division of labor and learning by doing, highlighting an important source of comparative advantage.

Empirically, learning by doing has been shown to have equal import, as scholars have frequently observed that improvements in the efficiency with which outputs are produced from existing technologies and inputs are an important source of total factor productivity (TFP) growth. One early example was described in Lundberg (1961), who describes the experience of the Horndal iron works plant in Sweden. Although the plant had no new investment over a period of 15 years, output per worker hour rose about 2 percent annually. Another early observation of such progress was in the aircraft industry. As Wright (1936) and Middleton (1945) note, labor inputs per airframe considerably declined as the total number of airframes produced increased. Progress of this sort has been found across scores of studies, often attributed to adaptation efforts by labor, and argued to occur independently of scale effects (see Argote and Epple (1990) and Thompson (2010) for surveys).

A recent body of studies within the economics literature (more recent work includes Jarmin (1994); Darr, Argote, and Epple (1995); Jovanovic and Nyarko (1995); Benkard (2000); Sinclair, Klepper, and Cohen (2000); Thompson (2001); Thornton and Thompson (2001); and Thompson (2007)) has examined learning by doing in specific production settings.¹ This research has focused primarily on measuring the overall dynamics of the learning process—i.e., how fast productivity gains accrue or whether those rates imply that whatever is learned might be "forgotten" over time—and in some cases, whether there are learning spillovers across

¹ Related research has explored the impact of changes in incentive structures on worker output (e.g., Lazear (2000), Hamilton et al. (2003), Bandiera et al. (2007 and 2009), Hossain and List (2009)). Krueger and Mas (2004) document how labor unrest at Bridgestone/Firestone tire factories led to high rates of defective tires. Bloom et al. (2011) conduct a field experiment to measure how changes in management practices at a plant can improve productivity.

production units. However, the existing literature has revealed less (largely because of data limitations) about the *specific mechanisms* through which learning takes place within the plant. Theoretically, this has led traditional learning models to assume that efficiency is a direct function of cumulative output from the plant—the so called progress function literature—with no scope for managerial policies to affect outcomes (see Dutton and Thomas (1984)). The practical implication of this theorized structure is that, because the causes of the underlying efficiency improvements are obscure, there is limited scope for policy enhancements.

We are able to go beyond past research and lend insights into the underlying sources of efficiency improvements primarily due to the richness of our data. This data comes from an assembly plant of a major automaker. The core data set covers in incredible detail the production of nearly 200,000 cars over the course of a year. For each vehicle the plant assembles, we observe information on several hundred different operations that occur on the assembly line as the vehicle is produced. We observe the specific time at which each of these operations takes place and any problems that arise with the operation. This allows us to compute the rate at which production problems (defects) occur either car-by-car or over aggregated time periods. In addition to this production data, we also have two other useful data sets. First, we have records of absent employees for each day of the production year. This allows us to test if absenteeism is linked to production difficulties. Second, over a specific time horizon within our sample we know the warranty claims made on the cars produced during our sample. This provides us with a measure of how production defects' impact the automaker's bottom line.

Our data span a period that provides ample opportunity for learning by doing to appear. It covers a production year in which the car model made at the plant had just gone through a major redesign, and in which the firm had introduced a completely new production process at the plant.

Other features of the assembly plant's operations prove fortuitous for answering the questions at hand. For example, the plant adds a second shift only after substantial improvements have been observed on the first shift, allowing us to observe the extent to which productivity gains by the first shift spill over to the second shift, or whether the workers on the new shift have to reinvent the wheel (so to speak). This is important in light of the fact that many organizations have units in which spillovers are important to capture.

Further, over the course of the year, the plant began producing three variants of the model, with staggered production ramp-ups for each. We can therefore also measure how much production experience transfers across product variants, and whether problems encountered with newer variants spillover to those already being made. Additionally, physical capital inputs are essentially fixed over our sample; the plant's physical configuration and tooling was set prior to commencement of production and remained stable throughout the model year. This reduces concerns that the addition of unobserved inputs creates spurious learning by doing productivity patterns (Thompson (2001)).

Because of the unique features of our data, we focus on quality improvements—i.e., reductions in defects, rather than decreases in unit costs because many of the observed changes in line speed (a primary determinant of unit costs) are driven by the state of demand or calendar effects (e.g., slowdowns around the Christmas-New Year period) instead of technology shifts.

Our analysis reveals a number of interesting findings. First, consonant with previous learning by doing studies, we find that the auto assembly plant quickly realizes large efficiency gains: defects per vehicle fall more than 80 percent in the first eight weeks of production. Yet, the learning process is not restarted when the second shift begins production. Remarkably, virtually from the very beginning the second shift experiences defect rates that are lower than those observed contemporaneously on the first shift, even though the first shift had a nearly two month head start in production. This data pattern suggests that many of the plant's productivity gains are embodied in the plant's physical or organizational capital rather than in specific workers. Alternatively, introducing a new *model variant* into production does cause productivity setbacks—not only do the new variants initially exhibit much higher defect rates, but starting production of a new variant also leads to increased errors for the other variants already in production.

Second, the distribution of defect rates across the plant's hundreds of assembly processes is highly skewed. The top quintile of assembly stations accounts for more than 80 percent of all defects. Furthermore, these station-specific defect rates are persistent both over time and across shifts. Interestingly, defects that occur on one car spill over into other cars following nearby on the assembly line. The spillover effect of a defect creates statistically significant increases in defect rates for the 15 cars following (albeit at a declining magnitude). Interestingly, though, there is no evidence that these spillover effects decline over the model year. Learning by doing

benefits do not result from improvements in the automaker's ability to prevent production problems on one car from affecting others on the line.

Third, worker absenteeism is related to defect rates, both directly and through the rate at which acquired learning-by-doing knowledge is retained. However, the impact is economically small, with even large variations in absenteeism rates across periods implying only modest changes in average defect rates. This finding further bolsters the implication that much of the plant's production experience stock is embodied in the physical or broader organizational capital of the plant rather than its individual workers.

Finally, the number of production defects that occur during a car's assembly is related to later warranty payouts by the firm, indicating a bottom-line impact to the automaker of quality improvements resulting from learning by doing.

In sum, we find considerable evidence of learning by doing. Yet, such improvements do not appear to be embodied in workers: absenteeism has only a small impact on defects, and when a second shift comes online with new workers it immediately achieves lower defect rates than the existing shift. Rather, the gains are embodied in the organizational capital of the plant. These results help us go beyond the progress function that simply relates efficiency gains to cumulative production. In this way, the plant's experience gains do not cause efficiency enhancements, but rather provides an opportunity for management to exploit. This finding begins us on a path to better understand not only how specific productivity gains accrue at the plant level, but also how firms might optimally expand production. In addition, it provides some empirical content suggesting how comparative advantage arises, and what leads to specialization.

The remainder of the paper is structured as follows. Section I describes the production setting from which our data are taken. Section II describes the data. Section III presents the results of our several analyses, and Section IV concludes.

I. The Production Setting

Our production data comes from an assembly plant of a major automaker. Assembly plants are factories that piece together the thousands of parts that make up an automobile for delivery to final customers, either directly to fleet buyers like car rental companies or indirectly

through retail dealerships.² Non-assembly operations in the plant mostly involve conducting an assortment of quality-control tests of finished vehicles.

During our sample model year, the plant assembled three model variants built on a common midsize car platform. The shared platform means the variants have similar body frames and powertrains. However, their cosmetic styles, exterior and interior, are different enough that it may not be obvious to the untrained eye that the vehicles are such close cousins. Two of the variants are sedans with differing makes and model names. The third is a specialized style of car sold under the same model name as one of the sedans, though its specialization makes it obvious that it is a different variant. Without going into proprietary detail, we can say that these car types are distinct enough that each requires variant-specific parts and assembly procedures.

A particularly useful feature of our setting is that our observation period follows immediately upon the heels of several significant changes at the assembly plant. First, the platform had just undergone its first major redesign in six model years. This redesign involved both mechanical and stylistic changes. The automaker even renamed one of the models to further signal the platform's novelty.³

Besides these changes in the products themselves, the automaker substantially reorganized the plant's production process before the model year began. This involved changes in the assembly line's physical layout, a large investment in physical capital, and a shift into team-based production. Team production dispenses with the traditional assembly line practice of having individual workers hold responsibility for a particular task on the line. Instead, a team of workers is jointly responsible for a set of related tasks. The team members rotate through these tasks during the day and help their teammates when needed. One worker is designated as a team leader and is paid a small wage premium. This worker (a line worker not considered as management within the plant's hierarchy) has particular responsibilities involving oversight, training, and information aggregation.

These changes mean that, in many ways, the plant and its workers are starting over, or what the literature refers to as entering the 'ramping up' process: they are making new products

² The extent to which assembly plants are vertically integrated backward into making the parts they assemble differs somewhat across the industry, though there has been a general trend to move more parts fabrication offsite. Our particular plant reflects that movement; engines are brought into the plant ready to install from a separate factory, for example. Body panels are also struck at a separate (though geographically close) stamping plant.

³ We do not name the specific model year here for proprietary data reasons, but it was within the previous decade.

in a new way. This provides us with a unique glimpse of the learning by doing process, since as has typically been found in other studies, learning is nonlinear: large gains are realized quickly, but the speed of progress slows over time. Therefore we are likely to observe changes in defect rates of considerably larger magnitude than if we were to follow production of a product that the plant had already been making for some time.⁴ The high learning rate over our sample permits us to observe details of the learning process with greater resolution than would be possible later in the product design cycle. This improves our ability to measure the specific mechanisms driving learning by doing.

Another distinct advantage of examining data during the ramp up stage relates to shrinking product life cycles. As the lifespan of manufactured goods has decreased, especially among high-tech goods, there are considerably more ramp up moments in the sector. Understanding the factors influencing the time it takes to reach full production are invaluable not only for plant-level profits, but economy-wide efficiency.

Production for the model year ran from August to July. (We leave the specific calendar years spanned by our data unspecified because of the proprietary nature of the data. We refer below to the August-December period as occurring in Year 1 and the January-July period as being in Year 2.) A small number of prototype vehicles, on the order of a few dozen per week, were produced in late July and early August of Year 1. These were built to find any major difficulties in the production process and to train line workers in their new tasks and orient them to the plant's new team production process. As one might expect, defect rates during these first few weeks were extremely high. While no doubt part of the learning process, the fact that there were so few vehicles involved and average defect counts so varied between them led us to leave these first weeks out of the sample. We begin our sample in the first week that over 100 cars were produced. (A shift running at capacity at this plant can produce about 2000-2500 cars per week.) This occurred in mid-August.⁵

A second shift began production seven weeks after the first shift. The shift's workers were trained by having them observe operations at their respective stations during the week prior to the startup of second shift production.

⁴ This is confirmed in our case. Defect rates were much lower at the end of the model year previous to our sample, and had changed only slightly over the prior year. These earlier models had been produced for over six years at the plant, and as such most of the production kinks had already been worked out.

⁵ In specifications using daily production data below, we require that the plant produces at least 20 cars per day.

The timing of the car-variant ramp ups are as follows. Production initially focused exclusively on Model 1. Model 2 was introduced 17 weeks later to both the first and second shift. By that time, the second shift had already been operating for 10 weeks. Production of Model 3 began another 13 weeks after the start of Model 2.⁶

II. Data

Our primary data set is taken from the assembly plant's Factory Information System (FIS), proprietary software that interfaces with production through multiple modes. These include direct links to the tools themselves. For example, the FIS can read and record the torque applied by a particular wrench to bolts.⁷ It can also interface with the line workers themselves; either prompting them to respond to a query about a particular operation, or alerting them to a defect that needs attention. Throughout the production process the FIS records information about the thousands of distinct assembly operations that occur in the plant. Not all FIS records are defects. For operations done on safety-critical parts, the point of FIS is to document that the task was done successfully. But notices that are defects are recorded with particular identifiers that let us measure production defects with a high degree of accuracy.

The car-level defect data is straightforward to construct. We simply follow the car through the production process, counting any defects that it experiences along the way. We can track a single car through production because its vehicle information number (VIN)—a unique ID number given to every car assembled—is assigned to its critical component parts before production begins. FIS starts tracking these parts as soon as one is in process, and continues to follow the VIN through various stages of assembly until it leaves the factory.

The more aggregate measures of defect rates, such as average defect rates per car per week, take an extra step to compute. The numerator of this rate is the total number of defective

⁶ For the sake of consistency, when describing the introduction dates of Models 2 and 3 we impose a threshold of 100 cars per week being produced—this time of the particular variant—for the cars' production data to be included in the sample.

⁷ This sort of information is of obvious use to the automaker regarding the quality and speed with which production operations take place. But it also used to conform to regulations requiring verification of certain production operations deemed critical for safety reasons, such as tightening the lug nuts that hold the wheels on the vehicle.

production operations over the course of a time period (or in some specifications, on a specific over the period).⁸

The worker absentee data is from an administrative database that the plant uses to track absenteeism. The record level is at the individual absence, stating the date of absence and the ID number of the employee. It also contains the employee's shift as well as which one of the plant's seven broad operations departments (e.g., body, paint, chassis, trim) to which the employee is assigned. We use these data to construct time series of absences at the plant, shift, and department levels.

The warranty claims data contain information entered by technicians at retail dealerships where buyers have brought cars to make warranty claims. This information includes each car's VIN, the date, a description of the problem, a diagnosis, and two types of costs: customer and warranty. Customer costs are the responsibility of the car buyer and as such are not out-of-pocket costs for the auto manufacturer. (Of course, they could involve loss of goodwill, but we do not attempt to measure that here.) The latter is a direct cost to the manufacturer of the warranty claim. We test whether this cost is tied to the frequency of manufacturing defects as recorded in our FIS data by linking vehicles by VIN.

III. Results

III.1. Overall Patterns

Figure 1 plots the average number of defects per car by week.⁹ In mid August, when production begins, average defect rates were on the order of 70 per car. Eight weeks later, they

⁸ Measurement of the denominator, the number of cars produced during a period, is complicated by the fact that cars do not always start and end their time on the production line in the same period. For example, some cars begin production on a Friday and are completed on a Monday. Our approach is to break up the production process into segments that are divided by benchmark operations that FIS records for every car. This divides every car's production run into a consistent set of segments (though the segments aren't all equally spaced). We then apportion the car to a production week by segment. For example, if the first three of seven segment-ending benchmarks occur on Friday and the final four on Monday, we assign three-sevenths of the car to Friday and four-sevenths of it to Monday. The sum of complete and partial cars produced within a period gives us the denominator of our defects-per-car measure. When computing both total defects and cars produced, we exclude any operations that take place on a Saturday or a Sunday, when the plant is not on its regular operating schedule. Weekend records in our data are unusual, but occasionally occur.

⁹ Data shown in Figure 1 are subject to the condition that the plant produces at least 100 cars in the week, as discussed above. Figure A1 shows for the sake of reference the total number of cars produced in each workweek; the production ramp up during August and September accompanying the quality improvement is evident in the figure. It's worth noting that this ramp-up pattern, which is common in many production settings (and we will see below is also present when the plant starts a new shift or when it begins producing new model variants), is consistent

had fallen by 70 percent, to roughly 20 defects per car. These strong initial learning effects are consistent with findings in the broader literature on learning by doing. The absolute pace of defect reductions noticeably declined after that, with the remainder of the model year seeing downward drift in defect rates to a level of around 10 per car.

This slowing of the absolute rate of productivity growth is consistent with literature-standard power law specifications of the relationship between productivity and production experience. These specifications assume $S_t = AE_t^\beta$, where S_t is productivity at time t (average quality in our case), E_t is production experience up to that point (cumulative production), and A and β are parameters. Because we use a measure of inverse average quality (average defective operations per car) in our analysis, learning by doing implies β is negative; defect rates fall with production experience.

Table 1 shows the results of estimating power law specifications with our data. Panel A contains the results from specifications using weekly data (average defect rates over the week and production experience at the week’s outset); panel B shows results obtained using daily observations.

Column 1 in both panels shows the estimated value of the learning rate β from the most basic specification, where logged average defect rates for the week or day are regressed on the log of production experience at the beginning of the period. In this specification, production experience is simply the cumulative number of cars produced in periods prior to t : $E_t = \sum_{\tau=1}^{t-1} q_{t-\tau}$. Estimates of the learning rate β are similar in both the sampling frequencies, -0.284 in the weekly data and -0.301 in the daily data. Both specifications fit the data very well, with the R^2 of the weekly and daily specifications at 0.962 and 0.935, respectively. This fit can also be seen in Figure 2, which plots the logged average defect rate against cumulative production in the daily data. Given the power-law form of the learning by doing function, a learning rate of $\beta = -0.3$ implies that average defects will be roughly halved for every ten-fold increase in cumulative

with a model where a firm allocates its time between a) producing output with a production function fixed by the current “knowledge” level, where greater knowledge implies higher productivity, and b) engaging in learning that raises knowledge (and productivity) in future periods. (See Lucas and Moll (2011) for an example.) When faced with this resource allocation tradeoff, it is optimal under many conditions for the firm to allocate a relatively large amount of time to learning rather than producing when starting to make a new product, and hence operate at a low output rate, and then to steadily allocate greater shares of time to producing—i.e., run at a greater line speed—in later periods. This structure would explain the observed coincident patterns of increasing rates of production and decreasing defect rates. Interestingly, it would also imply the relationship between productivity and cumulative production in standard use in the literature is actually incidental to a deeper causal mechanism tying a firm’s productivity to its actively accumulated knowledge stock.

production.¹⁰ Thus reductions in defects are particularly notable early in the production process. In our sample, after weekly production begins in earnest (that is, with at least 100 vehicles being produced a week), cumulative production increases by a factor of 100 in only 8 weeks.

To distinguish whether the quality improvement is tied directly to production experience or more simply reflects progress that accrues with the passage of time, we add a time trend to the regression specifications. Empirical results are contained in column 2 of both panels of Table 1. The time trends are actually positive rather than negative, though small in magnitude and marginally significant, and the estimated learning rates are slightly higher. The quality improvement, therefore, appears to be related to production activity per se, not the passage of time since production began. Indeed, conditioning on production experience, the passage of time may even slightly decrease quality. This last fact suggests, perhaps, that an organization’s knowledge capital depreciates over time—there is “forgetting” (e.g., Benkard (2000), Thompson (2007)).

To allow more explicitly for potential forgetting, we estimate a specification that allows for knowledge depreciation over time. Specifically, we assume experience is accumulated according to a perpetual-inventory process: $E_t = \delta(E_{t-1} + q_{t-1})$. Experience at the beginning of period t is a fraction δ of the sum of two components: the experience at the start of the prior period and production in the prior period, q_{t-1} . Thus δ parameterizes forgetting, or more precisely, retention.¹¹

Nonlinear least squares estimates of the “forgetting” model are in column 3 of both panels.¹² The estimated learning rates are statistically indistinguishable from those in the specification that controls for a time trend. Not surprisingly, the estimated retention rate in the weekly data of 0.965 is smaller than the 0.985 rate seen in the daily data, but when the latter is

¹⁰ The multiple to production k that drops error rates to one-half of any initial level is $k = 2^{(1/\beta)}$. Thus with $\beta = -0.3$, $k = 2^{(1/0.3)} = 10.1$. Another measure of learning speed common in the literature is the progress ratio, the productivity gain—in standard settings, percentage drop in unit costs—obtained from each doubling of cumulative output. Here, a doubling of cumulative output leads defect rates to fall by 18.8 percent ($2^{-0.3} = 0.812$), which would be defined as a (quality) progress ratio of 0.812.

¹¹ We could have alternatively specified $E_t = \delta E_{t-1} + q_{t-1}$, so that the full amount of the prior period’s production is added to the experience stock. We use our approach because it is more intuitively appealing for the specifications below where the retention rate is a function of worker absenteeism. In any case, the results obtained with the alternative specification were very similar to those reported here.

¹² We set the starting values for the intercept and the learning rate equal to those estimated from the no-forgetting model in column 1, and the starting value for δ of 0.9. The column 3 results were not sensitive to these starting values.

compounded over five-day production week, the implied retention rate is 0.927. Thus the models imply that about three to seven percent of the plant's production experience stock is lost every week. This depreciation adds up quickly; over the course of a 45-week model year, only about 3 to 20 percent of the initial experience stock would remain if not replenished by production activity.¹³ Nevertheless, explicitly modeling the forgetting process does not substantially improve the ability of the power-law specification to fit the data, particularly relative to simply controlling for a time trend, as the model's R^2 values increase by less than 0.001 and 0.012 in the weekly and daily data samples, respectively.

Supplementary Evidence from Quality Audits. Our FIS data is the only source of information on the number of production defects for every car made in the plant. However, the automaker does conduct quality audits on randomly selected cars throughout the production shift (typically, this amounts to roughly 15-20 audited cars per day). These very thorough audits examine hundreds of details on finished automobiles, ranging for example from measuring the gap between the hood and the front fender along the hood's length to verifying that the headlights come on when the proper switch is actuated. Any problems that are found are scored into one of four categories by severity: 1-point defects (the most minor), 5-point defects, 10-point defects, and 20-point defects (the most serious). By way of example, a small irregularity in the hood-fender gap would be a 1-point defect, while a failure of the headlights to turn on when the switch is actuated would be a 20-point defect. The scores for the defects found on any given car are then summed to get the car's audit score. Higher scores reflect a greater number of more serious defects.

We have data on all quality audits conducted at the plant through Week 15 of Year 1—essentially, the first two-thirds of the production year. The weekly average audit score per car over this period is plotted in Figure 3.¹⁴ The quality improvement we document in the FIS data is reflected in the audit data as well. Audit scores start at a high level and quickly decline in the first several weeks of the production year. Eight to ten weeks after the production year began,

¹³ The daily results are also consistent with observed systematic differences in average productivity within the week. If we regress logged defect rates on logged (undepreciated) cumulative production and a set of day-of-week fixed effects, the Monday fixed effect has the highest value. That is, defect rates are on average higher on Mondays than any other day of the work week. This is consistent with forgetting happening over the weekend (though the average six percent gap between the two days' defect rates also suggests other deleterious factors are also at work on Mondays.)

¹⁴ A relatively small number of cars were audited during the production ramp-up in Weeks 34-36 of Year 1, so we have aggregated audit data from those weeks together, labeling the aggregate in the figure as Week 36 of Year 1.

average audit scores have fallen by about 70 percent from their initial levels. Scores then very gradually and noisily fall by perhaps another 10 percent of their initial level after this point until the end of the available data.

The similarity between this independent production defect measure and our core FIS data is reassuring in multiple ways. First, it eliminates the possibility that the drop in defects in the FIS data is simply an artifact of workers being less likely to report production problems as the year goes along, even if the frequency of actual defects remains the same. Second, it indicates that the production defects we measure in the FIS data, even if they are sometimes corrected later in the assembly process, are correlated with defects that would nevertheless leave the factory with the car. The quality audits are conducted after all assembly processes are done; if not for the audits, the audited cars would have certainly been shipped to dealers with the defects found in the audits. Thus our FIS defects matter (or are correlated with other problems that matter) to the car's end consumer. This is further supported by the fact that, as we discuss below, initial warranty claims on the cars are correlated with our FIS defect levels. Third, if we look at where the improvement in the audit scores comes from—that is, look at the changes in the relative frequencies of defects by severity score—the fastest drop in relative frequency is seen among the most severe 20-point defects, and the slowest drop occurs in the much more minor 1-point defects. (The frequencies of each severity level of defect relative to their levels at the beginning of the production year are shown in Figure A2 in the appendix.) This is a sign that quality improvement at the plant is a directed process; the defects with the larger expected impact on the customer are addressed faster.

III.2. New Shift and Product Variant Introductions

The results discussed thus far indicate that learning by doing is an important factor in this production process, particularly for the first few months of production after the initial ramp up. Multiple production ramp ups of various sorts in our sample provide the opportunity to better understand the specific mechanisms driving learning by doing.

Adding a Second Shift. We begin exploring the nature of the learning mechanism by examining defect rates as the second shift begins. Figure 4 shows average defect rates (again on a per car

basis, averaged over the week) separately by shift. (Figure A3 in the appendix shows the number of cars produced by shift in each workweek of the sample.)¹⁵

Notably, the second shift does not start with the high initial defect rates experienced by the first shift. Indeed, second shift defect rates are in fact *lower* than the first shift's contemporaneous defect rates from the outset, significantly so in the case of the daily data. Furthermore, this pattern holds on average throughout the production year. Second-shift defect rates are on average about 5-10 percent lower than first-shift rates, a difference that is statistically significant at conventional levels in both the weekly and daily data. The average difference declines over time as the two series converge, however.

For comparison purposes, we estimate shift-specific learning by doing rates, where logged average error rates on a shift are regressed on the log of cumulative production from that shift. Empirical results are shown in panel A of Table 2. The first shift's estimated learning rates in both the weekly and daily samples are similar to those found above in the overall sample. However, the estimated rates for the second shift are considerably smaller. This is largely a result of the flat defect rates observed during the first several weeks of the second shift's operations. Because cumulative second shift production is rising at a fast rate during this period while error rates remain flat, the estimated learning rate is pushed toward zero.¹⁶

These patterns indicate that whatever is learned in the early production period is not embodied solely in the workers on station during that period. The efficiency/quality gains seem instead to be fully incorporated into second shift production immediately, despite having completely new workers on the job. As Epple, Argote, and Murphy (1996) note in their own study with a similar finding of across-shift learning transfer, this pattern points toward the productivity gains from learning being embodied in the broader organization rather than being retained within the human capital of line workers.¹⁷

¹⁵ Cars are not produced start-to-finish within a single shift. We apportion cars to shifts using the same procedure described above for apportioning them across time periods. The fraction of a car produced on a shift equals the share of the seven benchmark production stations that we observe occurring during each shift.

¹⁶ This argument is confirmed in (unreported) regressions using only later weeks of second shift production, which yield faster estimated learning rates.

¹⁷ Of course, we cannot completely rule out a worker-embodiment hypothesis if the first-shift workers are able to fully convey their information to the second-shift workers during the week that the second-shifters observe the first-shifters in operation. However, the transfer of this knowledge, which took several months of production to build, would have to occur within one week. Further, even complete transfer would imply that the second shift would

We conduct several complementary tests to explore potential spillovers across the two shifts. The first is to examine whether first-shift defect rates are higher during the period in which the second shift is ramping up production. (We define ramp up as the first three weeks of second-shift production, the time it took second-shift output to rise to the level of the first). We do so by adding an indicator for the second-shift ramp-up period to the first-shift-specific learning regression in panel A of Table 2. Empirical results are contained in panel B of the table. Estimated coefficients on the ramp-up period indicator are positive in both the weekly and daily data, though only statistically significant in the latter. They indicate that first-shift defect rates are roughly 15 percent higher during the weeks the second shift is ramping up, controlling for the learning gains that have accrued to that point. A possible explanation for this temporary increase in defects is that it reflects reallocation of resources formerly used to reduce first-shift errors toward increasing second-shift production.

The two shifts' defect rates behave similarly in levels and changes. They have a correlation coefficient of 0.88 in the weekly data, though daily movements are more idiosyncratic, leading to a correlation of 0.14. Further, as we will see below, defect rates at specific stations on the production line are correlated across shifts within a given week. Nevertheless, we find little direct evidence of other experience spillovers across the two shifts after the initial ramp up.¹⁸

achieve defect rates on par with the first shift, yet as we discuss below, observed second-shift defect rates are significantly lower than those of the first shift.

¹⁸ We estimated specifications that allowed accumulated experience to be a weighted sum of cumulative production in each shift, based on the notion that shift-specific production experience might have greater impact than that in the other shift. These specifications (unreported here for space reasons) were not well identified and often resulted in large but imprecise estimated weights on the opposite shift's production. This likely reflects the multicollinear nature of the two shifts' production levels once both are up and running. Even so, the estimated learning rate with respect to this combined experience remained in a range similar to that found above.

We also estimated shift-specific learning specifications that instrument for the shift's cumulative production using cumulative production on the other shift. These specifications help address concerns that production levels might be endogenously chosen based in part on defect rates. If this is the case, the negative correlation we observe between defect rates and production experience might not just reflect learning, but also the choices of plant management to produce more when defect rates are lower. If such allocations of production occur across shifts, this instrumenting strategy will address any bias driven by endogeneity by identifying learning effects from changes in defects correlated with systematic changes in production rates across both shifts. In all specifications—using both first- and second-shift defect rates as dependant variables and in both weekly and daily data—the results of these IV estimates are qualitatively and quantitatively quite similar to those presented in Table 2, suggesting that endogenous production rates do not play a substantial role.

The Introduction of Additional Product Variants. The start of a second shift did not necessitate an intense new learning period. We find different data patterns, however, when we analyze the other type of production ramp up in our data: the introduction of the new model variants. This is apparent in Figure 5, which plots weekly average defect rates per car by model variant. Weekly production levels of each variant are shown in Figure A4.

The overall defect patterns are qualitatively similar for each of the three model types. Each starts at a high defect rate that quickly falls as production ramps up. For Model 1, the high initial learning rate discussed above resulted in an 85 percent drop in average defect rates by the time production of Model 2 began in mid-December. Despite these quality gains for Model 1, initial average defect rates on the new variant were much higher for Model 2, though they are not as high as Model 1's defect rates at inception. Model 2's defect rates subsequently decline quickly, dropping to meet the level of Model 1 after four weeks. When production of Model 3 begins in March, again initial defect rates are high and drop quickly after production begins. In this case, though, they do not fall to the level of the other models' defect rates before the end of the production year—over the last 12 weeks of the model year, Model 3 has about 16 defects per car on average, while Models 1 and 2 both have about 8 per car. Thus, unlike the introduction of the second shift, introduction of a new variant involves a considerable amount of learning.

We estimate model-specific learning by doing rates analogous to the shift-specific regressions discussed above, where experience is measured as cumulative production of the specific model variant. Empirical results are contained in panel A of Table 3. Model 1's estimated learning rates in both the weekly and daily samples are similar to those found in the overall sample above. The estimated rates for Models 2 and 3 are somewhat smaller, on the order of -0.2. We also test for spillovers across model variants. These results are shown in panel B of the table. Because of the timing of the model variant production, the Model 1 defect rate regressions include indicators for ramp-up periods of Models 2 and 3, while the Model 2 defect rate regressions include only a Model 3 ramp-up dichotomous variable indicator.¹⁹

Empirical results presented in the first two columns of panel B indicate that ramp ups of Models 2 and 3 do coincide with increases in defect rates in Model 1 production. Estimated coefficients on both ramp-up indicators are positive and statistically significant at conventional

¹⁹ We define ramp-up periods as beginning when production of the model variant reaches 100 cars per week and ending when production first exceeds that of at least one of the models whose production began earlier. This was a five-week period for Model 2 and a three-week period for Model 3.

levels. The Model 3 ramp up period, which is tied to an almost 30 percent increase in defect rates in Model 1 production, has a notably larger quantitative impact than the less than 10 percent increase tied to Model 2's ramp up. In contrast, similar spillovers from Model 3 ramp up into Model 2 defect rates are not apparent. Here, the ramp-up indicator coefficient is actually negative, small (on the order of four percent), and insignificant. These contrasting results across variants are particularly interesting in light of the fact that Model 3 is a specialized version of Model 1. The results suggests that solving the production problems that arise as Model 3 production begins detracts more resources from production of its closer cousin in product space, Model 1, than from the more distant Model 2.

Comparing the results in Tables 2 and 3 further emphasizes that learning by doing knowledge stocks in the plant are not accumulated simply by the plant's workers producing *any* type of car. Workers who have already acquired experience producing one product variety cannot fully transfer this knowledge to the production of new varieties. This is inconsistent with the most general of organizational learning models where workers simply need to become acclimated to operating together. It also suggests that learned knowledge is not simply contained in the individuals employed at the plant. It might be embodied in the physical capital (e.g., tools are adjusted or workstations redesigned) or some more amorphous organizational capital beyond a collection of single workers.

III.3. Station-Level Patterns

In addition to being able to observe the product quality that emerges from the overall production process—that is, the number of production errors per car—we observe outcomes for each of the hundreds of individual stations on the production line. With this information, we can answer questions about the composition of productivity gains that are beyond the reach of most prior learning-by-doing studies.

Distribution of Defects. We first explore the station-level distribution of defect rates. In particular, we measure the skewness of defect rates across production stations and test for intertemporal changes in this skewness. This enables us to determine whether learning by doing reflects across-the-board defect reduction throughout the stages of the production process, or

whether the gains primarily reflect improvements in some key processes that initially had high defect rates.

Figure 6 shows the evolution over the sample of various quantiles of the station-specific defect rates. Each station's defect rate is computed as the number of defects recorded at that station in a given workweek divided by the total number of cars produced that week. In Figure 6, stations are sorted into quantiles based on that week's performance, i.e. a station will move from one quantile to another over time if its relative defect rate fluctuates.²⁰

Panel A of Figure 6 shows the 50th, 75th, and 95th percentiles of the absolute levels of station-level defect rates.²¹ Defect rates are highly skewed across stations, with a thick right tail. In the first full week of production, for example, the 95th percentile station defect rate is 20 times the median rate that week, and five times the 75th percentile rate. The large early-period productivity gains seen in the aggregate series come almost exclusively from defect reductions in the right tail of stations. However, the basic shape of the distribution stays the same throughout the production year. This can be seen in panel B of Figure 6, which shows the *ratio* of a quantile's defect rate in a particular week to that same quantile's defect rate during the first week of production. That is, it shows the percentage decline of the quantile's defect rate relative to itself. All quantiles follow essentially the same pattern, with considerable reductions in relative defect rates early in the production period that decelerate over time.²² Aggregate learning-by-doing at the plant therefore reflects a proportional tightening of the entire station-level defect rate distribution, with all quantiles experiencing similar percentage declines in defect rates.²³

Persistence. Figure 6 sheds no light on the question of whether the particular stations remain at the same quantile in the distribution throughout the course of the sample, or whether some stations are a major source of defects at one point of time, but are nearly defect free at other

²⁰ Below, we analyze stations as a function of their initial defect rates and follow the performance of a fixed set of stations over time.

²¹ We do not show lower percentiles because the skewness of the distribution makes them difficult to distinguish given the scale of the figure.

²² We have added the 25th percentile series to this panel, as it can now be distinguished from the others. It follows the same pattern.

²³ The quantiles of the distribution are computed while including stations that report no defects, which number between 5-10 percent of stations in any given week. Qualitatively similar data patterns remain in both absolute and relative defect levels among the corresponding quantiles of the distribution conditional on stations reporting at least one production defect during the week.

times. We investigate the persistence of station-level defects as follows. First, we split the model year production into three equal-length periods: beginning, middle, and end. We then calculate the station-level defect rate distribution within the *first* period, and assign a station to its quintile in the distribution during that period. Finally, we calculate the number of defects in each of the three periods by the first-period quintile of their station of origin. Any changes in the relative contribution of a particular quintile over time will be informative about the degree of persistence in station-level defect rates.

We present the results of this exercise in Figure 7, which can be interpreted as follows. Panel A shows the results for the total number of production defects during the period. The bar labeled “1” breaks the total number of defects that occurred during the first production period, about one million, into quintile-specific components. The top portion of the bar is the portion of defects that occurred in the stations in the top quintile of the distribution. This skewness of the defect distribution is obvious in the figure; roughly 90 percent of all defects occurred at stations in the top quintile. Further, almost all of the remaining 10 percent of defects occurred at stations in the second quintile of the defect distribution, as indicated by the second color in the bar. Stations in the bottom three quintiles account for a miniscule portion of the total number of defects during the period.

The bar labeled 2 repeats the exercise for the second production period, but it holds the quintile assignments constant. Note first that the height of the overall bar is lower in period two because there are fewer total defects in this period is lower because of learning by doing. Because quintile assignments are held constant across time in Figure 7, the top color in the bar shows the number of defects in the second production period that are accounted for by stations that were in the top quintile of the distribution *in the first period*. If defect rates within stations have no persistence and are completely random over time, then we should expect each first-period quintile will account for 20 percent of all errors in the later two periods. This is clearly not the case. Stations at the right tail of the distribution in the first period still account for the vast majority of production defects in the second period. The bottom three quintiles again account for only a tiny part of the total number of defects in the period. Defect rates at the station level are therefore quite persistent. Persistence isn’t perfect, however. The *share* of defects accounted for by top-quintile stations, while quite large, does shrink between the first and second periods. This can be seen more easily in the proportional results in panel B of Figure 7, which simply

normalizes the upper panel numbers so that values are expressed in terms of the share of defects in the period rather than the absolute number. These patterns continue in the period 3 results. While total defect rates fall, the most error-prone stations early on the model year still account for a large fraction of the defects at the end of the year, though this share declines over time as mean reversion in station-level error rates becomes more prominent.

A second, related exercise to investigate station-level persistence is to compute a transition matrix for each station in terms of its location in the defect rate distribution early and late in the production periods. Again, we assign each station to its quintile in the defect rate distribution, but this time separately for the first six weeks of full-scale production and the last six weeks of the model year.²⁴ We then compute the fraction of stations in each cell of the 25-cell matrix of early-by-late period quintiles. Perfectly random error rates would have each cell accounting for four percent of the distribution (i.e., the 20 percent of stations in each of the early-period quintiles will be split evenly five ways across the late-period quintiles). Greater persistence shows up as heavier entries along the diagonals of the matrix.

We show the results in Table 4. The general persistence of error rates is again evident: the diagonals are the largest elements by row, and values tend to get smaller as the distance from the diagonal grows. However, it is also clear that defect rates are not completely immutable; many stations shift one quintile and others more. Interestingly, persistence is greatest at both the very top and very bottom of the defect rate distribution. Just over half of stations in the lowest- and highest-defect rate quintiles in the first six weeks of full production are still in the same quintile at the end of the year, and another third are within two quintiles of their original positions. On the other hand, only about 30-40 percent of stations in the second through fourth quintiles early in the year are in the same quintile at the end. Stations are roughly equally likely to jump from very bad to very good as they are to move sharply in the other direction.

Besides looking at temporal persistence, we can investigate the correlation of station-level error rates across shifts. To do so, we group all stations by their quintile within the shift-specific defect rate distribution during a particular week, and compare a given station's quintiles across the first and second shifts that week. Empirical results are contained in Table 5. Each cell

²⁴ We use the first six weeks after the second shift begins—actually the seventh through twelfth week of production after first-shift ramp-up—for the sake of a closer comparison to the final six weeks of the production year, which of course also involves two-shift operations. Similar results are obtained if we compare the first six weeks of first-stage production to the year-end defect rates, however.

reports the fraction of the station-weeks in each quintile-by-quintile grouping; these numbers sum to 100 percent. Rows correspond to the stations' first-shift quintile; columns correspond to the second-shift quintile. For example, 11.1 percent of stations were in the first (lowest) defect rate quintile in both the first and second shifts. Another 6.3 percent were in the lowest quintile among the first-shift defect rate distribution but the second quintile of the second-shift distribution.

Table 5 indicates that, just as there is persistence in a station's relative defect rate across weeks in the production year, defect rates are persistent across shifts within a week. Again the heaviest elements tend to be along the diagonals, and the persistence, this time across shifts within a week rather than over the production year, is greatest for stations at the distributions' tails. Stations that are error-prone during the day tend to be error-prone at night as well, even though the personnel working at the station have changed. This further indicates that not all learned production knowledge is embodied in the workers themselves. Instead, an important component of it seems to be tied to the particular capital of a station, the organizational capital managing that station across shifts, or temporal fluctuations in the quality of the parts being used as inputs. That said, the correlation is again far from perfect; there is considerable mass off the diagonals, and a small fraction of stations have very high defect rates during the first shift but very low rates during the second, or vice versa.

III.4. Defect Spillovers across Cars

In this section, we drill even further down into the production process to see if defect rates for a given station are correlated *across cars produced in sequence*. That is, we see whether defects spillover across cars located next to one another on the production line. There are several mechanisms through which this could happen. If a defect occurs on one particular car, a worker may sometimes attempt to rectify the problem before the car moves to the next production process. This can lead to a longer time being spent on the operation than typical, reducing the time available for the operation on the following vehicle, which could in turn result in a new defect on that car. More broadly, if a worker is distracted or otherwise performing poorly, then s/he is likely to be distracted on consecutive cars. Further, when a defect is not addressed at the location it occurs, a worker further down the line might sometimes attempt a fix. This could in turn result in a defect occurring in an adjacent car at that later station.

We are interested in spillovers in our context because one mechanism through which learning-based productivity improvements can occur is if production experience reveals how to limit the extent of spillovers. To see how this might work, imagine a stylized world where a certain number of defects happen by random chance and the nature of these defects is to adversely impact every car until someone catches and fixes the mistake. Learning-based productivity gains in this world could come from two sources: a reduction in the probability that defects arise in the first place, or more expedient identification and rectification of the defects that arise.

We investigate the issue by estimating a model that relates defects on one car to those on adjacent cars on the line. Specifically, we regress the defect count of a given car on the defect counts for each of the 25 cars that preceded it along the line. The coefficients on the preceding cars' defect counts gives the marginal effect of a defect on that car on the defect count of the reference car. (This is an analog to an autoregressive model where a variable is regressed on its past values, and the coefficients reflect the marginal effect of a change in the value in period $t - k$ on the value in period t .) We also control for day fixed effects to account for changes over the sample period in average defect rates; the coefficients in this model are identified from *within day* variation, i.e. relative to other cars produced on that day, do defects on proximate cars lead this car to have more problems? The presence of spillovers will lead to higher defects on one car leading to more defects on others that follow it in the production sequence. We do this separately for three different production periods: early in the model year, in the middle of the year, and toward year's end, to see if any spillover patterns change over the production year.

We show empirical results in Figure 8. Panels A through C show the estimated regression coefficients for each of the three production periods, along with their 95-percent confidence intervals. In short, spillovers clearly exist. Production defects on one car raise the likelihood of defects on cars that come later in line. As one might expect, the magnitude of these spillovers falls as the distance between cars on the assembly line grows. Still, defects on one car have statistically significant spillovers on at least the next 15 cars on the line. The economic size of these spillovers is nontrivial. A one-defect "shock" on a given car leads, on average, to about one-tenth of a defect on the next car in the line. The same shock leads to roughly 0.07 defects on the car that comes two positions later. The tenth car down the line experiences 0.02 more defects on average.

It does not appear, however, that a decrease in spillover effects over time explains the learning by doing patterns we observe. This can be seen in panel D of Figure 8, which plots the coefficient estimates for the three time periods on a common set of axes for ease of comparison. The three sets of coefficients have basically the same shape, with no noticeable decline in the size of defect spillovers across the production year.

III.5. Absences and the Role of Worker-Embodied Learning by Doing

Several of our results imply that a considerable amount of the production experience stock is embodied in either physical or organizational capital, and not solely in the individual workers themselves. Nevertheless, the results do not strictly rule out the potential for worker-embodied learning. We push more on worker-embodied learning in this section. To do so, we exploit worker absenteeism data. If individual workers are the store of any productivity gains, then we should see slower learning or even productivity regression during times in which more workers are absent.

The overall patterns in absenteeism are seen in Figure 9, which plots average weekly absentee rates over the production year. Absenteeism is volatile, with particularly low rates occurring when the second shift trained and began production (weeks 40 and 41 of Year 1) and in January of Year 2; spikes occurred in mid-November of Year 1 (a combination of coming off Veteran's Day weekend and the beginning of the state's firearm deer hunting season) and the week of Presidents' Day. Average absenteeism rates rise over the production year.

The divergence in the absenteeism and defect rate trends over the production year indicates that learning-by-doing productivity growth is occurring in spite of trends in worker attendance rather than because of them. This is further reflected in log-linear learning rate regressions that add absenteeism rates as controls. In both weekly and daily data, the estimated learning rates are statistically unchanged from the benchmark specification, and the coefficients on absenteeism rates have opposite signs, depending on data frequency (negative in the weekly regression, positive in the daily specification) with very large standard errors. If there are any effects of absenteeism on productivity and learning, then, they must be from a more subtle source.

We follow two empirical approaches to deepen our inquiry. In the first, we disaggregate our production and absentee data to test if absenteeism and defect rates are correlated at finer

levels, controlling for overall trends. In the second, we revisit our “forgetting” specifications and allow the rate at which knowledge stock depreciates in a given period to vary with the fraction of workers who are absent. The notion behind this latter specification is that worker absences prevent any knowledge embodied within them from being applied to the production process in their absence, and limits the accumulation of new knowledge.

We compute defect and absences by department-shift-day cells, where department denotes a major portion of the line’s operations.²⁵ Combining these data, we have a panel of 3293 observations. We then regress the natural logarithm of defect rates on the natural logarithm of employee absences, controlling for department-shift and day fixed effects. The resulting coefficient on absences is 0.156 (s.e. = 0.029). The economic size of this relationship is modest, however: this elasticity of 0.156 implies a one-standard deviation increase in absences raises defect rates by about 1/7th of a standard deviation.

The absenteeism-augmented forgetting specification allows the retention rate δ for a period to vary with plant-level absentee rate in that period. Specifically, we let $\delta_t = \frac{1}{1+\gamma \cdot abs_t}$, where abs_t is a measure of absences in period t and γ is a parameter we estimate. Note that with this functional form, if $abs_t = 0$, then $\delta_t = 1$. As absences grow, the retention rate falls at a rate that depends on γ . Table 6 shows the results of this specification for weekly and daily data. We estimate the specification using nonlinear least squares, with the same starting values for intercept and slope as in the constant-retention-rate specification in Table 1, and the starting value for γ chosen so that at the average value of abs_t in the sample, the implied value of δ_t equals the estimated coefficient from the earlier specification.

In both the weekly and daily data, γ is positive and statistically significant; periods when more workers are absent have lower retention rates. However, the magnitude of the variation in δ_t is economically small. While the average δ_t implied in the weekly data is 0.974, the standard deviation across weeks is only 0.006, and the range is 0.964 to 0.984. Similarly, the average implied daily δ_t is 0.989, with a standard deviation of 0.003 and a range from 0.965 to 0.995. We simulated what these estimates imply average defect rates would be if absences were held to the

²⁵ There are seven departments. In the order in which they occur on the line, they are Body-in-White (the assembly of the car’s metal frame and major body pieces like fenders and the hood), Paint, Trim (parts of the car that are not part of the powertrain and steering systems, like seats, handles, and dashboard), Chassis (engine, transmission, and other major mechanical systems), Final (finishing details and parts), Reprocessing (addressing any stages not fully completed or needing further attention) and Quality-Control (operational and aesthetic inspection and testing).

level of their 25th percentile in the sample. Consistent with the small variation in the estimated δ_i , this reduction in absences would lower average defect rates by only 0.6 percent in the weekly specification and 1.3 percent in the daily specification.

Therefore absenteeism has statistically significant effects on defect rates, both directly and through the rate at which acquired learning-by-doing knowledge is retained. However, its impact is economically small. This further reinforces the findings above, suggesting much of the plant's knowledge stock built by learning-by-doing is embodied in the physical or broader organizational capital of the plant.

III.6. Implications for Warranty Payments

The profit benefits of learning-by-doing gains captured in lower unit production requirements, the standard output examined in the literature, are clear enough: productivity gains are reflected in lower production costs. Our focus here is on output quality measures instead. While many models allow a link between product quality and profits, this connection is harder to measure in our data. We cannot gauge how quality improvements are reflected in consumers' willingness to pay for the cars produced by the plant, for example. However, we do have data on one profit factor tied to quality: warranty claims. Specifically, we have data on car-level warranty payments for all vehicles produced at the plant. While the data has a limited time horizon—we can only follow claims for cars produced during our sample year for at most a year after their production date, limiting our investigation to quality problems that arise quickly after purchase—we are able to explore if reductions in defect rates affect one bottom-line profit factor.

To measure the quantitative relationship between defect rates and warranty costs, we regress the automaker's warranty payments for a particular car on the number of production defects that occurred when the car was assembled. Our sample consists of all of the roughly 190,000 cars produced during the model year. The car-level match of these warranty and production data sets is done using the cars' VINs. We also include week-of-production fixed effects, so that we are comparing cars produced contemporaneously. This is necessary because cars produced at different times in the year will have different possible warranty claims horizons due to the right-truncation of our claim data (e.g., a car assembled later in the production year will have lower claims, all else equal, because it had not been in service as long as a car produced early in the year).

The vast majority of cars produced do not have warranty claims made on them within nine months of production, of course, so much of our sample is comprised of zeros. Still, there remains a positive relationship between production defects on a car and the amount of warranty payments that the auto company makes on it. The regression of payments on defects yields a slope of 47.6 cents per production defect (s.e. = 7.0 cents). To benchmark the warranty savings due to learning-by-doing effects, if we apply this slope to the roughly 60 defect-per-car drop in average defect rates over the production year (see Figure 1), this is a savings of \$28.56 per car. Applied to the 190,000 cars made the year of our sample, this is \$5.4 million in warranty claims savings. This is, of course, a very loose lower bound for the profit gain due to reductions in defect rates, as it does not include reduced future warranty claims later in the cars' service lives nor any increases in consumers' willingness to pay for higher quality cars.

IV. Discussion and Conclusions

Increases in total factor productivity are the key driver of long-term economic growth. While technological innovation is an important source of TFP growth, another key cause of TFP gains are improvements in the efficiency with which outputs are produced from existing technologies and inputs. Scores of empirical studies have documented such efficiency gains over the past several decades, showing that a simple progress function can explain the efficiency gains. Theoretically, learning by doing is now a trademark in explaining features of modern economies, such as comparative advantage and key sources of innovation and technical change. What is often missing, however, is knowledge of the specific mechanisms through which cumulative production leads to greater efficiency levels.

Our study provides a step in this direction by empirically analyzing a unique data set from an auto assembly plant. We find improvements in product quality similar to those seen in the earlier literature on production costs, with high defect rates at the beginning of production that fall quickly as plant operations begin. The overall relationship between defect rates and cumulative production is log-linear, with a learning rate of -0.3, implying that each ten-fold increase in cumulative production halves average defect rates.

Interestingly, the learning process is not restarted when the plant's second shift begins operations. Instead, the shift actually begins at average defect rates below the first shift's contemporaneous rates. This is evidence that not all learning-by-doing knowledge gains are

embodied in the plant's workers. In support of this interpretation, worker absenteeism is shown to affect defect rates, but the impact is economically small, highlighting the importance of the plant's physical and organizational capital as depositories of learning-by-doing know-how. Furthermore, the most defect-prone of the hundreds of processes involved in assembling a car have defect rates that are highly correlated across shifts within the same time period, even though the workers completing these tasks are different. These combined patterns indicate that much of what is learned in the plant very quickly becomes embodied in the physical or broader organizational capital of the plant rather than remaining with workers. This is consistent with one of the key institutionalized learning mechanisms at the plant: a system where workers can note problems they are having with the production process on their team's whiteboard. Plant management periodically visits (on the order of once every two weeks or so) every team to address any issues on its board that can be dealt with cost-effectively. This mechanism allows any knowledge gained by the workers to be quickly embodied in the process itself.

Beyond their theoretical and empirical relevance, our findings have managerial implications. In modern economies, shortening development times and fostering process improvements represent key drivers of success. Understanding the sources of learning by doing, rather than assuming that they are exogenous, provides managers with an important tool to achieve both high production rates and a high level of utilization earlier in the ramp up process.

We note that our results reflect the learning process at one particular plant, in one particular industry (though an important one), and in one particular time. Thus our study shares the issue of generalizability with most of the rest of the learning-by-doing literature. Nevertheless, we believe the results here offer insights that can be cautiously extended directly to other production operations, particularly complex manufacturing processes. Further, even when direct extension is not warranted, the results can be used to direct future research on learning by doing.

References

- Argote, Linda and Dennis Epple. 1990. "Learning Curves in Manufacturing." *Science*, 247(4945): 920-4.
- Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies*, 29(3): 155-173.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2007. "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment." *Quarterly Journal of Economics*, 122(2): 729-73.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2009. "Social Connections and Incentives in the Workplace: Evidence from Personnel Data." *Econometrica*, 77(4): 1047-94.
- Benkard, Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review*, 90(4): 1034-1054.
- Bloom, Nick, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2011. "Does Management Matter? Evidence from India." NBER Working Paper 16658.
- Darr, Eric D., Linda Argote, and Dennis Epple. 1995. "The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises." *Management Science*, 41(11), 1750-62.
- Dutton, John M. and Annie Thomas. 1984. "Treating Progress Functions as a Managerial Opportunity." *Academy of Management Review*, 9(2), 235-47.
- Epple, Dennis, Linda Argote, and Kenneth Murphy. 1996. "An Empirical Investigation of the Microstructure of Knowledge Acquisition and Transfer through Learning by Doing." *Operations Research*, 44(1): 77-86.
- Foster, Lucia, John Haltiwanger and C.J. Krizan. 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." *NBER Studies in Income and Wealth, vol. 63: New Developments in Productivity Analysis*. Chicago and London: University of Chicago Press: 303-63.
- Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. 2003. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy*, 111(3): 465-497.
- Hossain, Tanjim and John A. List. 2009. "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." NBER Working Paper 15623.
- Jarmin, Ronald S. 1994. "Learning by Doing and Competition in the Early Rayon Industry." *RAND Journal of Economics*, 25(3): 441-54.

- Jovanovic, Boyan, and Yaw Nyarko. 1995. "A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves." *Brookings Papers on Economic Activity, Microeconomics*: 247-305.
- Krueger, Alan B. and Alexandre Mas. 2004. "Strikes, Scabs, and Tread Separations: Labor Strife and the Production of Defective Bridgestone/Firestone Tires." *Journal of Political Economy*, 112(2): 253–289.
- Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review*, 90(5), 1346-1361.
- Lazear, Edward P. and Kathryn L. Shaw. 2007. "Personnel Economics: The Economist's View of Human Resources." *Journal of Economic Perspectives*, 21(4): 91-114.
- Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1): 3-42.
- Lucas, Robert E., Jr. and Benjamin Moll. 2011. "Knowledge Growth and the Allocation of Time." NBER Working Paper 17495.
- Lundberg, Erik. 1961. *Produktivitet och Räntabilitet*. Stockholm: Norstedt & Söner.
- Middleton, K.A. 1945. "Wartime Productivity Changes in the Airframe Industry." *Monthly Labor Review*, 61(2): 215-225.
- Sinclair, Gavin, Steven Klepper, and Wesley Cohen. 2000. "What's Experience Got to Do with It? Sources of Cost Reduction in a Large Specialty Chemicals Producer." *Management Science*, 46(1): 28-45.
- Thompson, Peter. 2001. "How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study." *Journal of Political Economy*, 109(1): 103-137.
- Thompson, Peter. 2007. "How Much Did the Liberty Shipbuilders Forget?" *Management Science*, 53(6): 908-918.
- Thompson, Peter. 2010. "Learning by Doing." in *Handbook of Economics of Technical Change*, Bronwyn H. Hall and Nathan Rosenberg, eds. Amsterdam: North-Holland/Elsevier.
- Thornton, Rebecca Achee and Peter Thompson. 2001. "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding." *American Economic Review*, 91(5): 1350-68.
- Wright, Theodore Paul. 1936. "Factors Affecting the Cost of Airplanes." *Journal of Aeronautical Sciences*, 3(4): 122-128.

Yang, Xiaokai and Jeff Borland. 1991. "A Microeconomic Mechanism for Economic Growth."
Journal of Political Economy, 99(3), 460-82.

Figure 1. Average Defect Rates per Car

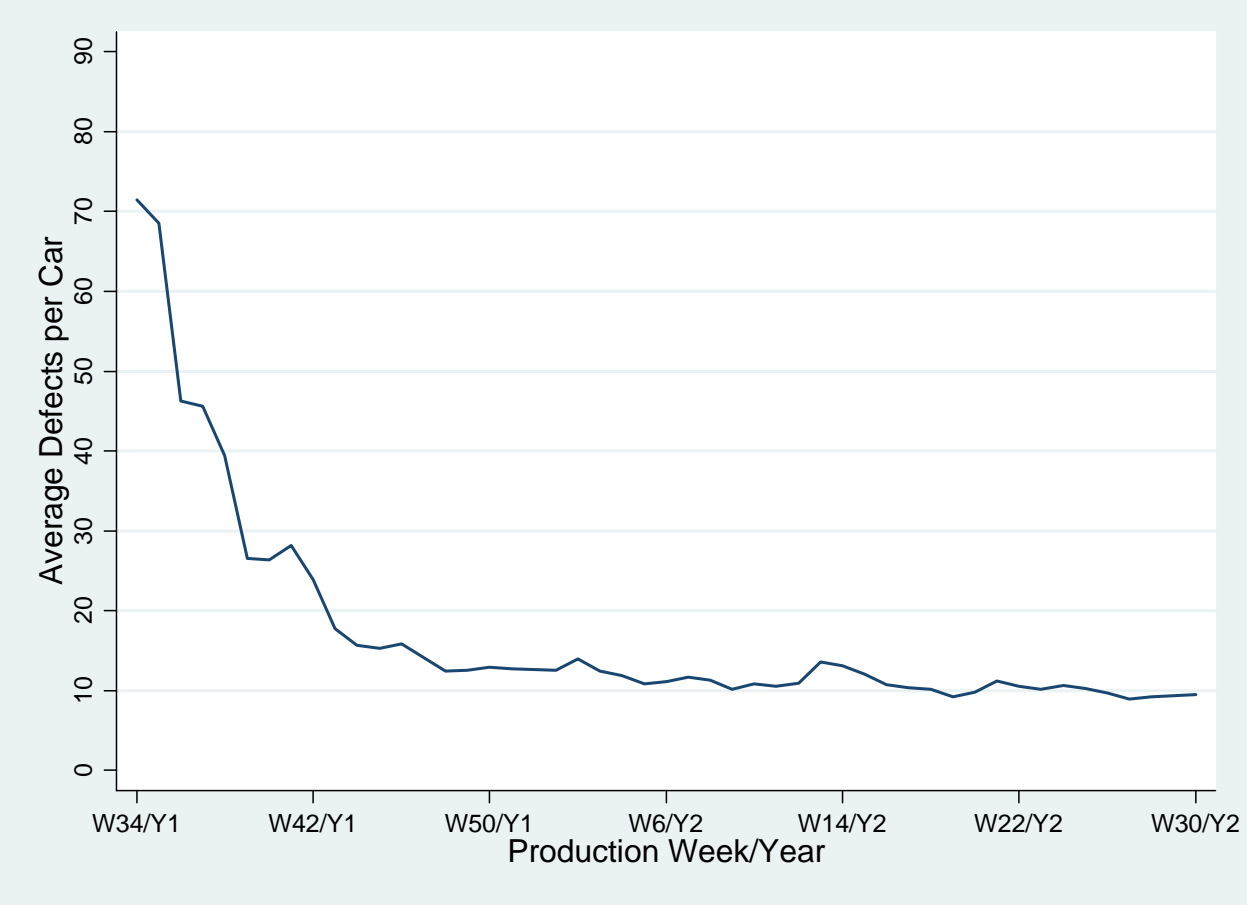


Figure 2. Log Defects per Car vs. Log Production Experience (Cumulative Output), Daily Data

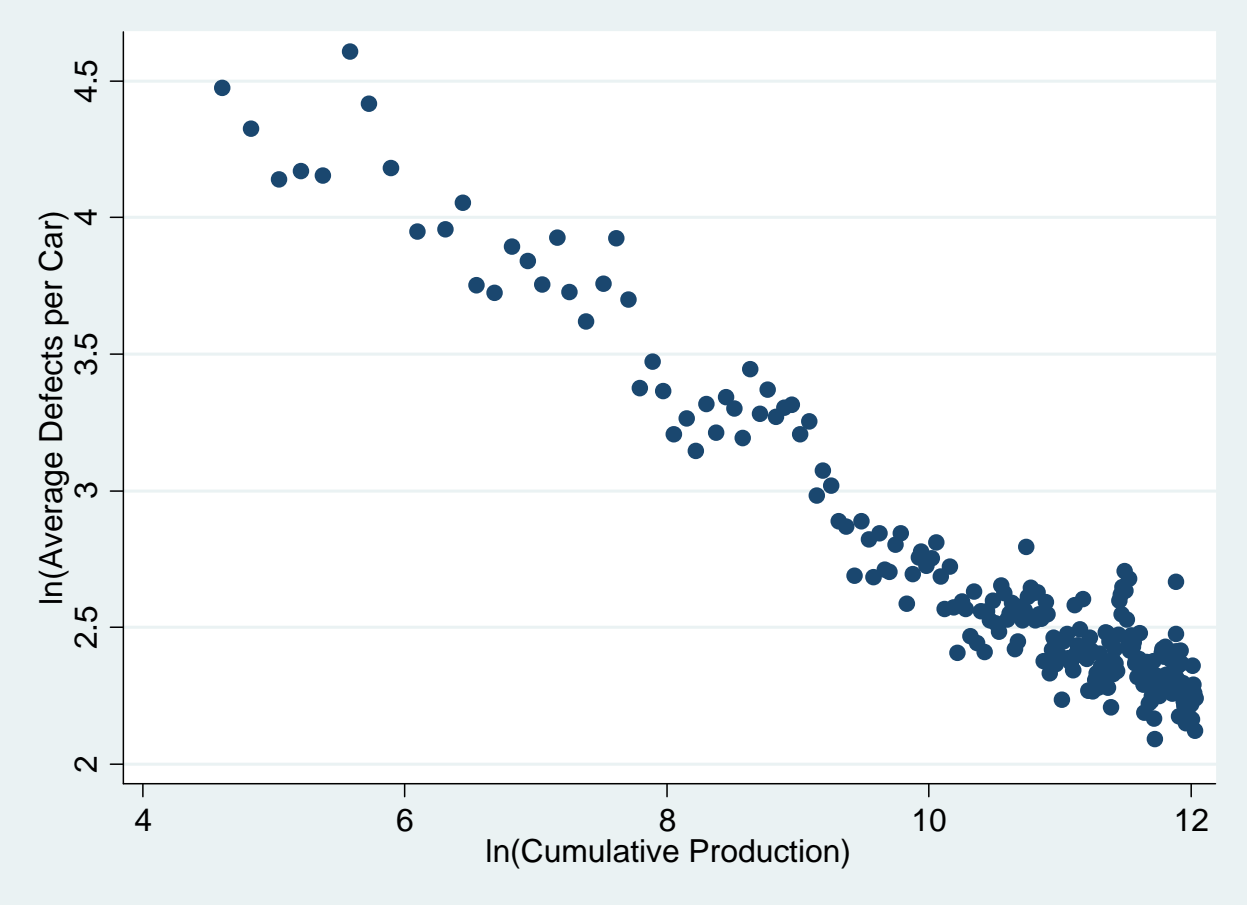


Figure 3. Average Quality Audit Scores per Car

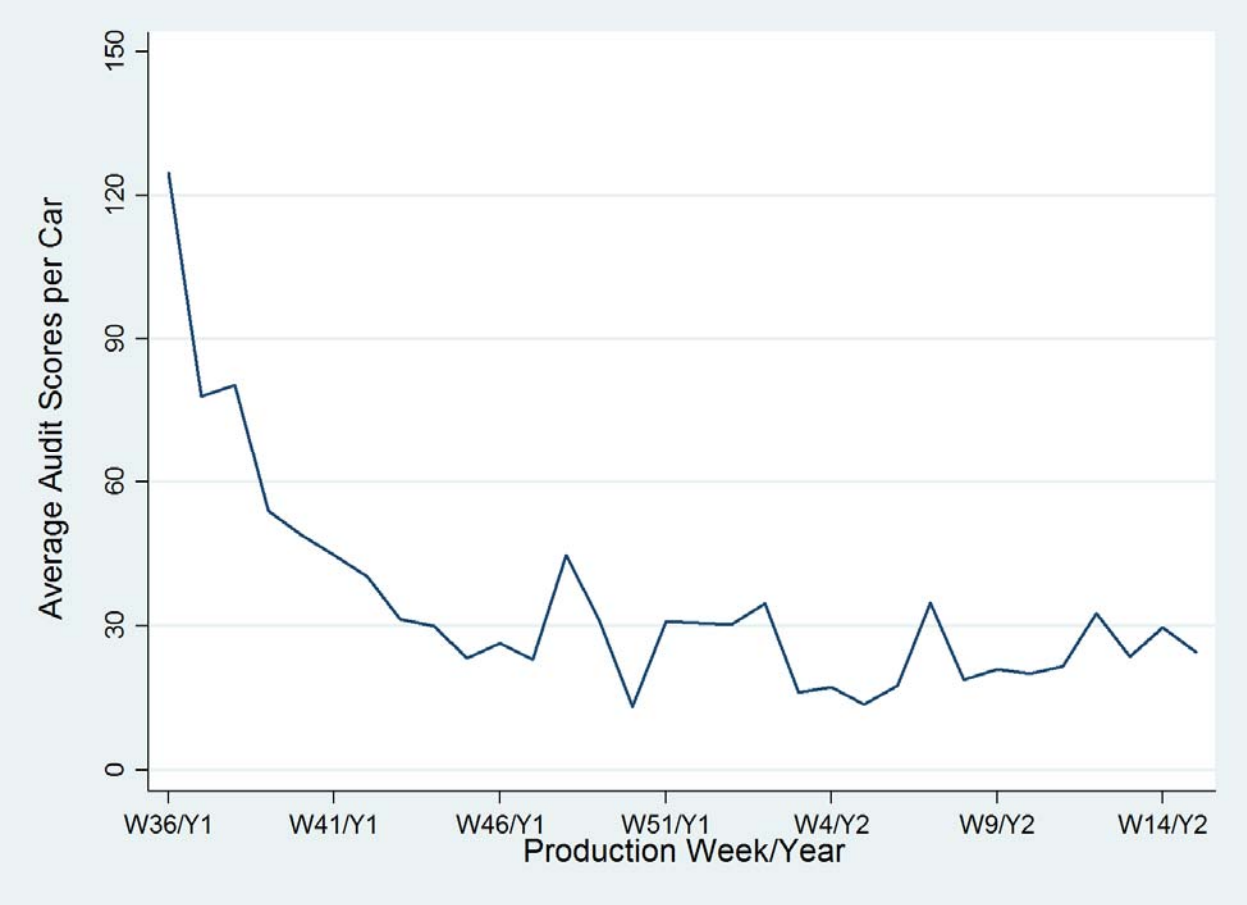


Figure 4. Average Defect Rates per Car, by Shift

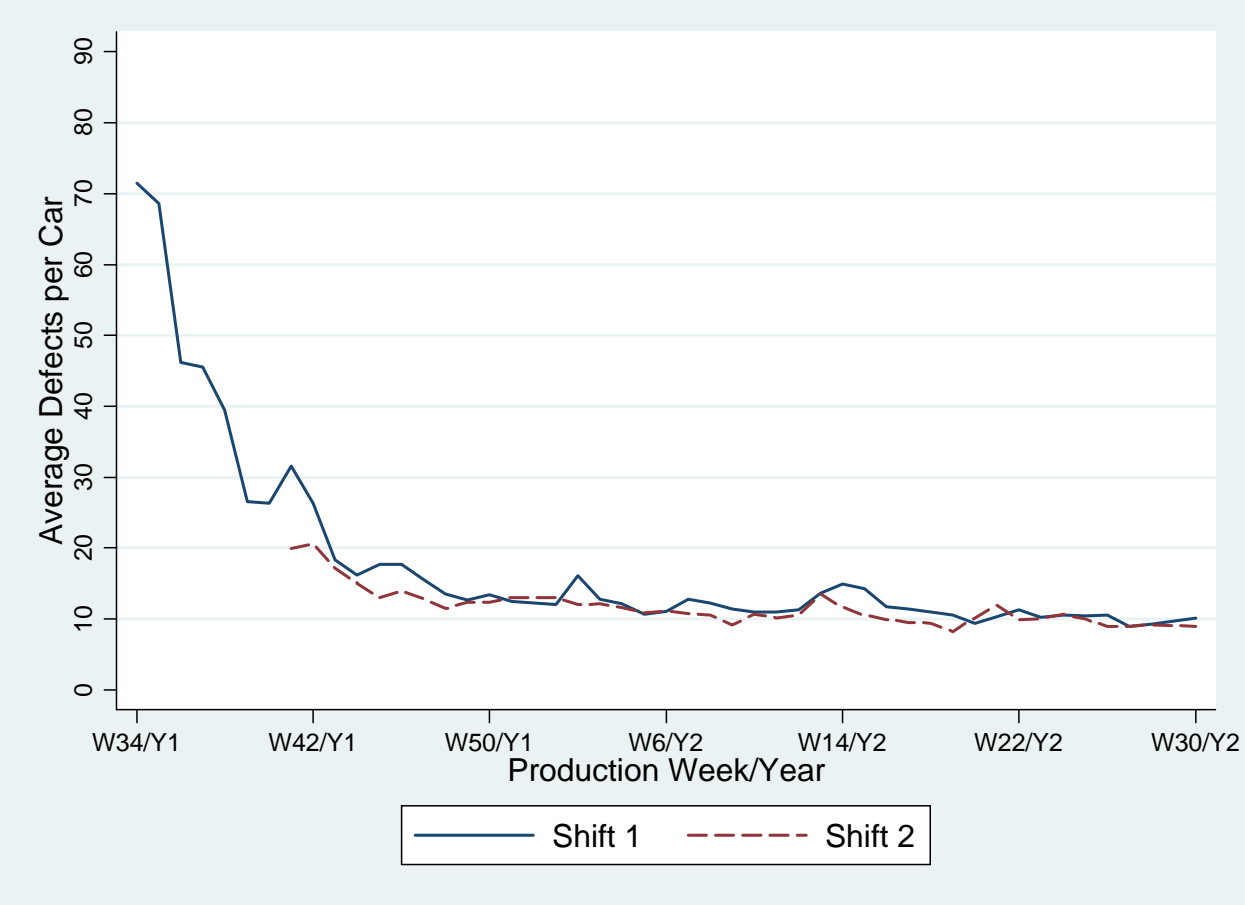


Figure 5. Average Defect Rates per Car, by Model Variant

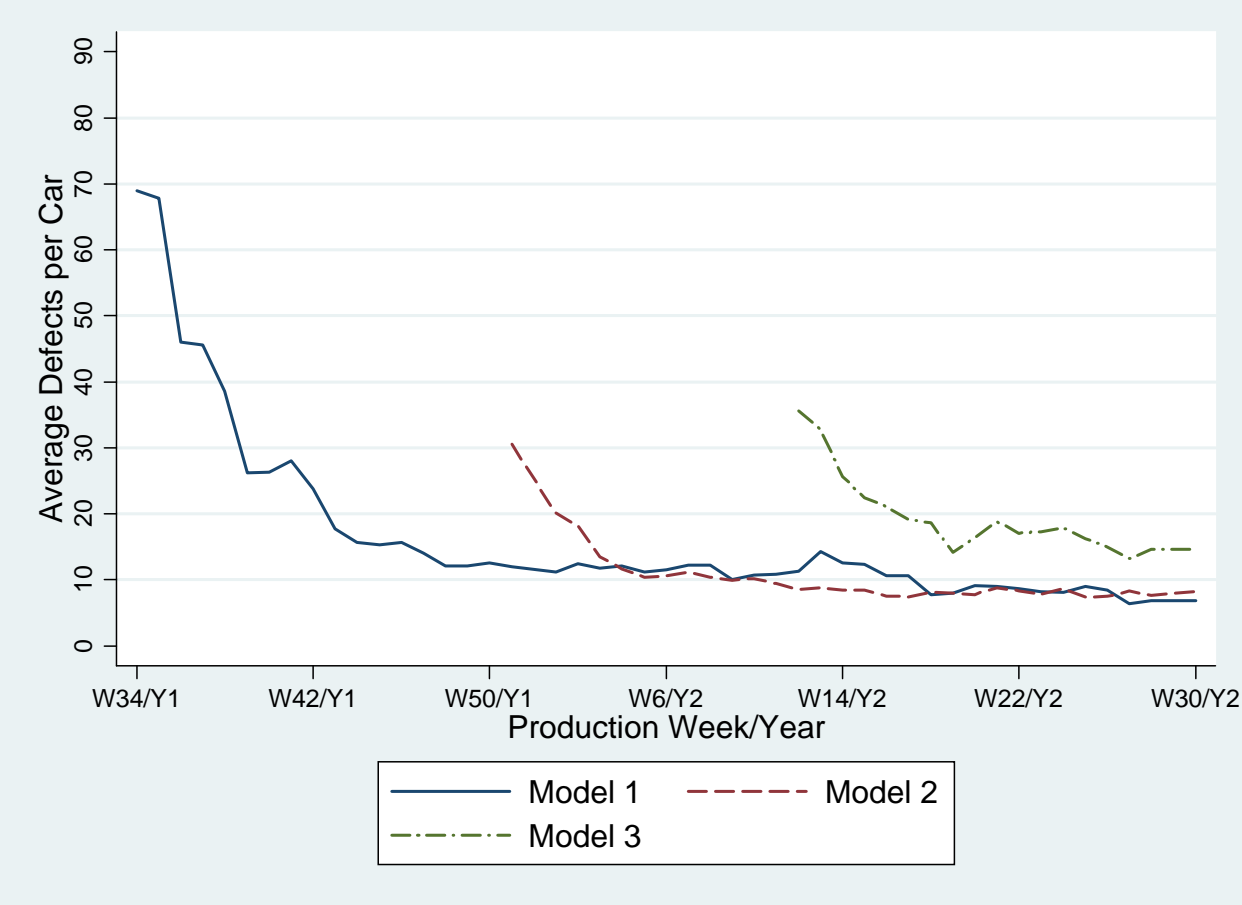
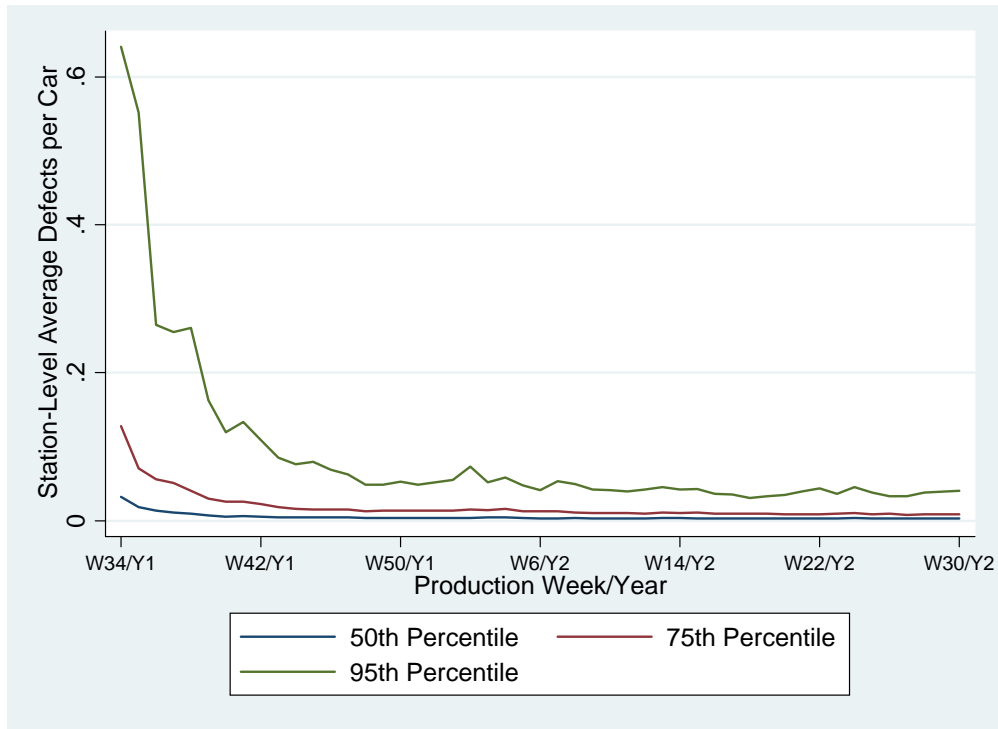


Figure 6. Distribution of Station-Level Defect Rates

A. Absolute Levels



B. Levels Relative to Quantile's First-Week Level

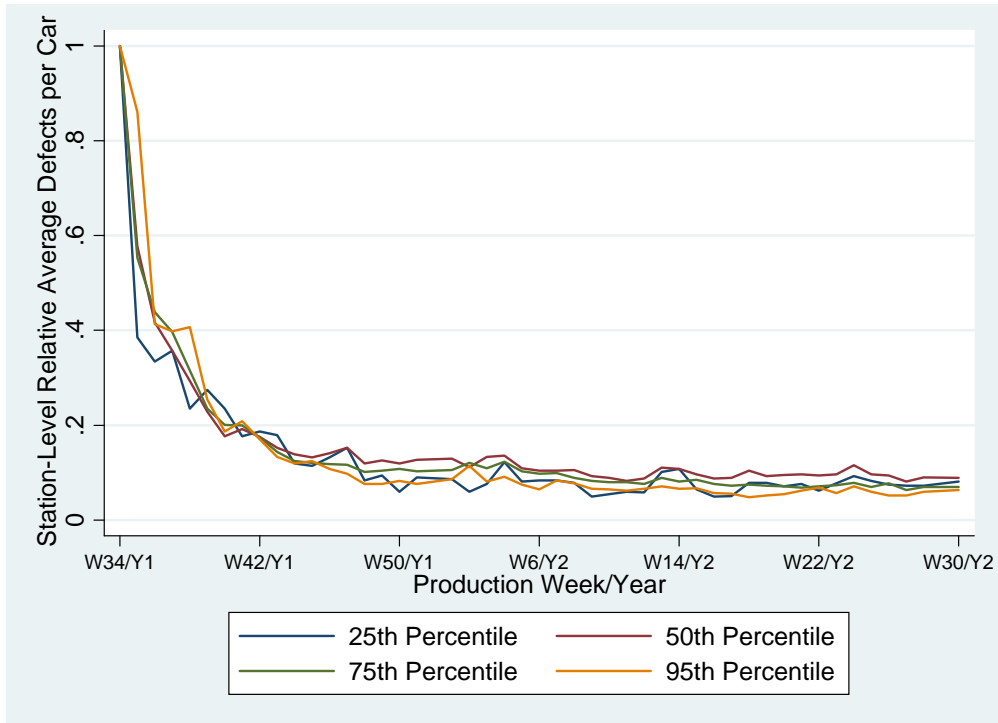
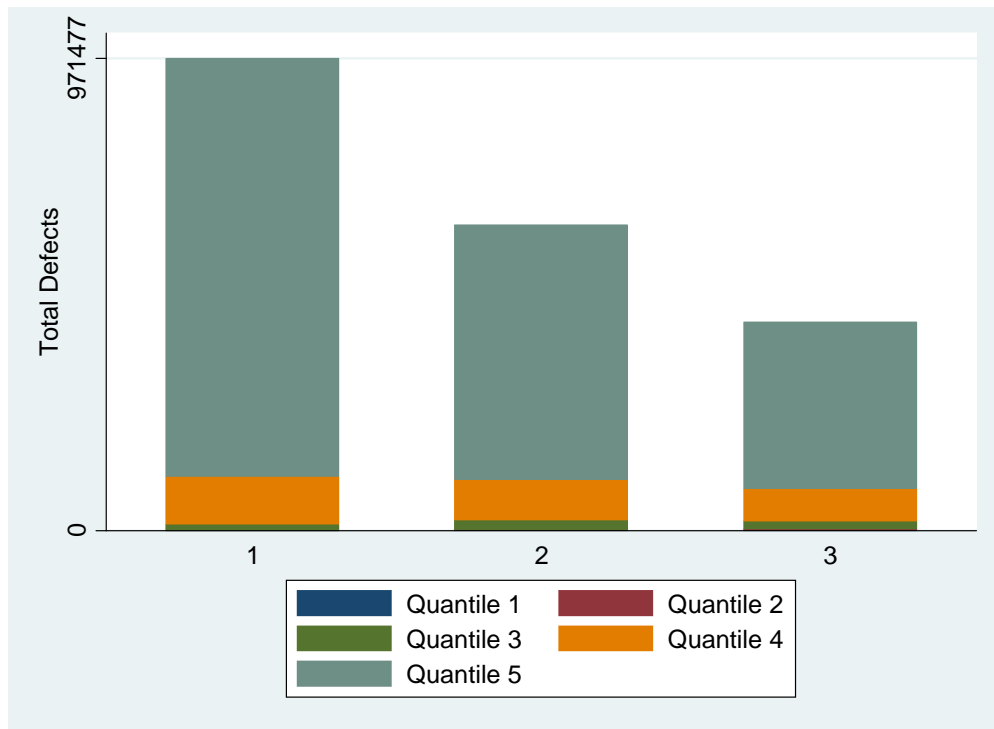


Figure 7. Defects by Station's Quintile in Defect Rate Distribution

A. Levels



B. Proportions

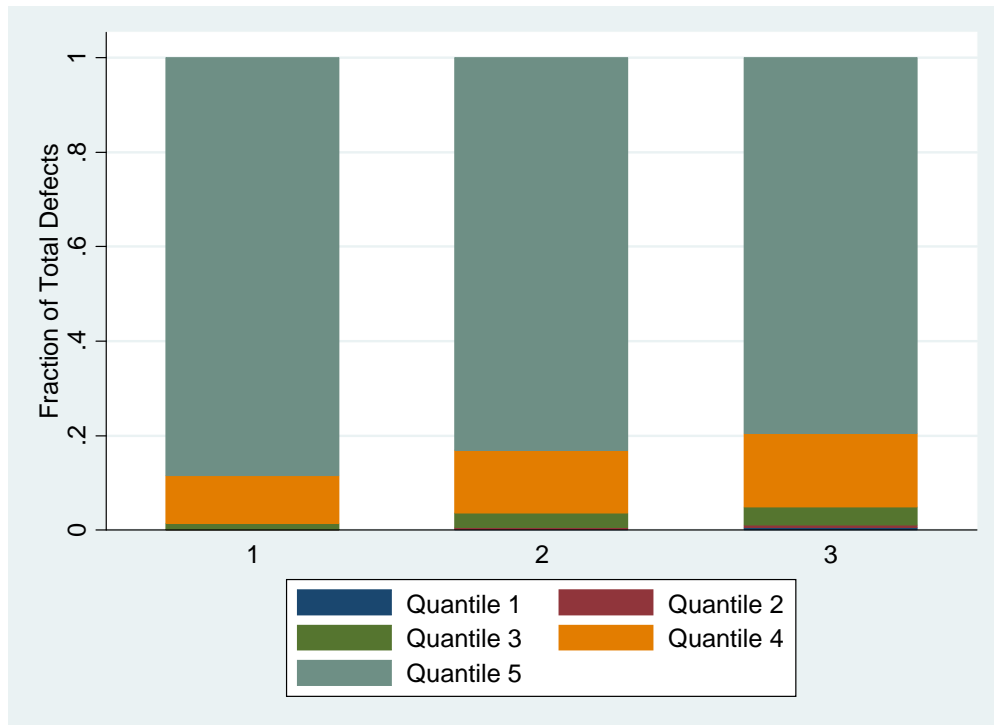
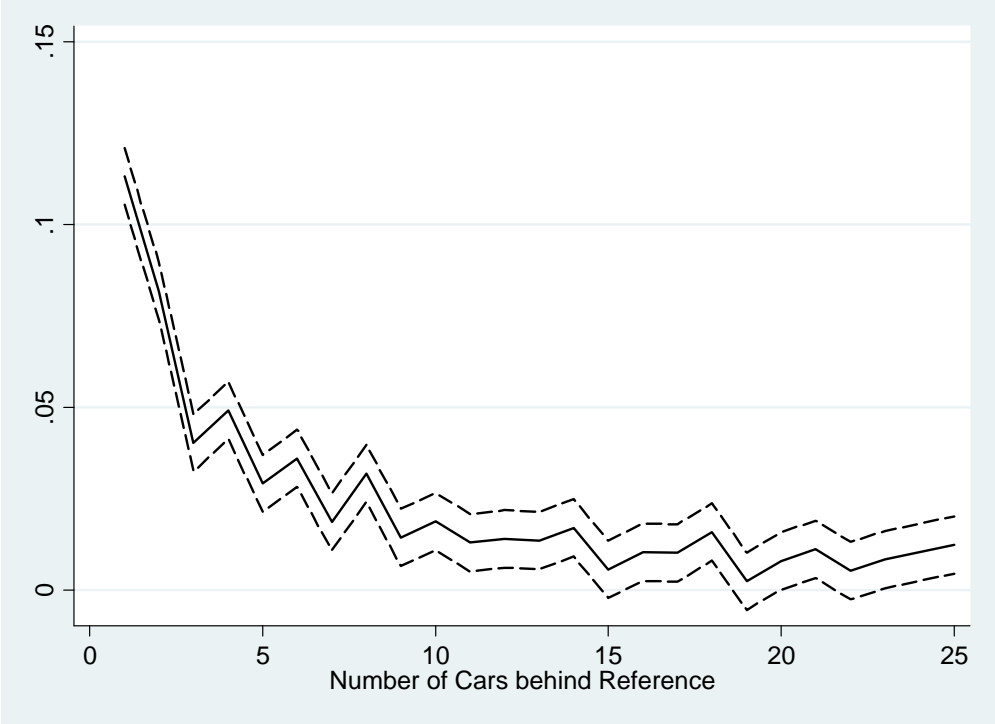


Figure 8. Defect Spillovers across Cars

A. Early Period



B. Middle Period

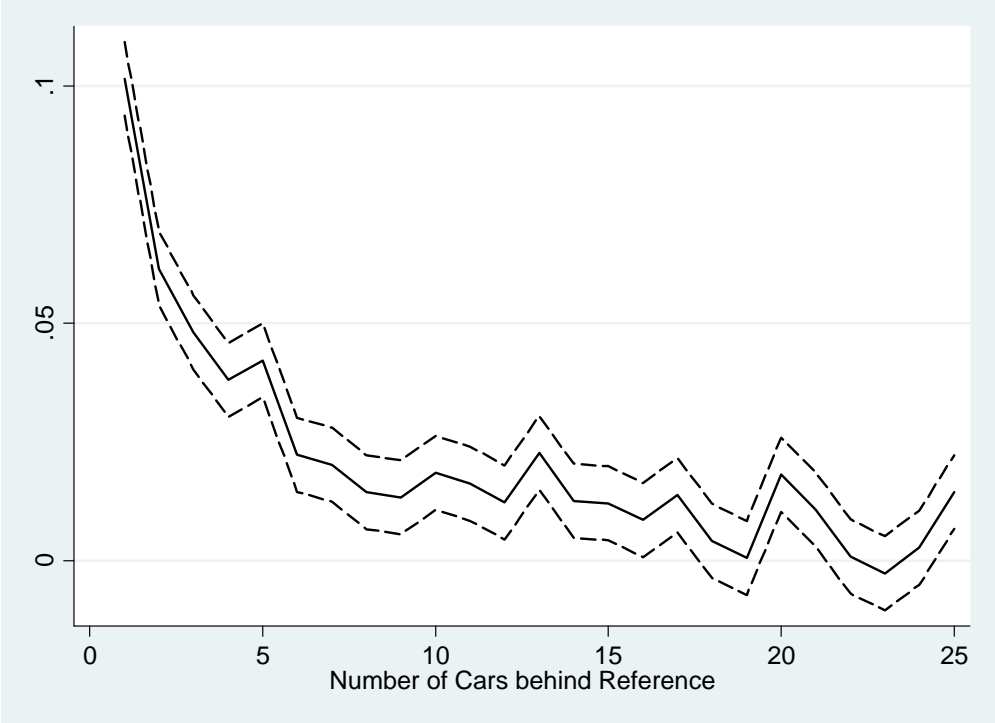
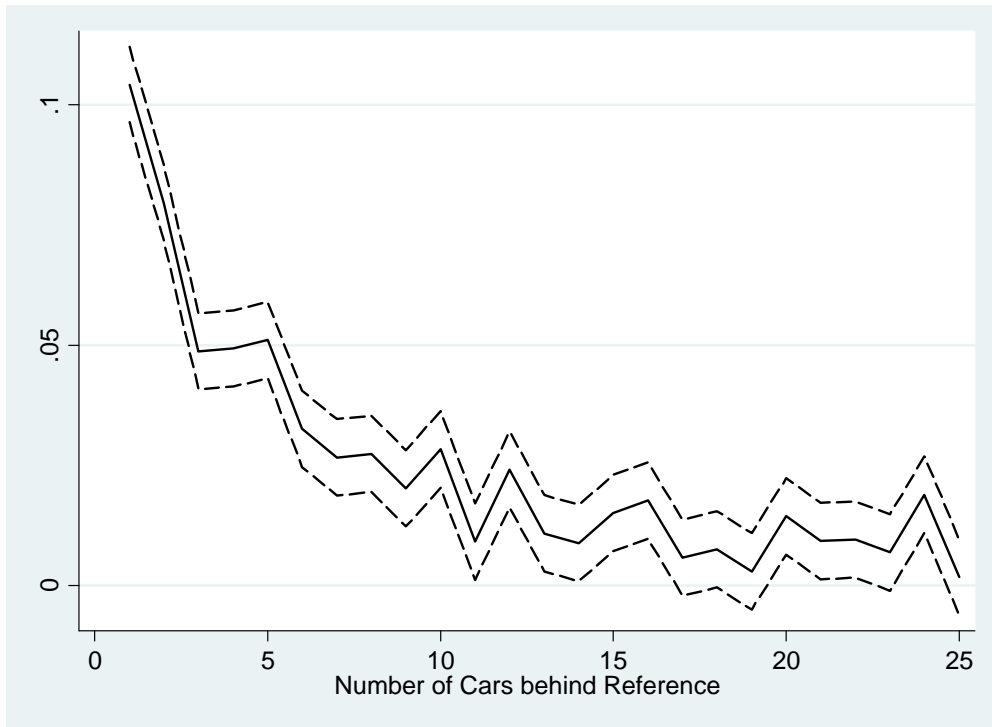


Figure 8 (cont.). Defect Spillovers across Cars

C. End Period



D. Comparison

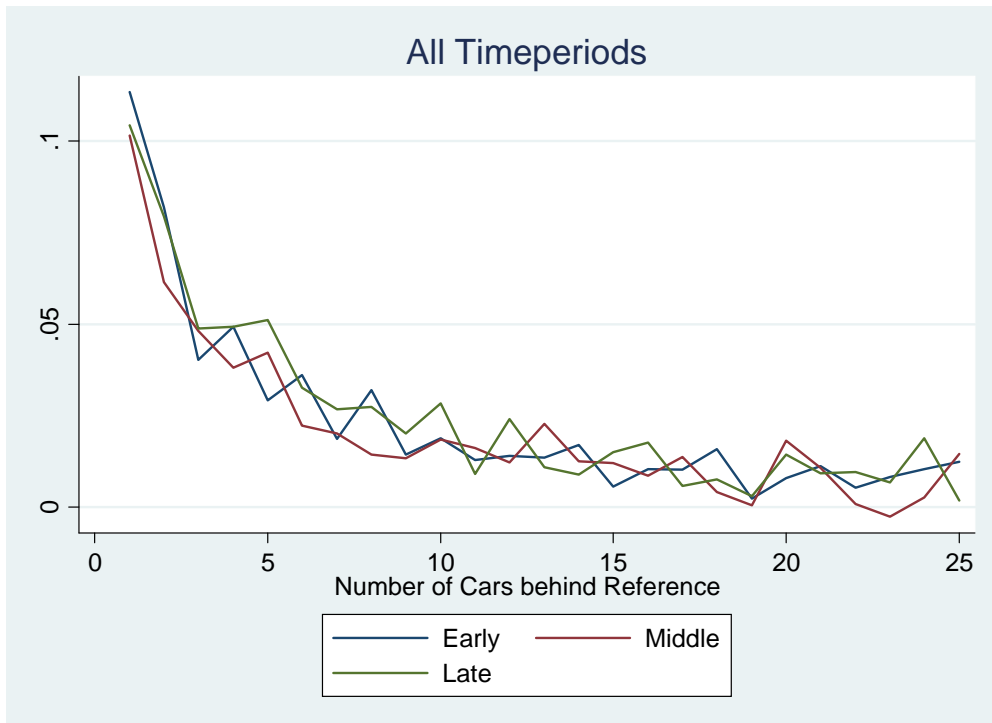


Figure 9. Average Weekly Absentee Rates

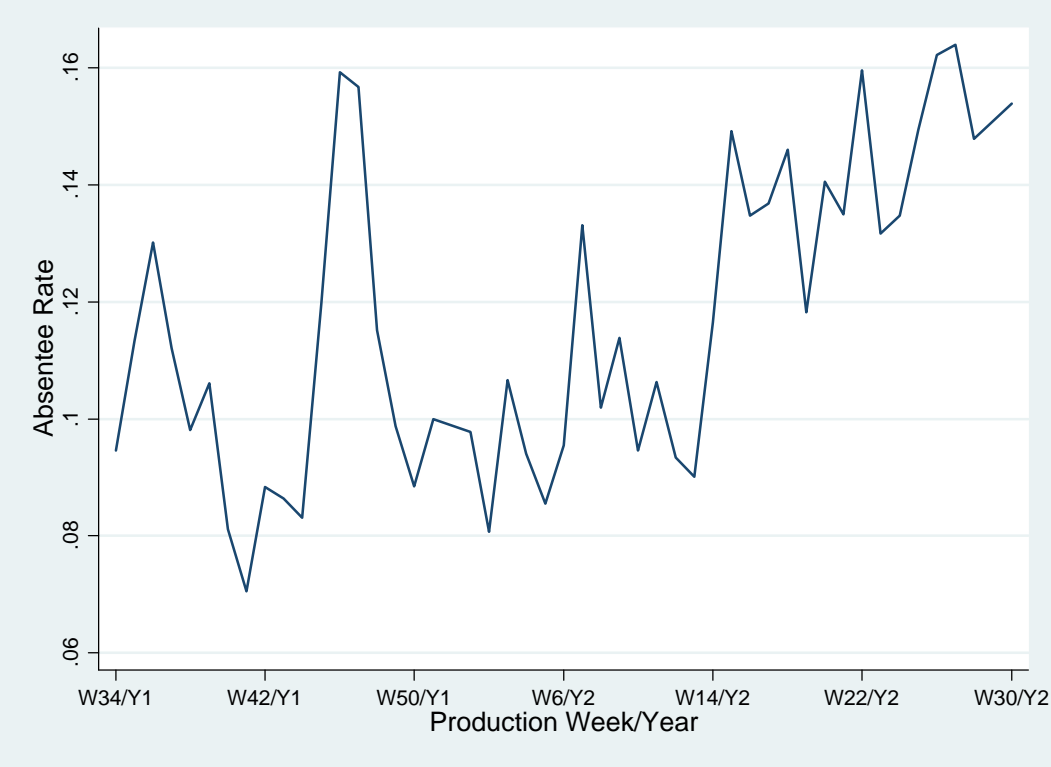


Table 1. Estimates of Learning by Doing Power Law Specification

A. Weekly Data

	[1]	[2]	[3]
Estimated learning rate (β)	-0.284* (0.008)	-0.329* (0.018)	-0.293* (0.026)
Time trend		0.007* (0.002)	
Retention parameter (δ)			0.965* (0.023)
N	47	47	47
R ²	0.962	0.969	0.962

B. Daily Data

	[1]	[2]	[3]
Estimated learning rate (β)	-0.301* (0.006)	-0.361* (0.013)	-0.356* (0.013)
Time trend		0.001* (0.0002)	
Retention parameter (δ)			0.985* (0.003)
N	224	224	224
R ²	0.935	0.946	0.947

Notes: These panels show the results of estimating power law learning by doing specifications of the type $S_t = AE_t^\beta$, where S is productivity at time t (average quality in our case), E_t is production experience up to that point, and A and β are parameters. In columns 1 and 2 of both panels, production experience is the cumulative number of cars produced before the current period. In column 3, the production experience stock is built up by a perpetual inventory process where experience at the beginning of the period is a fraction δ of experience at the start of the prior period plus production in the prior period. See text for details. Heteroskedasticity-robust standard errors in parentheses. An asterisk denotes significance at the five percent level.

Table 2. Shift-Specific Learning by Doing and Ramp-up Spillovers

A. Shift-Specific LBD

	Weekly First Shift	Daily First Shift	Weekly Second Shift	Daily Second Shift
Estimated learning rate (β)	-0.318* (0.011)	-0.324* (0.022)	-0.148* (0.010)	-0.026 (0.021)
N	47	224	39	190
R ²	0.943	0.906	0.795	0.019

B. Ramp-Up Spillovers

	Weekly First Shift	Daily First Shift
Estimated learning rate (β)	-0.314* (0.011)	-0.337* (0.008)
Second Shift Ramp-Up	0.174 (0.105)	0.151* (0.058)
N	47	224
R ²	0.949	0.911

Notes: These panels show the results of estimating power law learning by doing specifications of the type $S_t = AE_t^\beta$ separately for each shift and where experience is shift-specific. Panel B adds an indicator variable for the period in which second-shift production ramps up. See text for details. Heteroskedasticity-robust standard errors in parentheses. An asterisk denotes significance at the five percent level.

Table 3. Model-Specific Learning by Doing and Ramp-up Spillovers

A. Model-Specific LBD

	Weekly Model 1	Daily Model 1	Weekly Model 2	Daily Model 2	Weekly Model 3	Daily Model 3
Estimated learning rate (β)	-0.331* (0.014)	-0.355* (0.009)	-0.188* (0.015)	-0.204* (0.012)	-0.214* (0.012)	-0.236* (0.011)
N	47	224	30	143	18	86
R ²	0.933	0.897	0.939	0.811	0.910	0.824

B. Ramp-Up Spillovers

	Weekly Model 1	Daily Model 1	Weekly Model 2	Daily Model 2
Estimated learning rate (β)	-0.340* (0.015)	-0.363* (0.010)	-0.188* (0.015)	-0.203* (0.012)
Model 2 Ramp-Up	0.083* (0.036)	0.084* (0.025)		
Model 3 Ramp-Up	0.279* (0.066)	0.282* (0.045)	-0.031 (0.021)	-0.041 (0.022)
N	47	224	30	143
R ²	0.948	0.912	0.940	0.813

Notes: These panels show the results of estimating power law learning by doing specifications of the type $S_t = AE_t^\beta$ separately for each model variant and where experience is model-specific. Panel B adds indicator variables for periods in which production of later variants ramps up. There are no such regressions for Model 3 as no new variant is introduced after it. See text for details. Heteroskedasticity-robust standard errors in parentheses. An asterisk denotes significance at the five percent level.

Table 4. Station Transition Matrix from Early (First Six Weeks of Production Year) to Late (Last Six Weeks) Defect Rate Distributions

		Station's Late Period Quintile				
		1st	2nd	3rd	4th	5th
Station's Early Period Quintile	1st	12.0	5.3	2.0	1.1	0.4
	2nd	4.6	7.8	3.7	1.8	1.4
	3rd	1.0	4.8	6.7	5.3	2.5
	4th	1.1	1.2	4.9	7.9	4.8
	5th	1.4	1.0	3.0	3.8	10.9

Notes: This table shows the percentage of production line stations in each quintile of the defect rate distribution across stations during the first and last six weeks of production. For example, 12.2 percent of the stations in the sample were in the first quintile of the defect distribution (i.e., those with the lowest defect rates) in both the early and late production periods. Another 4.9 percent were in the first quintile during the early period and the second quintile during the late period, and so on.

Table 5. Across-Shift, Within-Week Station Defect Quintile Correlations

		Station's Second Shift Quintile in Week				
		1st	2nd	3rd	4th	5th
Station's First Shift Quintile in Week	1st	11.1	6.3	2.6	1.2	0.4
	2nd	8.7	4.8	3.5	1.8	0.6
	3rd	4.1	4.6	5.5	4.7	1.6
	4th	1.3	2.0	4.0	6.9	4.5
	5th	0.6	0.6	1.5	4.3	12.6

Notes: This table shows the percentage of production line stations in each quintile of the shift-week's defect rate distribution. For example, 11.1 percent of stations in the first quintile of the first shift's defect rate distribution during the week (i.e., those with the lowest defect rates) were also in the first quintile of the same week's distribution for the second shift. Another 6.3 percent were in the first quintile of the first shift but the second quintile of the second shift that week, and so on. (Row and column sums aren't always exactly 20 percent because of ties.)

Table 6. Worker Absenteeism Rates and “Forgetting”

	Weekly Data	Daily Data
Estimated learning rate (β)	-0.288* (0.012)	-0.345* (0.009)
Absences on retention (γ)	0.228* (0.017)	0.114* (0.021)
N	47	224
R ²	0.961	0.945

Notes: This table shows the results of estimating power law learning by doing specifications of the type $S_t = AE_t^\beta$, where S is productivity at time t (average quality in our case), E_t is production experience up to that point, and A and β are parameters. Production experience stock is built up by a perpetual inventory process with a production experience retention rate of δ that is a decreasing function of the absenteeism rate during the production period. See text for details. Heteroskedasticity-robust standard errors in parentheses. An asterisk denotes significance at the five percent level.

Figure A1. Weekly Production

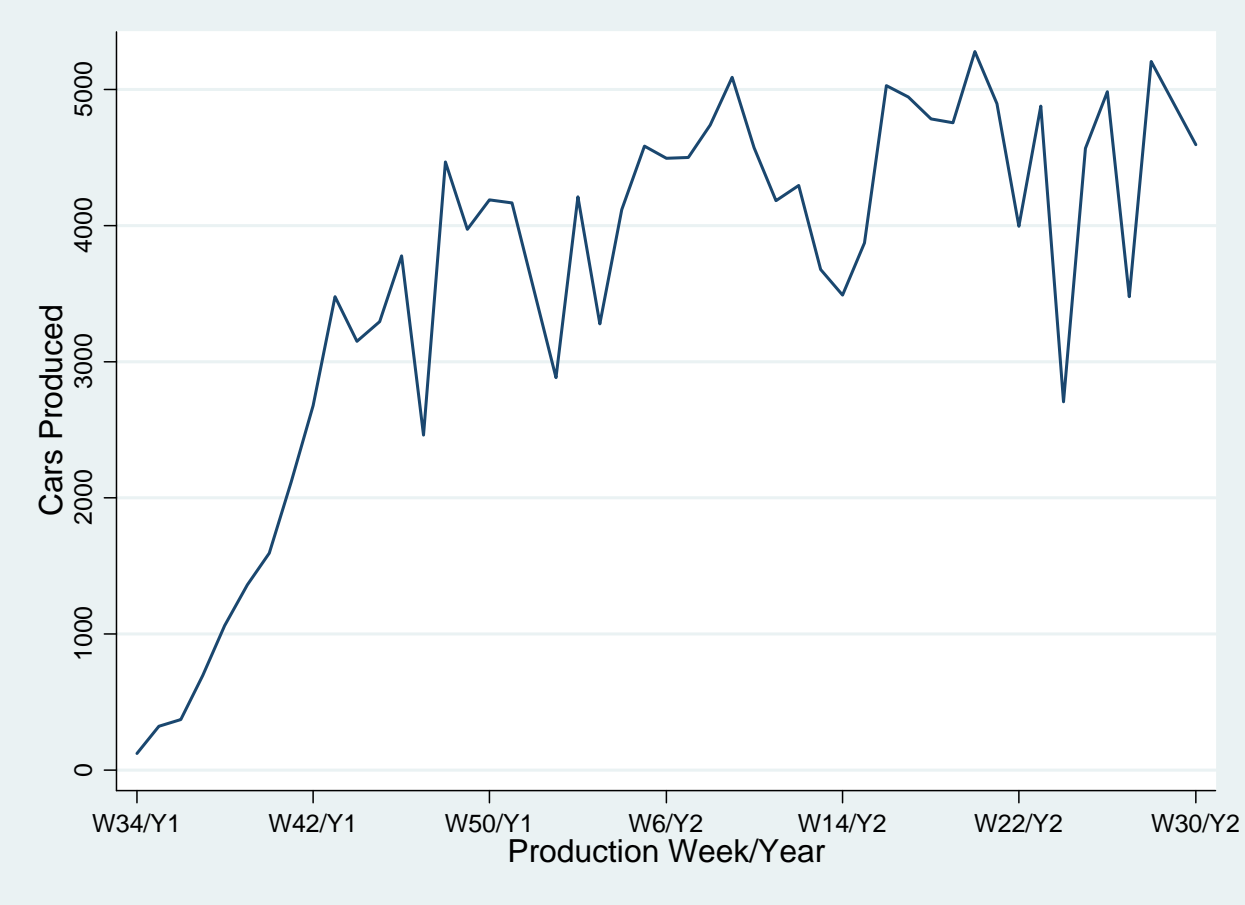


Figure A2. Audit Score Defects by Severity Level

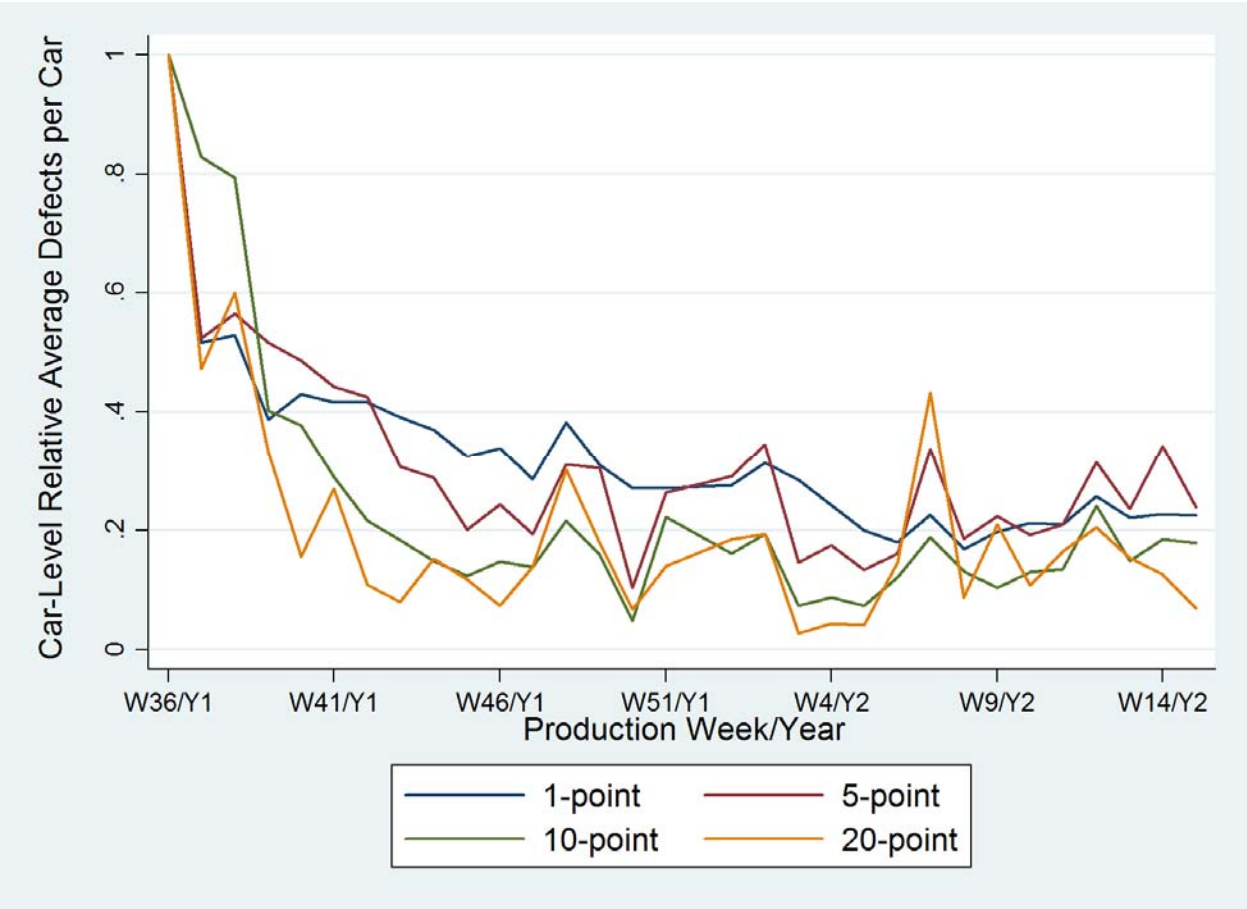


Figure A3. Weekly Production, by Shift

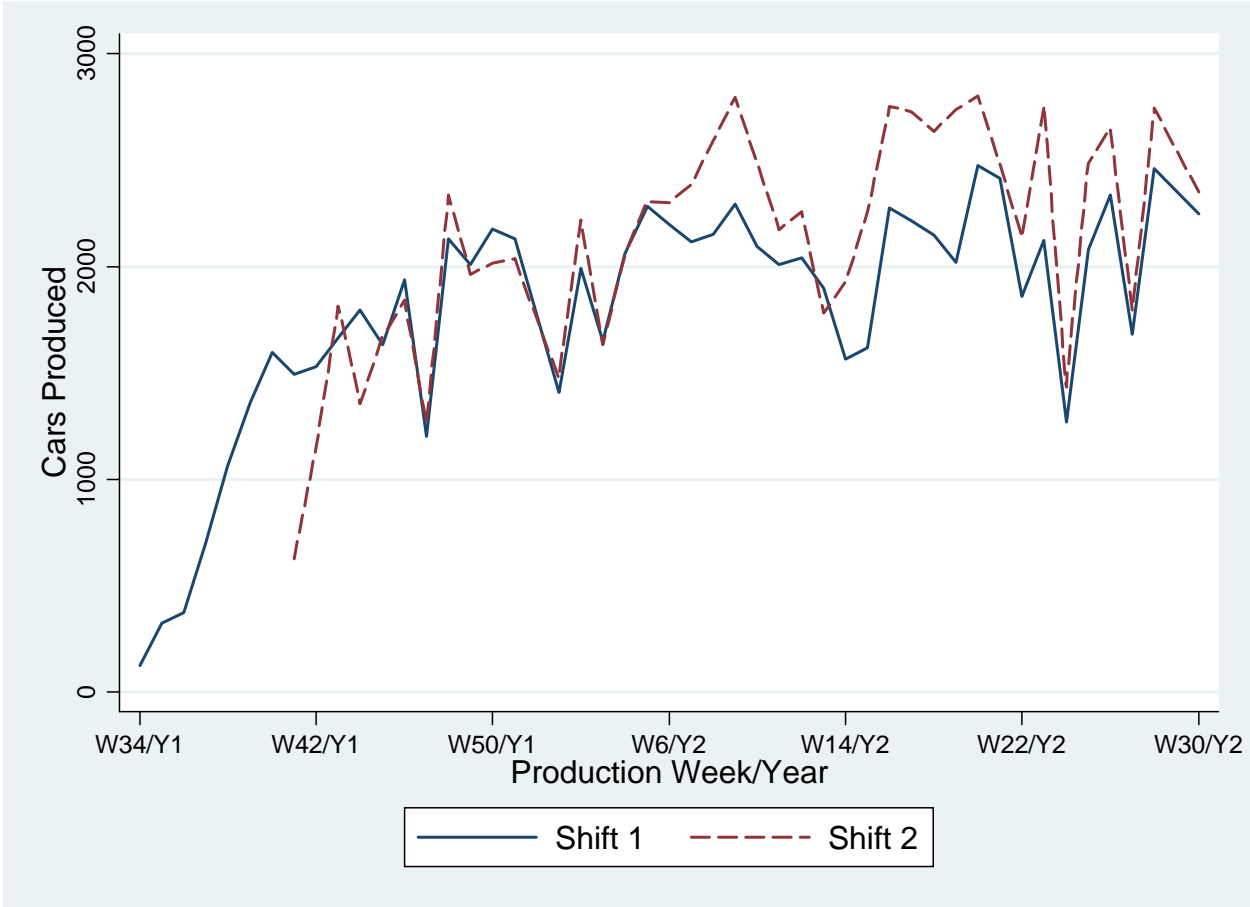


Figure A4. Weekly Production, by Model Variant

