The Variance Learning Curve

The Variance Learning Curve

Hessam Bavafa
Wisconsin School of Business, University of Wisconsin-Madison, hessam.bavafa@wisc.edu

Jónas Oddur Jónasson
MIT Sloan School of Management, joj@mit.edu

The expansive learning curve literature in Operations Management has established how various facets of prior experience improve *average* performance. In this paper, we explore how increased cumulative experience affects performance *variability*, or *consistency*. We use a two-stage estimation method of a heteroskedastic learning curve model to examine the relationship between experience and performance variability among paramedics at the London Ambulance Service. We find that for paramedics with lower experience, an increase in experience of 500 jobs reduces the variance of task completion time by 8.7%, in addition to improving average completion times by 2.7%. Similar to prior results on the *average* learning curve, we find a diminishing impact of additional experience on the *variance* learning curve. We provide an evidence base for how to model the learning benefits of cumulative experience on performance in service systems. Our findings imply that the benefits of learning are substantially underestimated if the consistency effect is ignored. Specifically, our estimates indicate that queue lengths (or wait times) might be overestimated by as much as 4% by ignoring the impact of the variance learning curve in service systems. Furthermore, our results suggest that previously established drivers of productivity should be revisited to examine how they affect consistency, in addition to average performance.

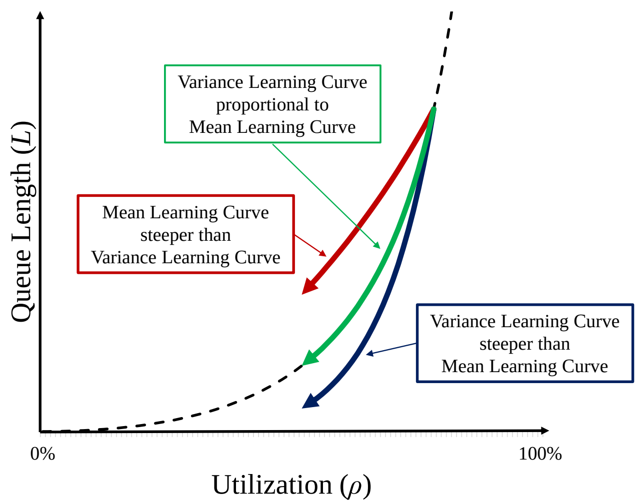
Key words: Learning Curves, Performance Consistency, Service Time Variability, Operational Performance, Service Operations, Ambulance Operations

History: June 25, 2020

1. Introduction

Learning curves–usually defined as productivity gains as a function of cumulative individual or organizational experience—have been a research topic in Operations Management (OM) and Organizational Behavior, for decades. Initial results established that unit production costs (or production time) decrease with the cumulative volume produced by an organizational unit (Argote and Epple 1990, Wright 1936). Subsequent work has further examined how various other aspects of experience can improve the productivity of individuals and teams (see discussion in §2). Most of this expansive literature has focused on identifying new experience-related explanatory variables that predict improvements in average performance.

Figure 1 **An illustration of the impact of the mean and variance learning curves on queue length in a single server service system.**



Notes: The dashed line plots queue length as a function of utilization (ρ) according to Kingman's G/G/1 formula (Kingman 1961); $E(L) \approx \left(\frac{\rho}{1-\rho}\right)\left(\frac{C_a^2+C_s^2}{2}\right)$, assuming fixed coefficients of variation for arrivals and service times. The solid lines illustrate how the performance impact of learning (through experience) in a service system depends on the relative magnitude of the mean and variance learning curves.

In this paper, our research question is whether there exists a learning curve for performance consistency. While the existing literature has clearly established that average performance improves with increased experience, we examine whether performance simultaneously becomes more consistent. Such a variance learning curve would have practical implications since service time variability is a key driver of performance in most service systems. Figure 1 illustrates how the variance learning curve could affect the overall performance impact of experience. The solid curves illustrate potential trajectories from the same starting point with the same improvement in average performance (captured here by a decrease in utilization on the x-axis). We observe from the figure that the impact of experience on service system performance (queue length or, equivalently, wait times) is difficult to determine in the absence of knowledge about whether experience affects consistency. In other words, if variability is reduced with experience (along with average service times) the benefits of the learning curve would be highly underestimated by assuming no change in consistency. To our knowledge, no prior work exists on this question.

Using a decade's worth of data from the London Ambulance Service (LAS), we examine how experience affects the mean and variance of paramedic performance, as they pick up patients and bring them to hospitals. We employ a two-stage estimation procedure. In the first stage, we fit a traditional learning curve. In the second stage, we use a transform of the residuals of the first stage regression as the dependent variable and estimate another learning curve for performance variance.[1] In both the first and second stages, we explore parametric and non-parametric specifications, all of which result in qualitatively the same results.

Unsurprisingly, we find strong evidence of the *average* learning curve among paramedics through our first stage analysis. We find that for paramedics with lower experience, an increase in experience of 500 jobs improves average performance (shortens the time it takes to deliver patients to hospitals) by 1.7%-2.8%, depending on the model specification. Subsequently, we observe diminishing returns to increased experience. These results are consistent with the existing literature. More importantly, our second stage regression reveals a significant impact of experience on performance consistency—the *variance* learning curve. According to our estimations, the variance of paramedic performance is initially reduced by 3.7%-8.7% with 500 additional jobs. We also find diminishing effects of increased experience on consistency.

Interestingly, in our setting the variance effect is initially stronger, resulting in a decreasing coefficient of variation (COV) for task completion times. However, for higher experience the impact of experience on average performance dominates the impact on performance consistency, resulting in a slightly increasing COV. Regardless, we find that assuming there is no impact of experience on performance variability would result in an overestimation of COV of 4%, which has implications for staffing, scheduling, and service system performance.

Our work makes academic and practical contributions. First, the operations management community has spent considerable effort identifying factors that contribute to productivity, usually captured in terms of average performance. As a result, there now exists a rich literature illustrating how various experience-related measures (e.g., diversity of experience or cumulative prior experience—at the individual, task, or team level) have a positive effect on performance. Our work is the first to examine the impact of experience on performance consistency. In itself, we believe this to be a fundamental insight about the operational learning curve. Furthermore, we believe that our results suggest that this still-growing literature can be revisited form the perspective of consistency. Second, despite the fact that the two-stage estimation procedure we use has been used in other contexts (e.g., finance and accounting), this useful technique has not been employed in OM empirical research, to our knowledge. Third, our results have practical implications since variability is a driver of system performance in most service systems (e.g., through queue lengths) and manufacturing systems (e.g., through quality control). Similarly, recent work has found that, through principles of behavioral economics, performance consistency can be a credible signal of

[1] Our methods are similar to Barth et al. (2008), Dechow and Dichev (2002), Sherman and Weiss (2017) and Lewis (2008) and trace back to the seminal work of Harvey (1976) on multiplicative heteroskedasticity.

competency (Falk and Zimmermann 2016). Therefore, our results suggest that the benefits of the learning curve may have been underestimated for practical purposes.

In the rest of the paper, we provide a brief summary of the learning curve literature in §2, describe our data, variables and methods in §3, and report the results in §4. We then summarize the various robustness checks in §5 and conclude with a discussion of our contributions in §6.

2. Literature Summary

Research on learning curves dates back over 100 years. Given the wealth of papers on the topic and numerous good reviews (Dar-El 2013, Argote 2012), we do not attempt to provide a comprehensive review but provide some background before focusing on more recent work in OM. Initial studies attempted to formalize the observation that *practice-makes-perfect*. In a very early study, Thorndike (1898) measured repeatedly how quickly animals were able to escape a maze, observing that trial-and-error strategies resulted in a learning curve. Soon after, researchers in applied psychology started using statistical methods to estimate learning curves (see e.g., Thurstone 1919). For a good overview of early theoretical and empirical findings on the learning curve, see Newell and Rosenbloom (1981).

The study of learning curves has continued to evolve recently in different streams of literature. Delaney et al. (1998) conclude that improvement of solution times is better explained by practice on a strategy than by practice on a task. Similarly, researchers have examined the role of retention (Ritter et al. 2013), relearning (Kim and Ritter 2015), and the length of the learning period (Papachristofi et al. 2016).

From an OM standpoint, researchers were initially interested in organizational learning curves. An oft-cited paper by Wright (1936) is credited with the first discussion of the phenomenon. In a study of air-frame production, he found that per-unit labor costs associated with a unit of production decrease with cumulative output. Much of the early literature focused on the production of airplanes or military equipment (see Alchian 1963, Yelle 1979, for an overview). The seminal paper by Argote and Epple (1990) summarizes organization learning curves and discusses reasons for observed differences in the rate of productivity improvements across various studies of manufacturing learning curves. In addition, Reagans et al. (2005) find that even when controlling for the individual experience and team experience of workers in a hospital, cumulative organizational experience has a distinct contribution to team performance.

More recently the OM community has examined the various aspects of cumulative experience of teams and individuals to further understand drivers behind performance. In terms of individual experience, Reagans et al. (2005) find that the cumulative experience of individuals improves completion times of total joint replacement procedures (controlling for organizational and team experience). Huckman et al. (2009) find that role experience is a better predictor of performance than total individual experience, while Huckman and Pisano (2006) and Clark et al. (2013) find that the experience accumulated at a particular firm or doing a particular task (respectively) is what affects subsequent performance. Similarly, KC and Staats (2012) distinguish between individual experience doing the focal task at hand or doing related task and find that experience doing the focal task has a greater effect on performance.

The literature on team experience (sometimes *team familiarity*) has similarly established that the cumulative number of times a team has jointly completed a task does impact performance (Reagans et al. 2005, Huckman et al. 2009, Akşin et al. 2020). In addition, Staats (2012) finds that the impact of team experience on performance can depend on whether the joint experience is acquired in the same geographic location and whether roles have been changed on the team.

Finally, a stream of literature has explored how diversity or variety in prior customers (Clark et al. 2013, Huckman and Staats 2011), prior tasks (Boh et al. 2007, Narayanan et al. 2009, 2014, KC and Terwiesch 2011), or prior partners (Akşin et al. 2020, Kim et al. 2018) can affect organizational, team, and individual performance.

The empirical literature, cited above, generally associates cumulative experience with output through regression analyses and hence establishes how experience affects performance, on average. Our work is distinct in that we focus explicitly on how performance variability changes with increased cumulative experience. To our knowledge, the work presented here is the first attempt to formalize that relationship and explore whether experience affects *consistency* of performance, in addition to these previously established effects on *average* performance. We provide empirical evidence for this important relationship and develop an evidence base for how to accurately model the impact of learning on performance in service systems.

3. Data, Variables, and Empirical Strategy

We now describe our empirical approach, starting with our data source (§3.1), then defining the main variables (§3.2), and finally discussing our estimation strategy (§3.3).

3.1. Setting and Organizational Context

Responsible for all emergency medical response in the greater London of the UK, the LAS provides care for almost 9 million people, making it one of the largest ambulance services in the world. The service employs over 5,000 people and is responsible for attending to over 2 million emergency calls per year. While the LAS responds to incidents in various ways, the vast majority of incidents is handled by ambulances.

Ambulance crews consist of two paramedics, who usually work 8, 10, or 12 hour shifts. Crews are formed at the base-level, in response to dynamic needs. Generally, new members are asked to cover the relief roster, which includes covering leaves and sick days for the more experienced members. After some point, certain members may work together frequently because they belong to the same shift at the same base.

The crews are dispatched to various incidents by a central dispatch center, based on their distance from the incident, the priority of the incident, and the crew's availability (i.e., whether they are occupied working on a previous incident). This dispatching process ensures that the ambulance-to-incident assignment is exogenous to crew experience. Only a subset of dispatches result in patients being brought to the ambulance, transported to the hospital, and handed over to the hospital staff in the Accident & Emergency (A&E) department. There can be various reasons for a dispatch coming to an end without a patient being transported to a hospital, including multiple ambulances arriving at the scene (i.e., the crew has no patient to transport), the dispatch being cancelled by the control room due to a higher priority dispatch, the patient being treated at the scene, or the patient having recovered by the time an ambulance arrives. We refer to dispatches which result in patient transportation as *activations*, a standard term used by LAS. In order to have a consistent set of observations, we focus the analysis on ambulance activations—i.e., dispatches in which the crew picked up a patient at the scene of an incident and brought him/her to an A&E in an ambulance.

Our final dataset includes 5,820,959 observations, by 10,137 paramedics, delivering patients to A&E departments in London during the years 2006-2015.[2] Table 1 provides summary statistics about our final data set.

3.2. Variables

3.2.1. Dependent Variable. In line with the previous literature on operational learning curves, we examine the impact of experience on task completion times. A full activation comprises 5 distinct components; driving to the scene, tending to the patient at the scene, driving the patient to the A&E, handing the patient over to the A&E, and preparing the ambulance for a subsequent dispatch (Akşin et al. 2020, Bavafa and Jónasson 2020). Unfortunately, the collection of time stamps for the last two components was inconsistent during our data collection period. We therefore focus our analysis on the first three components of an activation, all of which were consistently (automatically and electronically) collected throughout the period.

Time to hospital. For the main analysis, we define the crew's task as driving to the scene as quickly as possible, following dispatch; stabilizing and preparing the patient for ambulance transportation; and driving the patient to a receiving A&E. We refer to the time from dispatch to arrival at the A&E for activation a of crew c as $TimeToHospital_{a,c}$.

[2] For a full description of our inclusion criteria and data cleaning, see Appendix A1.

The three sub-components of $TimeToHospital_{a,c}$ have different task characteristics. Once a crew arrives at the scene of an incident, they must first locate the patient and ensure his/her safety, as well as their own. If possible, they stabilize the patient's health condition, conduct rudimentary diagnosis (sometimes including tests) and then decide whether to bring the patient to a hospital. In such cases, they must safely move the patient to the ambulance before starting the drive to the hospital. The aforementioned collection of actions requires a combination of clinical and non-clinical decisions (see examples of patient cases in Akşin et al. 2020). The clinical actions typically involve ensuring that the patient is able to breath normally, assessing circulation and disability, as well as deciding on a course of action for pain management and drug administration (all the while ensuring that their actions correspond to legal and ethical best practices). Non-clinical challenges can involve dealing with bystanders (often family of the patient), language barriers, or maneuvering in difficult settings (ranging from stressful scenes of car crashes to hard-to-reach locations from which the patient cannot be easily moved to the ambulance). As such, a crew's task at the scene is highly divergent, in that no single course of action applies to all settings (Shostack 1987).

In contrast, the task of driving to a scene or a hospital is an individual one, with a clear objective of reaching a destination as quickly as possible. While this objective is clear, it can be challenging to drive safely and fast, and some prior evidence (from very different experiments and settings) suggests learning effects can be observed in the context of driving (Da Silva et al. 2014, Larsson et al. 2014, van Leeuwen et al. 2015).

We denote the three sub-components of $TimeToHospital_{a,c}$ by $DriveToScene_{a,c}$, $TimeAtScene_{a,c}$, and $DriveToHospital_{a,c}$. As part of our robustness checks (§5), we examine the impact of experience on crew performance in the three sub-components and discuss the results in the context of the differences in task characteristics.

3.2.2. Main Independent Variables. In line with the prior literature (e.g., Reagans et al. 2005 and Huckman et al. 2009), our main independent variable is the average cumulative experience (number of dispatches) of the paramedics on crew c, prior to activation a, denoted by $AvgCrewExp_{a,c}$. As part of our robustness checks, we repeat the analysis replacing the average cumulative experience ($AvgCrewExp_{a,c}$) by the minimum team experience ($MinCrewExp_{a,c}$), the maximum team experience ($MaxCrewExp_{a,c}$), the average number of years working at LAS ($TimeSinceHired_{a,c}$), and the average activation (in contrast to dispatch) experience ($AvgCrewActExp_{a,c}$). Note that for teams including paramedics who joined LAS prior to 2006, our experience measures are censored because the data does not include their prior experiences. As a robustness check, we repeat the analysis for only those paramedics who joined the service after 2006, for whom we have a complete experience profile. The results are not affected by any of these robustness checks (see §5).

Table 1 Summary statistics.

	Mean	SD	N
Activations			5,820,959
Unique Teams			374,415
Unique Crew Members			10,137
Completion Time	49.37	16.47	5,820,959
To Scene	8.62	4.76	5,820,959
At Scene	27.42	13.14	5,820,959
To Hospital	13.33	7.13	5,820,959
Crew Experience	1,910	1,530	5,820,959

3.2.3. Control Variables

Long-term trends. Since our data spans a decade of ambulance activations in the city of London, we include flexible controls for any time-trends. We define $Time_a$ as the time passed (in days) since our first observation until activation a. In the analysis, we include a direct, squared, and cubed version of this term. This variable controls for long-term trends that might affect activation completion times and be correlated with experience.

Short-term seasonality. To account for short-term fluctuations in completion times, we include fixed effects for both the day of the week and the hour of the day in which the ambulance activation took place. This controls for average differences in completion times as a function of within-day and within-week seasonality.

Activation controls. For each ambulance activation we include fixed effect for whether the activation was a blue call, in which the paramedic crew alerts the receiving hospital that they are on the way to the hospital, carrying a patient who needs treatment as soon as they arrive at the hospital. This action is reserved for patients who must be fast-tracked through the patient handover process at the hospital. Similarly, we include fixed effects for each of the 98 primary illness codes, describing the main illness of the patient, which is recorded on a patient report form by the paramedic crew. Finally, we include fixed effects for the ambulance base the paramedic crew belongs to and the receiving hospital they are driving to.

Shift controls. For each activation a we include various controls for the specific crew and the specific shift the activation belongs to. We control for the average age of the crew members to account for any effects age might have on task completion times (Bavafa and Jónasson 2020). We also control for the workload the crew has experience as part of the shift (Akşin et al. 2020). Finally, we include controls for the timing of the activation within the shift, i.e., time since start of the shift (linear and squared). This controls for average differences in completion times due to fatigue and end-of-shift effects (Bavafa and Jónasson 2020, Deo and Jain 2019).

3.3. Empirical Strategy

Our main objective is to understand the impact of experience on performance consistency. To this end, we adopt a two-stage estimation procedure often used in accounting research to understand earnings quality (Barth et al. 2008, Dechow and Dichev 2002) and in economics to evaluate drivers of price dispersion (Sherman and Weiss 2017, Lewis 2008). Consider the following heteroskedastic regression model of performance;

$$TimeToHospital_{a,c} = \alpha + f(\boldsymbol{\beta}, AvgCrewExp_{a,c}) + \boldsymbol{Z^T}\boldsymbol{\phi_{a,c}} + u_{a,c}, \tag{1}$$

$$ln(Var(u_{a,c})) = \gamma + g(\boldsymbol{\pi}, AvgCrewExp_{a,c}) + \boldsymbol{\zeta^T}\boldsymbol{\phi_{a,c}} + \epsilon_{a,c}. \tag{2}$$

The role of (1) is in essence to de-mean the operational performance variable with respect to observable variables so that we can examine the variation in performance.[3] In particular, we de-mean with respect to the impact of cumulative experience on average performance according to some function $f(\cdot)$ with parameters $\boldsymbol{\beta}$. Subsequently, the residuals from (1) can be interpreted as the deviation from the predicted performance (given covariates), making them the basis for an empirical measure for performance consistency. The role of (2) is to evaluate whether performance variability changes as a function of cumulative experience.

In our analysis, we recover the parameters of (2) with a two-step procedure. We first evaluate (1) using ordinary least squares (OLS), calculate the residuals for each observation and denote them by $\hat{u}_{it}$. We then use the transformed residuals as a dependent variable in a second regression:

$$ln(\hat{u}^2_{a,c}) = \gamma + g(\boldsymbol{\pi}, AvgCrewExp_{a,c}) + \boldsymbol{\zeta^T}\boldsymbol{\phi_{a,c}} + \epsilon_{a,c}. \tag{3}$$

Squaring the residuals in (3) gives an estimate of the variance and ensures that all values are non-negative, which is appropriate for measuring dispersion. We adopt the multiplicative heteroskedasticity structure of Harvey (1976) to describe the impact of cumulative experience on performance variance.[4]

For our main analysis, we explore two main sets of models for $f(\cdot)$ and $g(\cdot)$ (with additional functional forms explored in §5).

[3] Alternatively, we could de-mean the observations using (1) and then simply calculate the standard deviation of the residuals for a reasonable time period, e.g., a shift, a month, or every 100 activations. However, the issue with such a specification is that it does not allow us to control for drivers of performance variability at the activation level. E.g. if certain activation types inherently have a higher variance, that would not be controlled for with an aggregation approach.

[4] We note that our choice of the multiplicative heteroskedasticity structure (as opposed to additive, power function, exponential, or other models of heteroskedasticity, e.g., Gaur et al. 2007 and Greene 2003) is consistent with prior literature (e.g., Western and Bloome 2009, Lewis 2008, Sherman and Weiss 2017), but the results are robust to not transforming the residuals and replacing the dependent variable of (3) with $\hat{u}^2_{a,c}$ (see Table A10 in §A3).

A parsimonious non-linear learning-curve. In our first main analysis, we estimate a parsimonious model of the average and variance learning curves. Most of the prior literature assumes some diminishing effect of experience.[5] To this end, we model the learning curve using a linear and squared term for experience, setting $f(\boldsymbol{\beta}, AvgCrewExp_{a,c}) = \beta_1 AvgCrewExp_{a,c} + \beta_2\ AvgCrewExp^2_{a,c}$ and $g(\boldsymbol{\pi}, AvgCrewExp_{a,c}) = \pi_1\ AvgCrewExp_{a,c} + \pi_2\ AvgCrewExp^2_{a,c}$.

A non-parametric learning-curve. In our second main analysis, we assume a non-parametric structure for $f(\cdot)$ and $g(\cdot)$. We define indicator variables for different levels of experience, rounding each observation by 500. This results in 15 indicator variables, denoted by $\mathbf{1}_{\{AvgCrewExp_{a,c} \sim n\}}$ for $n \in \{0,\ 500,\ \ldots\ ,\ 6{,}500\}$ and $\mathbf{1}_{\{AvgCrewExp_{a,c} \sim 7{,}000+\}}$ for crews with an average prior experience of more than 7,000 activations (99% of our observations have $AvgCrewExp_{a,c} < 7{,}000$). This second specification allows us to understand the functional form (e.g., in case of diminishing returns) of the impact of experience on the mean and the variance of performance. In addition, we fit a cubic spline with five knots at equally spaced percentiles, as recommended by Harrell (2001).

Finally, in all specifications of (1) and (3) we include all the controls described in §3.2.3 as part of $\phi_{a,c}$ and cluster standard errors at the shift level.

4. Results

In this section, we report the results of the two main analyses in §4.1 and §4.2. For a model-free depiction of the average and variance learning curves, see §A2.1. We discuss our findings, both in terms of the *average* learning curve and the *variance* learning curve (while our results on the average learning curve largely confirm previous findings, it is useful to include them as they contribute to changes in the COV with experience). We then describe the managerial implications of these results in §4.3.

4.1. A Parsimonious Non-Linear Learning-Curve

We start by discussing the results of the parsimonious model for the average and variance learning curves for $AvgCrewExp$ (scaled by 1/500 for ease of interpretation). For completeness, we include models with only a linear term in columns (1) and (2) of Table 2. However, we focus the interpretations on columns (3) and (4), which include a linear and squared term.

First, in column (3) we find evidence for the average learning curve—in line with the existing literature. We observe a steep initial learning curve ($\beta_1 = -0.872$, $p < 0.001$) with diminishing returns as paramedics gain more experience ($\beta_2 = 0.029$, $p < 0.001$). This indicates that the performance impact of each additional 500 activations is initially around 1.7% (almost one minute compared

[5] An alternative model that allows for diminishing returns is to set $f(\boldsymbol{\beta}, AvgCrewExp_{a,c}) = \beta\ ln(AvgCrewExp_{a,c})$ and $g(\boldsymbol{\pi}, AvgCrewExp_{a,c}) = \pi\ ln(AvgCrewExp_{a,c})$. We do this as a robustness check and obtain similar results (see discussion in §5).

Table 2 The impact of experience on mean and variance of performance: parametric models.

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$
$AvgCrewExp$	-0.511***	-0.018***	-0.872***	-0.040***
	(0.003)	(0.000)	(0.008)	(0.001)
$AvgCrewExp^2$			0.029***	0.002***
			(0.001)	(0.000)
Shift Controls	✓	✓	✓	✓
Activation Controls	✓	✓	✓	✓
Seasonality Controls	✓	✓	✓	✓
R-squared	0.207	0.016	0.207	0.016
Observations	5,820,866	5,820,866	5,820,866	5,820,866

Notes: For ease of interpretation $AvgCrewExp$ is scaled by 1/500. Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). All coefficient estimates are included in Table A2 in §A3. Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

to the average $TimeToHospital$ of 49 minutes) but decreases to about 0.2% once the crew has acquired an average experience of over 7,000 activations.[6] With each paramedic conducting 645 activations in a year in our data, they can be expected to bring patients to the hospital 2% faster after a year on the job.

Second and more importantly, our results in column (4) of Table 2 provide evidence for the previously unknown variance learning curve. We observe that the variance of performance (as measured by $\hat{u}^2_{a,c}$) decreases by 3.7% ($e^{\pi_1+\pi_2} - 1 = e^{-0.040+0.002} - 1$) with the first additional 500 activations.[7] Subsequently, a further increase in experience results in decreased reduction of variance. This implies that during the first year on the job (645 activations on average) the variance in a paramedic's performance is decreased by almost 5%.

These results indicate that performance variance is decreasing faster than average performance for at least the first 5,000 activations—resulting in decreased COV. However, in this analysis we have assumed a quadratic function for the impact of experience on both the average and variance of performance. With a negative β_1 (or π_1) and positive β_2 (or π_2) the functions $f(\cdot)$ and $g(\cdot)$

[6] As with much of the prior learning literature our data is censored in the sense that the sample does not include possible experience acquired prior to 2006. Therefore, our results should be interpreted as the average learning effect across paramedics with varying prior experience. We conduct a robustness check focusing only on paramedics who joined LAS in 2006 or later (see §5). We observe even stronger learning effects in that cohort, indicating that the learning curve estimates in Table 2 are conservative.

[7] We note that we take the ln transformation of the squared residuals (i.e., use $ln(\hat{u}^2_{a,c})$ as the dependent variable in the second stage regression) to follow the existing literature (Sherman and Weiss 2017, Lewis 2008). As a robustness check, we obtain very similar results without this transformation, using $\hat{u}^2_{a,c}$ as the dependent variable in the second stage regression (§5).

Table 3 The impact of experience on mean and variance of performance: non-parametric model.

	(1) Mean: $TimeToHospital$		(2) Variance: $ln(\hat{u}^2)$	
$\mathbf{1}_{\{AvgCrewExp\,\sim\,0\}}$	(Reference)		(Reference)	
$\mathbf{1}_{\{AvgCrewExp\,\sim\,500\}}$	-1.335***	(0.038)	-0.091***	(0.005)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,1000\}}$	-2.453***	(0.040)	-0.142***	(0.005)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,1500\}}$	-3.169***	(0.042)	-0.169***	(0.005)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,2000\}}$	-3.836***	(0.043)	-0.194***	(0.005)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,2500\}}$	-4.277***	(0.045)	-0.205***	(0.006)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,3000\}}$	-4.583***	(0.048)	-0.215***	(0.006)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,3500\}}$	-4.923***	(0.051)	-0.230***	(0.006)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,4000\}}$	-5.271***	(0.054)	-0.246***	(0.007)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,4500\}}$	-5.676***	(0.061)	-0.234***	(0.007)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,5000\}}$	-5.932***	(0.069)	-0.257***	(0.009)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,5500\}}$	-6.683***	(0.079)	-0.285***	(0.010)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,6000\}}$	-6.976***	(0.090)	-0.284***	(0.011)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,6500\}}$	-6.676***	(0.107)	-0.241***	(0.013)
$\mathbf{1}_{\{AvgCrewExp\,\sim\,7000+\}}$	-7.541***	(0.085)	-0.277***	(0.010)
Shift Controls	✓		✓	
Activation Controls	✓		✓	
Seasonality Controls	✓		✓	
R-squared	0.208		0.017	
Observations	5,820,866		5,820,866	

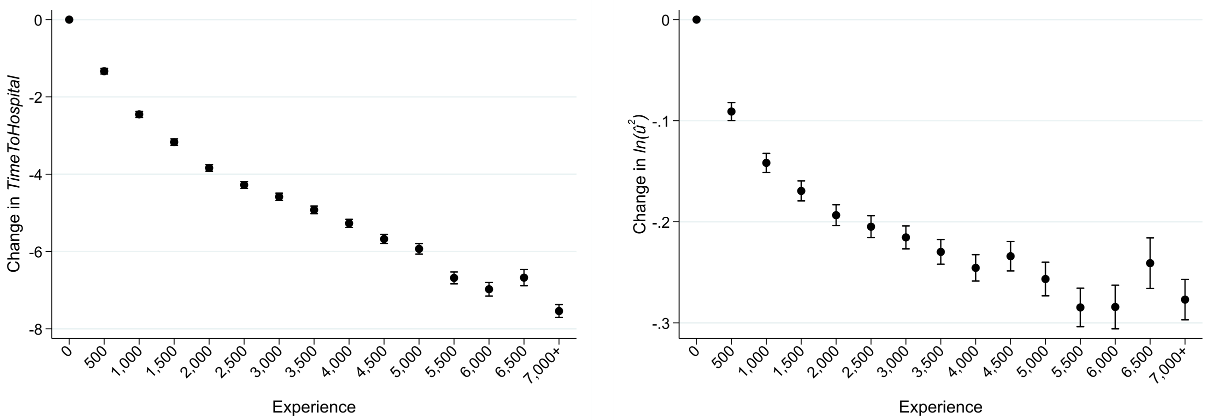
Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level.
* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

will theoretically have a vertex point beyond which increased experience would be associated with decreased performance. According to the coefficient estimates in Table 2, these vertex points are at $AvgCrewExp = 7{,}517$ and $AvgCrewExp = 5{,}000$ for $f(\cdot)$ and $g(\cdot)$, respectively. Since 99% of the observations in our data take values of $AvgCrewExp$ below 7,000, the vertex point for the average learning curve does not concern us. In contrast, the vertex point for the variance learning curve is well within the range of our data. However, this observation could represent a limitation of the functional form we have chosen for the parsimonious model. This motivates our second main analysis, in which we estimate non-parametric models for both the average and variance learning curves and examine the impact of experience on the COV of performance.

4.2. Non-parametric models

To relax the strict assumption of the functional form of the learning-curves in the parsimonious model discussed above, we repeat our analysis allowing for an arbitrary non-parametric effect of experience on the mean and variance of operational performance. We capture the learning-curve effect on the mean and variance of performance using the sequence of indicator variables

Figure 2 Coefficient estimates for the average and variance learning curves (from Table 3).



(a) Mean Effect (b) Variance Effect

($\mathbf{1}_{\{AvgCrewExp\,\sim\,n\}}$) defined in §3.3. We report the raw coefficient estimates in Table 3 and plot them, along with confidence intervals, in Figure 2.

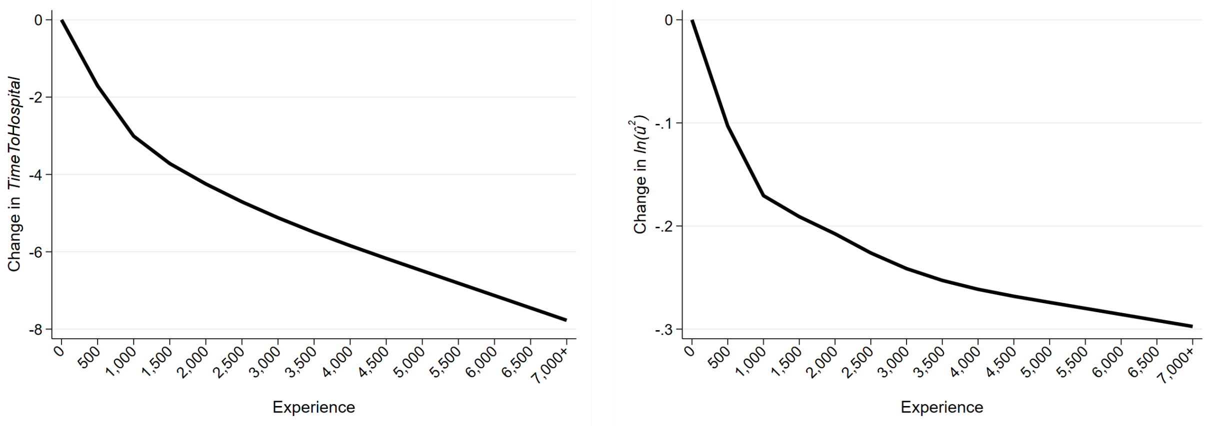
First, we examine the non-parametric estimates for the average learning curve. We observe from column 1 of Table 3 and Figure 2(a) that through increasing the average individual experience of the crew by 7,000 activations, $TimeToHospital$ is decreased by almost 7.5 minutes, controlling for shift, activation, and seasonality characteristics. The marginal improvement is largest in magnitude for the first 500 additional activations—shortening the time it takes to delivery patients to the hospital by 1.3 minutes (2.7% of the average of 49 minutes)—but is persistent throughout all levels of experience.

Second, we examine the corresponding estimates for the variance learning curve. Column 2 of Table 3 and Figure 2(b) illustrate the shape of this effect with the strongest impact of the first 500 additional activations—resulting in an 8.7% ($e^{-0.091}-1$) decrease in variance—and gradually decreasing marginal impact of increased experience. The variance of the crews who attain the highest level of experience in our data ($AvgCrewExp > 6,000$) has decreased by 21%-25%, as compared to when we first observe them in the data.

Third, since the main independent variable of interest, $AvgCrewExp_{a,c}$, has a long tail (the range is from 0 to 12,000), the coefficient estimates for the indicator variables $\mathbf{1}_{\{AvgCrewExp\,\sim\,n\}}$ become noisy for vales of n higher than 4,000. To generate a smooth prediction of the average and variance learning curves, we estimate a cubic spline with five knots at equally spaced percentiles (Harrell 2001). The results of this estimation are included in Figure 3.

Finally, we note that the impact of the performance improvements (average and variance) are likely to be meaningful, both for the operational performance of the system (through lowering

Figure 3 **Best fitting curves from spline model.**



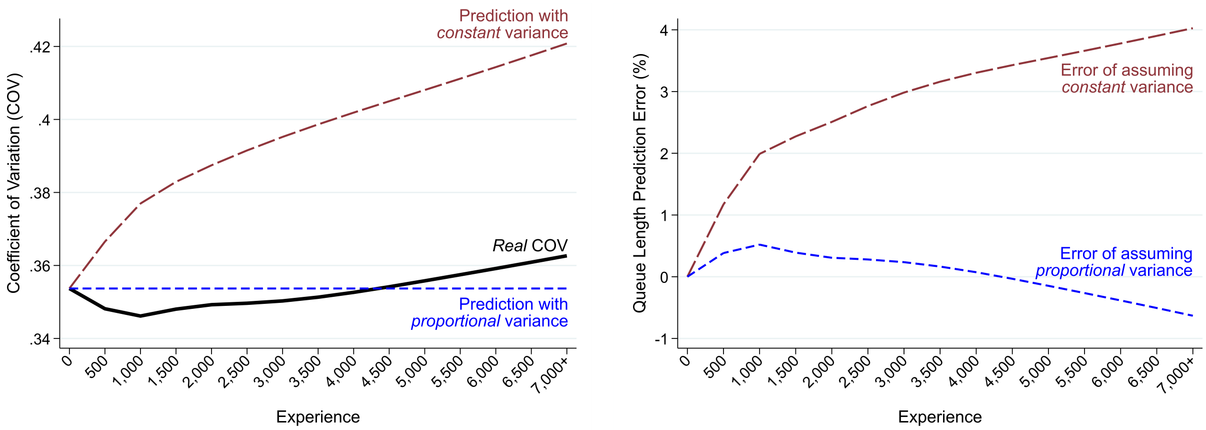
(a) Mean Effect (b) Variance Effect

utilization) and for patient outcomes (through minimizing out-of-hospital delays). Reduced out-of-hospital times have been associated with patient outcomes for various conditions. For cardiac arrests, survival rates have been shown to decline by over 5% for every minute of delay beyond 4 minutes, since onset (Gold et al. 2010). Unexpected survivors of trauma (i.e., patients with unexpectedly good outcomes, given their Trauma Injury and Injury Severity Score) tend to have shorter response times, on scene times, and transport times (Feero et al. 1995). Similarly, out-of-hospital times are an important predictor of outcomes for stroke patients, for whom minimal delays are an eligibility prerequisite of various treatments (Rossnagel et al. 2004, Evenson et al. 2001). Finally, while we do not observe specific severity scores in our data, previous work with LAS data has found that approximately 20% of activations are classified as *Category A: Immediately life-threatening*—indicating that swift and consistent transportation to a hospital will benefit a large fraction of the patients in our data (Akşin et al. 2020).

4.3. Managerial Implications

Through our empirical analysis, we find robust evidence of a variance learning curve. These results entail two main managerial implications. First, the straightforward managerial insight is for operations managers to consider the benefits of cumulative experience from the perspective of consistency as well as average performance. In settings where variability across multiple jobs is important—e.g., if a client or customer infers skill from consistency (more so than average performance) (Falk and Zimmermann 2016)—more experienced workers are more likely to achieve that objective. Second, our results have implications for capacity management and scheduling. In service systems, the benefits of the learning curve are manifested through the same service level being achieved with fewer resources. Alternatively, with the same workforce, average wait times or queue lengths can

Figure 4 **The implications of the variance learning curve on COV and service levels.**



(a) COV Predictions

(b) Queue Length Predictions

be reduced. These effects have been studied in the prior operations management literature (Gans et al. 2010, Batt and Gallino 2019, Arlotto et al. 2014, Robbins 2015). Due to the transient nature of such systems (in which parameters change with experience) the analysis of various scheduling strategies is usually conducted through simulation, which requires assumptions about the service time distribution of workers as they learn. While this literature has carefully estimated parameters for the rate of the average learning curve, it has made ad-hoc assumptions about the structure of the variance learning curve. However, since performance in service systems is often measured by queue lengths or delays, a key contributor to performance is service time variability. Our work provides a methodology and an initial evidence base for how to model variance in the context of learning.

Figure 4 summarizes the implication of our findings on the performance of service systems, using our cubic spline estimates (Figure 3) from the previous section. Panel (a) compares the empirical COV (solid line) with two types of erroneous estimates of the COV. The red line (long-dash, color online) plots the hypothetical evolution of COV under the assumption of a constant variance of performance as workers gain experience. We observe that such an assumption would result in a significant overestimation of COV (up to 16% in our sample) as workers gain experience. The blue line (short dash, color online) plots the hypothetical evolution of COV under the assumption of variance decreasing proportionally with average performance. The plot demonstrates how this assumption first results in an overestimate (up to 2.2%) and then an underestimate (up to 2.6%) of COV as workers gain experience.

Panel (b) converts these COV estimates into queue length predictions using Kingman's concise approximation for queue lengths in G/G/1 queues (Kingman 1961) (the same formula we use as the basis for Figure 1):

$$E(L) \approx \left(\frac{\rho}{1-\rho}\right)\left(\frac{C_a^2 + C_s^2}{2}\right),$$

where ρ, C_s, and C_a (set to 1) denote the system utilization, service time COV, and arrival rate COV, respectively.

As is evident from the approximation equation, the error in the queue length prediction is proportional to the square of the error in the COV estimation. Specifically, the mistake of assuming no impact of learning on performance variance can result in up to 4% error in the queue length (or service time), even when the average learning curve is taken into account. Similarly, the error of assuming proportional impact of learning on variance and average performance can result in an initial overestimation of queue length (up to 0.5%) and subsequently an underestimation of queue length (up to 0.6%). We note that an error of 0.6% can be economically significant or not, depending on the setting. Even for cases in which such error magnitudes are acceptable, our analysis provides the first empirical evidence for the validity of the assumption of constant COV. Furthermore, our results seem to indicate that such errors might increase with experience beyond what we observe in the data.

5. Summary of Extensions and Robustness Checks

In support of the main results, we conduct a number of extensions and robustness checks. We provide a summary of these analyses below and relegate the regression tables to Appendix A3.

5.1. More Granular Task Definition

In the main analysis, the outcome of interest is $TimeToHospital_{a,c}$. As we discuss in §3.2, this variable comprises the completion times of three distinct tasks: driving to the scene, patient pick-up at the scene, and driving to the hospital. To shed a light on which task component is the main contributor of the learning benefits observed in the main analysis—as well as to provide a preliminary understanding of whether the variance learning curve is observed for tasks with different characteristics (see §3.1)—we re-estimate the parsimonious model for each of the three components. The results of these robustness checks are in Table A3. We find that the qualitative results from the main analysis hold for each sub-component, but make three additional observations.

First, in terms of effect sizes, we observe that for paramedics with lower experience, an experience increase of 500 activations reduces the average task completion times by 2%, 2%, and 1% for the $DriveToScene_{a,c}$, $TimeAtScene_{a,c}$, and $DriveToHospital_{a,c}$, respectively. For the same outcomes, performance variance is reduced by 7%, 3%, and 1%, respectively. Second, we find it encouraging to see evidence of the variance learning curve for types of tasks that differ on at least two dimensions. Tending to patients at the scene and preparing them for transport is a collaborative, divergent task (Shostack 1987, Akşin et al. 2020) and could be considered *knowledge work* (in the sense described by Staats et al. (2011)—i.e., knowledge work having properties such as being dynamic, relying on invisible processes, and relying on a combination of exploration and exploitation). In contrast, the task of driving to a scene or a hospital is an individual one, with a much clearer objective, even if it can be challenging to do safely and fast. Third, while it is difficult to meaningfully contrast the effect sizes for different types of tasks such as driving and tending to patients at the scene—due to the differences described above—it is interesting to compare the magnitude of the effects for the two driving outcomes. The results indicate a larger performance improvement (in the average and variance of performance) for the drive to the scene than the drive to the hospital. A possible reason for this is that prior to arriving at the scene, there is considerable uncertainty about the condition of patients, so ambulances usually utilize emergency lights for this driving component. This is not the case for the driving to the hospital, since the lights are only used to transport patients who are in a critical condition. As a result, the second driving component may be more noisy, in nature, compared to the first one. We note that the results for average effects has qualitative similarities to prior work on the impact of critical incidents on paramedic performance, which found that critical incidents have a much higher negative effect on driving times to the scene than to the hospital (Bavafa and Jónasson 2020).

5.2. Maximum or Minimum Crew Experience

Our unit of analysis is an ambulance activation conducted by a crew of two paramedics. For the main analysis we used the average experience of the crew members ($AvgCrewExp_{a,c}$) as the main independent variable of interest. We repeat the analysis replacing this variable with $MinCrewExp_{a,c}$ and $MaxCrewExp_{a,c}$, which denote the minimum and the maximum experience of the two crew members, respectively. We observe that the results are not affected by changing the definition of the experience variable. Not only does crews' average performance improve with increase in the minimum or maximum experience on the team, but the variability in performance is also reduced. (See Table A4.)

5.3. Narrower Experience Definition

Our main independent variable of interest, $AvgCrewExp$, is defined as the average number of prior dispatches undertaken by the crew members. As a robustness check we repeat the analysis replacing this variable with $AvgCrewActExp$, defined as the average number of prior activations (i.e., dispatches resulting in a patient being transported to a hospital). We obtain qualitatively the same results, with larger coefficient magnitudes and lower p-values. (See Table A5.)

5.4. Alternative Experience Definition

An alternative measure of the main independent variable is the number of years since the crew members started working at LAS, denoted by $TimeSinceHired$. This variable is less granular as it does not measure the number of times a crew member has conducted an activation. However, it has the benefit of accurately describing the years on the job, even before our observation period (2006-2015). The results we obtain using this alternative measure are consistent with the main results (See Table A6).

5.5. Log Transformation of Main Independent Variable

Some of the prior learning-curve literature has used a log transformation of the main independent variable, to allow for diminishing returns of increased experience. In line with this, we repeat the analysis with $f(\boldsymbol{\beta}, AvgCrewExp_{a,c}) = \beta ln(AvgCrewExp_{a,c})$ and $g(\boldsymbol{\pi}, AvgCrewExp_{a,c}) = \pi ln(AvgCrewExp_{a,c})$. As before the coefficients in both the mean (β) and variance (π) regressions are negative and significant. (See Table A7.)

5.6. Excluding Paramedics with Prior Experience

As in many prior studies of experience effects, our data is censored since the sample does not include paramedics experiences prior to 2006. Therefore, the main results should be interpreted as the average learning effect for paramedics with varying prior experiences. We are not concerned about the aforementioned censoring of the data since, if anything, it would make our results conservative (the received wisdom on learning curves is that they are most steep early on). However, to address any concerns about the censoring, we repeat the analysis focusing only on ambulance crews involving paramedics who joined LAS in 2006 or later. For this subset of paramedics our data includes every ambulance activation they have conducted at LAS, so their learning curves are not censored. The results of this analysis show that, as expected, the linear coefficients of the parsimonious model are larger in magnitude (reflecting a steep initial learning curve) than in the main analysis. (See Table A8.)

5.7. Excluding Teams which do not Acquire High Experience

Similar to most observational studies of the learning curve, our sample includes workers who ultimately attained high on-the-job experience as well as workers who did not. As a result, some of the effect we observe in our main results could be due to a selection effect in which well performing crews achieve high experience whereas poor performing crews do not. We make two observations about this potential concern. First, there is no evidence of such selection effects. We have conducted a robustness check in which we eliminate all teams who do not ultimately acquire an average experience of at least 4,000 activations. Furthermore, we restrict this robustness analysis to teams for whom we observe performance at any experience level (see details in Appendix A3.7), to ensure that the observations used to estimate the learning effects at different levels of experience belong to a consistent set of teams. This exclusion criteria results in 93 crews who conducted almost 300,000 activations. We repeat the main analysis and find that the learning effects (for both mean and variance) are even stronger for this subset of crews (see Table A9). Second, even if there was a selection effect in which poorly performing paramedics quit the job, this would not affect the managerial insights since high experience crews would be the ones with faster and more consistent task completion times.

5.8. Alternative Control Variables

We conduct three sets of analyses to evaluate the robustness of the main results to alternative or expanded sets of control variables and include those results in Table A11. First, the main analysis uses a linear, squared, and cubed version of $Time_a$ to control for long-term average trends over the decade we study. As a robustness check, we repeat the analysis replacing this polynomial control for trends by month-year fixed effects (columns (1) and (2) of Table A11). This does not affect any of the results. Second, the main outcome, $TimeToHospital$, likely depends on the location of the scene and the hospital as well as driving conditions, which may vary across areas of London. To control for such potential effects, we conduct an analysis in which we control for each pair of paramedic base and receiving hospital (recall that in the main analysis, we control for base and hospital locations using separate indicator variables). This controls for the average distance a crew from a given base (a good proxy for the area in which an ambulance crew usually picks up patients) needs to travel to get to any given hospital. The results are presented in columns (3) and (4) of Table A11. The estimates are nearly identical to the ones in the main analysis in Table 2. Third, prior work on operational productivity has established how team familiarity—measured as the cumulative joint experience of team members—has a positive effect on performance, in addition to average individual experience (Reagans et al. 2005, Huckman et al. 2009). Columns (5) and (6) of Table A11 include the results of repeating the main analysis with $TeamFamiliarity_{a,c}$ included as a control. This additional control does not affect the results.

5.9. Experience Effects for Blue Calls

We limit the main analysis to activations (defined as dispatches resulting in a patient being transported to a hospital, see §3.1) for a consistent outcome definition. We note that while the need to bring a patient to a hospital is largely determined by the patient's condition, it is ultimately a crew decision whether to attend to the patient at the scene or bring him/her to an A&E. We conduct two additional analyses to examine whether there is any evidence that changes in crew decision making might be the source of the improvements. First, we provide summary statistics (mean and confidence interval) on the fraction of dispatches which turn into activations in Figure A2(a). We observe that regardless of experience, the proportion is consistently between 0.75 and 0.78. Second, we conduct a robustness check for the impact of experience on performance for blue calls (these are activations in which the patient is seriously ill and must get to a hospital without delay, see §3.1), since for these activations there is less crew discretion as to whether the patient needs to go to a hospital. Specifically, we repeat the main analysis with an added interaction between blue calls and the experience variables. As before, we observe strong learning effects for both the mean and variance of performance, for blue call activations as well as regular activations (see details in §A3.10 and results in Table A12).

6. Conclusion

Few operational phenomena are as well studied as the learning curve. Not only has it become a well understood concept in the popular vernacular, but it has also inspired a still-growing literature examining experience-related drivers of productivity. Our main objective in this paper is to demonstrate the existence of a *variance* learning-curve—that in addition to improving average performance, increased experience improves the consistency (or reduces the variability) in workers' performance. Using data on paramedic performance provided by the LAS and a two-stage regression approach, we find robust evidence of such an effect. We estimate various models of the mean and variance learning curves and find evidence that experience drives consistency in performance.

These results have academic and managerial implications. From the academic perspective, we establish a previously unknown effect of experience on productivity. We believe this to be a fundamental insight about the operational learning curve and that much of the operational productivity literature can be revisited from the perspective of consistency. Furthermore, our results motivate future analysis of how the operational environment of healthcare delivery affects consistency in clinical outcomes. In particular, it is of interest to understand how experience affects variability in patient outcomes, in addition to the operational outcomes we focus on. More broadly, other aspects of team formation, the operational scheduling of providers, and operational guidelines (Kent and Siemsen 2018) are likely to affect consistency in both operational and clinical outcomes.

From a managerial perspective, our results indicate that the benefits of the learning curve on service system performance would be underestimated by up to 4% by ignoring the *variance* learning curve.

Acknowledgments

We thank the London Ambulance Service for providing data, information about the study setting, and multiple opportunities to observe ambulance crews, dispatchers, and hospital staff in action. In particular, we thank Leanne Smith, Michael Damiani, and Patrick Brooks. We are grateful for excellent research assistance from Gabriel Chua.

References

Akşin, Z., S. Deo, J.O. Jónasson, K. Ramdas. 2020. Learning from many: Partner exposure and team familiarity in fluid teams. *Management Science* (forthcoming).

Alchian, A. 1963. Reliability of progress curves in airframe production. *Econometrica: Journal of the Econometric Society* 679–693.

Argote, L. 2012. *Organizational learning: Creating, retaining and transferring knowledge*. Springer Science & Business Media.

Argote, L., D. Epple. 1990. Learning curves in manufacturing. *Science* **247**(4945) 920–924.

Arlotto, A., S.E. Chick, N. Gans. 2014. Optimal hiring and retention policies for heterogeneous workers who learn. *Management Science* **60**(1) 110–129.

Barth, M.E., W.R. Landsman, M.H. Lang. 2008. International accounting standards and accounting quality. *Journal of Accounting Research* **46**(3) 467–498.

Batt, R.J., S. Gallino. 2019. Finding a needle in a haystack: The effects of searching and learning on pick-worker performance. *Management Science* **65**(6) 2624–2645.

Bavafa, H., J.O. Jónasson. 2020. Recovering from critical incidents: Evidence from paramedic performance. *Manufacturing and Service Operations Management* (forthcoming).

Boh, W.F., S.A. Slaughter, J.A. Espinosa. 2007. Learning from experience in software development: A multilevel analysis. *Management Science* **53**(8) 1315–1331.

Clark, J.R., R.S. Huckman, B.R. Staats. 2013. Learning from customers: Individual and organizational effects in outsourced radiological services. *Organization Science* **24**(5) 1539–1557.

Da Silva, F.P., J.A. Santos, A. Meireles. 2014. Road accident: driver behaviour, learning and driving task. *Procedia-Social and Behavioral Sciences* **162** 300–309.

Dar-El, E.M. 2013. *Human learning: From learning curves to learning organizations*, vol. 29. Springer Science & Business Media.

Dechow, P.M., I.D. Dichev. 2002. The quality of accruals and earnings: The role of accrual estimation errors. *The Accounting Review* **77**(s-1) 35–59.

Delaney, P.F., L.M. Reder, J.J. Staszewski, F.E. Ritter. 1998. The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science* **9**(1) 1–7.

Deo, S., A. Jain. 2019. Slow first, fast later: Temporal speed-up in service episodes of finite duration. *Production and Operations Management* **28**(5) 1061–1081.

Evenson, K.R., W.D. Rosamond, D.L. Morris. 2001. Prehospital and in-hospital delays in acute stroke care. *Neuroepidemiology* **20**(2) 65–76.

Falk, A., F. Zimmermann. 2016. Consistency as a signal of skills. *Management Science* **63**(7) 2197–2210.

Feero, S., J.R. Hedges, E. Simmons, L. Irwin. 1995. Does out-of-hospital EMS time affect trauma survival? *The American Journal of Emergency Medicine* **13**(2) 133–135.

Gans, N., N. Liu, A. Mandelbaum, H. Shen, H. Ye. 2010. Service times in call centers: Agent heterogeneity and learning with some operational consequences. *Borrowing strength: theory powering applications–A Festschrift for Lawrence D. Brown*. Institute of Mathematical Statistics, 99–123.

Gaur, V., S. Kesavan, A. Raman, M.L. Fisher. 2007. Estimating demand uncertainty using judgmental forecasts. *Manufacturing & Service Operations Management* **9**(4) 480–491.

Gold, L.S., C.E. Fahrenbruch, T.D. Rea, M.S. Eisenberg. 2010. The relationship between time to arrival of emergency medical services (ems) and survival from out-of-hospital ventricular fibrillation cardiac arrest. *Resuscitation* **81**(5) 622–625.

Greene, W.H. 2003. Econometric analysis 5th edition. *International edition, New Jersey: Prentice Hall* .

Harrell, F.E. 2001. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer.

Harvey, A.C. 1976. Estimating regression models with multiplicative heteroscedasticity. *Econometrica* **44**(3) 461–465.

Huckman, R.S., G.P. Pisano. 2006. The firm specificity of individual performance: Evidence from cardiac surgery. *Management Science* **52**(4) 473–488.

Huckman, R.S., B.R. Staats. 2011. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management* **13**(3) 310–328.

Huckman, R.S., B.R. Staats, D.M. Upton. 2009. Team familiarity, role experience, and performance: Evidence from indian software services. *Management Science* **55**(1) 85–100.

KC, D.S., B.R. Staats. 2012. Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing & Service Operations Management* **14**(4) 618–633.

KC, D.S., C. Terwiesch. 2011. The effects of focus on performance: Evidence from california hospitals. *Management Science* **57**(11) 1897–1912.

Kent, P., E. Siemsen. 2018. Production process moves: template use and the need to adapt. *Production and Operations Management* **27**(3) 480–495.

Kim, J.W., F.E. Ritter. 2015. Learning, forgetting, and relearning for keystroke-and mouse-driven tasks: Relearning is important. *Human–Computer Interaction* **30**(1) 1–33.

Kim, S., H. Song, M. Valentine. 2018. Cross-disciplinary temporary teams in health care organizations: The role of partner variety and shared experience. Working Paper.

Kingman, J.F.C. 1961. The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 57. Cambridge University Press, 902–904.

Larsson, A.F.L., K. Kircher, J.A. Hultgren. 2014. Learning from experience: Familiarity with acc and responding to a cut-in situation in automated driving. *Transportation research part F: traffic psychology and behaviour* **27** 229–237.

Lewis, M. 2008. Price dispersion and competition with differentiated sellers. *The Journal of Industrial Economics* **56**(3) 654–678.

Narayanan, S., S. Balasubramanian, J.M. Swaminathan. 2009. A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Science* **55**(11) 1861–1876.

Narayanan, S., J.M. Swaminathan, S. Talluri. 2014. Knowledge diversity, turnover, and organizational-unit productivity: An empirical analysis in a knowledge-intensive context. *Production and Operations Management* **23**(8) 1332–1351.

Newell, A., P.S. Rosenbloom. 1981. Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and their Acquisition* **1**(1981) 1–55.

Papachristofi, O., D. Jenkins, L.D. Sharples. 2016. Assessment of learning curves in complex surgical interventions: A consecutive case-series study. *Trials* **17**(1) 266.

Reagans, R., L. Argote, D. Brooks. 2005. Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science* **51**(6) 869–881.

Ritter, F.E., G. Baxter, J.W. Kim, S. Srinivasmurthy. 2013. Learning and retention. *The Oxford Handbook of Cognitive Engineering*. Oxford New York, NY, 125–142.

Robbins, T.R. 2015. Experience-based routing in call center environments. *Service Science* **7**(2) 132–148.

Rossnagel, K., G.J. Jungehülsing, C.H. Nolte, J. Müller-Nordhorn, S. Roll, K. Wegscheider, A. Villringer, S.N. Willich. 2004. Out-of-hospital delays in patients with acute stroke. *Annals of Emergency Medicine* **44**(5) 476–483.

Sherman, J., A. Weiss. 2017. On fruitful and futile tests of the relationship between search and price dispersion. *Economic Inquiry* **55**(4) 1898–1918.

Shostack, G.L. 1987. Service positioning through structural change. *The Journal of Marketing* 34–43.

Staats, B., D. Brunner, D. Upton. 2011. Lean principles, learning, and knowledge work: Evidence from a software services provider. *Journal of Operations Management* **29**(5) 376–390.

Staats, B.R. 2012. Unpacking team familiarity: The effects of geographic location and hierarchical role. *Production and Operations Management* **21**(3) 619–635.

Thorndike, E.L. 1898. Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements* **2**(4) i.

Thurstone, L.L. 1919. The learning curve equation. *Psychological Monographs* **26**(3) i.

van Leeuwen, P.M., R. Happee, J.C.F. de Winter. 2015. Changes of driving performance and gaze behavior of novice drivers during a 30-min simulator-based training. *Procedia Manufacturing* **3** 3325–3332.

Western, B., D. Bloome. 2009. Variance function regressions for studying inequality. *Sociological Methodology* **39**(1) 293–326.

Wright, T.P. 1936. Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences* **3**(4) 122–128.

Yelle, L.E. 1979. The learning curve: Historical review and comprehensive survey. *Decision Sciences* **10**(2) 302–328.

Appendix for "The Variance Learning Curve"

A1. Data cleaning and inclusion

Below is a summary of the inclusion criteria for the performance analysis discussed in §4. We limit the analysis to ambulance activations undertaken by a team of two paramedics and for which we have accurate time stamps. In addition, we clean the dataset of outliers by deleting the observations with the 1% longest and 1% shortest task completion times. Note that in calculating team members' prior experience we count the total number of dispatches they have participated in (regardless of whether a patient was conveyed to a hospital or not). However, we also conduct robustness checks in which only patient conveying activations contribute to the experience variable (see results in Table A5 in §A3.3).

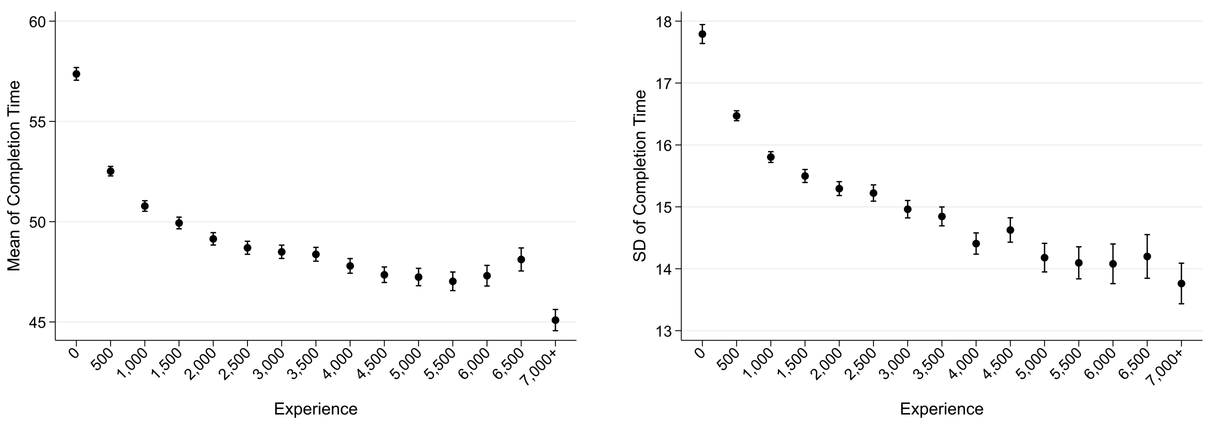
Table A1 Summary of data cleaning and inclusion/exclusion criteria.

	Activations	
	N	ΔN
Patient conveying activations by paramedic teams (with available time stamps and crew data)	6,433,352	
- Teams of 2 paramedics	6,260,185	2.69%
- Drop if timestamps are in the wrong order	6,194,393	1.05%
- Drop if errors in timestamps	6,168,038	0.43%
- Drop high outliers (top 1%)	5,987,256	2.93%
- Drop low outliers (bottom 1%)	5,820,959	2.78%
- Singletons dropped from analysis	5,820,866	0.00%

A2. Additional Material for Main Analysis

A2.1. Visual Analysis Using De-Trended Raw Data

To provide additional support for the validity of the two-step methodology, we use raw data to conduct a simple visual analysis. Figure A1, below, plots the average completion times (panel (a)) and standard deviation of completion times (panel (b)) for various levels of experience at the daily level. The only step of this analysis was de-trending the data, as London experienced increased congestion levels during the decade for which we have data. The patterns in the plots are nearly identical to those produced using the two-step procedure (Figure 2), which includes various controls and fixed effects.



(a) Mean Effect

(b) Variance Effect

Figure A1 Visual analysis using de-trended raw data.

A2.2. All coefficients for main analysis

Table A2 The impact of experience on mean and variance of completion times (all coefficients)

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$
$AvgCrewExp$	-0.511***	-0.018***	-0.872***	-0.040***
	(0.003)	(0.000)	(0.008)	(0.001)
$AvgCrewExp^2$			0.029***	0.002***
			(0.001)	(0.000)
$AvgCrewAge$	0.083***	0.005***	0.088***	0.006***
	(0.001)	(0.000)	(0.001)	(0.000)
$Utilization$	-1.666***	-0.039***	-1.658***	-0.041***
	(0.041)	(0.006)	(0.041)	(0.006)
$BlueCall$	-2.351***	-0.038***	-2.348***	-0.039***
	(0.026)	(0.004)	(0.026)	(0.004)
$TimeOnShift$	0.348***	0.026***	0.350***	0.027***
	(0.023)	(0.003)	(0.023)	(0.003)
$TimeOnShift^2$	-0.061***	-0.003***	-0.061***	-0.003***
	(0.002)	(0.000)	(0.002)	(0.000)
R-squared	0.207	0.016	0.208	0.017
Observations	5,820,866	5,820,866	5,820,866	5,820,866

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3. Robustness Tables

A3.1. Robustness to outcome definition: More granular task definitions

In addition to the main analysis, we examine the three sub-components of $TimeToHospital_{a,c}$ to understand whether the effects we observe are uniform across the different types of tasks. We define the three components as follows.

Driving time to scene. The first component of $TimeToHospital_{a,c}$ is the driving time from the ambulance's position at the time of dispatch to the scene of the incident, denoted by $DriveToScene_{a,c}$ for activation a of crew c.

Patient pick-up time at scene. The second component is the time the crew spends at the scene, stabilizing and attending to the patient. We denote the completion time of this component by $TimeAtScene_{a,c}$ and define it by the time from ambulance arrival at the scene until ambulance leaves the scene, heading for a hospital.

Driving time to hospital. The last component of the out-of-hospital delay is the driving time from the scene to the hospital, denoted by $DriveToHospital_{a,c}$ for activation a of paramedic crew c.

We note that for simplicity of exposition we denote the residuals for each regression by $\hat{u}_{a,c}$. However, the values of $\hat{u}_{a,c}$ are different for columns (2), (4), and (6).

Table A3 The impact of experience on mean and variance of sub-process completion times

	(1) Mean: $DriveToScene$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeAtScene$	(4) Variance: $ln(\hat{u}^2)$	(5) Mean: $DriveToHospital$	(6) Variance: $ln(\hat{u}^2)$
$AvgCrewExp$	-0.205***	-0.075***	-0.573***	-0.027***	-0.094***	-0.014***
	(0.002)	(0.001)	(0.005)	(0.001)	(0.003)	(0.001)
$AvgCrewExp^2$	0.009***	0.004***	0.017***	0.001***	0.003***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.079	0.031	0.194	0.028	0.138	0.044
Observations	5,820,866	5,820,866	5,820,866	5,820,866	5,820,866	5,820,866

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3.2. Robustness to experience definition: Maximum and minimum crew experience

Table A4 Robustness to alternative definition of experience variable

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$
Panel A				
$MaxCrewExp$	-0.358***	-0.014***	-0.620***	-0.028***
	(0.002)	(0.000)	(0.007)	(0.001)
$MaxCrewExp^2$			0.017***	0.001***
			(0.000)	(0.000)
R-squared	0.206	0.016	0.207	0.017
Observations	5,820,866	5,820,866	5,820,866	5,820,866
Panel B				
$MinCrewExp$	-0.398***	-0.011***	-0.664***	-0.026***
	(0.003)	(0.000)	(0.008)	(0.001)
$MinCrewExp^2$			0.025***	0.001***
			(0.001)	(0.000)
R-squared	0.205	0.016	0.205	0.016
Observations	5,820,866	5,820,866	5,820,866	5,820,866

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3.3. Robustness to experience definition: Narrow definition of experience

Table A5 **Robustness to alternative definition of experience variable: only patient conveying experience**

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$
$AvgCrewActExp$	-0.647***	-0.023***	-1.109***	-0.050***
	(0.004)	(0.001)	(0.010)	(0.001)
$AvgCrewActExp^2$			0.049***	0.003***
			(0.001)	(0.000)
R-squared	0.207	0.016	0.207	0.017
Observations	5,820,866	5,820,866	5,820,866	5,820,866

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3.4. Robustness to alternative experience definition

The main independent variable of interest, $AvgCrewExp$, is defined as the average number of prior dispatches undertaken by the crew members during our observation period (2006-2015). We use this measure as it aligns well with the existing learning curve literature and is a granular measure of the number of times each crew member conducted the task of bringing a patient to the hospital. However, our crew dataset includes the hiring date of each crew member, so an alternative measure is the time-on-the-job at LAS. We denote the average years of experience of each ambulance crew by $TimeSinceHired$ and conduct a robustness check to this alternative definition of experience. Table A6 includes the results of this analysis. Columns (1) and (2) show the impact of $TimeSinceHired$ on the mean and variance of performance, respectively. These results are consistent with the main analysis, i.e., years of experience improves performance in a convex fashion. In columns (3) and (4), we include both definitions of experience, to examine whether crew performance improves with more activations after controlling for years of experience. We observe that the estimates on the average crew experience remain similar to the main results in Table 2. The estimates on the number of years of experience become smaller compared to columns (1) and (2), but retain statistical significance.

Table A6 Robustness to alternative experience definition: Time since being hired as a measure of experience.

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$
$TimeSinceHired$	-0.284***	-0.010***	-0.160***	-0.003***
	(0.003)	(0.000)	(0.003)	(0.000)
$TimeSinceHired^2$	0.005***	0.000***	0.002***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
$AvgCrewExp$			-0.772***	-0.036***
			(0.006)	(0.001)
$AvgCrewExp^2$			0.026***	0.002***
			(0.000)	(0.000)
Shift Controls	✓	✓	✓	✓
Activation Controls	✓	✓	✓	✓
Seasonality Controls	✓	✓	✓	✓
R-squared	0.204	0.016	0.209	0.017
Observations	5,820,866	5,820,866	5,820,866	5,820,866

Notes: For ease of interpretation $AvgCrewExp$ is scaled by 1/500. Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3.5. Robustness to log transformation of experience variable

Table A7 The impact of experience on mean and variance of completion times: Logged Experience

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$
$Ln(AvgCrewExp)$	-1.638***	-0.069***
	(0.010)	(0.001)
Shift Controls	✓	✓
Activation Controls	✓	✓
Seasonality Controls	✓	✓
R-squared	0.207	0.017
Observations	5,820,866	5,820,866

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3.6. Robustness to excluding paramedics with prior experience

As in many prior studies of experience effects, our data is censored since the sample does not include paramedics experiences prior to 2006. To address any concerns about the censoring, we repeat the analysis focusing only on ambulance crews involving paramedics who joined LAS in 2006 or later. For this subset of paramedics, the data includes every ambulance activation they have conducted at LAS, so their learning curves are not censored. Table A8 includes the results of this analysis and shows that, as expected, the linear coefficients of the parsimonious model are even more negative than in the main analysis.

Table A8 The impact of experience on mean and variance of completion times: Only 2006 onward

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$
$AvgCrewExp$	-0.797***	-0.023***	-0.979***	-0.041***
	(0.009)	(0.001)	(0.022)	(0.003)
$AvgCrewExp^2$			0.024***	0.002***
			(0.003)	(0.000)
R-squared	0.179	0.016	0.180	0.016
Observations	1,576,081	1,576,081	1,576,081	1,576,081

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3.7. Robustness to excluding teams that do not acquire high experience

To address potential concerns about selection (e.g., if well performing crews tend to achieve high $AvgCrewExp$, whereas poor performing crews do not), we repeat the main analysis for only the crews that eventually accrue an average experience of over 4,000 activations and for whom we have at least one observation at each level of experience (i.e., with $\mathbf{1}_{\{AvgCrewExp\ \sim\ n\}} = 1\ \forall n \in \{0,\ 500, \ldots,\ 4{,}000\}$). Table A9 reports the coefficient estimates for this cohort, which includes 93 crews who conducted a total of 289,706 ambulance activations.

Table A9 Analysis repeated for teams for whom we observe all levels of experience.

	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$
$AvgCrewExp$	-0.726***	-0.030***	-1.683***	-0.085***
	(0.026)	(0.004)	(0.063)	(0.009)
$AvgCrewExp^2$			0.046***	0.003***
			(0.003)	(0.000)
R-squared	0.205	0.023	0.206	0.023
Observations	289,706	289,706	289,706	289,706

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. $^{*}\ p<0.10$, $^{**}\ p<0.05$, $^{***}\ p<0.01$.

A3.8. Robustness to second stage regression specification

Table A10 Analysis repeated without log transformation of residuals for second stage regression.

	(1) Mean: $TimeToHospital$	(2) Variance: $\hat{u}^2$	(3) Mean: $TimeToHospital$	(4) Variance: $\hat{u}^2$
Individual Experience	-0.511***	-3.338***	-0.872***	-7.979***
	(0.003)	(0.068)	(0.008)	(0.168)
Individual Experience Sq.			0.029***	0.377***
			(0.001)	(0.012)
R-squared	0.207	0.021	0.208	0.021
Observations	5,820,866	5,820,866	5,820,866	5,820,866

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A3.9. Robustness to inclusion of alternative control variables

Table A11 Analysis repeated with alternative sets of control variables.

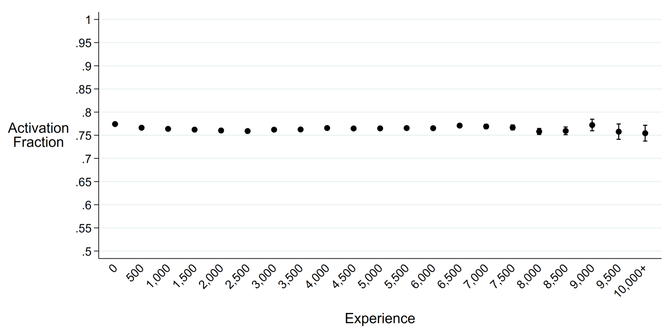
	Month-Year FEs		Location Interactions		Team Familiarity	
	(1) Mean: $TimeToHospital$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeToHospital$	(4) Variance: $ln(\hat{u}^2)$	(5) Mean: $TimeToHospital$	(6) Variance: $ln(\hat{u}^2)$
$AvgCrewExp$	-0.873***	-0.038***	-0.828***	-0.037***	-0.875***	-0.040***
	(0.006)	(0.001)	(0.006)	(0.001)	(0.006)	(0.001)
$AvgCrewExp^2$	0.029***	0.002***	0.027***	0.002***	0.035***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.209	0.017	0.225	0.019	0.208	0.017
Observations	5,820,866	5,820,866	5,820,866	5,820,866	5,820,866	5,820,866

Notes: Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Each column modifies these sets of controls: column (1) replaces the time trend with month-year fixed effects; column (2) add interactions between base and hospital location fixed effects; column (3) adds a control for team familiarity defined as the number of activations that the two paramedics have conducted together. Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

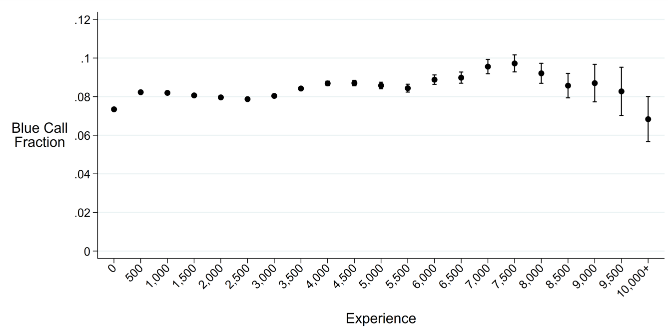
A3.10. Robustness to the crew's decision of transporting a patient to a hospital

Figure A2(a) provides a mean and confidence interval for the fraction of all dispatches that are activations (i.e., a patient was transported to the AE), as a function of crew experience (the main independent variable). We observe that regardless of experience, the proportion is consistently between 0.75 and 0.78. Furthermore, even within that narrow interval, we cannot identify a pattern of changes in decision making as a function of experience. Similarly, Figure A2(b) demonstrates the fraction of blue calls is around 7%-10% for all values of experience in the data. These figures suggest that there is no relationship between crew experience and the likelihood of being dispatched to scenes which either result in a patient being brought to the hospital or a blue call.

Figure A2 **Fraction of dispatches which turn into activations (a) and activations which are blue calls (b).**



(a) Activations (fraction of dispatches)



(b) Blue Calls (fraction of activations)

Table A12 presents the results for the robustness check in which there are interactions between blue calls and the experience variables. As before, we observe strong learning effects for both the mean and variance of performance, for blue call activations as well as regular activations. We note that while there is a positive coefficient for the interaction $AvgCrewExp \times BlueCall$ for the mean learning curve, the overall effect of experience needs to be calculated by adding up the linear and squared terms for all the experience terms. This reveals a convex learning function (due to the large coefficient of the $AvgCrewExp^2 \times BlueCall$ interaction), much like in the main results.

Table A12 The impact of experience on mean and variance of performance: parametric models with blue call interaction.

	(1) Mean: $TimeAtScene$	(2) Variance: $ln(\hat{u}^2)$	(3) Mean: $TimeAtScene$	(4) Variance: $ln(\hat{u}^2)$
$AvgCrewExp$	-0.363***	-0.013***	-0.580***	-0.026***
	(0.002)	(0.000)	(0.005)	(0.001)
$AvgCrewExp^2$			0.018***	0.001***
			(0.000)	(0.000)
$AvgCrewExp \times BlueCall$	-0.077***	-0.011***	0.093***	-0.006**
	(0.006)	(0.001)	(0.016)	(0.003)
$AvgCrewExp^2 \times BlueCall$			-0.014***	-0.000**
			(0.001)	(0.000)
$BlueCall$	1.896***	0.111***	1.593***	0.103***
	(0.030)	(0.006)	(0.040)	(0.008)
Shift Controls	✓	✓	✓	✓
Activation Controls	✓	✓	✓	✓
Seasonality Controls	✓	✓	✓	✓
R-squared	0.194	0.028	0.194	0.028
Observations	5,820,866	5,820,866	5,820,866	5,820,866

Notes: For ease of interpretation $AvgCrewExp$ is scaled by 1/500. Activation controls include fixed effects for the illness code, whether it was a blue call, the base and receiving hospital. Shift controls include the minutes since first dispatch (linear and quadratic terms) and workload since the first dispatch. Seasonality controls include fixed effects for the hour of the day, day of the week, and month of the year as well as a time trend (linear, squared, and cubed). Standard errors are robust and clustered at the shift level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.