Essays on the counterintuitive consequences of labor policies in service industries

By

MohammadMahdi Hashemian

B.S. Electrical Engineering
Sharif University of Technology, 2009
MSC. Sharif University of Technology, 2013
S.M. Massachusetts Institute of Technology, 2016

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2020

Signature of Author:____

Department of Management
08/07/2020

Certified by:____

Hazhir Rahmandad
Mitsubishi Career Development Professor
Associate Professor, System Dynamics
Thesis Supervisor

Accepted by: ____

Catherine Tucker

Sloan Distinguished Professor of Management
Professor, Marketing
Faculty Chair, MIT Sloan PhD Program

Essays on the counterintuitive consequences of labor policies in service industries

By

MohammadMahdi Hashemian

Submitted to the Sloan School of Management on August 7, 2020 in Partial Fulfillment of the Requirements or The Degree of Doctor of Philosophy in Management

ABSTRACT:

In essays one and two, I examine how unstable schedules affect financial performance. In essay one, using 52 weeks of data from over 1,000 stores and more than 15,000 employees of a specialty retailer, I estimate the effect of unstable schedules on store productivity. I use an instrumental variable approach and a natural experiment to partially address the possible endogeneity of scheduling decisions. I find evidence that increasing the adequacy and consistency of employees' hours improves employee and store productivity and find partial support for the positive effect of predictability. To study the policy impact of these findings, I build a behavioral agent-based model of scheduling in essay two. My model provides a platform to conduct counterfactual analyses and thus increases the external validity of my findings. Results suggest that standard scheduling practices, under certain conditions, may have negative, direct labor cost consequences despite their intended rationale for aligning service capacity and demand. Findings highlight the unintended consequences of a narrow focus on matching labor supply to customer demand; designing more employee-friendly schedules could not only create better jobs but also improve firm performance. In essay three, I build a simulation model to explain why Startups play a major role in establishing many new markets when existing firms have more resources and the relevant core and peripheral capabilities. I explore how the strong link between startups' past performance and the resources available for their future capability building conditions their growth prospects. I show that this reinforcing loop leads to entrepreneurial financial markets rapidly focusing on more promising startups. The strength of this mechanism can allow startups to over-take projects within incumbent firms that are initially better endowed. Using an online experiment, I test the key requirement for our mechanism, showing that the strength of the reinforcing loop is larger for start-ups than in-house projects.

Thesis Supervisor: Hazhir Rahmandad

Title: Mitsubishi Career Development Professor
Associate Professor, System Dynamics

Acknowledgement

First and foremost, I feel profoundly grateful to my advisor Hazhir Rahmandad. He patiently supported me through every stage from the day I set foot on MIT to the last days. He has been very generous with his time, and understanding when I was struggling with the Ph.D. life. He always pushed me to be creative and to not compromise rigor. At the same time, he allowed me to pursue what I love and to disagree with him in order to find who I am as a researcher.

I owe a special debt of gratitude to John Sterman and Zeynep Ton. John accepted me to MIT, was always supportive and available when mattered. He provided invaluable feedback on all my work. Zeynep introduced me to the research on the quality of jobs and ignited a passion for it within me. She was very patient, understanding, and generous throughout the process of completion of my papers.

I am also grateful to my parents and my brother Yahya for mentally supporting me during this time. I am thankful to my daughter Zeynab for her presence have been filling my heart with calm and light. Without her knowing, I thought of her whenever I needed strength and I looked at her eyes whenever I needed hope. I am also thankful to my ex-wife Homa, my friends Mehdi Kazemi, Vahid Razavi, Mohammad Rashidian, Davoud Ebrahimi, Vahid Rashidian, Mohammad Siadat, and Sadeq Salehi for being there for me when they could.

I am thankful to Cat Turco and Ezra Zuckerman for they contributed significantly to my development as a scholar. They helped me improve the way I think and the way I approach research. Cat, has been a great role model for excellence in teaching and research.

I feel strongly about thanking Ali Mashayekhi and Babak Alavi of The Sharif University of Technology because they believed in me and made me ready for the Ph.D. journey.

Finally, I thank the SD community. David Keith and Nelson reprenning contributed to my work with their feedback and support. I thank Ozge Karanfil, Jad Sassine, and Sergey Naumov for their feedbacks and stimulating conversations during my time at MIT and various stages of research. I am also grateful to all other members of the SD community including my fellow Ph.D. students.

Table of Contents

Essay 1: Unproductively "Optimal" Schedules: The Effect of Unstable Schedules on Employee and Unit Productivity

Abstract: Unstable schedules—hours that are inadequate, unpredictable (i.e., deviate from scheduled hours), or that vary from week to week—hurt employees. We examine whether they also affect financial performance. Using 52 weeks of data from over 1,000 stores and more than 15,000 employees of a specialty retailer, we estimate the effect of unstable schedules on employee and store productivity. We use an instrumental variable approach and a natural experiment to partially address the possible endogeneity of scheduling decisions. We find evidence that increasing the adequacy and consistency of employees' hours improves employee and store productivity and find partial support for the positive effect of predictability. Our estimates suggest that improving adequacy has the largest productivity effect; having a typical employee work 24 hours a week instead of 13, without changing overall staffing levels, can increase her productivity by 10% to 29% in different estimations. Our findings highlight the unintended consequences of a narrow focus on matching labor supply to customer demand; incorporating hour adequacy, consistency, and predictability in scheduling decisions could not only create better jobs but also improve productivity.

Keywords: Hourly Employees, Schedules, Store Productivity, Hours, Schedule Consistency, Schedule Predictability, Instrumental Variables, Natural Experiment

"They have tons of us [in a single store] … they hire a lot of people but don't have enough hours to give [us]." – Part-time employee

Introduction

In 2018, 8.8 million retail sales workers and 6.5 million food and beverage serving workers together accounted for 8.5 percent of US employment (BLS 2018). These service-sector jobs are characterized not only by their low wages and paucity of benefits, but also by unstable schedules. Hourly employees often work different hours week to week and are given their schedules with short notice—sometimes only three days in advance (Kantor 2014). Even those schedules can change at the last minute and disrupt employees' lives. Many hourly employees find their hours inadequate and have to work two or three jobs (Kasperkevic 2014). While the negative consequences of unstable schedules on employee well-being are well documented (e.g. Lambert 2008; Schneider and Harknett 2018), there is limited research on their effect on company performance (Williams et al. 2018).

Unstable schedules arise from trying to maximize profits by matching labor supply with the highly variable workload driven by fluctuating customer traffic (Chuang, Oliva, and Perdikaki 2016; Perdikaki, Kesavan, and Swaminathan 2012) and other non-customer facing work. Having more labor-hours than workload raises labor costs, but having fewer undermines service and sales and increases mistakes and operational disruptions (Chuang and Oliva 2015; Oliva and Sterman 2001; Ton 2009). Companies therefore try to match labor supply to the workload in increments as short as one hour, but this requires accurate workload forecasts and a flexible labor supply. Announcing employee schedules as late as possible helps reduce the time horizon of forecasts and increase their accuracy. Having on-call employees and making last-minute schedule changes helps businesses react to changes in customer traffic, last-minute merchandising changes, store deliveries, or headquarters visits (Ton 2017; Williams et al. 2018). Operating with a lot of part-time employees who can work in short chunks of time helps with volume flexibility (Goyal and Netessine 2011; Kesavan, Staats, and Gilland 2014; Oliva 2001).

Unstable schedules may arise from efficiency concerns, yet they may also undermine performance (Williams et al. 2018). Unstable schedules may reduce employee focus and motivation and create attendance and turnover issues, which would make it more challenging to match labor supply to demand. Furthermore, stores might lose expertise. Finally, with instability in labor managers' ability to lead and develop employees can be challenged. Scheduling algorithms that do not account for these costs may lead to schedules that hurt both the employees and the productivity.

We use 52 weeks of data from over 1,000[1] stores and more than 15,000 hourly employees of a specialty retailer, NRC (a pseudonym), to test the effect of unstable schedules on employee and store productivity. We examine three dimensions of schedule stability (Williams et al. 2018)- Adequacy: Do employees get enough hours to earn a decent living? Consistency: Do their number of hours keep changing from week to week? And predictability: Do they work different number of hours than scheduled? In our sample, half of employees work less than 13 hours a week, earning less than $8,320 a year; 80% work less than 31 hours a week, earning less than $20,800 a year. On average, there is a 3.5 hours difference between the scheduled and actual weekly working hours (21%) and the actual fluctuates 5.6 hours from the mean (34%).

In addition to OLS estimates, we use an instrumental variable approach and a natural experiment and find strong evidence for the positive effect of adequacy and consistency and partial support for the positive effect of predictability on employee and store productivity. In particular, we find that increasing adequacy has the largest effect on productivity: having a

[1] We disguise the number of stores and the number of workers to protect the anonymity of NRC.

typical employee work 24 hours per week rather than 13 can increase her productivity by 10% to 29% in different estimations.

Our findings highlight the unintended consequences of a narrow focus on one dimension of store performance—supply-demand mismatch costs—which may compromise productivity from a more *systemic view of operations*. Therefore, we find thoughtful job design and finding ways to reduce self-inflicted variability in workload (e.g., last minute changes to deliveries, promotions, long delivery windows) may offer win-win opportunities in scheduling domain (e.g. Caro, Kök, and Martínez-de-Albéniz 2019). Finally, our findings point to having fewer employees but each working more hours as a *better labor strategy* to improve productivity.

Hypothesis Development

Our field observations and conversations with employees and managers suggest unstable schedules may harm retailers through both individual- and store-level dynamics:

Effects on Employee Productivity

At the employee level, unstable schedules hurt productivity by undermining both ability and motivation.

Reduced Ability

The financial insecurity arising from inadequate, inconsistent, and unpredictable hours reduces attention and focus. One interviewee said, "*You don't know from week to week how many hours you're going to have but your bills are still the same amount…. you can't even make enough money to pay your internet fee. I need a second job, but the way they schedule me, every time I get a second job, I can't hold on to it.*" Transportation costs alone can be a problem: a study of welfare recipients, many working in retail, found that transportation accounted for 14 percent of their monthly budgets (Halpern-Meekin et al.

2015). We also heard many retail workers complain about not being able to hold a second job due to their fluctuating schedules.

Financial insecurity has been associated with diminished focus (Herzberg 1968; Lavie 2005) and loss of cognitive capacity equivalent to 13 IQ points (Mani et al. 2013). Focus and attention are especially important in customer-facing roles in which employees need to understand customers' needs, communicate effectively, and handle issues rapidly. Employees working limited hours don't make as much progress learning the store's procedures, products, layout, and promotions, so they are more likely to make mistakes, work slowly, or have ineffective customer interactions.

Reduced Motivation and Commitment

Inconsistent schedules cause work-life conflict, which, in turn, reduces work effort (Wayne, Musisca, and Fleeson 2004) and performance (Wood and de Menezes 2010). The difficulty—or impossibility—of making plans, such as arranging child-care, can reduce job satisfaction and commitment (Beauregard and Henry 2009; Kelly, Moen, and Tranby 2011). One interviewee said, "*My life is always in a turmoil. Every day. Because you can't sleep. You need a set schedule to be able to sleep. .... as far as trying to have a life and doing things with people, you can't*." With these effects in mind, we hypothesize:

H1.1 Employees with more hours per week have, on average, higher sales per hour.

H1.2 Employees with more consistent hours have, on average, higher sales per hour.

H1.3 Employees with more predictable hours have, on average, higher sales per hour.

Effect of Unstable Schedules on Store Productivity

Unstable schedules put significant time constraints and considerable cognitive and psychological burden on managers, who find themselves constantly fighting fires rather than developing employees, boosting store performance, and improving the customer experience.

In such an environment, there is a greater number of employees, each having less time with the manager. Having more employees working shorter shifts amplifies the impact of absenteeism and tardiness; someone has to take up the slack and often it is the manager herself. That, in turn, leads to burnout (Mowday, Porter, and Steers 1982). Rushed employees may take shortcuts, reducing service quality and process conformance (Oliva and Sterman 2001). Unstable schedules reduce team continuity, undermining team coordination and learning (Huckman and Pisano 2006).

Although the point of unstable schedules is to match labor supply to demand, they also complicate matching by creating attendance uncertainty and increasing scheduling frictions (Fisher, Gallino, and Netessine 2019). By fostering low commitment and motivation, unstable schedules can increase tardiness, absenteeism, and turnover. Retail employees often do not inform managers in a timely way—or even at all—that they will be late or absent or even that they have quit; managers often waste time following up with employees repeatedly only to find out they have quit. With these effects in mind, we hypothesize:

H2.1 Stores with employees who have more hours per week have, on average, higher sales per hour.

H2.2 Stores with employees who have more consistent hours have, on average, higher sales per hour.

H2.3 Stores with employees who have more predictable hours have, on average, higher sales per hour.

Research Setting and Data

We use a novel dataset from "NRC," a national specialty retailer (in 2015) with over 1,000 small to medium-sized stores and more than 15,000 hourly (including temporary) employees in 2015. The median hourly wage at NRC in 2015 was $10.10, close to the national median of $10.47 for retail sales workers (BLS 2015). NRC stores have, on average, fewer than four employees working at any given time. Employees shelve, clean, make price changes, manage displays, and help customers and are expected to cashier quickly and accurately. Stores average over 3,300 customer visits per week. When busy, there is relatively strict division of labor. When not busy, an employee will often help a customer all the way through to check-out. Employees are encouraged to engage with customers to increase sales and customer loyalty, although many customers shop without much assistance. All customers, however, are affected by store cleanliness, the organization of displays and price tags, and the speed of the check-out lines.

NRC managers work full-time. They are responsible for improving sales, managing costs, running store operations smoothly, and hiring and developing high-performing teams. They face significant pressures in creating employee schedules. They are expected to keep labor costs down by staffing their stores just enough to meet the customer demand in hourly increments. Making staffing decisions is complex. Managers do so with the help of the corporate office algorithmic schedule suggestions alongside customer traffic forecast, and labor budget goal. This, as we argued, leads to unstable work schedules. Managers may also prefer stable schedules, not only because they care about their staff, but also because unstable schedules require managing more employees, handling more complaints and requests, taking on employees' tasks to handle staff shortages, and spending more time on the scheduling itself. Balancing these competing goals requires tradeoffs. For example, improving predictability (through earlier, fixed, schedules) would limit managers' ability to fine tune schedules later.

Dependent Variables

Our dependent variables are store productivity, defined as net sales per total associate hours (*StoreSPAH*), and employee productivity, defined as sales per hour (*AssociateSPAH*) and used by NRC to evaluate employees. Productivities are reported to employees, but no policy connects them to compensation or schedules.

While measuring *StoreSPAH* is straightforward and accurate, measuring *AssociateSPAH* is complicated. Cashiers are responsible for attributing transactions to employees, but if the cashier wasn't the employee who made the sale, correct attribution may not always happen. Moreover, many customers shop without help. Therefore, many transactions are not attributed. Closer examination of data suggests the attribution issues arise mostly during the busy shopping season, November to December. Our results are robust to excluding this period. See appendix A.

Independent Variables

Schedule adequacy (*HOURS*) is measured by the number of hours employees work in a given week. *Schedule consistency* can be operationalized in several ways. A simple measure might be the reverse of the coefficient of variation of hours throughout the year *(CV).* Since *CV* is time-invariant, we also use a time-variant measure of schedule consistency, *SC,* defined in Equation (1), which compares the number of hours in the current week and its rolling average over the prior four weeks.[2]

(1) $SC_{it} = -\left| HOURS_{it} - \frac{\sum_{t-4}^{t-1} HOURS_{it}}{4} \right|$

[2] See the appendix B for robustness checks using other operationalizations.

Schedule predictability (*SP*) is defined as the negative absolute difference between the announced and actual schedules. Note that we are measuring predictability at the weekly level and cannot capture schedule changes within a day or when days or times of the day in which employees work changed but the number of hours did not. Moreover, we cannot distinguish when the difference between scheduled and actual hours is voluntary vs. involuntary. Employees suffer from predictability issues only if the deviation is involuntary, but stores suffer regardless. Therefore, our measure is conservative, and we expect our results to underestimate the true impact of involuntary changes to schedules.

Table *1*. Variable definitions and summary statistics

	Variable	Mean	Median	SD
	Individual-Level			
Productivity	*AssociateSPAH*	79.41	68.60	98.49
Adequacy	*HOURS*	16.28	13.00	11.39
Consistency	*SC*	-5.63	-4.25	4.99
	CV	-0.47	-0.48	0.177
Predictability	*SP*	-3.50	-2	4.44
	HourlyWAGE	10.71	10.11	3.23
	ManagerTENURE	248.04	198.50	195.25
	Store-Level			
Productivity	*StoreSPAH*	173.43	172.20	63.25
Adequacy	*WHOURS*	23.34	23.54	4.56
Consistency	*WSC*	-5.13	-4.38	3.05
	WCV	-0.41	-0.41	0.07
Predictability	*WSP*	-3.53	-2.99	2.5
	TRAFFIC	3335.22	2409.00	3142.85
	SKUs	2929.28	2861.00	447.90
	TPAH	10.64	10.50	4.92

At the store level, we average the corresponding individual-level variables, weighted by employees' hours to capture the impact of each individual proportionally. Table 1 provides summary statistics.

Empirical Strategy

We start with multivariate regressions to estimate the impact of schedule stability on individual and store productivity, and then consider instrumental variables and natural experiments to partially address endogeneity issues. We start with a random-effects specification, presented in Equation (2). Then we estimate a fixed-effect model in Equation (3). In these equations, τ_t represents the time fixed effects, δ_i the individual fixed effects, ϑ_j the store fixed effects, and $\sum_k \beta_k C_{kt}$ the controls.

(2) $AssociateSPAH_{it} = a + \beta_1 HOURS_{it+} \beta_2 CV_i + \beta_3 SP_{it} + \sum_k \beta_k C_{kt} + \vartheta_j + \tau_t + \varepsilon_{it}$

(3) $AssociateSPAH_{it} = a + \beta_4 HOURS_{it} + \beta_5 SC_{it} + \beta_6 SP_{it} + \sum_k \beta_k C_{kt} + \vartheta_j + \delta_i + \tau_t + \varepsilon_{it}$

At the store level, we estimate models using Equations (4) and (5). Note that *WCV* (the weighted average of *CV*) is time-variant since different employees work different number of hours each week.

(4) $StoreSPAH_{jt} = a + \beta_7 WHOURS_{jt} + \beta_8 WCV_{jt} + \beta_9 WSP_{jt} + \sum_k \beta_k C_{kt} + \vartheta_j + \delta_i + \tau_t + \varepsilon_{it}$

(5) $StoreSPAH_{jt} = a + \beta_{10} WHOURS_{jt} + \beta_{11} WSC_{jt} + \beta_{12} WSP_{jt} + \sum_k \beta_k C_{kt} + \vartheta_j + \delta_i + \tau_t + \varepsilon_{it}$

We include the following controls (C_k's):

1. *TPAH* (traffic per store's total associate hours) controls for the possible impact of traffic and the employee-customer ratio.
2. *HourlyWage*[3] controls for the possible correlation of hourly wage with ability and schedule stability.

[3]At the store level we use *WHrlyWage* or the weighted average of wages.

3. *Salesplan* (weekly store sales goal set by the corporate office) controls for the effect these plans have on how managers run their store and schedule employees.
4. *ManagerTenure* (in weeks) controls for the effect of manager experience on how well stores are managed, which may affect scheduling practices and their impact.
5. *CoworkerExperience*[4] (average experience of coworkers in the store, in weeks) controls for the impact of coworkers' experience and skill on each other's performance and schedules.
6. *SKUs* (number of Stock Keeping Units) controls for the impact of store complexity in terms of number of products on employee hours and productivity.

As with all OLS estimations, we face identification challenges, mainly from possible reverse causalities. For example, to increase sales, managers may give more and better hours to more productive employees, even though hours are not officially connected to productivity. Moreover, managers in high-performing stores may be under less pressure to cut hours. In additional analyses we partially address these concerns by including employees' past month sales productivity (*PMSPAH*) as another control and past month hours (PMHOURS) as an alternative dependent variable. We also use an instrumental variable approach and a natural experiment. We use time, individual, and store fixed effects to address unobserved heterogeneities.

Instrumental Variable

Tensions between the various pressures managers face partly explain variations in scheduling. Thus, exogenous events that temporarily elevate managers' awareness of schedules marginally change the priority of stable schedules. We therefore identify exogenous US-state–level events that could make scheduling more salient to store managers without affecting other

[4] At the store level we use weighted average of all team-members experience

aspects of store operations or customers' decisions. To the extent that we see marginal scheduling changes following those events we can partially identify the causal impact of scheduling on productivity.

We use newspaper articles as a proxy for those events, assuming that the articles accurately reflect the event's time and location. We conduct a comprehensive search using three databases that cover local and legal news, press releases, and newswires; namely, Factiva, LexisNexis, and ProQuest global news stream. We look for articles with implications for retail schedules but not for other topics, such as wages.

We find 17 news articles from various sources such as *The New York Times*, *The Washington Post*, and *Buzzfeed*. These articles cover events at multiple locations nationally, or states such as New York, Oregon, and California. From these 17 articles, 5 are public or policy actions and the rest are private company actions. We use the first 5 for our main analyses since they have a clear direction for improving schedule stability. Our results are robust to including all. [5] Examples include a letter from the New York attorney general to 13 retailers and advertisement for better schedules by activist groups in California.

Store managers in our sample could be affected both by the events themselves and by the news coverage of those events. Focusing on a narrower pathway, we analyze the extent to which managers are affected by the event and not the (potentially broader) impacts of national news coverage of scheduling issues. We therefore look for effects only in the states in which the events took place, not expanding to all the states in which the news source's readership lives. We construct our instruments by generating place-time pair dummies using the article publication date and the two following weeks, arguing that the relevant events affect managers' decisions to update hours in the same week and to finalize the schedules for the next two. Note

[5] See Appendix C for the full list of sources and analyses.

that since most articles mention schedules only broadly, they are hard to categorize based on stability dimensions and thus we use the same three instruments for all.

Natural Experiment

A San Francisco labor ordinance introducing employee scheduling constraints provides an opportunity to study the impact of unstable schedules by comparing NRC stores in San Francisco and elsewhere. Put into effect on July 3, 2015, the "Formula Retail Employee Rights" ordinance required retailers to (a) offer more hours to existing employees instead of hiring new ones, (b) give employees their schedules two weeks in advance and pay employees a penalty if they change the schedules later, and (c) pay the on-call employees even if they end up not working.

The ordinance, arguably, affects stores only through changing employee schedules. While the ordinance was intended to help with predictability of employees' schedules, we observe that its actual impact was more complex. The introduction of additional constraints in a complex scheduling process may change schedule stability in unanticipated ways. We therefore first investigate the ordinance's impact on our explanatory variables. Surprisingly the ordinance did not statistically improve any of the measures of schedule stability. To the contrary, it is associated with a decrease in schedule consistency. A closer look at alternative levers managers may use in scheduling provides an explanation. Managers may provide a smoother initial schedule, knowing that later they can fine-tune it by last minute deviations from the original schedule. The San Francisco ordinance made such last-minute adjustments more costly, reducing the flexibility of managers. We conjecture that in response San Francisco stores more aggressively matched supply and demand in the original schedule (decreasing schedule consistency) to compensate for the added cost of unpredictability. Ironically, it seems that San Francisco stores changed their initial planning but still had to change employee hours late which kept predictability at the same level. Therefore, in practice the natural experiment reduced schedule consistency without changing other measures. In the results section we focus on how this change affected the productivity outcomes and discuss the influence of the schedule changes on labor cost and its impact on our estimations.

Results

The individual-level OLS results reported in Models 1 and 2 of Table 2 show that schedule adequacy, consistency, and predictability have a significant positive impact on employee productivity. Particularly, we find the magnitude of the impact of *HOURS* striking. Having a typical employee work 24.39 rather than 13 hours (one standard deviation change), with constant staffing, can increase her sales between $8.08 and $12.75 per hour; that is, by 10% to 16% of the average. Improving all stability dimensions by one standard deviation increases employee productivity by 11% to 24%.

Table *2*. Impact of unstable schedules on Individual and store productivity

		AssociateSPAH		**StoreSPAH**		**AssociateSPAH**		**StoreSPAH**	
		Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
Adequacy	*HOURS*	1.12***	0.71***				0.58***		
		(0.01)	(0.01)				(0.02)		
	WHOURS			0.32***	0.41***				
				(0.01)	(0.01)				
	PMHOURS					0.43***			
						(0.02)			
	WPMHOURS							0.11***	
								(0.02)	
Consistency	*CV*	29.17***							
		(0.77)							
	WCV			17.76***					
				(0.67)					
	SC		0.14***				0.12***		
			(0.02)				(0.02)		
	WSC				0.25***				0.38***
					(0.01)				(0.01)
	PMSC					0.04***			
						(0.01)			
	WPMSC							0.15***	
								(0.02)	
Predictability	*SP*	0.17***	0.10***				0.02		
		(0.02)	(0.03)				(0.03)		
	WSP			0.27*	0.02~				0.43***
				(0.11)	(0.01)				(0.01)
	PMSP					0.10*			
						(0.05)			
	WPMSP							0.46***	
								(0.02)	
Controls	*PMSPAH*						0.28***		
							(0.00)		
	STOREPMSPAH								0.02***
									(0.00)
	HourlyWAGE	0.69***	0.25***			0.49***	-0.05		
		(0.04)	(0.05)			(0.07)	(0.05)		

WHrlyWAGE			0.15***	0.26***			0.48***	0.39***
			(0.02)	(0.02)			(0.02)	(0.02)
TPAH	0.88***	0.67***	2.01***	2.01***	0.54***	0.66***	2.05***	2.15***
	(0.04)	(0.04)	(0.01)	(0.01)	(0.05)	(0.05)	(0.01)	(0.01)
ManagerTENURE	-0.00~	0.00	0.00	0.00	0.00*	0.00	0.00~	0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
CoworkerEXPERIENCE	0.02**	0.02*			0.02~	0.02*		
	(0.01)	(0.01)			(0.01)	(0.01)		
AverageExperience			0.00***	0.00***			0.00***	0.00***
			(0.00)	(0.00)			(0.00)	(0.00)
SalesPlan	15.21***	15.83***	27.38***	27.25***	12.52***	14.02***	28.27***	27.83***
	(0.55)	(0.63)	(0.14)	(0.14)	(0.60)	(0.72)	(0.12)	(0.12)
SKUs	-2.72*	-5.10***	0.99**	0.99**	-5.62***	-3.74*	0.14	0.14
	(1.18)	(1.34)	(0.33)	(0.33)	(1.49)	(1.58)	(0.29)	(0.28)
TimeFE	YES	YES	YES	YES	YES	YES	YES	YES
EmployeeFE	NO	YES	YES	YES	YES	YES	YES	YES
StoreFE	YES	YES	YES	YES	YES	YES	YES	YES
R2	0.17	0.18	0.77	0.77	0.18	0.20	0.85	0.85

Note. Cell entries are estimated regression coefficients and (Robust standard errors). $^{\sim} p < 0.10$, $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Store-level results in Models 3 and 4 reveal significant coefficients consistent with our predictions, with predictability significant at the p=0.1 level. We find that improving all stability dimensions (at store level) by one standard deviation increases store productivity by 1.5% to 1.9%.

We replace our original independent variables in Equations 2 and 4 with their respective rolling prior four weeks average (e.g., *PMHOURS* and *PMWHOURS*), and report the results in Models 5 and 7 in Table 2. These variables represent residual effects of unstable schedules over prior weeks on the current week productivity, and partially address the possible endogeneity within a given week. Results are statistically significant and consistent with our predictions and show the lasting impact of unstable schedules.

In Models 6 and 8 of Table 2, we add past performance to control for the longer-term endogeneity. We use the rolling prior four weeks productivity at the individual and store level (*PMSPAH* and *PMStoreSPAH*), which does not change our results except for the individual-level schedule predictability that loses its statistical significance.

Instrumental-variable Analysis Results[6]

Results for instrumental-variable analysis are presented in Table 3. In Models 9, 10, and 11, we report the individual-level estimates. We find that adequacy has a positive and predictability has a negative correlation with our instrument WeekofNews (the time-state dummy). Adequacy and predictability have a positive correlation with WeekAfter. Consistency and adequacy have a positive correlation with 2WeeksAfter. Cragg-Donald Wald, Sargan, and Anderson Canonical Correlation Statistics show that our instruments are not weak, and endogenous variables are not over or under-identified. Results are consistent

[6] See Appendix D for Instrumental Variable analysis using lagged independent variables. Results are consistent with our predictions.

with our hypotheses (except for predictability, for which we do not see a significant effect). Our results suggest that one standard deviation increase in adequacy and consistency yield a 32% improvement in employee productivity, an effect consistent with (and somewhat larger than) OLS estimates.

Models 12–15 report store-level results. Expectedly, first-stage results are similar to employee-level results. Second stage results are consistent with our hypotheses. Based on these estimates, one standard deviation increase in all stability dimensions yields a 7% improvement in store productivity, again comparable with, and slightly larger than, the OLS results.

Table 3. Instrumental Variable Results

		Individual-Level				Store-Level			
		First-Stage			*Second-Stage*	*First-Stage*			*Second-Stage*
		Hours	**SC**	**SP**	**AssociateSPAH**	**WHOURS**	**WSC**	**WSP**	**StoreSPAH**
		Model9	Model10	Model11	Model12	Model13	Model14	Model15	Model16
Instrumental Variables	*WeekofNews*	0.41***	-0.09	-0.21***		0.31***	-0.02	-0.18***	
		(0.11)	(0.08)	(0.06)		(0.04)	(0.03)	(0.03)	
	WeekAfter	1.03***	0.13	0.36***		0.90***	0.43***	0.20***	
		(0.08)	(0.08)	(0.05)		(0.03)	(0.02)	(0.02)	
	2WeeksAfter	0.34***	0.42***	-0.08		0.30***	0.06*	-0.01	
		(0.09)	(0.07)	(0.05)		(0.03)	(0.03)	(0.02)	
Adequacy	*HOURS*				2.03***				
					(0.11)				
	WHOURS								1.32***
									(0.30)
Consistency	*SC*				0.42***				
					(013)				
	WSC								0.73***
									(0.19)
Predictability	*SP*				0.11				
					(0.76)				
	WSP								1.57*
									(0.67)
Controls	*HourlyWage*	0.17***	0.02***	0.05***	-1.89*				
		(0.00)	(0.00)	(0.00)	(0.75)				
	WHrlyWage					0.21***	0.02***	0.07***	-1.11
						(0.00)	(0.00)	(0.00)	(1.45)
	TPAH	-0.17***	-0.03***	-0.03***	2.89**	-0.18***	-0.04***	-0.03***	2.51***
		(0.00)	(0.00)	(0.00)	(0.88)	(0.00)	(0.00)	(0.00)	(0.38)
	Mtenure	0.00***	-0.00*	-0.00~	-0.01*	-0.00***	-0.00**	-0.00***	0.00~
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	CoworkerExperience	-0.02***	-0.00	-0.00**	0.26**				
		(0.00)	(0.00)	(0.00)	(0.09)				
	TeamExperience					-0.00***	-0.00***	-0.00***	0.00
						(0.00)	(0.00)	(0.00)	(0.01)
	SalesPlan	4.08***	1.62***	0.24***	-47.08*	1.69***	1.79***	0.01	36.83**

		(0.04)	(0.03)	(0.03)	(21.41)	(0.01)	(0.01)	(0.01)	(13.20)
	SKUs	-0.44***	-0.22**	-0.14*	2.20	-0.22***	-0.31***	-0.06*	-0.04
		(0.10)	(0.08)	(0.06)	(2.70)	(0.04)	(0.03)	(0.03)	(2.23)
	Fixed-Effects	YES	YES	YES	YES	YES	YES	YES	YES
	FStatistics	43.70***	17.21***	18.71***	13.33***	52.26***	13.95***	17.95***	12.72***
	R2	0.23	0.09	0.07	-0.97	0.26	0.41	0.34	0.43

Note. Cell entries are estimated regression coefficients and (Clustered standard errors). ~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Natural Experiment Results

NRC has 23 stores in San Francisco. Figure 1 reports a summary of how our dependent variables changed after the introduction of the labor ordinance. As we discussed earlier, the only statistically significant effect on schedule stability is the unintended decrease in consistency.

Figure 1. Effect of San Francisco's "Formula Retail Employee Rights" ordinance on Schedules and productivity

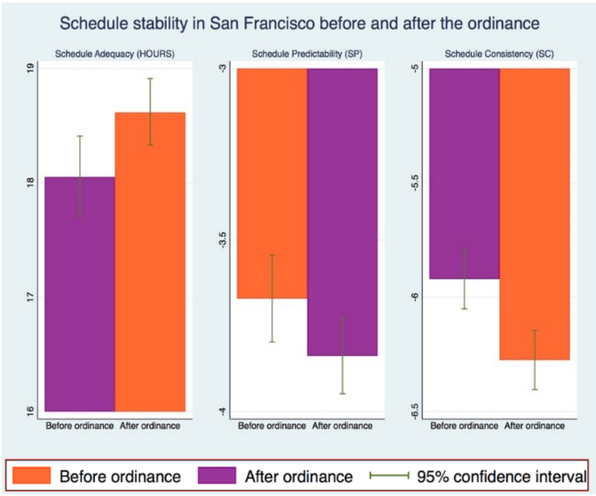


Figure 1.a Changes in schedule stability dimensions

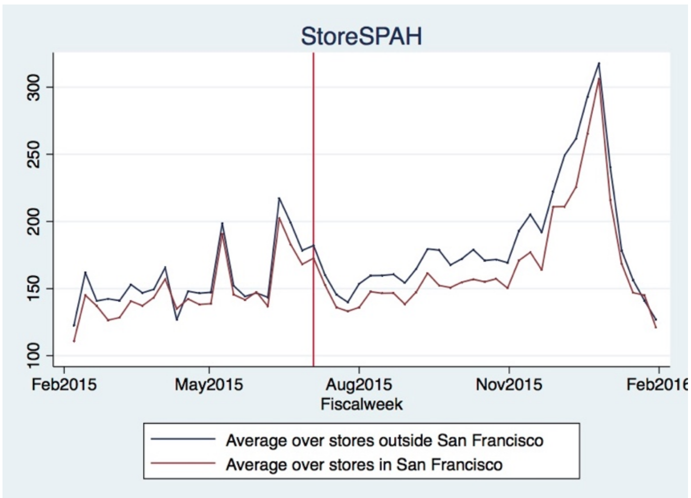


Fig1.d Comparing the productivity of NRC stores in and outside San Francisco

Comparing stores in and outside San Francisco, Figures 1.d and 1.e depict the dynamics of AssociateSPAH and StoreSPAH. We estimate the Equations (6) and (7), clustering errors by US-States, and report the result in Table 4:

(6) $AssociateSPAH_{jt} = a + \beta_{13} SanFrancisco_j * AfterOrdinance_t + \beta_{14} AfterOrdinance_t + \sum_k \beta_k C_{kt} + \vartheta_j + \delta_i + \tau_t + \varepsilon_{it}$

(7) $StoreSPAH_{jt} = a + \beta_{13} SanFrancisco_j * AfterOrdinance_t + \beta_{14} AfterOrdinance_t + \sum_k \beta_k C_{kt} + \vartheta_j + \delta_i + \tau_t + \varepsilon_{it}$

$SanFrancisco_j$ is a dummy equal to 1 for San Francisco stores and 0 otherwise. $AfterOrdinance$t is a dummy equal to 1 for observations after the ordinance and 0 otherwise. On average, stores all around the country have higher productivity after the summer when ordinance was introduced. Comparing San Francisco stores to others suggest, in line with our hypothesis, that lower schedule consistency reduces employee productivity.

In interpreting results from this natural experiment, we should bear in mind a separate San Francisco minimum wage law that went into effect on May 1st, 2015. Our results are robust to including this legal change in our analyses. See online appendix E. The increased minimum wage could actually increase employee productivity due to the pressure from higher labor cost. Since we observed a net reduction in productivity, the actual impact of schedule consistency might be even larger than estimated here.

Overall, this natural experiment provides additional evidence that a reduction in schedule consistency hurts employee productivity. The fact that the reduction in consistency resulted from an intervention originally designed to enhance schedule predictability would also raise a cautionary note about designing effective interventions in complex scheduling contexts.

Table 4. Difference in Difference estimates of Store SPAH

	AssociateSPAH	**StoreSPAH**
	Model17	Model18
SanFranciscoXAfterOrdinance	-31.89***	-4.80***
	(1.74)	(0.49)
AfterOrdinance	48.65***	5.74***
	(1.01)	(0.28)
HourlyWAGE	0.28***	
	(0.04)	
WHrlyWAGE		0.37***
		(0.02)
TPAH	0.40***	1.93***
	(0.04)	(0.01)
ManagerTENURE	0.00~	0.00
	(0.00)	(0.00)
CoworkerEXPERIENCE	-0.01	
	(0.01)	
TeamExperience		0.00***
		(0.00)
SalesPLAN	12.01***	27.54***
	(0.48)	(0.13)
SKUs	-2.22~	0.96**
	(1.19)	(0.33)
TimeFE	YES	YES
EmployeeFE	YES	YES
StoreFE	YES	YES
R2	0.13	0.77

Note. Cell entries are estimated regression coefficients and (Robust standard errors). ~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Conclusion

Many retailers try to match labor supply to demand by creating a flexible workforce. They staff their stores largely with part-time employees, most of whom are assigned too few hours to make a decent income, and have different hours each week, hours that may deviate significantly from original schedules. Using three complementary empirical methods we find significant negative effects of this strategy on both individual and store productivity.

These findings are especially relevant to today's retail environment. Increased competition, higher minimum wages in some states, and adoption of digital strategies and new technologies all necessitate a more productive and motivated workforce. Focused primarily on matching labor hours to demand, the common scheduling approaches do not often account for the full impact of scheduling practices. We show that the productivity effects of schedules are large and may justify some change in common scheduling practices. For example, an alternate labor strategy, relying on fewer employees who each work more hours with less variability, can be increasingly promising. Some case studies provide supportive evidence for viability of this alternative. At Mercadona, Spain's largest supermarket chain, 85% of employees work full-time in shifts of 6.6 hours/day (Ton and Simon 2010). At Costco, one of the world's largest retailers, most part-time employees work a minimum of 24 hours a week. Operating with fewer employees, each working more hours, provides stability at the stores and improves productivity, which in turn may enable companies to pay higher wages. Indeed, both Mercadona and Costco pay better than their competitors do. In 2019, the average hourly wage in Costco's U.S. warehouses is slightly over $23, more than double the median wage for retail store workers in 2018.

However, in order to provide more stable schedules, retailers will have to simultaneously change other aspects of their operating systems to smooth the workload; for example, at Costco product introductions are arranged so that new items are brought out at staggered times which smooth workloads. Stores will also have to alter their merchandising and logistics strategies. For example, unstable schedules may result as much from self-inflicted variability such as last-minute changes to promotions, deliveries, or floor sets, as it is due to demand variability (Williams et al. 2018). Actively scheduling non-customer facing activities to complement demand turns a source of instability into an effective lever to enable stable schedules.

Our study focused on productivity effects of scheduling. Accounting for other systemic implications, from effect of schedules on worker turnover to negative externalities imposed on the community in the form of reduced employee health and well-being, may further justify significant shifts in firms' practices, as well as regulatory interventions. In fact, in the United States, in addition to San Francisco, several other cities such as New York, Seattle and the state of Oregon now penalize unstable scheduling practices. However, as our study of the San Francisco ordinance highlights, finding effective policy constraints to impose on firms' complex scheduling systems may lead to unintended consequences, at least in short term before companies figure out how to adjust scheduling appropriately. We hope that highlighting and measuring the systemic costs of current scheduling practices can better align the incentives of the firms and those of policy makers, enabling closer collaboration in developing more effective regulatory frameworks.

References

Beauregard, Alexandra T. and Lesley C. Henry. 2009. “Making the Link between Work-Life Balance Practices and Organizational Performance.” *Human Resource Management Review* 19(1):9–22.

Caro, Felipe, A. Gürhan Kök, and Victor Martínez-de-Albéniz. 2019. “The Future of Retail Operations.” *Manufacturing & Service Operations Management* 1–20.

Chuang, Howard Hao Chun and Rogelio Oliva. 2015. “Inventory Record Inaccuracy: Causes and Labor Effects.” *Journal of Operations Management* 39–40:63–78.

Chuang, Howard Hao Chun, Rogelio Oliva, and Olga Perdikaki. 2016. “Traffic-Based Labor Planning in Retail Stores.” *Production and Operations Management* 25(1):96–113.

Fisher, Marshall, Santiago Gallino, and Serquei Netessine. 2019. “Retailers Are Squandering Their Most Potent Weapons.” *Harvard Business Review* 97(1):72–79.

Goyal, Manu and Serguei Netessine. 2011. “Volume Flexibility, Product Flexibility, or Both: The Role of Demand Correlation and Product Substitution.” *Manufacturing & Service Operations Management* 13(2):180–93.

Halpern-Meekin, Sarah., Kathryn. Edin, Laura. Tach, and Jennifer. Sykes. 2015. *It’s Not like I’m Poor: How Working Families Make Ends Meet in a Post-Welfare World.* Univ of California Press.

Kantor, Jodi. 2014. “Working Anything but 9 to 5.” *New York Times*.

Kasperkevic, Jana. 2014. “Faces of Part-Time Workers: Food Stamps and Multiple Low-Paid Jobs.” *The Guardian*.

Kelly, Erin L., Phyllis Moen, and Eric Tranby. 2011. “Changing Workplaces to Reduce Work-Family Conflict: Schedule Control in a White-Collar Organization.” *American Sociological Review* 76(2):265–90.

Kesavan, Saravanan, Bradley R. Staats, and Wendell Gilland. 2014. “Volume Flexibility in Services: The Costs and Benefits of Flexible Labor Resources.” *Management Science* 60(8):1884–1906.

Lambert, Susan J. 2008. “Passing the Buck: Labor Flexibility Practices That Transfer Risk onto Hourly Workers.” *Human Relations* 61(9):1203–27.

Lavie, Nilli. 2005. “Distracted and Confused?: Selective Attention under Load.” *Trends in Cognitive Sciences* 9(2):75–82.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. 2013. “Poverty Impedes Cognitive Function.” *Science* 341(6149):976–80.

Mowday, Richard, Lyman Porter, and Richard Steers. 1982. *Employee-Organization Linkages: The Psychology of Commitment, Absenteeism, and Turnover*. New York, NY.: Academic Press.

Oliva, Rogelio. 2001. “Tradeoffs in Responses to Work Pressure in the Service Industry.” *California Management Review* 43(4):26–43.

Oliva, Rogelio and John D. Sterman. 2001. “Cutting Corners and Working Overtime : Quality Erosion in the Service Industry.” *Management Science* 47(7):894–914.

Perdikaki, Olga, Saravanan Kesavan, and Jayashankar M. Swaminathan. 2012. “Effect of

Traffic on Sales and Conversion Rates of Retail Stores." *Manufacturing & Service Operations Management* 14(1):145–62.

Schneider, Daniel and Kristen Harknett. 2018. *How Work Schedules Affect Health and Wellbeing : The Mediating Roles of Economic Insecurity and Work-Life Conflict*.

Ton, Zeynep. 2009. "The Effect of Labor on Profitability: The Role of Quality." *Harvard Business School* 1–33.

Ton, Zeynep. 2017. "The Case for Good Jobs." *Harvard Business Review*.

Wayne, Julie Holliday, Nicholas Musisca, and William Fleeson. 2004. "Considering the Role of Personality in the Work-Family Experience: Relationships of the Big Five to Work-Family Conflict and Facilitation." *Journal of Vocational Behavior* 64(1):108–30.

Williams, Joan C., Susan J. Lambert, Saravanan Kesavan, Peter J. Fugiel, Lori Ann Ospina, Erin Devorah Rapoport, Meghan Jarpe, Dylan Bellisle, Pradeep Pendem, Lisa McCorkell, and Sarah Adler-Milstein. 2018. *Stable Scheduling Increases Productivity and Sales: The Stable Scheduling Study*.

Wood, Stephen J. and Lilian M. de Menezes. 2010. "Family-Friendly Management, Organizational Performance and Social Legitimacy." *International Journal of Human Resource Management* 21(10):1575–97.

Essay2: Understanding the Role of Labor Scheduling Strategies in Service Organizations

Abstract: Labor flexibility practices and the resulting instability of schedules are common in the service industry. The economics and operations management literature has traditionally focused on designing optimal schedules to minimize the mismatch costs of employee hour supply and demand. Common managerial practices treat employees as a cost to be minimized, suggesting that workers and firms are increasingly at odds. We build on the literature that suggests labor flexibility practices hurt store productivity and financial performance. We contribute by building a behavioral agent-based model to study the role of scheduling in service organizations. We estimate our model using novel approaches adopted from econometrics of structural models utilizing 52 weeks of data from over 1,000 stores and more than 15,000 employees of a specialty retailer. Accounting for systemic implications of scheduling creates a platform to assess the feasibility and financial impact of alternative scheduling strategies. We find that identifying high-performing employees and providing them with more stable schedules may not only improve productivity but also systematically reduce labor costs and ultimately result in up to a 20% decrease in labor percentage of costs.

1. Introduction

Job design in the service industry, particularly the design of employee work schedules, have a significant impact on employee quality of life, firm profits, employment, and the economy in general. Retail and the fast-food and beverage industries have been the top two employers in the U.S. for many years. For example, in 2019, 4.3 million people worked as retail salespersons, nearly 4 million worked as fast-food workers, and 3.6 million worked as cashiers (BLS, 2019). Moreover, technological changes in the economy, particularly the increasing trends in automation, are moving employees to the service sector (Acemoglu & Autor, 2011; Autor, 2014). Therefore, it is crucial to understand the tradeoffs and complexities of designing better schedules for service workers.

Brick and mortar retailers face significant pressure from online competition and are compelled to aggressively reduce costs. Labor costs span 10–20% of a retailer's revenue and are one of the more flexible parts of their cost structure (Adhi et al., 2020), thus becoming a major focus for cost reduction. As a result, retailers frequently offer jobs that pay little and require little training, create few career opportunities, and rely heavily on flexible and unpredictable schedules to better match staff to demand changes. In this paper, we focus specifically on labor flexibility practices and the complexities of designing better work schedules. Flexible scheduling practices provide employers with more control over the number of employees, their hours, the match between demand and available labor, and, ultimately, labor costs. This often results in a system in which retailers hire many hourly employees and give each of them fewer hours that change from week to week and might be different from the announced, formal schedules (e.g., Lambert, 2008).

The economics and operations management literature has traditionally focused on designing optimal schedules to minimize the supply and demand mismatch costs (e.g., Läubli et al., 2015). Research shows that employing many part-time workers and scheduling them in short shifts creates the flexibility that retailers need to match the supply and demand of employee hours (Goyal & Netessine, 2011; Kesavan et al., 2014; Oliva, 2001). In contrast, the industrial relations literature has focused mostly on the impact scheduling practices have on the service employees' quality of life (Kelly et al., 2011; Lambert, 2008, 2009; Schneider & Harknett, 2018). For example, Lambert et al. (2019) show that these scheduling practices are widespread and a source of income insecurity and institutional distrust among workers. Similarly, Halpern-Meekin et al. (2015) report, based on interviews, that low-income working households want, more than anything, stable jobs with stable schedules that offer a reliable paycheck. Moreover, the volatility of hours creates additional uncertainty across various aspects of life and generates significant risks of personal bankruptcy and poverty traps (e.g., Ehrenreich, 2001). Additionally, research suggests that fluctuating schedules increase work–life conflict, mainly due to the impact on childcare (Ben-Ishai et al., 2014; Henly & Lambert, 2005). Finally, the existing literature suggests that the reduction of financial security and the increase in work–life conflicts that stem from changing work schedules jeopardize employees' health and wellbeing (Cho, 2018; Schneider & Harknett, 2018).

In contrast to the growing research on the impact scheduling practices have on employees, there is limited research about their effects on company performance (Hashemian et al., 2020; Williams et al., 2018), and no previous research has focused specifically on whether and how changes in scheduling practices might improve *both* profitability and employee experience (Osterman, 2018).

Common managerial practices that treat employees as a cost to be minimized suggest that workers and firms are increasingly at odds. On the other hand, many suggest that the two could be aligned more effectively. Even in low-cost services, work can be complex and might benefit significantly from the enhanced customer service and higher productivity enabled by more motivated and capable employees (Fisher et al., 2019; Ton, 2014). Job quality affects an employee's motivation, ability to focus, and participation in improving organizational processes, thus affecting productivity, service quality, and customer satisfaction (Ton, 2009, 2012). Moreover, unstable schedules demotivate workers, increase turnover, put employees and management in a mindset of conflict rather than collaboration, and increase the uncertainty associated with absenteeism in the work environment.

Proponents of aligning firm profitability with employee outcomes argue that common schedule design practices ignore the above relationships between the schedule quality and performance, productivity, or turnover, implicitly assuming that organizational labor costs are only linearly associated with total employee hours and that worker productivity is constant. Consequently, scheduling models optimize the allocation of hours to ensure the minimum number of necessary employees required to meet forecasted customer traffic (Van Den Bergh et al., 2013). While selected research has started to quantify some of the costs of unstable schedules (Hashemian et al., 2020; Williams et al., 2018), it remains unknown if firm-level profit-maximizing scheduling choices would differ from the current practices if those costs were accounted for in firms' scheduling. Specifically, one may empirically quantify the costs associated with unstable schedules, but even if those costs are significant, that does not imply that improving scheduling practices is feasible. Any change in scheduling practices intended to improve employee outcomes may also impose other costs to the firm that should be considered against the potential benefits. It is only with modeling the impact of changes in scheduling practices on various costs and benefits that one can answer the question of concern: Are there changes in existing scheduling practices that could improve financial performance and employee outcomes?

In this paper, we tackle this question directly. We build a behavioral simulation model of stores with boundedly rational agents (i.e., employees and managers) to explore the impact of alternative scheduling strategies on firm performance. Few organizations practice counterfactual scheduling policies to inform alternative designs and their value, making it difficult to use traditional empirical analyses (see Williams et al., 2020 for a notable experiment). Moreover, even if different practices are found in one firm, it is not clear whether those practices may transfer to other companies. Therefore, a research design is needed that not only informs the relationship between productivity and improved schedules but also addresses how the changes in scheduling practices may impact overall firm outcomes.

Without a counterfactual analysis, the ceteris paribus assumption in reduced-form studies may lead to misleading policy suggestions. Changing a single aspect of store operations might alter its other features and negatively affect profits, limiting the applicability of suggested policies. In the same vein, alternative policies may not be feasible if other systemically connected aspects of store operations cannot be changed. Given the lack of prior work on this question, we focus here on creating a simulation model based on the physics of scheduling systems. We estimate the model using data from a large retail chain. This detailed model informs the necessary counterfactuals that enable extrapolation from existing data to what a given service chain could do and would experience under alternative scheduling practices.

Overall, the model allows us to quantify and integrate a few important tradeoffs impacted by scheduling heuristics: How well are employees matched to customer demand? How experienced and productive are the employees? How does turnover affect store productivity? Are there feasible scheduling practices that can simultaneously improve store performance and stability for employees? Empirically estimating these relationships and integrating them into a single model of a service chain is at the heart of our analysis.

We use 52 weeks of data from over 1,000 stores and more than 15,000 hourly employees of a national specialty retailer, NRC (a pseudonym). We estimate the model using novel methods adopted from the econometrics of structural models. We then conduct sensitivity analyses using the scheduling policy parameters to assess whether promising alternatives may exist. We are particularly interested in the parameters that managers use to set schedules. We use our estimated model to explore how changing the scheduling strategies (or allocation decisions) affects stores' sales and labor costs. We also explore the extent to which stores can offer more stable schedules while keeping the labor percentage of sales down.

Our results suggest that common scheduling practices, under certain conditions, may have negative, direct labor cost consequences despite their intended rationale of aligning service capacity and demand. While cost-savings based on matching the supply and demand of employee hours are generally significant, there are other possible ways to considerably improve the stability of employee schedules while improving a store's financial performance in terms of labor percentage of sales. Our findings suggest that giving higher scheduling priority to more productive employees (i.e., giving them more hours, more consistently) increases store productivity (up to 10%). We find that this scheduling strategy may also reduce labor costs by up to 3% because it creates a reduction in total employee hours needed at the store. Finally, we find that reducing both the stochasticity in schedules and the late schedule updates during the week can increase the labor percentage of sales by up to 37%, harming store performance.

2. A Behavioral Model of Scheduling

We create a dynamic behavioral feedback model of work scheduling and its downstream implications in terms of costs and revenues. In our model, managers stochastically allocate hours to employees using heuristics that are informed by actual scheduling practices and data from an employee's past performance, status, and demand for hours. Employees then sell products based on customer traffic, time of year, skill level, and other events treated as random variations. Moreover, employees may decide to leave the company, partially due to scheduling concerns. More employees will then be hired to keep staffing at the required level.

The aim of scheduling procedures at NRC, like most other retailers, is to match employee hours to customer demand, based mostly on (forecasted) customer traffic to stores and other non-customer-facing work. At NRC, schedules are often announced two weeks in advance. These schedules are created based on a corporate-level algorithm that has employee availability and customer traffic forecast as inputs and the first version of a schedule as an output. Store managers then change the schedule based on their preferences, often adjusting for information unobservable at the corporate level, then announce the resulting schedule to employees. However, actual customer traffic can vary considerably from the forecast on an hourly or daily basis. Managers use two levers to adjust employee hours to match actual traffic. First, they schedule some employees as "on-call" to have the freedom to bring more employees to the store with short notice. Second, they send some employees home if not needed, or they update the schedules for the rest of that week. Such scheduling strategies and processes further exacerbate the unpredictability, inconsistency, and inadequacy of hours for employees (Chuang et al., 2016; Perdikaki et al., 2012).

In the next parts of this section, we describe our scheduling model in detail. Our model of the scheduling system includes (1) determining the demand for total employee hours based on both the actual and forecasted customer traffic data, (2) allocating hours to individual employees based on the demand for total employee hours, and (3) determining output variables (i.e., product and turnover). Figure 1 provides an overview of the scheduling system. In this figure, we use light blue ovals to denote the use of empirical data, orange rectangles for store-level variables, and blue rectangles for individual-level variables. We use diamonds to denote hiring and turnover, emphasizing store personnel changes.

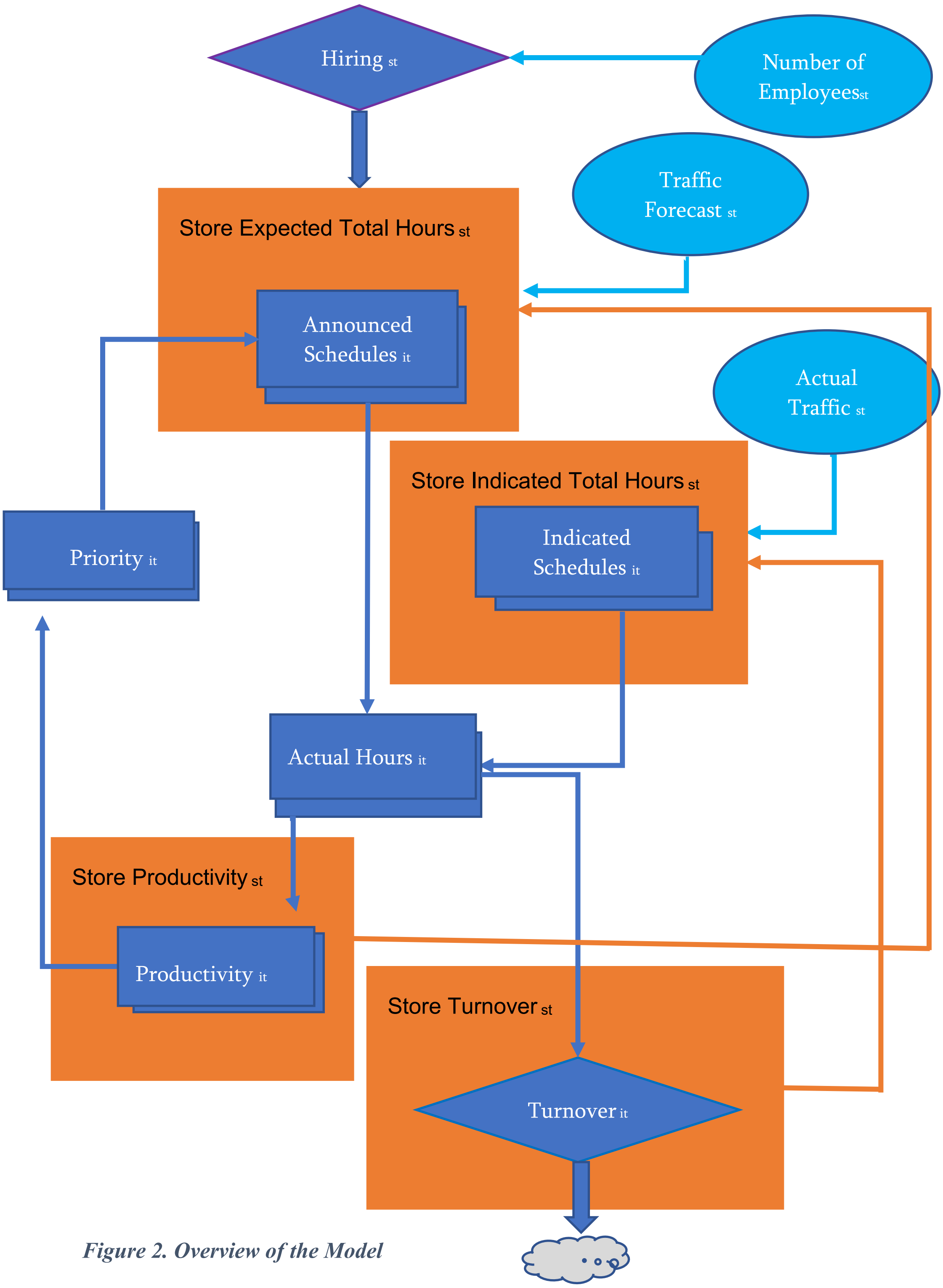


Figure 2. Overview of the Model

2.1 Determining the Demand for Employee Hours

Store managers, with the help of the corporate office, determine the formal, announced employee schedules by first calculating the *expected total employee hours* at the store level for each week. The expected total of employee hours is a function of forecasted customer traffic data and how many employees are needed per customer. As we explained above, actual employee hours often vary from the announced schedules because of the difference between actual and forecasted customer traffic. To account for this variation, we calculate the counterfactual *indicated total employee hours* using the actual traffic data.

Expected and indicated total employee hours are calculated using Equations 1, 2, and 3. The script s in these equations identifies each store, and script t reflects the specific week of the year. Since at least one employee needs to be present per each open hour, in both cases, we have the minimum total weekly hours of 77. Store managers then add more employee hours as customer traffic increases. The relationship between the number of customers and employee hours is nonlinear since the value of subsequently added employees diminishes as customer traffic grows. We hypothesize that managers determine the employees needed per customer based on two frequently overlooked factors. First, managers may marginally increase the total number of hours needed if, based on store productivity, they believe more employees are needed to help the customers. For example, if an employee is slow at the register, there might be a need to open another one. Similarly, if an employee takes longer to help a customer, more employees are needed to help other customers. Therefore, we compare the average of the previous two weeks of store productivity to the store's overall average to adjust the schedule. Second, managers marginally increase the total number of hours needed as the uncertainty about employee attendance grows. In this industry, it is quite common for an employee to leave a company with no advance notice. It is also likely for workers to be late or absent, especially given the great work–life conflicts they face. We do not have absenteeism data and, so, use employee turnover as a proxy. Explicitly, we use the average of the previous two weeks' turnover, compare it to the store average, and adjust the schedule accordingly in Equation 3. We estimate parameters β_1, β_2, and β_3.

(1) $$Expected\ Total\ Employee\ Hours_{st} = 77 + \ Traffic\ Forecast_{st}{}^{B_{st}}$$

(2) $$Indicated\ Total\ Employee\ Hours_{st} = 77 + \ Actual\ Traffic\ _{st}{}^{B_{st}}$$

(3) $$B_{st} = \beta_1 + \beta_2 * \frac{\frac{1}{2} * \sum_{(t-2)}^{(t-1)} Store\ Productivity_{s\tau}}{Average\ Store\ Productivity_s} + \beta_3 * \frac{\frac{1}{2} * \sum_{(t-2)}^{(t-1)} Turnover_{s\tau}}{Average\ Turnover_s}$$

2.2 Allocating Hours to Employees

To determine *announced schedules,* we allocate the expected total employee hours at the store level to individual employees. In the same fashion, we determine counterfactual individual-level *indicated schedules* using the indicated total employee hours. We calculate the *actual hours* by taking a weighted average between the announced and indicated schedules at the individual level to reflect the extent to which managers can and are willing to update employee hours during the week. The process of calculating announced schedules reflects the effort to implement an optimal schedule. The process of determining actual hours based on indicated schedules reflects the managers' efforts to maintain the optimality of the schedules in real-time, given the changing circumstances.

The allocation of total hours to individual employees is based on an individual worker's *Priority*, as assigned by that individual's manager. Priority is a dynamic, (partially) random variable that determines a worker's place in the queue for receiving hours. Hours are allocated first to those with the highest priority. Therefore, lower priority individuals receive hours only if there is still a need for additional hours after higher priority individuals receive theirs. If the priority differences are not significant, employees will receive a similar share from the total employee hours.

We first create a weekly matrix for every store that reflects the employee hour supply for each priority level. We then intersect the manager's demand for employee hours based on customer traffic (i.e., total hours needed) with the supply of available employee hours to determine a store-specific priority level that "clears" the market for hours (i.e., equates the employee hour supply with the manager's demand for hours). The number of hours each person receives is based on their position in the priority matrix and their distance from the designated priority. Those who stay below the designated priority do not receive any hours, and those who are much higher receive a full 40 hours. A hypothetical store with three employees is presented in Figure 3. Employee 1 has a priority of two, meaning she receives hours only if the designated priority is lower than two. If the designated priority were at zero, Employee 1 would get a full 40 hours. If the designated priority were to increase, her number of hours would decline linearly. For example, a designated priority of one means that Employee 1 can only get 20 hours. Based on this rule, we could calculate the matrix for the employee hour supply at each priority level. In our hypothetical store, Employee 2 has a priority of three, and Employee 3 has a priority of four. Therefore, the maximum supply is set at 120 hours at a priority of zero. At a priority of two, the supply of hours is 60, out of which 40 hours go to Employee 3, and 20 hours to Employee 2.

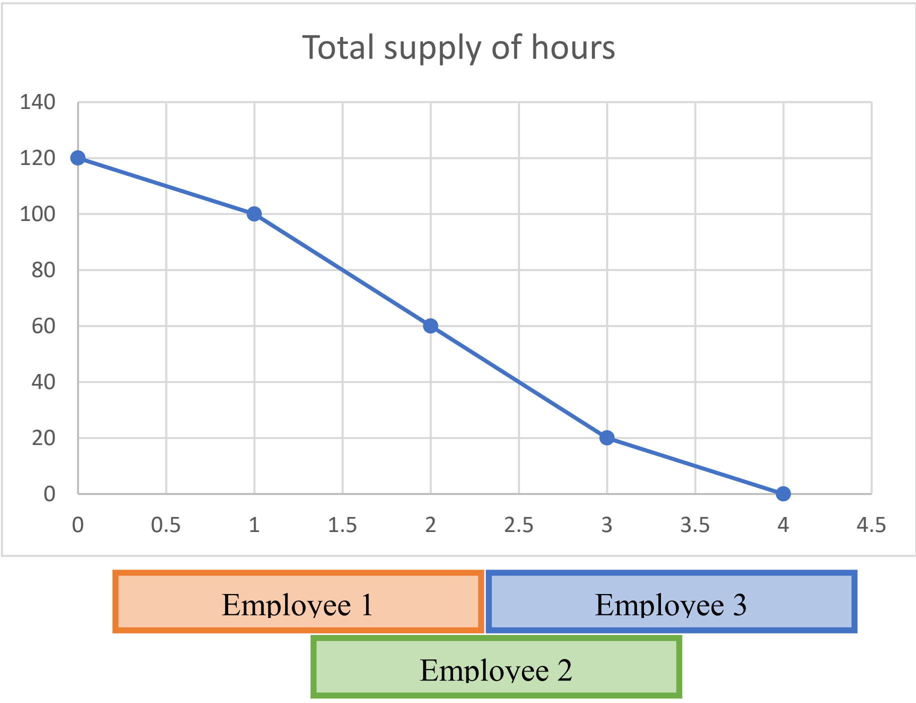


Figure 4. Hypothetical Priority Structure for a Store's Employees

Priority is a random variable with a standard deviation of (σ_1) that initially correlates with an individual's inherent *Capability* to sell products (defined below) and then changes with that individual's *Productivity* (also defined below). The distribution of priority (σ_1) and the degree to which it updates based on productivity (ϑ) are parameters that are estimated. As the variance in the distribution of priority (σ_1) increases, individuals are treated differently, and those with higher priority receive more hours (see Equation 4). As the correlation between priority and productivity (ϑ) increases, more capable employees receive higher priority. Therefore, using σ_1 and ϑ, we can generate a spectrum of allocation strategies. At one end of the spectrum, all employees are treated equally and receive equal hours, regardless of their capability or productivity. At the other end, managers are good at identifying more capable, highly productive employees and consistently give them more hours.

$$(4)\quad Priority_{ti} = \mathcal{N}\left(\mu, \sigma_1^2\right) + \int Priority_{(t-1)i} * \left(\frac{Productivity_{(t-1)i}}{Average\ Productivity} - 1\right) * \vartheta$$

When determining the announced schedules, managers make use of customer traffic forecasts, employee availability, and requests. However, allocating hours to reach the indicated number of total hours for a particular week is more challenging, given that the manager has to work both with (changing) employee availability and unpredictable circumstances, such as logistical changes and variability in customer traffic. Therefore, we add a random variable to each individual's priority, $X_{ti} = \mathcal{N}(0, \sigma_2^2)$, which affects scheduling when distributing the indicated total number of hours. We estimate the standard deviation of this random variable (σ_2). This deviation causes few changes to those employees with the highest and lowest priorities; however, those closer to the designated priority potentially receive significantly more or fewer hours as a result of this random component. Actual simulated hours are a weighted average of the announced schedules and indicated schedules (see Equations 5, 6, and 7). We estimate ω, which represents the degree to which managers can or are willing to change employee schedules.

(5) $$Scheduled\ Hours_{ti} = f_A(Priority_{ti}, Expected\ Total\ Employee\ Hours_{st})$$

(6) $$Indicated\ Hours_{ti} = f_A(Priority_{ti}, X_{ti}, Indicated\ Total\ Employee\ Hours_{st})$$

(7) $$Actual\ Hours_{ti} = Required\ Hours_{ti} * \omega + (1 - \omega) * Scheduled\ Hours_{ti}$$

2.3 Determining Output Variables

In this section, we explain how our main outputs, employee productivity, store productivity, turnover, and labor cost, are calculated

Employees who receive nonzero hours interact with customers and sell products. *Productivity* is defined as sales (in dollars) per employee hour. We determine employee productivity as a linear function of employee work hours, hour *Consistency*, *Predictability*, employee tenure, wages, and inherent *Capability* as well as store-level variables that we extract from the data. The store-level data affecting productivity are customer traffic, the store's sales plan, and the number of products in the store. We use a linear function because it allows us to use panel data regression coefficients to determine the parameters of the model (see Equation 8). Hour consistency is calculated by looking at the absolute difference between an employee's hours in the current week and their average from the prior four weeks (see Equation 9). Predictability is determined by calculating the absolute difference between an individual's announced schedule and actual hours. Inherent capability is an individual fixed effect (ς_i) in Equation 8. After calculating individual productivity, we add up all of the sales to determine the total store sales. In our data and, thus, similarly in the model, some sale transactions are not attributed to any employee. For example, when a store is busy, many customers do not receive personal help from any individual employee and, thus, these sales are not attributed to any single individual. Therefore, to calculate the store-level sales or productivity, we need to keep in mind that some percentage of the sales is not calculated at the individual level. Therefore, we derive the ratio of total sales over attributed sales for each week (θ) from the data and use this ratio to calculate total store sales (see Equation 10).

(8) $$Productivity_{it} = a + \kappa_1 Actual\ Hours_{it+}\ \kappa_2 Consistency_i + \kappa_3 Predictability_{it} + \kappa_3\ Tenure + \sum_k \kappa_k C_{kt} + \varsigma_i + \tau_t + \varepsilon_{it}$$

(9) $$Hour\ Consistency_{it} = -\left| Actual\ Hours_{it} - \frac{\sum_{t-4}^{t-1} Actual\ hours_{it}}{4} \right|$$

(10) $$Store\ Sales_{st} = \theta_t * \sum_i Productivity_{it} *\ Actual\ Hours_{it}$$

Similar to the process of determining the productivity formula, we use a logit model, presented in Equation 11, to calculate individual turnover and then be able to estimate the parameters using logit regression. We then stochastically remove employees from their stores based on the probability calculated from Equation 11. The turnover count at the store level is the sum of individuals who left the store that week. A store manager hires new employees if needed to match the number of employees to the desired number (i.e., the number of employees in the dataset for any given week).

(11) $logit\ (Turnover_{it}) = a + \kappa_4 Actual\ Hours_{it+}\ \kappa_5 H\ Consistency_i + \kappa_6 Predictability_{it} + \kappa_7\ Tenure + \sum_k \kappa_k C_{kt} + \varsigma_i + \tau_t + \varepsilon_{it}$

3. Model Estimation

We estimate our model using 52 weeks of data for more than 15,000 employees working in more than 1,000 NRC stores in the U.S. and Puerto Rico in 2015. Our data provides a unique opportunity for policy analyses because we have access to three key elements of the scheduling process. First, we have weekly traffic forecasts, actual traffic, and sales plan data. This information provides critical inputs in scheduling algorithms or important schedule determinants. Second, we have individual-level data on announced schedules and actual employee hours. Third, our data includes productivity, sales, and labor costs, which allow in-depth analyses of the impact of alternative scheduling strategies.

To estimate our model, we use a novel approach based on the econometrics of structural models (Hansen, 1982; Imbens, 2002). The economics literature on structural models is particularly helpful for two related reasons. First, structural models are based on theoretical models, with nuances that explore causal mechanisms beyond single, reduced-form relationships (Heckman, 2000). In this view, instead of merely having control variables, different variables play roles in an explicit model (Wolpin, 2013). Second, this theory-based feature of structural models allows for creating counterfactuals that are critical for extrapolating from the model and conducting policy analyses (Heckman, 2000; Nevo & Whinston, 2010). In other words, with the help of a model grounded in the literature and available empirical data, researchers can more carefully and explicitly distinguish between policy variant and invariant parameters (Heckman, 2000). This differentiation helps clarify the extent to which the model has external validity and in which situations the policy suggestions would be more helpful.

Following the tradition of the econometrics of structural models, we use the generalized method of moments (GMM) to estimate our parameters (Imbens, 2002). GMM is appropriate for estimating the parameters of a theory-based model because it provides the flexibility to use known functions and observed random variables to extract the value of unknown parameters with few assumptions, particularly without needing assumptions of normality or independence (Hansen, 1982; Imbens, 2002).

We use GMM to estimate model parameters related to scheduling allocation. These include seven parameters, three of which are structural, determining the total hours needed at the store level (i.e., $\beta_1, \beta_2, \beta_3$), and four of which are policy parameters, determining the allocation of total hours to individual employees based on details of priority function explained above (i.e., $\sigma_1, \vartheta, \sigma_2, \omega$). See Table 1 for a summary of our parameters. We follow Hansen's (1982) two-step process to estimate the scheduling parameters in our simulation model by minimizing the distance between simulated and empirical moments. The goal of this two-step process is to first calculate the weight matrix or the covariance matrix among moments for our second and final estimation, based on parameters extracted from the initial model optimization.

Table 1. List of Model Parameters

Parameter	Explanation	Type	Estimation
β_1	The constant that determines the total number of hours from customer traffic	Structural	Structural (GMM)
β_2	Determines the relationship between past productivity and the total number of hours needed	Structural	Structural (GMM)
β_3	Determines the relationship between turnover and the total number of hours needed	Structural	Structural (GMM)
σ_1	Determines the heterogeneity in employee priority	Policy	Structural (GMM)
ϑ	Determines the degree to which store managers update employee priority based on their productivity	Policy	Structural (GMM)
σ_2	Determines the uncertainty in employee priority	Policy	Structural (GMM)
ω	Determines the degree to which store managers update employee schedules	Policy	Structural (GMM)
$\kappa_1 \ldots . \kappa_n$	Determine productivity and turnover based on schedules	Structural	First step (OLS)

To create GMM error function to be minimized, we select 14 moments that represent both the scheduling process and the relevant organizational outcomes. Importantly, we include moments related to store-level productivity and labor costs to be able to analyze the impact of various policies on these outcomes. These 14 moments include the means and standard deviations of the variables specified in Table 2. Having both the means and the standard deviations ensures that our model reproduces results with distributions similar to the empirical data.

Table 2. Empirical and Simulated Moments

No.	Moment		Empirical Value	Simulated Value
1	Store productivity	Mean	173.43	162.9
2		S.D.	63.25	81.65
3	Total number of hours (actual)	Mean	281	297.01
4		S.D.	151	172.86
5	Turnover	Mean	0.56	0.48
6		S.D.	1.28	1.41
7	Employee productivity	Mean	79.41	60.19
8		S.D.	98.49	89.78
9	Employee hours	Mean	16.28	21.34
10		S.D.	11.39	9.53
11	Employee schedule consistency	Mean	−5.63	−8.32
12		S.D.	4.99	5.06
13	Employee schedule predictability	Mean	−3.50	−5.34
14		S.D.	4.44	4.53

Table 2 reports the empirical and simulated moments produced as a result of the two-step GMM process explained above. We present the parameter estimates in Table 3; all parameters are statistically significant. The effect of store turnover from the past two weeks and productivity on total employee hours needed is particularly interesting. Increasing store productivity by 10% can decrease the total hours needed by 18% on average. Similarly, decreasing store turnover by 10% can decrease the total hours needed by 33%. Results suggest that NRC stores update their schedules often (parameter $\omega = 0.492$) and have medium-level uncertainty in the scheduling process (parameter σ_2 = 3.715). Their scheduling strategies are not based on identifying high-capability, high-priority individuals and giving them more and better hours (parameter σ_1 = 1.815).

Table 3. GMM Parameter Estimation

Parameter	Estimate	Confidence Interval	
β_1	0.669	0.665	0.673
β_2	−0.241	−0. 242	−0.240
β_3	0.596	0.595	0.598
σ_1	1.815	1.806	1.824
ϑ	0.104	0.104	0.104
σ_2	3.715	3.695	3.734
ω	0.492	0.490	0.493

Moreover, as explained above, we use reduced-form models in Equations 8 and 11 to determine our output variables (i.e., productivity and turnover) and use panel data fixed-effect regressions to estimate the corresponding parameters (i.e., $\kappa_1 - \kappa_n$). We report the results in Table 4. Results in Model 1 (corresponding to Equation 8) suggest that schedule adequacy, consistency, and predictability all have a significant positive effect on employee productivity. Improving all the three dimensions of scheduling stability by one standard deviation increases employee productivity by 12%. Similarly, in Model 2 (corresponding to Equation 11), we find significant results suggesting that improving schedule adequacy, consistency, and predictability by one standard deviation reduces the probability of turnover by 57%.

Table 4. Impact of Unstable Schedules on Employee Productivity and Turnover

	Model 1	Model 2
	Employee Productivity	Employee Turnover
Hours	0.71***	−0.05***
	(0.01)	(0.00)
Schedule Consistency	0.14***	−0.02***
	(0.02)	(0.00)
Schedule Predictability	0.10***	−0.04***
	(0.03)	(0.00)
Controls	YES	YES
TimeFE	YES	YES
EmployeeFE	YES	YES
R2	0.18	

4. Policy Analysis

Holding our structural parameters constant at estimated values, we conduct policy analyses by changing our four policy levers from Table 1. First, we analyze the effects of σ_1 and ϑ on model behavior. Higher σ_1 means that store managers differentiate between employees more aggressively, consistently giving those with higher priority (often the more productive and experienced employees) more hours. Lower σ_1 means that store managers give similar hours to more employees, resulting in fewer hours per employee. As mentioned in the previous section, employees who have worked at the store from the beginning have priorities that correlate with their inherent capabilities or skill levels. However, those who are hired later have an average priority initially uncorrelated with capability since the store managers do not have any indications about their capability levels. Higher ϑ means that priorities are updated more aggressively with employee productivity, resulting in higher correlations between priority, allocated hours, and capability. See Figures 3a and 3b for the impact of the policy levers σ_1 and ϑ on schedules and store turnover, respectively.

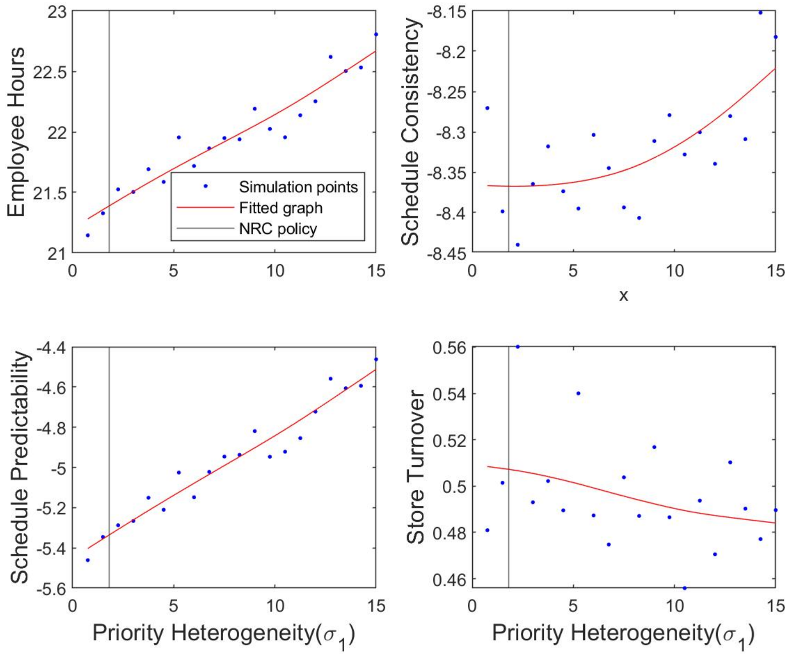


Figure 5.a. Impact of Priority Heterogeneity on Employee Schedules and Store Turnover

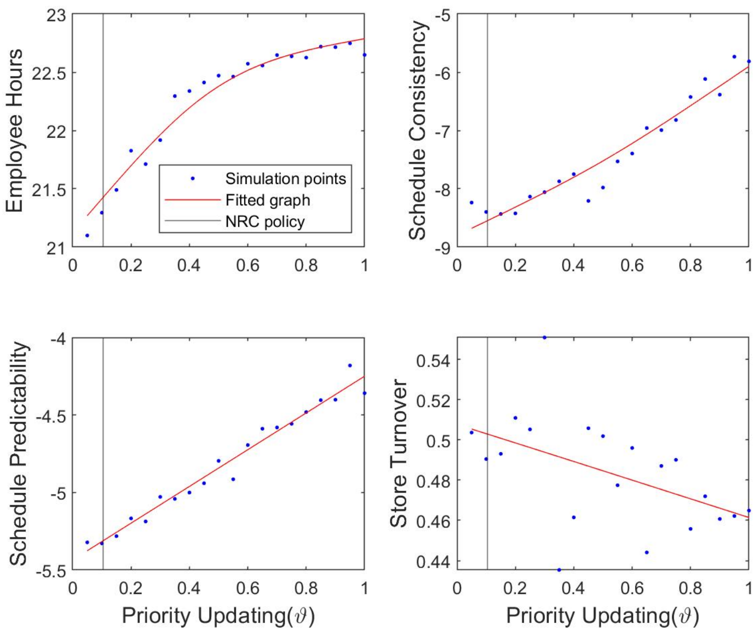


Figure 3.b. Impact of Priority Updating on Employee Schedules and Store Turnover

We expect to see an increase in store performance as σ_1 and ϑ increase for two reasons. First, in our model, those with higher productivity often have higher wages. Second, increasing the average employee hours correlates with total store hours (unless hours are distributed among fewer employees) and, thus, might be costly for stores. However, based on our model, we expect to see a decrease in labor costs and, more importantly, a decrease in labor percentage of sales. More productive stores require fewer employee hours in total,

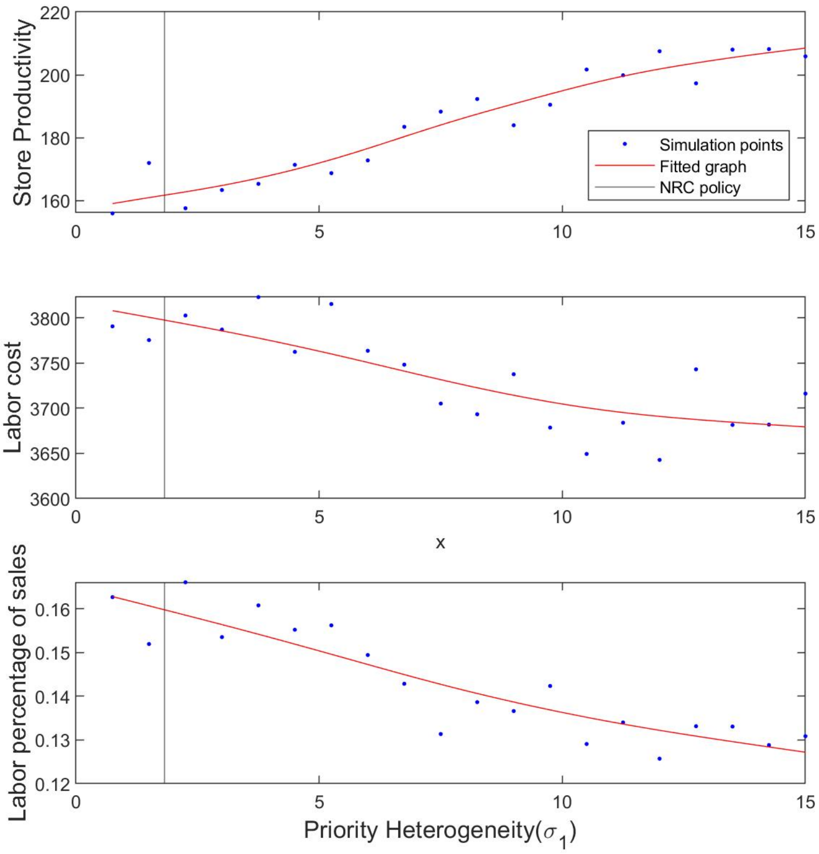


Figure 6.a. The Impact of Increasing Priority Heterogeneity on Store Financial Performance

and employees with better schedules are less likely to leave the store, providing lower turnover levels and a decrease in the total employee hours needed per customer traffic. Also, high turnover levels decrease the average employee experience, reducing productivity and introducing more employees with lower priority who, in turn, work few, inconsistent, and unpredictable hours and might leave the store quickly. The results of the policy analyses are presented in Figure 4. Moving from NRC's current policy to the maximum σ_1 and ϑ, we observe up to a 26% and 7% increase in productivity, respectively. Moreover, we observe 3% and 7% decreases in labor costs as we increase each dimension of priority differentiation from NRC policy to maximum values. Moving the values to their maximum may not be feasible, but the qualitative results remain consistent with smaller changes as well.

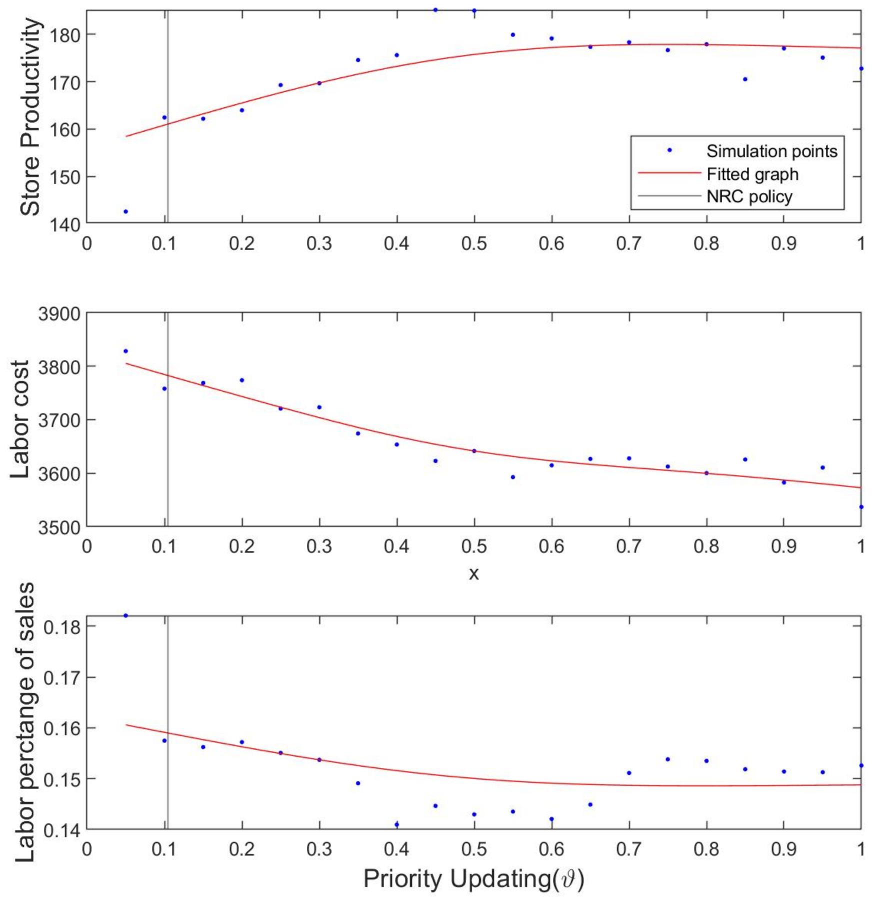


Figure 4.b. The Impact of Increasing Priority Updating on Store Financial

Next, we study the effects of various levels of σ_2 and ω on productivity and labor costs. Parameter σ_2, or schedule stochasticity, represents how much the distribution of hours during the week is affected by random events and unforeseen circumstances. The higher the σ_2, the more the store manager deviates from the original priority distribution in allocating hours. While this parameter may not be fully under manager control, the manager can certainly change it by making the store operation smoother and ensuring that the distribution of hours is based on higher priority and on the philosophy of giving more consistent hours to employees. Similarly, higher ω means the store manager updates the employee schedules more aggressively toward the required hours and away from announced schedules.

An increase in both σ_2 and ω ensures a better match between the supply and demand of employee hours and, thus, on average, should increase productivity and reduce labor costs. However, as shown in Figure 5, in most cases, increases in σ_2 and ω also result in less consistent, less predictable schedules and higher levels of turnover. The effect on employee hours is, however, mixed.

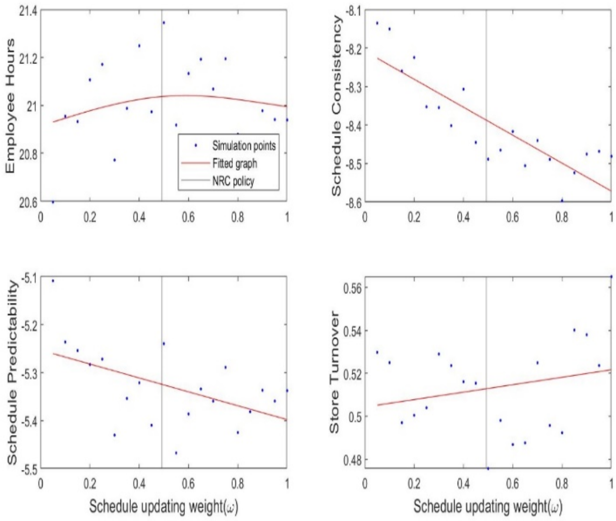


Figure 7.a. Impact of Increasing Schedule Updating Weight on Employee Schedules and Store Turnover

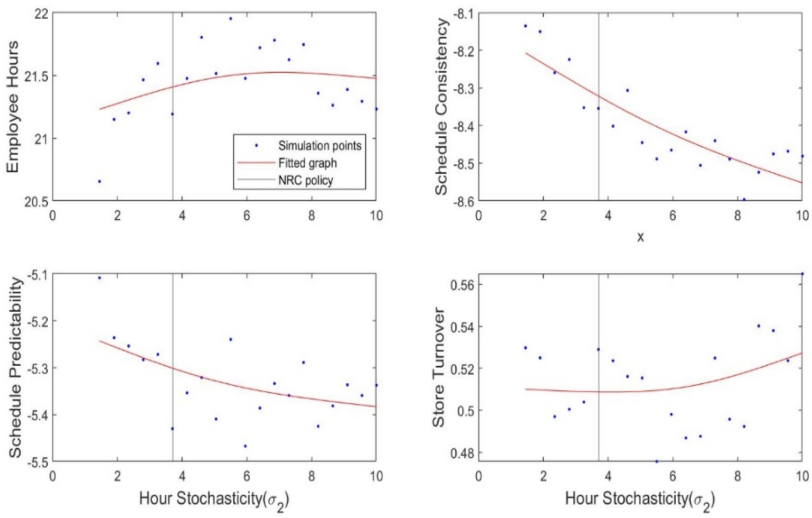


Figure 5.b. Impact of Increasing Hour Stochasticity on Employee Schedules and Store Turnover

In contrast to the traditional perspective on scheduling, we see a decrease in productivity as σ_2 and ω increase. This is partly because of the negative impact that inadequate, inconsistent, and unpredictable schedules have on employee productivity. Moreover, higher levels of σ_2 and ω reduce the correlation between priority and employee capability and hours received. We see a moderate increase in labor costs with higher ω. We observe a positive correlation between deviating from initially announced, priority-based schedules and labor percentage of sales. Similar to our previous analyses on the first two policy levers, we see that the negative productivity and turnover consequences of unstable schedules on the total hours needed might be larger than the benefits of optimized allocation. This is, of course, considering the current structural parameters and policies of NRC. The results are presented in Figure 6. Not updating hours could increase NRC productivity by 22%, decrease labor cost by 7%, and improve labor percentage of sales by 33%. Finally, removing schedule stochasticity could increase NRC productivity by 2.5% and decrease the labor percentage of sales by 4%.

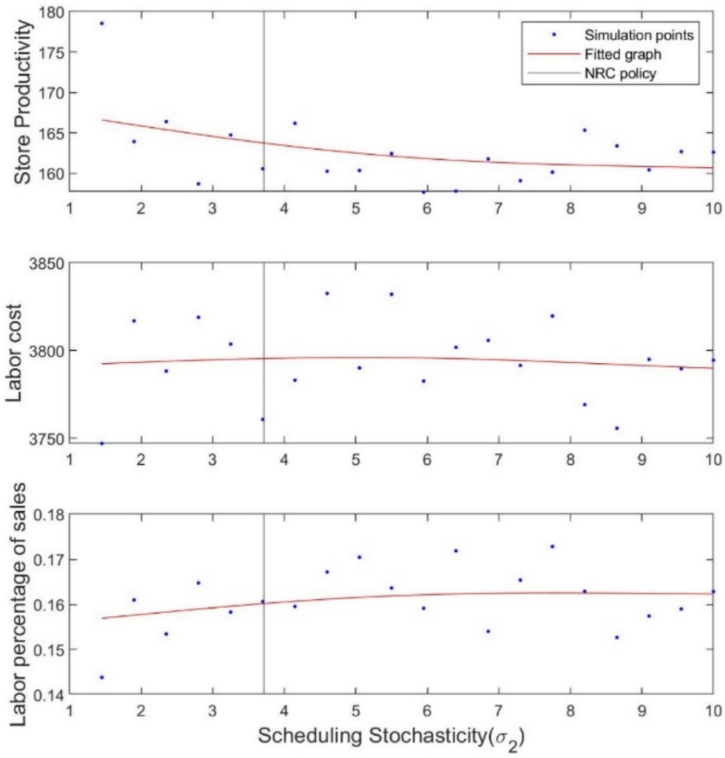


Figure 6.a. Impact of Increasing Schedule Updating Weight on Store Financial Performance.

In summary, we find that NRC is operating far from optimum in its scheduling strategy. Based on our results, NRC does not have a strategic focus on identifying and prioritizing employees with better skills and giving them better hours and seems not to have an effective system for doing so. Moreover, NRC store managers work in a system with a high level of scheduling stochasticity and limited focus on selecting high performers. Based on the results on schedule stochasticity and updating weight, it seems that NRC stores merely try to match labor supply and demand without much attention to productivity and quality of employee schedules and the downstream impacts on overall profitability.

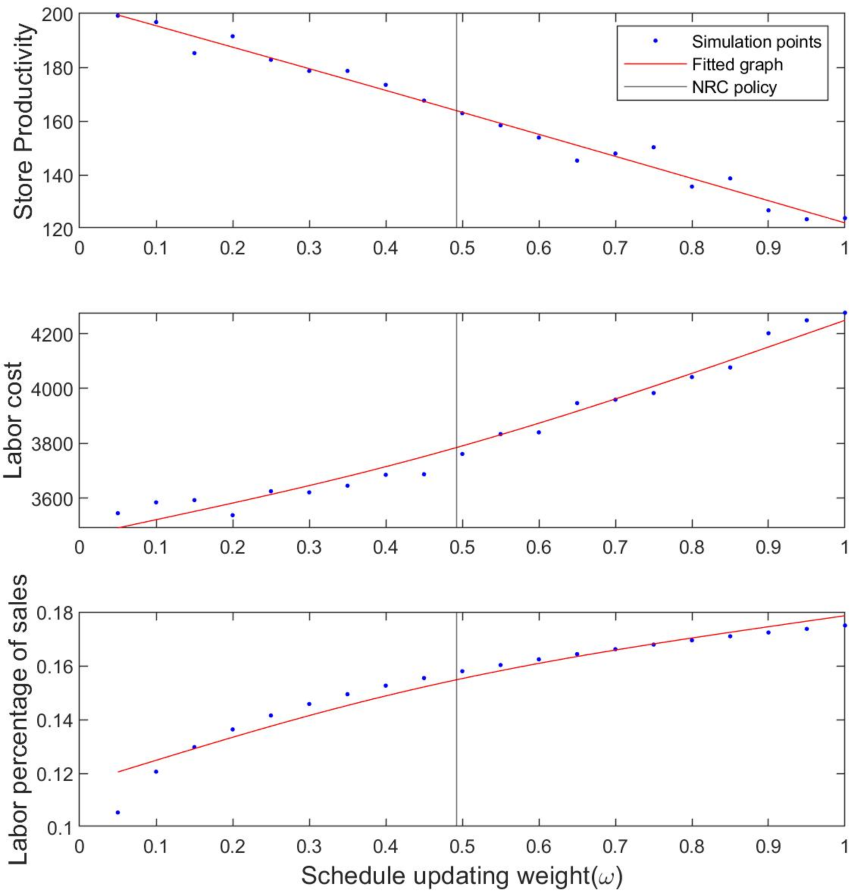


Figure 6.b. Impact of Increasing Scheduling Stochasticity on Store Financial Performance.

To put NRC's scheduling policy in perspective, we compare the financial performance of an average store in the base case and under four alternative strategies. In Strategy 1, we assume medium priority heterogeneity (from $\sigma_1 = 1.81$ to $\sigma_1 = 5$) and a medium priority updating (from $\vartheta = 0.1$ to $\vartheta = 0.5$), and we keep other parameters constant at their estimated values. This strategy represents an increased focus on identifying more skilled employees and giving them more and better hours. In Strategy 2, we increase the priority updating ($\vartheta = 0.9$), keep the priority heterogeneity medium, and do not change the scheduling stochasticity. This strategy represents an improvement in identifying the best talent. Changing the scheduling updating weight and scheduling stochasticity might be more difficult for the store managers; nevertheless, we consider variations in those factors in the next two strategies. In Strategy 3, we replicate the Strategy 2 and improve the scheduling updating weight (from $\omega = 0.49$ to $\omega = 0.3$), and in Strategy 4, we also improve scheduling stochasticity (from $\sigma_2 = 3.71$ to $\sigma_2 = 2$). Table 5 presents the results. Strategy 1 improves the labor percentage of sales but also reduces store productivity. The other three strategies all enhance productivity *and* reduce labor costs. The magnitude of impacts is significant, with store productivity increases that are in the 7% to 16% range while labor costs are reduced by 11% to 20%. We find that in Strategy 4, compared to NRC's current policy, the labor percentage of sales decreases by 20%. In short, our analysis suggests that NRC can potentially change scheduling practices in ways that both improve the quality of life for employees and enhance financial performance.

Table 5. The Impact of Scheduling Strategies on Financial Performance

	Description	Store Productivity	Labor Cost	Labor Percentage of Sales
Base	NRC scheduling strategy	162.9	3790.3	16.5%
Strategy 1	Increased focus on priority-based scheduling	175.55	2991.3	14.8%
Strategy 2	Great improvements in identifying talented employees	176.65	2895.2	14.4%
Strategy 3	Increases focus on priority + mild improvements in reducing stochasticity and last-minute changes	187.37	2869.8	13.8%
Strategy 4	The most aggressive improvement scenario	189.12	2770.5	13.3%

5. Conclusion

Many service organizations, particularly brick and mortar retailers and fast-food chains, use labor flexibility practices to ensure a better match between the supply and demand of employee hours. These practices focus mostly on cost reduction, limiting the need for building a more rewarding work environment on productivity grounds. The use of these practices suggests that workers and firms are seen as increasingly at odds. We build on the literature suggesting that better labor practices create considerable productivity benefits, and we contribute to the literature by creating a counterfactual simulation to study in an empirically and operationally rigorous manner the impact of labor flexibility practices on productivity and labor cost. Our simulations allow for policy analyses that explore the feasibility and desirability of better labor practices. We find that the lower turnover rates and better store productivity that stem from employee-friendly schedules reduce the total employee hours needed. Therefore, we find that a strategy in which managers identify highly skilled employees, consistently give them more hours, and reduce schedule changes and randomness provides not only significant productivity gains but also labor cost reductions and, ultimately, sees up to a 20% decrease in the labor percentage of costs. Future research can focus in more detail on finding the best strategies by optimizing the labor percentage of sales or profit.

The current study only quantifies a subset of considerations relevant to the scheduling tradeoffs. Several other mechanisms are likely at play and should be studied in future research. For example, more productive employees and better customer service most likely increase customer traffic at a store. Satisfied customers are more likely to come back. Stores that handle customers more efficiently are often less crowded and have lower wait times. Therefore, it is less likely for those stores to lose visitors. On the human resource management side, there might be other feedback loops in play. Better schedules not only reduce turnover but also decrease absenteeism, which results in less uncertainty at the store, lower total hours needed, and lower chances of employee burnout.

Moreover, high turnover levels create a consistent hiring cost for the store. Finding and training employees is costly, regardless of the level of training. Hiring can also take a considerable amount of time and energy from the managers. Finally, better labor practices might result in a pool of applicants with higher skills and commitment levels. Given that we do not account for any of these schedule instability costs, our results, that more stable scheduling practices could also enhance performance in this setting, should be seen as conservative.

Another method for future research is using indirect inference. For example, one can use turnover's relationship with store productivity and total employee hours as two moments in the estimation process and, thus, a basis for ensuring similar structural relationships in the model. Moreover, it is possible to dig deeper into employee productivity by distinguishing reported and actual productivity. This analysis can shed light on the importance of an accurate reporting system and how it might affect inferences by store managers and corporations about productivity gains.

Our study has multiple limitations. Most importantly, we allocate schedules on a weekly level and thus may not be able to capture the nuances and costs that surface on an hourly level. We opt for weekly level analyses because of the exponential computational cost of simulation at the hourly level. However, we believe this choice has a minimal impact on our results. The possible discrepancy occurs most if there are sudden, sharp increases in customer traffic. For example, a store manager might not be able to easily redistribute hours as desired if customer traffic triples just for two hours within a day. However, customer traffic during the week usually follows a pattern in which most of the traffic and sales happen on Fridays, Saturdays, and Sundays.

This creates a possibility of around 31 hours of work for an employee who only works on these days. Therefore, there is often enough flexibility for a manager to distribute hours as needed. Moreover, other logistics practices help smooth the workload during the week. For example, handling shipments during the week instead of on the weekends can reduce employee demand surges or peaks.

Moreover, there are some considerations for extrapolation from this study. NRC is a specialty retailer and is different from general merchandise retailers like Walmart, fast-food stores like McDonald's, and wholesale stores like Costco in a variety of ways. Most NRC stores are small and part of a bigger mall. NRC sells small items and has high product promotion rates. Although these differences exist, our main mechanisms are most likely at play in various kinds of retail stores. Many stores have implemented labor flexibility practices under various names, such as "smart schedules" or "just-in-time" scheduling. Many of these stores use promotions extensively, and many have high turnover levels, constantly losing employees and hiring new ones, with inadequate training and unproductive employees.

Our study focused on the role of scheduling in service organizations. Accounting for the systemic implications of scheduling creates a better platform to assess the feasibility and financial impact of alternative scheduling strategies. We find that identifying high performing employees and providing more stable schedules not only improves productivity but also systematically reduces labor costs. We hope that highlighting and measuring the systemic costs of scheduling practices can better align the incentives of the firms, employees, and policymakers, opening up new possibilities for major improvements.

References

Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics* (Vol. 4, Issue PART B). Elsevier Inc. https://doi.org/10.1016/S0169-7218(11)02410-5

Adhi, P., Magnus, K.-H., & Sanger, F. (2020). Future of retail operations: Winning in a digital era. In *McKinsey & Company* (Issue January).

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent." *Science*, *344*(6186), 843–851. https://doi.org/10.1126/science.1251868

Ben-Ishai, L., Matthews, H., & Levin-Epstein, J. (2014). Scrambling for Stability: The Challenges of Job Schedule Volatility and Child Care About Our Work. In *Washington, DC: Center for Law and Social Policy*. www.clasp.org

BLS. (2017). *National Occupational Employment and Wage Estimates*. https://www.bls.gov/oes/2017/may/oes_nat.htm

Cho, Y. (2018). The effects of nonstandard work schedules on workers' health: A mediating role of work-to-family conflict. *International Journal of Social Welfare*, *27*(1), 74–87. https://doi.org/10.1111/ijsw.12269

Chuang, H. H. C., Oliva, R., & Perdikaki, O. (2016). Traffic-Based Labor Planning in Retail Stores. *Production and Operations Management*, *25*(1), 96–113. https://doi.org/10.1111/poms.12403

Ehrenreich, B. (2001). *Nickel and dimed : on (not) getting by in America*. Henry Holt & Co.

Fisher, M., Gallino, S., & Netessine, S. (2019). Retailers Are Squandering Their Most Potent Weapons. *Harvard Business Review*, *97*(1), 72–79. https://scholar.google.com/scholar?as_q=&as_epq=&as_oq=&as_eq=&as_occt=any&as_sauthors=Fisher&as_publication=harvard+business+review&as_ylo=2018&as_yhi=2019&hl=en&as_sdt=0%2C22#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3AqOd_A-N2Y7UJ%3Ascholar.google.com%2F%26outp

Goyal, M., & Netessine, S. (2011). Volume Flexibility, Product Flexibility, or Both: The Role of Demand Correlation and Product Substitution. *Manufacturing & Service Operations Management*, *13*(2), 180–193. https://doi.org/10.1287/msom.1100.0311

Halpern-Meekin, S., Edin, K., Tach, L., & Sykes, J. (2015). *It's not like I'm poor: How working families make ends meet in a post-welfare world.* Univ of California Press.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators Author. *Journal of The Econometric Society*, *50*(4), 1029–1054.

Hashemian, M., Ton, Z., & Rahmandad, H. (2020). *Unproductively " Optimal " Schedules : The Effect of Unstable Schedules on Employee and Store Productivity*.

Heckman, J. (2000). Causal parameters and policy analyses in economics: A twentieth century retrospective. *The Quarterly Journal of Economics*, *February*, 45–97.

Henly, J., & Lambert, S. (2005). *Nonstandard Work and Child-Care Needs of Low-Income Parents.* https://psycnet.apa.org/record/2005-06809-030

Imbens, G. W. (2002). Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statistics*, *20*(4), 493–506. https://doi.org/10.1198/073500102288618630

Kelly, E. L., Moen, P., & Tranby, E. (2011). Changing workplaces to reduce work-family conflict: Schedule control in a white-collar organization. *American Sociological Review*, *76*(2), 265–290. https://doi.org/10.1177/0003122411400056

Kesavan, S., Staats, B. R., & Gilland, W. (2014). Volume Flexibility in Services: The Costs and Benefits of Flexible Labor Resources. *Management Science*, *60*(8), 1884–1906. https://doi.org/10.1287/mnsc.2013.1844

Lambert, S. J. (2008). Passing the buck: Labor flexibility practices that transfer risk onto hourly workers. *Human Relations*, *61*(9), 1203–1227. https://doi.org/10.1177/0018726708094910

Lambert, S. J. (2009). Making a difference for hourly employees.". In *Work-life policies* (pp. 169–195).

Lambert, S. J., Henly, J. R., & Kim, J. (2019). Precarious Work Schedules as a Source of Economic Insecurity and Institutional Distrust. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *5*(4), 218. https://doi.org/10.7758/rsf.2019.5.4.08

Läubli, D., Schlögl, G., & Silén, P. (2015). Smarter schedules, better budgets: How to improve store operations. In *McKinsey & Company*.

Nevo, A., & Whinston, M. D. (2010). Taking the dogma out of econometrics: Structural modeling and credible inference. *Journal of Economic Perspectives*, *24*(2), 69–82. https://doi.org/10.1257/jep.24.2.69

Oliva, R. (2001). Tradeoffs in responses to work pressure in the service industry. *California Management Review*, *43*(4), 26–43. https://doi.org/10.1109/EMR.2002.1022405

Osterman, P. (2018). In Search of the High Road: Meaning and Evidence. *ILR Review*, *71*(1), 3–34. https://doi.org/10.1177/0019793917738757

Perdikaki, O., Kesavan, S., & Swaminathan, J. M. (2012). Effect of Traffic on Sales and Conversion Rates of Retail Stores. *Manufacturing & Service Operations Management*, *14*(1), 145–162. https://doi.org/10.1109/AIEEPAS.1959.4500355

Schneider, D., & Harknett, K. (2018). *How Work Schedules Affect Health and Wellbeing : The Mediating Roles of Economic Insecurity and Work-Life Conflict*.

Ton, Z. (2009). The Effect of Labor on Profitability: The Role of Quality. *Harvard Business School*, 1–33. https://doi.org/10.2139/ssrn.1269523

Ton, Z. (2012). Why “good jobs” are good for retailers. *Harvard Business Review*, *90*(1–2).

Ton, Z. (2014). *The good jobs strategy: How the smartest companies invest in employees to lower costs and boost profits.* Houghton Mifflin Harcourt.

Van Den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., & De Boeck, L. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, *226*(3), 367–385. https://doi.org/10.1016/j.ejor.2012.11.029

Williams, J. C., Lambert, S. J., Kesavan, S., Fugiel, P. J., Ospina, L. A., Rapoport, E. D., Jarpe, M., Bellisle, D., Pendem, P., McCorkell, L., & Adler-Milstein, S. (2018). *Stable Scheduling Increases Productivity and Sales: The stable scheduling study*. http://worklifelaw.org/publications/Stable-Scheduling-Study-Report.pdf

Wolpin, K. I. (2013). *The limits of inference without theory*. MIT Press.

Essay 3: Startup Advantage: The Role of Endogenous Capability Building in Creating New Markets

Abstract: Startups play a major role in establishing many new markets. This is theoretically puzzling because existing firms have more resources and the relevant core and peripheral capabilities, which should provide them with an advantage when diversifying into new markets. Here, we explore how the strong link between startups' past performance and the resources available for their future capability building conditions their growth prospects. Using a simulation model, we show that this reinforcing loop leads to entrepreneurial financial markets rapidly focusing on more promising startups. Despite their initial resource disadvantage, the strength of this mechanism can allow startups to catch up, and over-take projects within incumbent firms. Using an online experiment, we test the key requirement for our mechanism, showing that the strength of the reinforcing loop is larger for start-ups than in-house projects when individuals play the role of a venture capitalist vs. a CEO.

Keywords:

Organizational capabilities, disruption, simulation modeling

1. Introduction

The birth of new markets has a significant impact on firms' survival, growth employment trends, and the economy in general. Those companies that lead in creating new markets can shape how the market is structured and perceived, and they can also benefit from various advantages that accrue for the market leader. Thus, both startups and existing firms compete to create new markets: the former to establish the foundation for a new and successful enterprise, and the latter to expand their boundaries and thrive in the face of competition in existing markets. One theoretically important and practically relevant question is whether startups or existing firms are better placed to successfully create a new market.

According to the resource-based view of strategy, existing firms are likely to have a significant advantage over newcomers because diversifying entrants[7] are well endowed, compared to startups. For example, existing firms have more access to vital resources, particularly in the crucial early stages, during which they can obtain greater human capital and higher financial resources. Moreover, these firms, in many cases, have experience in prior or neighboring technologies. For example, Klepper and Simons (2000) show that firms in the radio market had greater experience and thus had advantage over newcomers in the TV receiver manufacturing market. Based on their greater resources and prior experience, existing firms can be expected to develop relevant core and peripheral capabilities faster. For instance, while new technological capabilities should be developed in a new market, existing firms may leverage their established brand and well-functioning distribution channels to aid their nascent project. Existing firms also have a network of relevant customers and suppliers that can potentially provide a good basis for exploring and

[7] Since our focus is on creating new markets, we use the terms entrant, diversifying entrants and existing firms interchangeably.

spearheading the new market. To summarize, diversifying entrants have access to potentially relevant resources, customers, and capabilities, all of which are expected to give them a significant advantage against startup firms when creating a new market.

Thus, one may expect existing firms to dominate the launch of new markets. Yet, there is plenty of evidence that highlights the importance of startups. Companies such as Uber, Dropbox, and Airbnb, among others, are example of startups that have recently created large new markets, in which existing firms abound. Similarly, we analyzed the commercialization of 26 major innovations[8] from 1979 to 2009, identified Forbes Magazine in 2009. We found that, in at least 50% of new markets, startups first commercialized a notable product. Microprocessors (Intel), DNA sequencing (e.g., Illumina), online shopping (e.g., eBay and Amazon) and social networking (e.g., Myspace) are a few examples from our sample of new markets that were led by startups. Additionally, startups are often more successful at introducing existing innovations into new geographical areas (Neffke, Hartog, Boschma, & Henning, 2016). On the national scale, US economic data[9] shows that startups are a major driver of growth and new jobs. Startups are responsible for around 70% of gross job creation, but they also have very high failure rates (and corresponding job losses).

This raises the question of why startups succeed in establishing many new markets, despite the resource-based view suggesting they have little chance of succeeding. Existing firms have access to potentially relevant resources, customers, and capabilities, all of which are expected to give them a significant advantage over startup firms. Yet, existing firms'

---

[8] Forbes (2009). Top 30 innovations of the last 30 years. *Forbes*. http://www.forbes.com/2009/02/19/innovation-internet-health-entrepreneurs-technology_wharton.html

[9] OECD (2015). Young SMEs, growth, and job creation. *OECD*. http://www.oecd.org/sti/young-SME-growth-and-job-creation.pdf

track records suggest that those advantages do not always translate into successful market creation. Understanding the mechanisms that promote startups in competitive markets are thus central to understanding sources of innovation, structures of emerging markets, and the competitive dynamics around new technological opportunities.

Prior research presents two distinct sets of arguments regarding this question. One strand of literature builds on psychological research by suggesting that entrepreneurs are prone to overconfidence and an escalation of commitment (e.g., Cooper, Woo, & Dunkelberg, 1988; Dosi & Lovallo, 1997; McCarthy, Schoorman, & Cooper, 1993). Therefore, startups often enter new markets against the odds, which results in the vast majority exiting in failure. However, by chance, a few startups stumble upon effective new products ahead of existing firms and come to lead these new markets. In this view, sheer luck and the large number of startups explain their widespread success.

A second organizationally focused perspective highlights how cognitive frames limit existing firms (e.g., Kaplan & Henderson, 2005). Those frames, which have evolved through adaptive processes of capability building and routine formation (March & Simon, 1958; Nelson & Winter, 1982), guide information collection and processing by organizational decision-makers. For example, existing firms may only explore incremental improvements to existing platforms, thus missing platforms' more promising radical changes or underestimating the future value of new markets, given their need for large revenue streams that are not satisfied in early markets (Henderson, 1993; Utterback, 1996; Christensen, 2000). Therefore, when new opportunities arrive or potentially disruptive technologies emerge, existing firms are late in recognizing those developments, which offers startups the first-mover advantage.

It is likely that both these mechanisms are at work. Experimental evidence of overconfidence among entrepreneurs is strong (e.g., Busenitz & Barney, 1997; Koellinger, Minniti, & Schade, 2007; Palich & Bagby, 1995), while data from several markets supports the idea that biases and inertia slow down incumbents more than diversifying entrants (Henderson & Clark, 1990; Tripsas & Gavetti, 2000; Utterback, 1996). However, it is less clear if these mechanisms fully explain startup advantage in *new markets*. For example, explaining startups' success based on the abundance of entrepreneurial entry and overconfidence implicitly assumes that only a few existing firms, who have the relevant core or peripheral capabilities, compete for the new market. However, for every new opportunity, there are scores of firms with potentially relevant capabilities that could be leveraged in the new market. Similarly, in the absence of incumbents, theories of organizational inertia and constraining cognitive frames require diversifying entrants to miss new opportunities that mostly align with their existing businesses. However, empirical evidence suggests that many existing firms actually compete in new markets (Busenitz & Barney, 1997; Dunne, Klimek, & Roberts, 2005; King & Tucci, 2002; Klepper & Simons, 2000). This is consistent with theories that suggest diversifying entrants may not be bound by the same inertia that holds incumbents back, resulting in successful radical innovation in many cases (Sosa, 2013; Utterback, 1996). Therefore, we suspect that complementary mechanisms work in favor of startups, which counter the resource and capability advantages of existing firms in terms of shaping new markets.

In this study, we explore one such mechanism that relies on differential rates of learning with endogenous growth across startups and existing firms. We view each competitor in a new market as engaged in a process of searching within a complex landscape of technological and business-model configurations. We capture an important endogeneity in the search process: namely, that startups and diversifying entrants search more or less rapidly, depending on their access to resources, which in turn is a function of their past performance as it is perceived by resource holders. For example, venture capitals, as well as stock markets, reward promising startups with additional rounds of funding, and firms allocate more resources to more promising research and development projects.

Focusing on the same opportunity for a new market, we analyze the competition among projects within existing firms and startups. By focusing on this competition, we exclude the mechanisms of inertia and lack of entry, which relate to existing firms in new markets, and identify the mechanisms that matter when startups compete head-to-head with projects in existing firms. Internal projects are distinguished from startups using two features. First, following the resource-based literature, we allow projects in existing firms to have access to more resources than startups. This includes capabilities, network, talent, and financial resources that existing firms offer their internal projects with a discount or at no cost. Second, due to their organizational coupling and portfolio logic, existing firms follow a more egalitarian approach to allocating resources to internal projects, compared to how financial markets tightly couple startups' resources to their perceived promise. While the first feature represents the existing theory, we hypothesize that the second feature activates an unexplored mechanism that enables startups to benefit from a stronger reinforcing feedback loop for exploration, outcomes, and resources. The startups that, by chance, arrive at better configurations earlier are proportionally rewarded with more resources for further exploration and the refinement of their promising idea. Parallel projects inside an incumbent firm may initially receive more resources, but they receive a weaker boost in resources when they find a promising path. Therefore, promising startups can break out of competition faster and be the first to establish new markets. This mechanism is dynamically complex and rooted in differential learning rates, in the presence of endogenous resources, various technological opportunities, and the inherent uncertainty of learning and capability building. We therefore utilize simulation modeling to formalize and explore this mechanism in depth and to establish its boundary conditions. Finally, we conduct an online experiment to test the strength of our core assumption. We randomly assign participants the role of a venture capitalist or CEO of a diversifying firm, in order to observe their investment patterns, across different startups vs. internal projects (that are otherwise identical). We find strong support linking the framing of roles (CEO vs. venture capitalist) to the strength of the relationship between performance and resource allocation.

2. The Dynamics of Capability Building

Most new markets are launched when a firm develops both a technological solution for an unmet need and a business model that can potentially realize and scale-up the new solution. There is no general prescription for this process, and existing literature suggests that learning and experimentation is at the heart of finding a product design that starts a new market (Dosi & Marengo, 2000; Helfat & Lieberman, 2002; Nelson & Winter, 1982). The result of this learning process is more effective organizational capabilities in the form of routines (Winter, 2000), which enhance firms' performance and help them meet the needs of an emerging customer base (Helfat & Peteraf, 2003). Therefore, understanding how different firms learn and build their capabilities differently can help explain performance heterogeneity, in both mature markets and in terms of firms' success and survival in the face of waves of creative destruction (Gibbons & Henderson, 2012).

Firms competing to start a new market can include both startups and projects within existing firms. Both types of players seek resources to search for and develop effective capabilities, in order to position themselves as the first to offer a viable product and business model, and subsequently attract customers and launch a new market. This first-mover advantage can help startups become profitable and thus handsomely repay entrepreneurs and early investors (Lieberman & Montgomery, 1988; Markides & Sosa, 2013). Conversely, existing firms that succeed in starting new markets enhance their chances of survival and strategically renew their capabilities, while also retaining a high leverage position in their industry (Lieberman & Montgomery, 1988). Therefore, startups and existing firms often find themselves in direct competition, in terms of establishing new markets. This raises the question of which type of firm is more likely to succeed in building the requisite capabilities more rapidly, as well as why this is the case.

2.1 Endogenous Capability Building

Extant literature offers two relevant insights into the process of capability building. The first insight states that the relationship between cumulative investment in capabilities and performance follows an S-shape curve (Foster, 1988). Return on investment is low in the initial phases, during which time distant exploration is pursued, the needs and tastes of customers are assessed, and multiple alternative solutions are sampled, but no promising technological path is established. As capabilities build, firms' knowledge base grows and early uncertainties are resolved, which reads to performance gains speeding up as a function of capability learning. Once a technological platform is finalized, a more focused process of search and learning-by-doing is initiated, which involves typical learning curve dynamics (Argote & Epple, 1990). In this regime, the return on learning slows down as low-hanging improvement opportunities are first discovered and the learning organization approaches the "fundamental limit of the technology." When performance arrives near that fundamental limit, further investment in capabilities yields few improvements, and these improvements matter less to customers (Christensen, 2000). The specifics of this process are a function of the underlying technology and market, as different platforms have different learning-growth rates and limits.

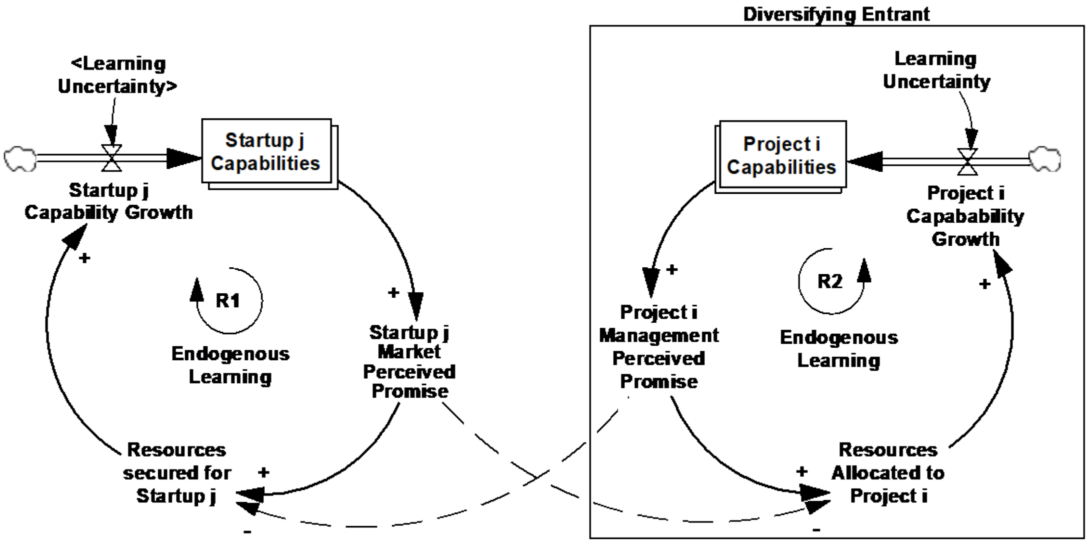
The second insight presented by the extant literature is the uncertainty of the capability-building process. Some investments prove fruitful, while others offer little improvement in the underlying capabilities or with the resulting promise of the firm or project. This probabilistic feature emerges from two sources of uncertainty that are separate but related.

First, some explorations, while helpful in terms of learning about the problem at hand, never pay off in terms of actual efficiency gains. Moreover, there is considerable variation in terms of realizing those returns, both for the potential returns of different improvement activities and organizational effectiveness. Building on these two features of capability-building dynamics, we view firms as engaged in an adaptive learning process that is uncertain and bounded by the technological trajectory they choose to follow.

Our model of organizational learning is distinguished from existing models in the literature because it captures endogeneity in the speed of the search. Specifically, the resources available to startups and internal projects for search and capability investment partially depend on their past performance. The more promising the progress of a startup, the better its prospects for securing the next round of funding. This consequently enables further capability building and refinement. Similarly, how managers perceive the promise of internal projects for future investment depends on projects' past performance. Managers approve higher budgets and allocate more organizational resources to projects that have shown higher promise.

The endogeneity of the resources for capability investment is due to resource allocation, both by financial markets to multiple startups and by existing firms to multiple projects. Specifically, comparing the perceived promise of multiple startups that are active in a new market, venture capitals and other investors must determine their allocation according to the perceived promise of each contender. A similar mechanism determines the allocation of organizational budgets to multiple research and development projects. Thus, markets and internal decision-makers allocate resources based on the promises of other alternatives, as well as on the focal firm's or project's promise. As a result, when past investments have resulted in higher capabilities and perceived promise for one alternative, compared to its competitors, new resources are more likely to flow in the direction of that alternative. This creates a reinforcing loop, which we call endogenous learning. Figure 1 provides a stylized causal-loop diagram of our model (Sterman, 2000).

Figure 1. Summary of feedback loops in dynamic competition among startups and existing firms.



2.2 Securing Resources in a Competitive Environment

In order to focus on endogenous learning as the core mechanism of capability building and performance in new markets, we need to specify how the two types of players differ in their allocation of resources. Specifically, we focus on the differences between projects in existing firms and startups, in terms of securing resources to build their respective capabilities. Startups acquire much of their resources from financial markets, such as angel investors and venture capitals. In contrast, projects inside existing firms rely on the parent firm for their resources. Usually, resource holders (either venture capitals or higher managers) rely on many similar cues to assess the perceived promise of both startups and internal projects, such as technology maturity, market size, business-model coherence, team quality, intellectual property, and financial projections. However, there are important differences between the two. First, the levels of resources that startups secure may be less than projects inside well-endowed firms, especially in the pre-commercialization stage. On the other hand, existing firms often have significant financial resources at their disposal, which can give their internal projects an initial boost. Moreover, existing firms own various resources and capabilities that could benefit new projects with limited costs, including technological expertise, test equipment, market research, human resource systems, and supply chains. This could increase the return on investments in internal projects' capabilities.

Second, there is a significant difference in how resources are allocated inside a firm, with regards to the market. Different startups are largely independent of each other, and thus they are decoupled in the eyes of financial markets. Early on, markets may allocate resources to multiple startups with varying levels of promise, due to uncertainties in technologies and the assessment of that promise. However, as the market matures, venture capitals quickly cut their losses and focus on the most promising platforms.

We expect decoupling among internal projects to be significantly less for at least four reasons. First, projects inside an organization share expertise, systems, capabilities, and resources with each other, which prohibits full decoupling (Bresnahan, Greenstein, & Henderson, 2011). For example, an investment in a firm's human resource systems impacts all internal projects, as it is costly to design and build separate systems for each internal project.

Second, the power to commit resources and formulate organizational strategy is distributed (Bower & Doz, 1979). Middle managers can use existing organizational resources to make progress in new projects without official consent from top management, especially in earlier stages that require fewer resources(Bower & Gilbert, 2007; Burgelman, 1983). This increases the chance of multiple projects working on similar ideas and continuing to use limited organizational resources to advance their respective projects. Additionally, extant research suggests that top-management resource-allocation decisions are based on a combination of project attractiveness and the organizational credibility of the manager that is proposing the project (Bower, Doz, & Gilbert, 2005). Therefore, more time and organizational resources might be needed to identify the best projects in an existing firm. For example, if a highly credible manager proposes a project with medium promise, another less credible manager with a superior project needs to show further results (and therefore requires more time), in order to persuade top management to allocate more resources.

Third, organizational and psychological pressures work against full decoupling. The members of different internal projects see themselves as parts of the same organization, so they subsequently expect to be rewarded based on their effort and overall organizational performance, rather than their luck in establishing a new market, which is very uncertain. Thus, these members will likely feel mistreated when their rewards and resources are tied to the perceived performance of their project, rather than the efforts that they have put into it. This can create pushback against such decoupling in a large firm.

Finally, a diversifying entrant that invests in multiple projects within a new opportunity space is likely to draw on the logic of portfolio management to keep investing in multiple internal projects. With less regard for immediately perceived promise, they hope that, with more eggs in their basket, they will ultimately have a winning project in the market. Such an investment policy not only spreads the inherent risk of investing in new markets, but also builds the firm's absorptive capacity for potential future acquisitions or expansion in the emerging dominant design when the market is created. In summary, we expect financial markets to aggressively sift through startups to find those with the highest potential.

However, we expect existing firms to continue investing in multiple projects for longer, rather than narrowing down their focus to the most promising project early on.

3. **Analyzing Competition in Creating New Markets**

In order to capture the qualitative mechanisms that are discussed in Section 2, we model the competition among N startups (N = 5 in the reported results) and N projects inside a diversifying entrant firm. Each startup or project has a stock of capabilities that accumulate through investment in search and adaptation. Capability levels inform the perceived promise of each alternative in the eyes of relevant resource holders (financial markets or management), using an S-shape function. Financial markets allocate resources to various startups by comparing their promise against other startups. Equation 1 formalizes this decision and allows for different levels of aggressiveness in the market (parameter g). Similarly, managers in the diversifying entrant allocate resources to their internal projects based on the relative promise of each project (Equation 2). We capture the relative decoupling among internal projects, compared to the market's decoupling among startups, using parameter α**.** Therefore, when α approaches 1, the decision process within the organization comes to resemble that of the market more closely. Smaller values of α reflect managers' decisions to diversify their investments, which more cautiously links a project's resources to its past performance.

$$(1)\quad Resources\ Secured\ for\ Startup_j = Base\ Investment * \frac{e^{g*Perceived\ Promise\ Startup(i)}}{\sum_{k\,=\,1}^{N} e^{g*Perceived\ Promise\ Startup(k)}}$$

$$(2)\quad \begin{aligned} &Resources\ Allocated\ to\ Project_i \\ &= r * Base\ Investment * \frac{e^{\alpha*g*Perceived\ Promise\ Project(i)}}{\sum_{k\,=\,1}^{N} e^{\alpha*g*Perceived\ Promise\ Project(k)}} \end{aligned}$$

We capture the endowment difference (both financial and non-financial) between startups and internal projects by varying the level of resources available to diversifying entrants' projects, compared to startups. Specifically, parameter *r* reflects the ratio of total resources allocated to the portfolio of internal projects, compared to that allocated by the market to the group of competing startups. Theoretical arguments often suggest *r* is higher than 1; that is to say, existing firms have more resources to allocate to their projects or can offer various capabilities and assets to their internal projects at discounted costs. This parameter can also capture the differences in productivity of those investments. For example, r values that are lower than 1 could be justified in settings where diversifying entrants have no relevant capabilities or resources, and when startups use their existing resources in more agile and productive ways. Finally, we capture the uncertainty of the search-and-learning process using an auto-correlated noise process that regulates capability growth. Table 1 summarizes the main parameters of the model and their values in base-case simulations.

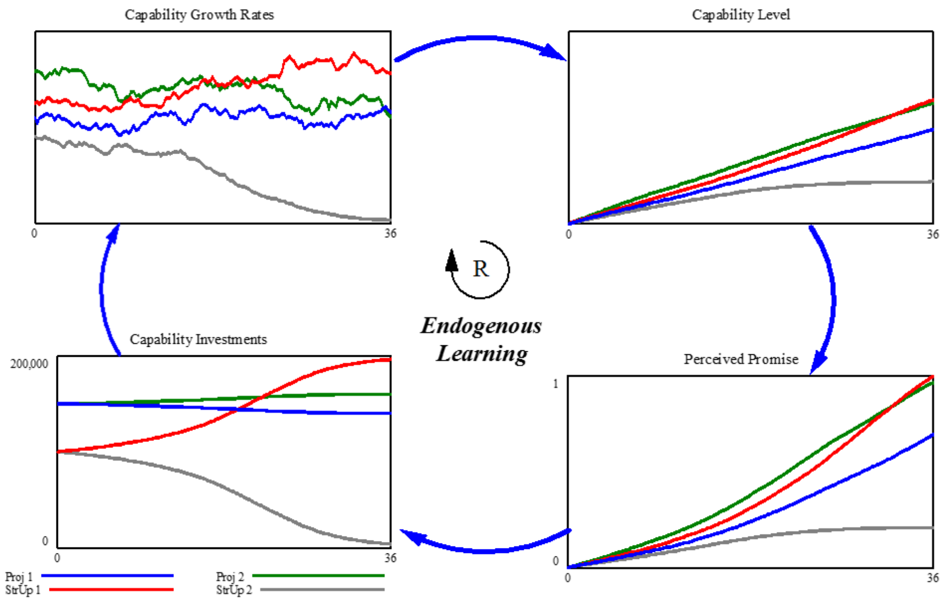
Table 1. Model parameters

Parameter		**Value/Range**	**Units**
N	Number of startups and projects within the diversifying entrant	5	Scalar
α	Entrant's allocation decoupling	[0 1]	Scalar
r	Entrant's extra internal resources	[1 2]	Scalar
g	Market aggressiveness	5	Scalar
SD	Uncertainty (noise standard deviation)	0.2	Scalar

3.1. Basic Dynamics

To build intuition about the core mechanism of our model, we first provide a sample simulation, with N = 2 startups and two parallel projects within a diversifying entrant. Figure 2 summarizes these results. Here, the projects within the diversifying entrant are endowed with 50% more resources (r = 1.5). Moreover, it is assumed that the managers in the firm are reluctant to decouple the internal projects (α = 0.1). As a result, internal projects (shown as blue and green lines) have faster capability investment rates early on. However, the randomness of the capability growth, due to the uncertain returns on investment, bolsters one of the startups (shown in red). Thus, this startup gains even more traction in the financial market early on, which leads to an increasing shift in investment resources toward this startup and away from the others (shown in gray). The internal projects also see a similar decoupling in resource allocation, but much slower than in the startups. As a result, after a while, the red startup catches up with the better performing internal project, which further enhances its capability and ultimately helps it beat all the other players to establish the new market first[10]. Despite the startup's resource disadvantage, its key mechanism of promotion here is the reinforcing loop of endogenous learning that is stronger for the startup, compared to the internal projects, due to $\alpha < 1$. Yet, this core mechanism is also moderated by the uncertainty of capability building, the extra resources available to the internal projects, and the shape of the technological landscape that underlies the competition. We explore these factors using large sample simulations in the following sections.

Figure 2. Sample simulation with N = 2, A = 0.1 and R = 1.5. (Startups are depicted in gray and red, while projects within the existing firm are in blue and green.)



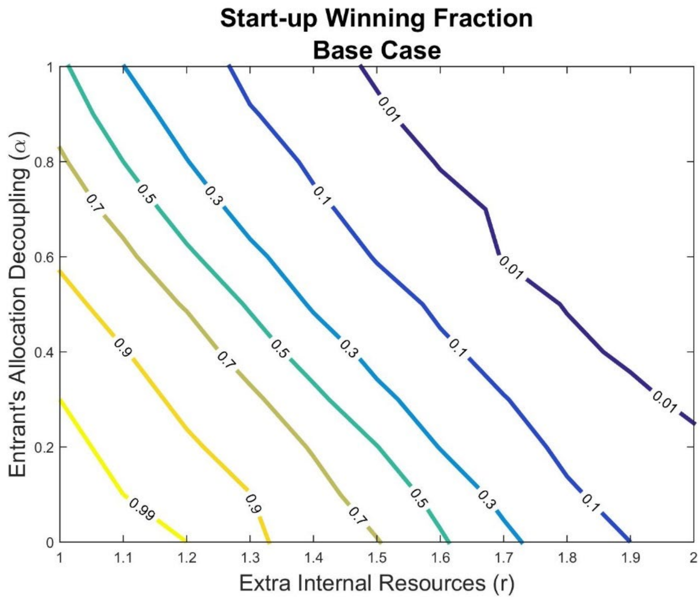
[10] We identify the successful competitor as the one that first reaches a promising performance threshold that can sustain positive cashflow in the market (which is defined as the promise level of 1 in the model).

3.2. Startup Advantage in Creating New Markets

In this section, we report the ways in which a startup's chances of establishing a new market depend on the resource advantage of existing firms (*r*) and the level of decoupling (α). Keeping all other parameters, the same across startups and entrants, we increase the two parameters of interest by increments of 0.1, simulate 200 random markets in each setting, and report the startups' winning fractions (across those 200 simulations) using a contour plot that summarizes the 24,200 resulting simulations (Figure 3).

When the diversifying entrant has no resource advantage (r = 1) and is able to imitate the market's allocation by fully decoupling internal projects (α = 1), we expect no difference between startups and entrants' internal projects. This leads to a startup success fraction of 0.5. When the diversifying entrant has twice more resources as the startups (r = 2) and is able to fully decouple its projects, we expect the vast majority of winners to come from the diversifying entrant. However, as α decreases from 1 to 0, we expect to see increasing opportunities for the startup's success. This advantage works against the entrants' resource advantages. For example, with no decoupling (α = 0), the diversifying entrant requires over 1.6 times more resources to compensate for the stronger reinforcing loops that startups can activate.

Figure 3. Simulation results of the dynamic competition between startups and a diversifying entrant for the base case.

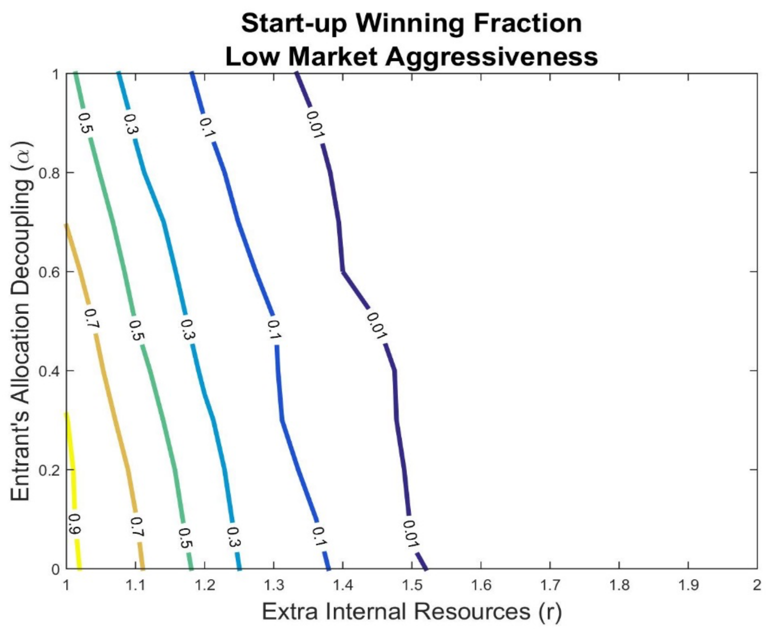


3.3. Impact of Market Aggressiveness

In this section, we report the impact of market aggressiveness on startup advantage. In our model, market aggressiveness captures the speed with which financial markets rally around the emerging top startup. It is reasonable to expect that more transparency in the market, higher discount rates, and stronger competition among different investors triggers more aggressive treatment of the pool of existing startups. Therefore, it is instructive to assess the sensitivity of the results within this parameter. Specifically, we compare a case of a less-aggressive market with the base case and track changes in the impact of extra resources (r) and decoupling (α) on startup advantages. We change the market aggressiveness parameter (g) from 5 to 2, but everything else remains identical to the base case.

The results depicted in Figure 4 show a similar tradeoff to the base case, but the mechanism that promotes startups has a weaker impact. When the diversifying entrant has no resource advantage (r = 1) and similar decoupling in the allocation process (α = 1), we still see a 50% startup success rate. Similarly, we still observe that when α = 1 and the diversifying entrant has a significant resource advantage (r = 2), it is able to clear out the competition. However, compared to the base case, the relative impact of decoupling has faded here. In low-aggressive markets, existing firms only need 1.2 times more resources to compensate for the lack of decoupling (α = 0). This result is due to the weakening of reinforcing loops of endogenous learning across the board, when we reduce g. In the extreme, when g = 0, neither financial markets nor internal managers consider perceived promise when allocating resources. With a fixed share of investment guaranteed, the outcome of searching and capability building is solely a function of the total resources allocated, as well as luck. More generally, when a market is less aggressive, the diversifying entrant has sufficient time to improve projects' performance by using its resource advantage before the most promising startup can differentiate itself from the rest and obtain comparable resources. As a corollary, this mechanism puts even more pressure on investors in startups to be aggressive in their resource allocation since this increases their chances of funding a successful startup.

Figure 4. Simulation results in the case of low market aggressiveness (G = 2).



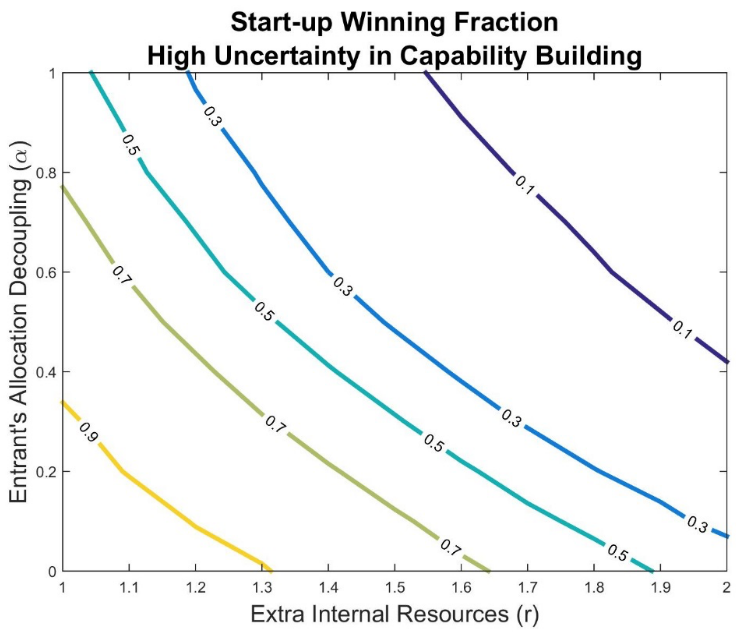
3.4. Competition in Highly Uncertain Markets

In this section, we report how the level of uncertainty in capability building impacts a startup's chance of success. The maturity stage of new technologies, how close a new market is to existing markets (and thus how clearly the needs of its potential customers is understood), and the extent to which the technology and market are subject to influences outside of the model's boundary, among others, all regulate the level of uncertainty in terms of searching and capability building. To assess the sensitivity of results at different levels of uncertainty, we double the noise standard deviation from the base case (SD = 0.2 to SD = 0.4). Everything else remains the same as the base case, in order to isolate the impact of increased uncertainty on startup advantage.

The results are presents in Figure 5. As expected, the same basic dynamics of increased startup chances occur when the entrant has less decoupling and resource advantage. Compared to the base case, we see increased chances for startups' success when r and α are higher. For example, the fraction of startups winning reaches 0.1 when $\alpha = 1$ and $r = 1.6$, compared to the base case's much steeper decrease in startups' chances (from 50% to 10% when $r = 1.3$). On the other hand, the entrant's chances of success increase when it has less resource advantage and is not able to implement much decoupling between projects. For example, in the base case, the entrant has almost no chance when $r < 1.1$ and $\alpha < 0.2$. However, in the case of highly uncertain markets, entrants' chances never pass below 1%, even when there are no more resources ($r = 1$) and no decoupling ($\alpha = 0$).

Two mechanisms explain these results. First, by increasing the uncertainty of searching and capability building, success becomes more a matter of chance than in core model mechanisms. Therefore, the probabilities of success approach 50% across the board, reducing the sharp distinction in the extremes (e.g., $\alpha = 0$ with $r = 1$ and $\alpha = 1$ with $r = 2$). On the other hand, the startup advantage provided by the endogenous learning loop is only observed when the symmetry among startups is broken by randomness in investment returns. Increasing uncertainty thus triggers this mechanism earlier and helps startups win a larger fraction of simulated markets (thus the shift from $r = 1.6$ to $r = 1.9$ for the breakeven point under $\alpha = 0$).

Figure 5. Simulation results in the case of highly uncertain markets (SD = 0.4).



3.5. Competition in Rugged Technological Landscapes

In this section, we report the systematic manipulation of the S-shape function that connects capabilities to each startup's or project's promise. Different technologies may have widely different trajectories. For example, in the alternative-fuel vehicle domain, hydrogen, electric, and hybrid technologies promise different ideals of fuel efficiency, carbon footprints, and costs. They also have very different trajectories of investment, with hybrid vehicles promising faster early improvements but the likelihood of lower maximum benefits (Keith, 2012). So far, we have used the same S-shape curve to connect the level of capability to the perceived promise, but different startups and projects may actually follow different architectures with various underlying capability-promise curves. This S-shape function represents the inherent features of each technology, including its fundamental limit or maximum, its level of promise at the inflection point (where the maximum marginal gain from capability investments occur), and the slope of its maximum marginal gain (which regulates how fast that technology reaches its maximum). These features are highlighted in Figure 6.a.

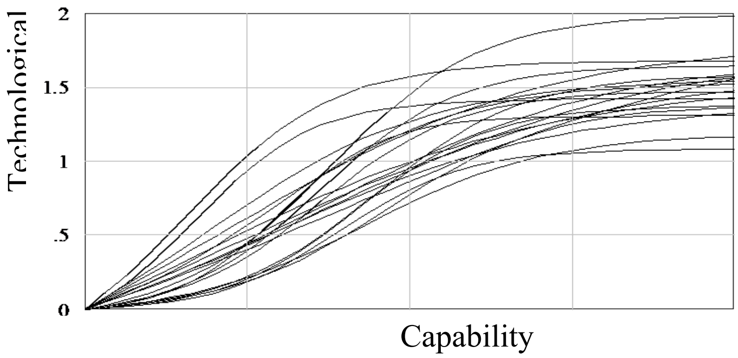
In the base case, we assume that every startup and project within a diversifying entrant follows the same S-shape path. This S-shape function has a maximum of 1.5, an inflection point of 1, and a slope of 1. Here, to simulate more rugged technological landscapes, we randomly assign a maximum to each startup or project from a uniform distribution in the range of 1 to 2. Similarly, we draw the inflection point and slope randomly from a uniform distribution that ranges from [0.5 1.5] to [0.5 1].

The results depicted in Figure 6.b show that, while we see similar patterns in the base case, startups' winning fractions have increased across the board. For example, with no decoupling ($\alpha = 0$), the diversifying entrant requires more than two times the resources to compensate (compared to $r = 1.6$ in the base case). Similarly, when the entrant is able to fully decouple projects ($\alpha = 1$), the startup's chance of success decreases to around 1%, when the entrant has 80% more resources ($r = 1.8$). In the base case, startups' chances reach the same 1% point with just 50% more resources ($r = 1.5$).

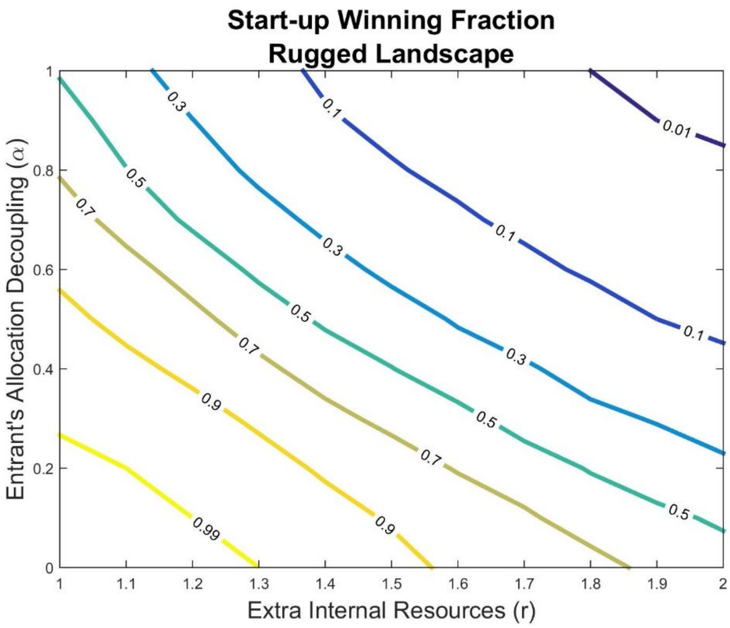
These results are regulated by the heterogeneity in the strength of endogenous learning loop that is introduced in rugged technological landscapes. Specifically, increasing both the maximum potential and the slope speeds up the reinforcing loop, while the operation of that loop in favor of startups is brought forward in time by the earlier inflection point. Thus, by introducing heterogeneity into these parameters, we are more likely to see startups that disproportionately benefit from the endogenous learning loop and that are thus more likely to take over the otherwise identical internal project. Therefore, randomness in the shape of technological landscape, as well as randomness in the search process, can enhance startups' chances of establishing new markets.

Figure 6. Simulation results of competition in the rugged technological landscape**.**

6.a. Randomness in the technological S-curves.



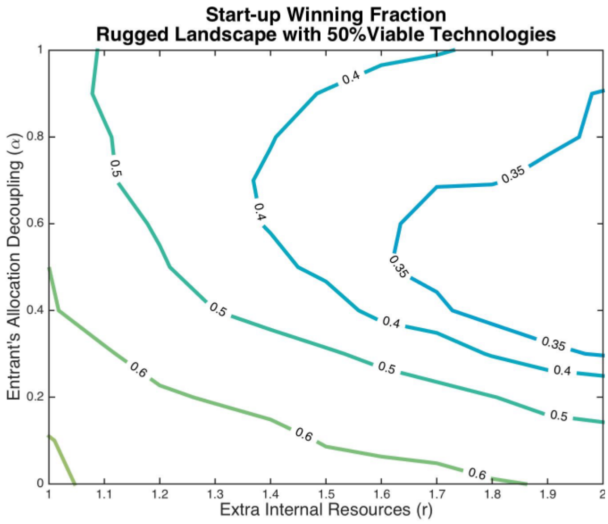
6.b. Startup winning fraction in the rugged technological landscape**.**



One important assumption in the analyses presented above is that all startups and projects have the potential to successfully create a new market. Specifically, in these previous analyses, we set the minimum fundamental technological limit to always be greater than or equal to the threshold for winning the (market creation) competition, which is 1. In this part, we relax this assumption by randomly assigning a maximum to each startup or project, using a uniform distribution from 0.5 to 1.5 (instead of [1 2]; Figure 6.a). Figure 7 depicts the simulation results for the rugged landscape with just 50% of viable technologies.

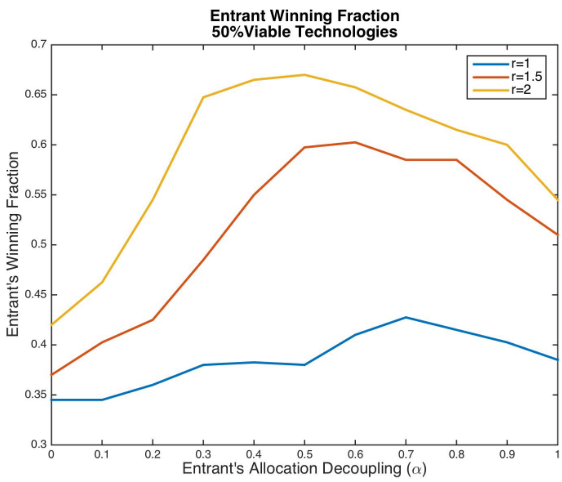
Figure 7. Simulation results of competition in the rugged technological landscape with 50% viable projects.

7.a. Startup winning fraction in the rugged technological landscape with 50% viable



projects.

7.b. Change in the optimal level of aggression without the 100% viability assumption.



While the results are similar for low levels of investment and aggressiveness, increasing aggressiveness can become counterproductive for the entrant. To illustrate this point, Figure 7.b depicts the optimal level of aggressiveness for three different levels of investment. Aggressively investing resources into projects that show early promise and success can therefore be very risky. To explain this counterintuitive result, it is important to note that unviable trajectories increase the uncertainty and risk that are associated with focusing on one project. Therefore, when rugged landscapes increase the chance of failure, high levels of decoupling can actually hurt chances of success, due to prematurely abandoning promising alternatives.

3.6. Robustness of the Simulation Results

In addition to the results reported above, we conducted sensitivity analysis on all the model parameters and key structural features. We found no significant qualitative change in the results. Various nuances were revealed in the sensitivity analysis process. For example, we investigated how access to information about the other side of the competition affects results. Interestingly, knowing about the promise of the other side (e.g., startup funders knowing about the promise of internal projects and vice versa) can actually discourage investments when the other side has an early advantage, either by chance or due to extra investments. If this visibility is symmetric (i.e., both sides can observe the promise of the other), the net result supports internal projects that often start ahead because of extra resources. However, asymmetric visibility, where startups are observable by all parties, but internal projects are opaque to outsiders, is potentially more realistic and can actually enhance startup advantage. Additionally, we also replicated the results using the NK model structure, which is commonly applied in strategy research (Ganco & Agarwal, 2009; Levinthal, 1997), to model the same phenomenon. As we observed qualitatively similar results, we opted for the simpler, more flexible, and computationally more efficient model structure that is reported here. Overall, our results robustly inform key structural assumptions and parameter values in sensible ranges. The most important sensitivities have been discussed in detail in the previous sections.

4. Empirical Study

Our model and propositions rest on a central assumption: there is a significant difference in resource allocation aggressiveness between diversifying entrants and the venture capital market. We assess the empirical validity of this assumption using an experiment. An experiment is particularly helpful here because of data-availability issues for an observational study. It is quite difficult to gather data about multiple products and multiple organizations, especially in such a way that firms and startups compete synchronously. Firms are particularly hesitant about sharing any detailed financial data about internal projects, many of which may never lead to actual products.

Our experimental approach is however limited in that it only partially captures the drivers of the gap in resource allocation between startups and entrant projects. In Section 2, we discussed four pathways through which diversifying entrants may be less inclined to fully decouple their projects. First, projects inside an organization share expertise, systems, capabilities, and resources with each other, which prohibits full decoupling. Second, the power to commit resources and formulate organizational strategy is distributed among project managers. Third, organizational and psychological pressures work against full decoupling. The members of different internal projects see themselves as parts of the same organization, so they subsequently expect to be rewarded based on their effort and overall organizational performance, rather than results in market place that may be as much about luck as their efforts. Finally, a diversifying entrant that invests in multiple projects within a new opportunity space is likely to draw on the logic of portfolio management to keep investing in multiple internal projects. In our experiment we focus only on the third pathway.

As such, we see our results as offering a lower bound on the magnitude of the decoupling expected in the real world.

4.1 Experimental Research Design

We developed an online experiment with two conditions for the between-subjects design. Subjects were randomly assigned to play either the role of a company CEO or a venture capitalist in a market. All subjects played the same task, but with different cover stories that depended on their assigned roles. Those playing as a CEO were asked to allocate a total amount of resources to five of their projects led by managers, with whom they had been working for a significant amount of time. Those playing the role of venture capitalists were asked to allocate funds to five startups in their portfolio.

We recruited 287 potential participants from the Amazon's Mechanical Turk (MTurk), and 159 who qualified in our initial training and testing participated in the online experiment. Participants were paid based on their level of effort and their success at being the first to successfully introduce a product to the market. This compensation scheme incentivized the participants to take the task seriously and exert effort (measured in time taken between decisions) while doing their best to "win".

The median time to complete the experiment was 11.6 minutes. Participants were paid a maximum of $4. The median compensation was $3.0 (equivalent to $15.50 per hour).

4.1.1 Procedures and Measures

After being recruited, subjects were randomized into one of two experimental groups and then completed an instruction and qualification survey. Participants read the instructions for the task, which were specifically designed for their treatment condition. This was followed by a set of questions that assessed their understanding of the goal, main concepts, and mechanics of the task. Subjects that scored higher than a designated threshold were allowed to complete the main experimental task.

Selected individuals (159 out of 287 people) then completed a demographic survey. They were 44.7% female. The majority identified themselves as white (74.2%), following by Asian (10.1%) and Latino/Hispanic (9.4%). The median participant was aged 35–44 years old, had a four-year college degree, and had an income between $30,000 and $40,000.

After completing the survey, subjects were directed to a website to complete the experimental task. Subjects completed the online task in two rounds with different sets of startups (or projects). Playing the task in two rounds reduced the role of luck and allowed for measuring the impact of learning (see Appendix A for the task interface and information provided to participants). The task was based on the simulation model described above.

The task was a single-player simulation in which participants competed against a simulated agent with five projects, aggressiveness (g) of 10, and with 0.8 to 1.5 times the resources of participant (drawn from a uniform distribution). Subjects were unaware of the specifics of the competition. Each quarter, participants made decisions on what percentage of resources to allocate to their startups (or projects). Participants observed the promise of each of their startups (or projects) and whether they are leading the market, but not the detailed promise of their simulated competition. The task ended when one startup or project from the competition or the participants reached the winning threshold (promise=1). This setup was known to all.

We recorded all decisions and the resulting promises. We measured and analyzed the relationship between the position of each startup (or project) among the five in terms of fraction of promise and the fraction of resources allocated to those startups (or projects).

4.1 Experiment Results

To measure possible differences in resource allocation, we conducted an ordinary least squares estimation. Since participants made repeated allocation decisions over time, we used fixed effects models (results are robust to this choice). The full fixed effects model is shown in Equation 3. Our dependent variable is the allocation ratio (*AR*) that each startup (or project) receives. The first independent variable is the ratio of total promise (RTP_{it}) for the corresponding startup (or project). More importantly, we are interested in the relationship between *RTP* and *AR* depending on our treatment conditions. Therefore, we used the interaction effect between the dummy startup (S_i) and *RTP*. Moreover, we added interaction effects first between *RTP* and Time (t_i) and second, between *RTP* and the task round (R_i). We also added the interaction of *RTP* and dummy variables D_k, which represent demographic data. υ_i represents participant fixed-effects.

(3) $$AR_{it} = a + \beta_1 RTP_{it} + \beta_2 RTP_{it} * S_i + \beta_2 RTP_{it} * t_i + \beta_2 RTP_{it} * R_i + \sum_k \beta_k D_k * RTP_{it} + \upsilon_i + \varepsilon_{it}$$

We present the regression results in Table 2. Our results indicate a significant difference in resource allocation behavior between the two treatment groups. Specifically, we find that those who played the role of a venture capitalist tied the allocation ratio (*AR*) to promise more strongly than those that played the role of a company CEO. When $\beta_1 = 0$, it reflects that participants allocated resources independent of their projects' promises. When $\beta_1 = 1$ it reflects that resources are allocated "fairly", or linearly based on promise. Larger β_1 reflects more aggressive resource allocations. In our models, we find β_1 to be close to 1. β_2 reflects how much more aggressive those who played the role of venture capitalist were. We find that being in the venture capitalist role increased the relationship between performance and resource allocation by 17% (in model 4). We also find that participants were more aggressive in their resource allocation in the second round and as time progressed and in each round. In summary, our results suggest that an intervention as simple as labeling one's role as CEO (vs. venture capitalist) strongly moderates the relationship between promise and resource allocation.

Our experiment only manipulated one of the mechanisms moderating the relationship between promise and allocated resources, leaving out shared-resources, distributed decision making, and the logic of portfolio management, all of which could further distinguish allocations within a firm vs. those among startups. Results suggest that the effects of organizational and psychological pressures toward fairness and against full decoupling can significantly affect resource allocation in a way that gives startups an advantage. In fact, our manipulation in this experiment is rather weak, compared to actual psychological pressure managers face in actual organizational settings, face to face interactions, and long-term relationships. Therefore, we believe our estimates should be interpreted as conservative, only setting a lower bound for the relevance of the mechanism we explored in this paper.

Table 2. Regression Models of Resource Allocation

	(1) Model1	(2) Model2	(3) Model3	(4) Model4
Ratio of Total Promise	1.06^{***} (0.01)	0.92^{***} (0.01)	0.79^{***} (0.01)	1.14^{***} (0.05)
Ratio of Total Promise *Startup	0.14^{***} (0.01)	0.13^{***} (0.01)	0.12^{***} (0.01)	0.19^{***} (0.01)
Ratio of Total Promise *Round		0.26^{***} (0.01)	0.25^{***} (0.01)	0.21^{***} (0.01)
Ratio of Total Promise *Time			0.02^{***} (0.00)	0.02^{***} (0.00)
Ratio of Total Promise *Demographics	No	No	No	Yes
Participant Fixed-effects	Yes	Yes	Yes	Yes
N	90420	90420	90420	90420
R2	0.56	0.56	0.57	0.58

Note. Cell entries are estimated regression coefficients and (robust standard errors).

$^{\sim} p < 0.10$, $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

5. **Discussion and Conclusions**

Startups regularly succeed in creating new markets, despite the presence of and investments from well-endowed and capable existing firms, who can benefit from the launch of a new market for their strategic renewal. According to the resource-based view of strategy, existing firms should have a significant advantage, which should drastically reduce the odds of any startup succeeding in such a competition. Existing research provides key insights into why incumbents might not invest enough in new opportunities or may not do so in time. However, we know much less about why diversifying entrants often fail in cases in which they do invest and do so in time, especially in new markets, where the risk of cannibalization and organizational inertia are lower than in established markets.

We propose a novel mechanism to resolve this puzzle. This mechanism captures the endogeneity of the search process, as greater investment not only results in building more capabilities, it also leads to higher performance, based on those capabilities, which thus attracts and enables more investment. Building on this feedback loop, we argue that startups can speed up their learning processes because existing firms cannot fully decouple their internal projects. Better performing projects within a diversifying entrant receive additional attention and resources, but much less so than startups showing promise. As such, they start to fall behind the fastest-growing startups, with the latter fully utilizing the endogenous learning feedback loop to their advantage.

In summary, our results show a distinct opportunity for startups to succeed, even when competency traps and other sources of inertia do not weigh down existing firms. We explored the strength of our proposed mechanism under various levels of complexity in the technological landscape, varied aggressiveness of external markets in funding startups based on their perceived promise, and different incumbent strategies to internally provide resources for various projects. While we found interesting nuances based on these sensitivity analyses, our core results are robust in a wide range of parameter settings and assumptions.

We corroborated our central assumption, that existing firms cannot fully decouple their internal projects, using an online experiment. This assumption is motivated both by functional and psychological mechanisms. Prior literature shows that existing firm's projects share resources, that the power to commit resources is distributed among project managers, and that top managers may decide against decoupling based on the logic of portfolio management. Besides these functional arguments for decoupling between promise and allocated resources, our experiment shows organizational and psychological processes can alone significantly limit decoupling in existing firms.

Our analyses suggest that the endogenous learning mechanism is most salient when the development of a new market is complex, uncertain, subject to increasing returns, and contested by multiple startups and entrants. Specifically, our analysis provides a set of propositions about market conditions that favors startups. For example, we propose that complex technological landscapes limit the applicability of a diversifying entrant's existing knowledge about a new opportunity, thus reducing the baseline advantage those entrants may benefit from. Similarly, we hypothesize that startups have an advantage in markets that have strong reinforcing loops, including the increased availability of external funding mechanisms, and in tightly coupled administrative systems within existing firms.

Some of these predictions coincide with those of the existing theories, but others offer opportunities to empirically tease out alternative mechanisms. For example, while spreading risk across a portfolio is a cornerstone of conventional finance and resource-allocation strategies, our model predicts that the benefits of portfolios may be diminished when resources for learning depend on past performance. Our model also suggests that entrepreneurial spin-offs succeed by combining the benefits of being both a startup and having roots in a well-endowed firm (Christensen, 1993; Dosi, 1984; Klepper, 2001; Klepper & Sleeper, 2005). On the other hand, existing firms may be able to improve their odds of succeeding in new markets by decoupling their projects, such as through internal venture capitals (Chesbrough, 2000).

Our model, while providing interesting insights, is relatively stylized and limited in scope. Other dynamics can also influence startups' and existing firms' competition. For example, we assume all the startups and projects inside the existing firms are launched at the same time. This assumption provides us with a platform to focus on our main mechanism. However, in many cases, how soon existing firms recognize and engage in new-market creation plays a significant role (Christensen, 2000; Henderson & Clark, 1990; Utterback, 1996). Moreover, we imply that fast learning and capability building are always desirable. However, existing literature suggests that get-big-fast strategies might backfire and create significant problems in execution, as well as a loss of credibility for a startup (Sterman, Henderson, Beinhocker, & Newman, 2007). For example, some firms may try to enhance their perceived promise without building the necessary capabilities, which could lead to cheating, corner cutting, and other risks that we have not explored.

Moreover, we limited our analysis to the competition among startups and existing firms. However, the relationship between these two types of entity is very complex. Some existing firms do not enter new markets early because they hope to acquire a promising startup later. Although laggard firms find fewer highly profitable acquisition opportunities when the fog of uncertainty clears, this strategy indicates a very complex relationship between the two types of entity. Indeed, some existing firms may internally nurture a few projects, in order to enhance their environmental scanning and identify promising startups early on. This also helps them in building the absorptive capacity to acquire a new startup. These nuances suggest higher-level competition among existing firms that invest in new technological areas, not only by counting on the likelihood of building a successful new product internally, but also by strengthening their chances of acquiring promising startups before their competitors. This potential mechanism offers one answer to the question of why existing firms invest in new markets even if they can see a significant startup advantage, due to the endogenous learning mechanism we explored. Formalizing this intuition and exploring its implications offer opportunities to extend the current study.

Future research can also assess the impact of endogenizing both markets' and firms' resource allocations. Specifically, over time, it is reasonable to expect that venture capitals and other funders of early stage businesses will adjust their level of emphasis on past performance, as an indication of a firm's chances of success, in order to maximize their own expected return on investment. Similarly, managers within existing firms may increase investments in projects or may attempt more decoupling in later stages, if and when they observe signs of startups surpassing their projects.

References

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

Bower, J. L., & Doz, Y. L. (1979). Strategy formulation: A social and political process. *Strategic Management: A New View of Business Policy and Planning*, 152–188.

Bower, J. L., Doz, Y. L., & Gilbert, C. G. (2005). Linking resource allocation to strategy. *From Resource Allocation to Strategy,* 3–25.

Bower, J. L., & Gilbert, C. G. (2007). How managers' everyday decisions create or destroy your company's strategy. *Harvard Business Review*, *85*, 72–79.

Bresnahan, T. F., Greenstein, S., & Henderson, R. M. (2011). Schumpeterian competition and diseconomies of scope: Illustrations from the histories of Microsoft and IBM. *The Rate and Direction of Inventive Activity Revisited*, 203–271: University of Chicago Press.

Burgelman, R. A. (1983). Corporate entrepreneurship and strategic management: Insights from a process study. *Management Science*, *29*, 1349–1364.

Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, *12*(1), 9–30.

Chesbrough, H. (2000). Designing corporate ventures in the shadow of private venture capital. *California Management Review*, *42*(3), 31–49.

Christensen, C. M. (1993). The rigid disk-drive industry: A history of commercial and technological turbulence. *Business History Review*, 67(4), 531–588.

Christensen, C. M. (2000). *The innovator's dilemma: When new technologies cause great firms to fail* (1st HarperBusiness ed.). New York: HarperBusiness.

Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing*, *3*(2), 97–108.

Dosi, G. (1984). Industry in France - French - Bellon,B, Chevalier,Jm. *Futures*, *16*(1), 99–101.

Dosi, G., & Lovallo, D. (1997). *Rational entrepreneurs or optimistic martyrs? Some considerations on technological regimes, corporate entries, and the evolutionary role of decision biases*. Cambridge, UK: Cambridge University Press.

Dosi, G., & Marengo, L. (2000). Some elements of an evolutionary theory of organizational competences. *The Theory of the Firm: Critical Perspectives on Business and Management*, *4*, 50.

Dunne, T., Klimek, S. D., & Roberts, M. J. (2005). Exit from regional manufacturing markets: The role of entrant experience. *International Journal of Industrial Organization*, *23*(5–6), 399–421.

Foster, R. N. (1988). *Innovation: The attacker's advantage*. Summit Books.

Ganco, M., & Agarwal, R. (2009). Performance differentials between diversifying entrants and entrepreneurial start-ups: A complexity approach. *Academy of Management Review*, *34*(2), 228–252.

Gibbons, R., & Henderson, R. (2012). Relational contracts and organizational capabilities. *Organization Science*, *23*(5), 1350–1364.

Helfat, C. E., & Lieberman, M. B. (2002). The birth of capabilities: Market entry and the importance of pre-history. *Industrial and Corporate Change*, *11*(4), 725–760.

Helfat, C. E., & Peteraf, M. A. (2003). The dynamic resource-based view: Capability lifecycles. *Strategic Management Journal*, *24*(10), 997–1010.

Henderson, R. (1993). Underinvestment and incompetence as responses to radical innovation: Evidence from the photolithographic alignment equipment industry. *The RAND Journal of Economics,* 248–270.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, *35*(1), 9–30.

Kaplan, S., & Henderson, R. (2005). Inertia and incentives: Bridging organizational economics and organizational theory. *Organization Science*, *16*(5), 509–521.

Keith, D. R. (2012). *Essays on the dynamics of alternative fuel vehicle adoption: insights from the market for hybrid-electric vehicles in the United States* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology, Engineering Systems Division.

King, A. A., & Tucci, C. L. (2002). Incumbent entry into new market niches: The role of experience and managerial choice in the creation of dynamic capabilities. *Management Science*, *48*(2), 171–186.

Klepper, S., & Simons, K. L. (2000). Dominance by birthright: Entry of prior radio producers and competitive ramifications in the US television receiver industry. *Strategic Management Journal*, *21*(10–11), 997–1016.

Klepper, S. (2001). Employee startups in high-tech industries. *Industrial and Corporate Change*, *10*(3), 639‐674.

Klepper, S., & Sleeper, S. (2005). Entry by spinoffs. *Management Science*, *51*(8), 1291–1306.

Koellinger, P., Minniti, M., & Schade, C. (2007). "I think I can, I think I can": Overconfidence and entrepreneurial behavior. *Journal of Economic Psychology*, *28*(4), 502–527.

Levinthal, D. A. (1997). Adaptation on rugged landscapes. *Management Science*, *43*(7), 934–950.

Lieberman, M. B., & Montgomery, D. B. (1988). 1st-mover advantages. *Strategic Management Journal*, *9*, 41–58.

March, J. G., & Simon, H. A. (1958). *Organizations*. New York: Wiley.

Markides, C., & Sosa, L. (2013). Pioneering and first mover advantages: The importance of business models. *Long Range Planning*, *46*(4–5), 325–334.

McCarthy, A. M., Schoorman, F. D., & Cooper, A. C. (1993). Reinvestment decisions by entrepreneurs: Rational decision-making or escalation of commitment? *Journal of Business Venturing*, *8*(1), 9–24.

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2016). Agents of structural change: The role of firms and entrepreneurs in regional diversification.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Belknap Press of Harvard University Press.

Palich, L. E., & Bagby, D. R. (1995). Using cognitive theory to explain entrepreneurial risk-taking: Challenging conventional wisdom. *Journal of Business Venturing*, *10*(6), 425–438.

Sosa, M. L. (2013). Corporate structure, indirect bankruptcy costs, and the advantage of de novo firms: The case of gene therapy research. *Organization Science*, *25*(3), 850–867.

Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Boston: Irwin/McGraw-Hill.

Sterman, J. D., Henderson, R., Beinhocker, E. D., & Newman, L. I. (2007). Getting big too fast: Strategic dynamics with increasing returns and bounded rationality. *Management Science*, *53*(4), 683–696.

Tripsas, M., & Gavetti, G. (2000). Capabilities, cognition, and inertia: Evidence from digital imaging. *Strategic Management Journal*, *21*(10–11), 1147–1161.

Utterback, J. M. (1996). *Mastering the dynamics of innovation*. Harvard Business Press.

Winter, S. G. (2000). The satisficing principle in capability learning. *Strategic Management Journal*, *21*(10–11), 981–996.