



2018

Exploring The Multi-Trillion Dollar Question: How Resource Dynamics Shape Long-Term Profit Patterns

Phebo Derk Wibbens

University of Pennsylvania, phebowibbens@gmail.com

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Business Administration, Management, and Operations Commons](#), [Economics Commons](#), and the [Management Sciences and Quantitative Methods Commons](#)

Recommended Citation

Wibbens, Phebo Derk, "Exploring The Multi-Trillion Dollar Question: How Resource Dynamics Shape Long-Term Profit Patterns" (2018). *Publicly Accessible Penn Dissertations*. 2997.
<https://repository.upenn.edu/edissertations/2997>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/2997>
For more information, please contact repository@pobox.upenn.edu.

Exploring The Multi-Trillion Dollar Question: How Resource Dynamics Shape Long-Term Profit Patterns

Abstract

Only a few dozen firms across different industries have been able to appropriate a large share of the trillions of dollars of total net value appropriation in US markets over the past decades. This dissertation examines, using formal modeling, how resource dynamics in competitive markets shape such long term profit patterns.

Chapter 1 introduces the main questions and concepts, including a general formal framework that can be used to represent RBV-based theories in terms of stochastic state space models, as well as a five-step program to do so. In chapter 2 a new measure of long-term firm performance is developed: LIVA (long-term investor value appropriation). This measure helps to address a disconnect between the common theoretical assumption that managers optimize firm value, and the widespread empirical practice of measuring performance using short-term ratios such as ROA (return on assets). Chapter 3 describes a dynamic model and empirical inference of the relation between resource competition and the decomposition of variations in returns into firm-specific and industry components. The model shows that the investment dynamics for scarce resources amplify any idiosyncratic shocks to their resource positions, thus providing a very general mechanism for the empirical regularity that variations in return have a high firm-specific component. Empirical results from a Bayesian hierarchical analysis of stock market returns support the model predictions. Chapter 4 describes a formal model to analyze how profit persistence is affected by higher order resources, which are an abstract representation of dynamic capabilities. The model in this chapter indicates that higher order resources should lead to the presence of a second order autoregressive term in profit time series. Empirical estimation of the model using both classical and Bayesian hierarchical methods indeed provides evidence for the existence of such a second order autoregressive term.

The models in these final two chapters are implementations of the formal RBV framework introduced in chapter 1, and thus exemplify the value of such a framework to the strategy field.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Management

First Advisor

Nicolaj Siggelkow

Keywords

Bayesian inference, Competitive strategy, Game theory, Long-term profit, Resource based view, Stochastic dynamics

Subject Categories

Business Administration, Management, and Operations | Economics | Management Sciences and Quantitative Methods

EXPLORING THE MULTI-TRILLION DOLLAR QUESTION:
HOW RESOURCE DYNAMICS SHAPE LONG-TERM PROFIT PATTERNS

Phebo D. Wibbens

A DISSERTATION

in

Management

For the Graduate Group in Managerial Science and Applied Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

Nicolaj Siggelkow, David M. Knott Professor, Professor of Management

Graduate Group Chairperson

Catherine Schrand, Celia Z. Moh Professor, Professor of Accounting

Dissertation Committee

Daniel A. Levinthal, Reginald H. Jones Professor of Corporate Strategy

David H. Hsu, Richard A. Sapp Professor, Professor of Management

EXPLORING THE MULTI-TRILLION DOLLAR QUESTION:
HOW RESOURCE DYNAMICS SHAPE LONG-TERM PROFIT PATTERNS

© COPYRIGHT

2018

Phebo Derk Wibbens

Hypotheses non fingo

—Sir Isaac Newton

ACKNOWLEDGEMENT

I have been extremely fortunate to write this dissertation at the Wharton Management Department, an amazing group of smart, knowledgeable, and collaborative people. I particularly wish to thank my dissertation committee, consisting of Nicolaj Siggelkow, who has the uncanny ability to always ask the exact single most important question; Dan Levinthal, who spots and polishes the good part of any paper, idea, or random thought, like finding a rough diamond in no matter how big a pile of sand; and David Hsu, who really thinks about what a message evokes with different audiences, and how to get across what you want to say. Other Wharton professors who have been very generous with their input are (in alphabetical order) Ulrich Doraszelski, Emily Feldman, Vit Henisz, Shane Jensen, Katherine Klein, Anoop Menon, Arkadiy Sakhartov, Paul Shaman, Harbir Singh, and Natalya Vinokurova. Also my fellow PhD students have provided amazing support, both as friends and with their help on papers and presentations; I particularly wish to thank Sarath Balachandran, Andrew Boysen, Andrea Contigiani, John Eklund, and Kate Odziemkowska.

In addition, the broader community of strategy scholars has been wonderful with support on many different levels. In particular I wish to thank for their helpful comments on various projects Vikas Aggarwal, Olivier Chatain, Karel Cool, Felipe Csaszar, J.P. Eggers, Sendil Ethiraj, Phanish Puranam, Bart Vanneste, Maciej Workiewicz, and Peter Zemsky. I also thank for their constructive input the participants at the Consortium on Competitiveness and Cooperation in Milan, the SMS annual conferences in Berlin and Madrid, several AoM conferences, several Wharton-INSEAD doctoral cosortia, two Theoretical Organizational Models (TOM) conferences, two Transatlantic Doctoral Conferences (TADC) at London Business School, as well as seminar participants at Wharton, BI Oslo, London School of Economics, INSEAD, Copenhagen Business School, Tilburg University, Southern Denmark University, and the Frankfurt School for Finance & Management.

Moreover, I am indebted to Chris Zook and Mathijs Robbens at my previous employer Bain

& Company, without whose encouragement I probably would never have embarked on the academic journey. In hindsight I cannot imagine a better career decision; I am very grateful for them showing me the way.

Finally, I want to thank my family. My mother, who has always provided a warm, stimulating, and supportive home; my brother Biem, a companion throughout my life, unimaginable to do without; my children Ole, Berit, and Bo, providing great joy and meaning; and my wife Truke, love of my life, providing happiness, balance, and energy to jointly explore the world.

ABSTRACT

EXPLORING THE MULTI-TRILLION DOLLAR QUESTION: HOW RESOURCE DYNAMICS SHAPE LONG-TERM PROFIT PATTERNS

Phebo D. Wibbens

Nicolaj Siggelkow

Only a few dozen firms across different industries have been able to appropriate a large share of the trillions of dollars of total net value appropriation in US markets over the past decades. This dissertation examines, using formal modeling, how resource dynamics in competitive markets shape such long term profit patterns. Chapter 1 introduces the main questions and concepts, including a general formal framework that can be used to represent RBV-based theories in terms of stochastic state space models, as well as a five-step program to do so. In chapter 2 a new measure of long-term firm performance is developed: LIVA (long-term investor value appropriation). This measure helps to address a disconnect between the common theoretical assumption that managers optimize firm value, and the widespread empirical practice of measuring performance using short-term ratios such as ROA (return on assets). Chapter 3 describes a dynamic model and empirical inference of the relation between resource competition and the decomposition of variations in returns into firm-specific and industry components. The model shows that the investment dynamics for scarce resources amplify any idiosyncratic shocks to their resource positions, thus providing a very general mechanism for the empirical regularity that variations in return have a high firm-specific component. Empirical results from a Bayesian hierarchical analysis of stock market returns support the model predictions. Chapter 4 describes a formal model to analyze how profit persistence is affected by higher order resources, which are an abstract representation of dynamic capabilities. The model in this chapter indicates that higher order resources should lead to the presence of a second order autoregressive term in profit time series. Empirical estimation of the model using both classical and

Bayesian hierarchical methods indeed provides evidence for the existence of such a second order autoregressive term. The models in these final two chapters are implementations of the formal RBV framework introduced in chapter 1, and thus exemplify the value of such a framework to the strategy field.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	iv
ABSTRACT	vi
LIST OF TABLES	xi
LIST OF ILLUSTRATIONS	xii
CHAPTER 1 : Introduction	1
1.1 LIVA and its distribution	2
1.2 A state space model for resource dynamics	5
CHAPTER 2 : Introducing LIVA to measure long-term firm performance	8
2.1 Introduction	8
2.2 Literature review	10
2.3 LIVA and its properties	14
2.4 LIVA compared to other measures	20
2.5 Drivers of M&A performance	26
2.6 Case study: Circuit City vs. Best Buy	37
2.7 Discussion: LIVA vs. other measures	43
CHAPTER 3 : The dynamics of resource competition: Implications for firm and industry components of returns	48
3.1 Introduction	48
3.2 Theory	52
3.3 Model output	59
3.4 Empirical evidence	69
3.5 Discussion	83

CHAPTER 4 : Performance persistence in the presence of higher order resources	87
4.1 Introduction	87
4.2 Theory	89
4.3 Methods	99
4.4 Results	105
4.5 Discussion	115
CONCLUSIONS	118
APPENDIX	123
BIBLIOGRAPHY	136

LIST OF TABLES

TABLE 1 :	LIVA distribution	3
TABLE 2 :	LIVA example	17
TABLE 3 :	Top 10 companies 1995-2014 by LIVA and other common performance measures	22
TABLE 4 :	Biggest LIVA creating and destroying industries 1995-2014	25
TABLE 5 :	Analysis of companies with negative LIVA	27
TABLE 6 :	Regressions of LIVA and excess return on M&A characteristics	32
TABLE 7 :	Characteristics of different size groups in regressions	34
TABLE 8 :	LIVA decomposition for Circuit City and Best Buy	40
TABLE 9 :	Comparison of LIVA with other common performance measures	44
TABLE 10 :	Literature overview	54
TABLE 11 :	Base case parameter values	60
TABLE 12 :	Bayesian posterior inference for main parameters	75
TABLE 13 :	Firm-specific and industry-wide volatility by industry	77
TABLE 14 :	Analysis by decade	79
TABLE 15 :	Summary statistics and correlations	81
TABLE 16 :	Regression results	82
TABLE 17 :	Numerical examples.	94
TABLE 18 :	Summary of the effects of operating and higher order resources on profit persistence.	99
TABLE 19 :	Percentile distribution of excess profits y	106
TABLE 20 :	(Partial) autocorrelation functions.	107
TABLE 21 :	Model comparison for MLE.	108
TABLE 22 :	Model comparison for Bayesian inference.	110

TABLE 23 : Posterior inference for ARMA(2,2) model.	111
TABLE 24 : Alternative specifications for ARMA(2,2) model.	134
TABLE 25 : Bayesian inference on synthetic data.	135

LIST OF ILLUSTRATIONS

FIGURE 1 : LIVA distribution	3
FIGURE 2 : Performance measures used in <i>SMJ</i> in 2016 and number of studies that use them	12
FIGURE 3 : Cumulative LIVA over time for Circuit City and Best Buy	41
FIGURE 4 : Overview of variance component studies	49
FIGURE 5 : Time evolution of state variables and investment decisions in the model	56
FIGURE 6 : Investment efficiency function	57
FIGURE 7 : Base case equilibrium output	61
FIGURE 8 : Amplification of resource asymmetries in profit evolution	63
FIGURE 9 : Equilibrium without resource competition ($d = 0$)	65
FIGURE 10 : Firm and industry components for various parameter changes	66
FIGURE 11 : Equilibrium with scale economies and high resource cost	68
FIGURE 12 : Excess return patterns in four industries	74
FIGURE 13 : ARMA time series simulations	100
FIGURE 14 : Inference of λ -parameters by industry	112
FIGURE 15 : Inference of λ -parameters by decade	114

CHAPTER 1 : Introduction

The fundamental quest of the strategy field is to understand why some firms perform better than others, and how managers can influence those outcomes. The resource based view (RBV) is one of the core theories to explain the origins of performance variations: given that some firms perform persistently better than others, apparently the better performing firms must have access to certain resources that are not readily available on strategic factor markets (Wernerfelt, 1984; Rumelt, 1984; Barney, 1986; Dierickx and Cool, 1989; Barney, 1991; Peteraf, 1993). However, the largely verbal nature of the RBV has made this theory vulnerable to several fundamental critiques (Priem and Butler, 2001; Kraaijenbrink et al., 2010). One of the most pressing critiques is that the RBV would be tautological in nature; in short, the accusation is that resources need to explain firm value, but are also defined in terms of value to the firm, and thus do not explain much at all.

Providing a more formal fundament to the RBV would help to alleviate such critiques, as formal theories help provide rigorous and internally consistent explanations. For these reasons, several scholars have called for more formal theory in management research (e.g., Adner et al., 2009; Oxley et al., 2010). Given the importance of the RBV in the academic management literature, it is all the more surprising that only a handful of papers have attempted to formalize some of its features (e.g., Makadok and Barney, 2001; Pacheco-de Almeida and Zemsky, 2007; Jacobides et al., 2012; Knudsen et al., 2014), and even fewer have linked such formal foundations directly to empirical data (Knott et al., 2003).

With this dissertation I aspire to address this gap, and thus lay a new piece of the puzzle to understand firm performance. Specifically, I use formal modeling techniques from several academic fields, including finance (performance appraisal), industrial organization economics (dynamic games), engineering (state space models), and statistics (Bayesian inference and time series analysis) to formalize core tenets of the RBV. Moreover, I use empirical performance data to test the formally derived propositions, and thus make in-

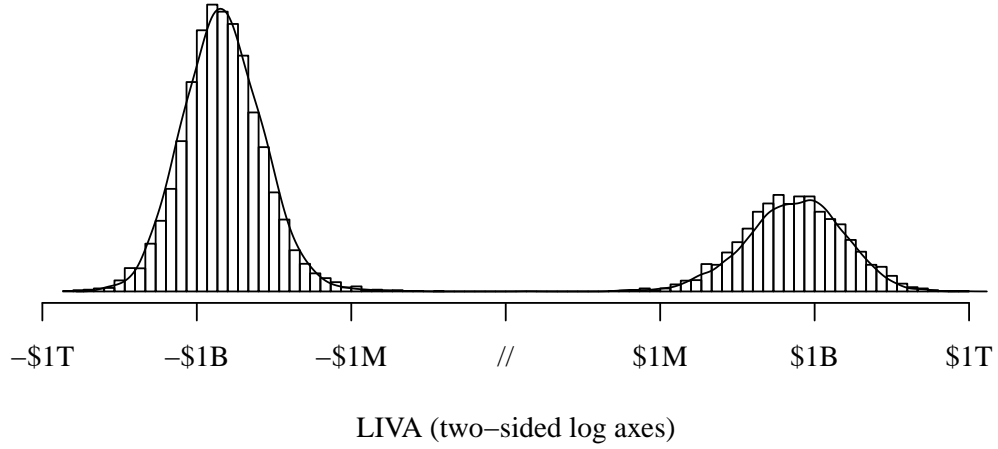
ferences about the underlying resource dynamics that drive long-term firm performance. Finally, I show how the types of models used are part of a more general formal framework that can be used as a program to develop and investigate formal models based on the RBV.

1.1. LIVA and its distribution

In chapter 2, based on a joint paper with my advisor Nicolaj Siggelkow, we introduce LIVA (long-term investor value appropriation) as a new measure of long-term firm performance. We introduce this measure to address a gap in the literature between the theoretical tenet that firms ought to optimize long-term value, in absolute terms (e.g., in dollars), and the empirical practice to measure performance using short term relative measures (e.g., return on assets, total shareholder return, or Tobin's q , expressed as a percentage or ratio instead of dollar figure). In order to develop a measure consistent with the theoretical managerial goal of value maximization, LIVA operationalizes the long-term *ex post* net present value (NPV) at the firm level, based on stock market data. Hence, LIVA is measured in monetary amounts (e.g., dollars), and its total across all firms is equal to zero—the latter property follows immediately from the definition of NPV, and the fact that the cost of capital is based on the average firm performance (see section 2.3.3 on page 18 for details).

The bars in Figure 1 show the empirically observed distribution of LIVA for all US listed firms over the 20-year period 1995 to 2014. Note that both the negative and positive x -axes of this chart are on a logarithmic scale; the distribution of LIVA is thus heavily skewed on both sides of the distribution. In fact, a mere 10 firms appropriated 2.9 trillion dollars over this period (see Table 3 Panel a on page 22 for an overview of these companies). Table 1 shows the same distribution in numbers. This table shows more detail of the asymmetry between positive- and negative-LIVA firms. Clearly, many more firms have a negative LIVA than a positive one, but this is offset by more firms having an extremely positive than an extremely negative LIVA. In chapter 2 we discuss further operational details of LIVA, how to interpret it, its potential uses in strategic management, as well as some caveats.

Figure 1: LIVA distribution



Note. Observed distribution of LIVA per firm over the period 1995-2014 (bars) and distribution function of simulated dynamic model (line). Both negative and positive x -axis are on a logarithmic scale, i.e. LIVA is transformed using the function $x \mapsto \text{sign}(x)(\log|x| + c)$ for a suitable offset c such that $\log|x| + c > 0$.

Table 1: LIVA distribution

Absolute LIVA per firm (\$B)	Negative LIVA		Positive LIVA	
	No. of firms	Total LIVA (\$B)	No. of firms	Total LIVA (\$B)
0-0.1	2,465	-109	1,288	51
0.1-1	5,295	-2,098	1,838	727
1-10	2,310	-6,413	1,235	4,106
10-100	255	-6,071	314	8,182
>100	14	-2,551	21	4,176
Total	10,339	-17,242	4,696	17,242

A more fundamental question, though, is the origin of the process that drives the at first maybe surprising distribution in Figure 1 and Table 1. It turns out that such a distribution can be derived from a surprisingly simple process, based on Gibrat’s law (Gibrat, 1931; Knudsen et al., 2017). This law states that firm growth rates do not depend on their size, and thus that the growth process follows a geometric random walk:

$$dx_t = \sigma x_t dz_t \tag{1.1}$$

For simplicity, this is a process without drift ($\mu = 0$); x_t is the size variable, σ the volatility parameter, and dz_t a Wiener process—the continuous-time analogue of a Brownian motion of accumulating independent random shocks ϵ_t with a standard normal distribution. This leads to a size distribution of x_t that is log-normal at any time $t > 0$. The LIVA distribution can be generated by assuming an initial log-normal size distribution of firms $\exp(x_0) \sim \mathcal{N}(\mu, \sigma)$, an initial LIVA of $y_0 = 0$, and a LIVA growth process equal to the size growth process $dy_t = dx_t$.¹ The line shows the distribution function for y_t generated by such a process with suitable parameters—it matches strikingly well with the observed data.

The process in equation (1.1) also explains why there are so many more firms with a negative than with a positive LIVA. Due to the geometric random walk, a firm that has had a negative LIVA ($\Delta y_1 = \Delta x_1 > 0$) over a certain period, in a future period will have a lower x_t , thus a smaller shock size $|\Delta y_2|$, and thus a lower probability to catch up on its negative negative Δy_1 (note that this process only leads to a negative median and mode for y_t ; the mean always remains zero). On the other hand, a firm that has had a positive LIVA, will in a future period receive a larger shock size, and thus get a higher probability to get to an extremely positive LIVA than an extremely negative LIVA, consistent with the empirical data in Table 1. Thus, the LIVA distribution is at a very fundamental level shaped by an evolutionary process in which more profitable firms, on average, grow faster

¹Note that the “size” variable x_t does not in general correspond to an observable size in terms of, for instance, market capitalization or balance sheet assets; rather it parameterizes the size of the standard deviation of future shocks $dy_t = dx_t$, which will certainly be related to, for instance, the market capitalization, but in general not equal or even proportional to it.

and subsequently have larger variances in profit.

1.2. A state space model for resource dynamics

What does it mean that the distribution of long-term value appropriation can be described by a very simple geometric Brownian motion? It certainly does not mean that because everything is just a random process, that no-one can influence it. Rather, it signifies that it is fruitful to view performance as driven by a dynamic stochastic process for which we need to understand what firms do to get “good shocks” dz_t .

In very general terms, one can describe the state of a firm i at a certain time t in terms of a long vector $(X_{i1t}, X_{i2t}, \dots, X_{iKt})$. This state vector should encode anything that can affect competitive positions (and hence performance) currently or in the future. Then the cumulative LIVA y_{it} of any firm i must be some function of the states of all economic actors $k = 1, \dots, K$ at that time, denoted by $\mathbf{X}_t = (X_{11t}, \dots, X_{IKt})$:

$$y_{it} = f_i(\mathbf{X}_t) \tag{1.2}$$

The evolution of the state can then be described by a stochastic process:

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t) dt + \Sigma(\mathbf{X}_t) dz_t \tag{1.3}$$

In this equation, $\boldsymbol{\mu}$ and Σ are respectively vector and matrix valued functions of the state, and dz_t is a vector of independent Wiener processes. Because y_{it} must be a zero drift process² (by definition of LIVA), it follows from Itô’s lemma that the rate of value appropriation dy_{it} is described by:³

$$dy_{it} = f'_i(\mathbf{X}_t)^T \Sigma(\mathbf{X}_t) dz_t \tag{1.4}$$

The framework in equations (1.2), (1.3), and (1.4) can provide a more formal foundation

²The requirement that y_{it} is a zero drift process puts certain constraints on the functions f_i , $\boldsymbol{\mu}$, and Σ that can be derived directly from Itô’s lemma.

³ f'_i is the gradient vector of the function f_i .

to the RBV. The firm-specific state vector $(X_{i1t}, X_{i2t}, \dots, X_{iKt})$ can be seen as a formal definition of a resource bundle. This definition corresponds to the broad concept of a resource in for instance Helfat et al. (2007) as “something that the organization can draw upon to accomplish its aims” (p. 4). Indeed, in the above framework, the resource bundles of firms both determine value appropriation (equation (1.2)) as well as the probability distribution of future resource bundles (equation (1.3)). Ultimately, these two equations determine the stochastic evolution firm profits over time. In fact, according to equation (1.4), if one wants to understand the evolution of value appropriation, one “just” needs to know the functions f_i and Σ , which describe, respectively, how firm profits relate to resource bundles, and how the standard deviations and correlations of the future evolution relate to the current resource bundles.

Of course, in practice there is no hope to of finding such functions that will be valid for all firms under all circumstances. However, these equations do provide a sufficiently general framework to develop models that focus on a specific theoretical mechanism of resource dynamics, and relate it to empirically observable performance data. Chapters 3 and 4 of this dissertation provide two such models, investigating, respectively, the effects of resource competition, and the effects of higher order resources on performance.

The model in chapter 3 describes how endogenous investments in competitive resource markets shape the emergence of firm-specific vs. industry-level components of profit heterogeneity. In the language of above framework, the state space \mathbf{X}_t consists of two firms’ resource positions in a single market q_i and q_j and an industry state a_k (which can be thought of as the collective impact of other economic actors on the industry in which the two focal firms play). The Bellman equation (3.5) is the analogue of equation (1.2), describing how economic value relates to the state space.⁴ The equations (3.2), (3.3), and (3.4) determine the state space evolution of equation (1.3)—the details are described in

⁴The naming conventions in chapter 3 are different from above: s denotes the state \mathbf{X}_t , and $V(s)$ the value function $f_1(\mathbf{X})$. Note that as required $V(s)$ is indeed a zero drift process by construction. Also note that the model in chapter 3 is formulated as a discrete-time model, but by taking a small Δt it approaches a continuous-time model to arbitrary precision.

appendix A.3. Because this model cannot be solved analytically, I use iterative econometric techniques to estimate the evolution of value. Of primary interest in this chapter is to what extent rents are firm-specific—or in the language of this chapter: to what extent dy_{1t} and dy_{2t} are uncorrelated for two actors playing in the same industry.

The model in chapter 4 describes how higher order resources shape profit persistence patterns. Higher order resources are an abstract representation of dynamic capabilities, and are defined as resources that do not directly affect current profits, but do affect profit evolution in the future. The state space \mathbf{X}_t of the model in this chapter consists of two types of resources for a single firm: an operating resource, and a higher order resource. The operating resource directly affects expected profit in the presence (equation (4.1)), while the higher order resource affects the evolution of operating resources (equation (4.3)), and thus only affects expected future profits.⁵

⁵Also the model in this chapter is formulated as a discrete time model. Note moreover that the dependent variable y_t in the model is a measure of operating profits (instead of LIVA), which will in general not be a zero drift process.

CHAPTER 2 : Introducing LIVA to measure long-term firm performance

This chapter is based on a joint paper with Nicolaj Siggelkow.

2.1. Introduction

Core theories in the strategy field, such as the resource based view, the positioning school, transaction cost economics, and value-based strategy, assume that firms and their managers seek to maximize appropriation of long-term value (Barney, 1986; Brandenburger and Stuart, 1996; Porter, 1980; Williamson, 1979). Likewise, formal models in strategy usually assume some form of maximization of profit or long-term value appropriation (e.g., Gans and Ryall, 2017; Jacobides et al., 2012; Levinthal and Wu, 2010). This core theoretical assumption contrasts with how performance is usually measured in empirical studies: an analysis of all 2016 issues of *Strategic Management Journal (SMJ)* shows that a significant majority of studies use short-term relative accounting metrics (such as quarterly or annual return on assets), instead of the long-term value appropriation that managers are supposed to maximize, even though previous literature has shown that optimizing such short-term relative accounting metrics in general does not correspond to long-term value maximization. For instance, firms that create economic value for their investors can actually have an accounting return on capital (ROC) below the cost of capital and vice versa (Fisher and McGowan, 1983). Likewise, firms that invest in initiatives that have a positive net present value might decrease their ROC (Levinthal, 1991).

To address this apparent disconnect between the theoretical performance goal of firms and the empirical measurement of it, we develop an empirical performance metric that measures whether firms created value for their investors over long time periods: LIVA (long-term investor value appropriation). We investigate its properties and derive how it relates to common performance measures in the literature, such as total shareholder return (TSR), economic profit and ROC.

In order to develop a measure that is consistent with the theoretical notion of value maximization, we define LIVA in terms of the backward looking net present value (NPV) of returns over time. We show that the definition of LIVA based on NPV is equivalent to the discounted sum of absolute excess stock market returns, and moreover that in the long run, LIVA becomes approximately equal to the sum of discounted economic profits as calculated from accounting statements. These measures converge because short-term differences between stock returns and accounting profits tend to cancel out in the long run when properly discounted: essentially, the use of different measures shifts profits forward or backward over time, but in the long run this does not affect the total.

We perform three analyses to show how LIVA can bring new strategic insights. First, we analyze rankings of the best and worst performing companies using LIVA and other common performance measures. We find that top performers in terms of LIVA correspond to often lauded companies, such as Apple, Microsoft and Wal Mart. These companies end up much lower in lists that are based on relative measures such as excess return and ROA. These latter lists are dominated by relatively small companies, because only a small base allows a stellar performance in terms of measures relative to size. Additionally, we show that LIVA provides a useful measure of corporate decline, particularly in cases of bankruptcy, when measures such as excess return are hard to define meaningfully.

Second, a regression analysis of the performance of mergers & acquisitions (M&A) shows that using LIVA typically produces more stable and more economically significant results, because regression analyses using relative measures such as ROA, ROC or TSR as dependent variable are strongly skewed towards smaller firms, with more volatile performance and less economic relevance. We show how this effect can potentially explain some of the conflicting findings relating deal frequency and deal experience to acquirer performance (King et al., 2004; Laamanen and Keil, 2008). Moreover, our LIVA analysis suggests that deal size rather than frequency is a statistically and economically more significant variable that has been largely overlooked: both very small and very large deals underperform deals that are

approximately half of the acquirer market capitalization. Moreover, mega-mergers perform particularly badly: deals over \$10B are associated with an average negative LIVA of \$26B.

Third, we perform a case study of Circuit City and Best Buy which shows how LIVA can be decomposed into specific sources of value creation and destruction. The LIVA decomposition indicates, for instance, that despite Circuit City's bankruptcy, it destroyed relatively little investor value due to its successful spin-off of Car-Max.

Finally, we discuss the advantages and disadvantages of LIVA. Advantages of LIVA are that it increases with NPV-positive investments; it accounts well for major corporate events such as mergers and bankruptcy; it measures magnitude of economic impact; and it can be decomposed into different sources. Potential disadvantages of LIVA are that it depends on stock market expectations; and it measures returns absolutely instead of relative to size. We discuss how the advantages and disadvantages compare under various conditions, and thus when LIVA can lead to new strategic insights as a complement to existing performance measures.

2.2. Literature review

Explaining organizational performance is at the heart of the strategy literature. Though there are many ways to measure performance, financial performance for investors plays a central role. The main theories of the strategy field, including the resource based view (Barney, 1991), the positioning school (Porter, 1980), transaction cost economics (Williamson, 1979), and more recently value based strategy (Gans and Ryall, 2017), all assume some form of profit maximizing behavior of firms, and seek to explain differences in the profit appropriated across firms, and ultimately their investors. Though some scholars have argued for measuring performance of other stakeholders too (Coff, 1999; Lieberman et al., 2017), measuring financial performance (for investors) has remained important in the empirical literature. For instance, out of all 119 empirical papers published in *SMJ* in 2016 using quantitative methods, 55 used some measure of financial performance, almost half of them

as a dependent variable. These numbers are consistent with an earlier review of accounting metrics in the strategy literature by Richard et al. (2009, Table 1).¹

We also looked more specifically which performance measures were used in *SMJ*. Figure 2 shows the results, split by accounting versus stock market measures, as well as by absolute measures (in dollar amounts) versus ones that are relative to size (ratios). Clearly, relative measures based on accounting returns are by far the most widely used, with return on assets (ROA) leading the pack (34 out of 55 studies²). In contrast, only four studies used an absolute measure based on accounting returns, and six studies used a relative measure based on stock market measures. Not a single paper used an absolute measure based on stock market returns. The key reasons for the ubiquity of relative measures based on accounting returns are presumably that they are easy to obtain, easy to use, and have been widely used in the past. At the same time these relative accounting measure are not without long-standing and well-known critiques.

For instance, in a widely cited American Economic Review classic, Fisher and McGowan (1983) use a formal model to investigate the theoretical relation between accounting returns and the underlying economic returns (which they define as the cash returns from investments). They find that “there is no way in which one can look at accounting rates of return and infer anything about relative economic profitability” (p. 90). They base this conclusion on several example firms with realistic underlying economic assumptions exhibiting very high accounting returns despite earning an economic return well below the cost of capital, and vice versa. Though the conclusion of this paper might be overly stark, the examples should still provide a serious exhortation to interpret accounting returns with great care.

¹Additionally, several papers in the accounting literature review and compare measures of firm performance (e.g., Goetzmann and Garstka, 1999; Ittner and Larcker, 1998). These papers mainly assess short-term measures based on accounting data that can inform managerial decision making to optimize shareholder value, while our goal is to find a long-term backward looking measure that is consistent with the managerial goal of value optimization.

²Note that the total number of studies that use a performance measure does not equal the total of all numbers in Figure 2, because some studies use multiple measures, while other studies use different measures from the ones described in the figure, such as Tobin’s q.

Figure 2: Performance measures used in *SMJ* in 2016 and number of studies that use them

Absolute	EBIT(DA): 4	<i>none</i>
Relative	ROA: 34 ROS: 6 ROE: 4 ROC: 2	TSR: 3 CAR: 3
	Accounting	Stock market

Note. Analysis of all quantitative empirical studies in *Strategic Management Journal* volume 37 (2016) that use a measure of financial returns, split by accounting and stock market based measures, as well as by relative (i.e., ratios), and absolute (i.e., in dollar amounts) measures. EBIT(DA) = earnings before interest, taxes, (depreciation and amortization); ROA = return on assets; ROS = return on sales; ROE = return on equity; ROC = return on capital; TSR = total shareholder return; CAR = cumulative abnormal returns.

Another long-standing critique is the use of relative measures (or ratios) in regressions, summarized specifically for a strategic management audience by Wiseman (2009) though the original critique is more than a century old (Pearson, 1897). Wiseman (2009) shows that the use of ratios such as ROA and ROC is likely to lead to biases and unstable results primarily due to confounding correlations in the numerator (e.g., absolute profits) and the denominator (e.g., assets). Moreover, he provides specific guidance on using absolute measures instead of ratios in regressions, and how issues that might result from using absolute measures such as the fact that bigger firms in general will have bigger absolute profits can be resolved, for instance by careful inclusion of additional independent variables (or controls) or using a weighted least squares regression.

Finally, Levinthal and Wu (2010) argue that firms ought to optimize absolute value and not relative returns. For instance, a firm that has a cost of capital of 10% and a return on capital (ROC) of 20%, will appropriate additional value for its investors if it invests in a

project earning 15% returns, despite the fact that the investment will decrease the ROC for the firm. Thus, a firm with a lower ROC might in fact appropriate more value than a firm with a higher ROC, because the latter might artificially forego positive NPV opportunities that would lower its ROC. Indeed, theoretical papers almost never use accounting ratios such as ROA: of the six theoretical models in *SMJ* in 2016 that use a performance measure, five use absolute profit and one total shareholder return (TSR) as a dependent variable.

A few empirical studies in our *SMJ* sample use measures other than accounting ratios. Four studies use Earnings Before Interest, Taxes (Depreciation and Amortization) (EBIT(DA)), an absolute measure of operating profits, sometimes citing similar reasons as above for using absolute instead of relative measures. Six studies use stock market based measures. Total shareholder return (TSR) is the buy-and-hold return of an investment: the percentage return from dividend yield and share price appreciation, with dividends reinvested. Cumulative abnormal return (CAR) is a market-adjusted TSR-based measure, indicating how well a stock has done relative to the overall market. CAR is typically used in event studies, for instance to judge the performance of mergers, acquisitions or alliances around an announcement date. An advantage of using TSR or CAR is that it immediately relates to economic returns that are realizable by investors, and thus is not prone to the Fisher and McGowan critique.

We did not include Tobin's q and related measures such as market to book ratio in our analysis of performance measures, because they are valuation measures at a certain time, while we are mainly interested in measures of profitability over a certain time period (to address the question of long-term firm performance). Note though that Tobin's q would be prone to similar critiques as relative accounting returns, because its denominator is usually based on some accounting measure of assets. Recent work in finance including intangible capital in the denominator of q (Peters and Taylor, 2017) alleviates some of the issues raised by Fisher and McGowan (1983), but is still prone to the ratio issue (Wiseman, 2009) and the relative vs. absolute optimization issue (Levinthal and Wu, 2010).

While the strategy literature is mainly concerned with measuring past performance, there exists a significant literature on investment appraisal – the evaluation of future investment opportunities. The gold standard is appraisal based on net present value, NPV (Brealey et al., 2006; Koller et al., 2010). The basic idea is that an investment should be made if and only if it is expected to return more than the cost of capital, reflected by a positive NPV. The NPV is defined as the sum of expected cash flow (CF) from an investment, discounted by a cost of capital r :

$$NPV = \sum_t \frac{CF_t}{(1+r)^t}$$

Jacobides et al. (2012) highlight the importance of NPV for the strategy field: “[...] a cautionary point is that apparently high profits (accounting net income) can derive from sunk investments that are actually failing to yield a normal rate of return. To address this issue, we view financial performance in terms of the potential gains of *investors in the firm* over longer periods, focusing on net present value (NPV) of cash flows [...]” (p. 1386; emphasis in the original). In other words, NPV measured over a long time period captures the extent to which strategic investments have been (financially) successful.

2.3. LIVA and its properties

The literature review highlights an interesting disconnect between the theoretical and empirical strategy literature regarding performance measurement: while theoretical models assume that firms maximize absolute profits, and ultimately long-term NPV, empirical papers usually assess performance using short-term ratios such as ROA, which might not properly reflect underlying economic performance, even over long time periods. In this section, we want to fill this gap by developing an empirical measure of long-term firm performance based on a backward-looking measure of NPV, using data that are publicly available for listed companies. We call this measure *long-term investor value appropriation* (LIVA).

2.3.1. Defining LIVA

To define LIVA based on NPV, we want to estimate the *ex post* discounted value of all cash flows to and from investors between two time points $t = 0$ and $t = T$. This value is equivalent to assuming that at time 0 investors invest V_0 , the value of the firm at that time; in subsequent periods t , investors receive the free cash flow generated by the firm FCF_t ; and at time T investors sell the firm for its market value V_T .³ (The free cash flow (FCF) is the cash that is available to all investors (debt and equity holders), which is equal to the cash flow from operations net of capital expenditures.) Taking the sum of the present value of these cash flows at time $t = T$ yields, with r denoting the cost of capital:

$$LIVA = V_T - V_0(1 + r)^T + \sum_{t=1}^T \frac{FCF_t}{(1 + r)^{t-T}} \quad (2.1)$$

In appendix A.1, we show that when taking the starting and ending value V_t equal to the enterprise value of the firm (i.e., the market value of equity plus the book value of debt), the above equation is equal to:

$$LIVA = \sum_{t=1}^T \frac{ER_t}{(1 + r)^{t-T}} \quad (2.2)$$

In this equation, ER_t is the excess return (the total shareholder return above the cost of equity) over period t , and MC_{t-1} is the market capitalization (the market value of all shares) at the beginning of period t . In words: LIVA is equal to the sum of the discounted absolute excess returns to shareholders over a given period.

The excess return (ER_t) per period can be operationalized in several ways. All methods are based on total shareholder return (TSR), also called buy-and-hold returns, which are equal to the dividend yield plus stock price appreciation with dividends reinvested. The

³Throughout the chapter, we use the convention that state variables at a certain point in time (such as market capitalization) are measured at the end of a period, and that flow variables (such as FCF) start in period 1. Thus, $t = 0$ reflects the moment before any flows have started.

simplest way to calculate excess return is to take the difference between the company TSR and the value-weighted market returns. More involved methods would use the capital asset pricing model (CAPM), based on which the excess returns are the residuals of a regression without intercept of TSR on the market return. Additionally, other factors than just the market return could be included; common methods in the finance literature include a three factor (Fama and French, 1993) or four factor model (Carhart, 1997).

We prefer to use the simplest method (TSR minus market average), because it has a clear interpretation, and is least prone to errors and dependencies on methodological choices (CAPM or multi-factor models, for instance, require choosing rolling time windows over which to perform the regressions, the choice of which will affect the results). Moreover, Fama and French (1997) show that in practical applications the factor loadings (i.e., the betas) usually have such high error margins that differences across companies and industries are usually not statistically significant. Indeed, we did not find material differences in outcomes across the various methods for calculating excess return, and thus for most analyses use the simplest one (details are available upon request).

2.3.2. Numerical example

Table 2 shows a numerical example of a LIVA calculation. The hypothetical firm has its initial public offering (IPO) of 1 million shares of \$10 each in year 0, acquires another firm by issuing 1 million shares at a price of \$100 each in year 10, and is acquired at a share price of \$80 in year 20. For simplicity, we assume that the shares do not pay dividends, and that the cost of capital throughout the entire period is constant at 8% per year. Thus, there are three cash flows over the entire 20-year period: investors pay \$10M in year 0, pay another \$100M in year 10, and receive \$160M in year 20. In present values (as of year 20), these are -\$47M, -\$216M and +\$160M, respectively. Adding these cash flows leads to a total LIVA of -\$103M using the discounted cash flow formula from equation (2.1).

Alternatively, one can calculate LIVA with equation (2.2) using shareholder returns. In the

Table 2: LIVA example

	Year 0 (IPO)	Year 10 (pre-M&A)	Year 10 (post-M&A)	Year 20 (acquired)	Total
Share price	\$10	\$100	\$100	\$80	
Shares outstanding	1M	1M	2M	2M	
Market capitalization	\$10M	\$100M	\$200M	\$160M	
Discounted market cap	\$47M	\$216M	\$432M	\$160M	
Discounted cash flow	-\$47M		-\$216M	\$160M	-\$103M
TSR [†]		25.9%		-2.2%	11.0%
Cost of capital [†]		8.0%		8.0%	8.0%
Excess return [†]		16.6%		-9.5%	2.7%
LIVA		\$169M		-\$272M	-\$103M

Note. Numerical example of LIVA and shareholder return calculations for a firm that has its initial public offering (IPO) for 1 million non-dividend paying shares of \$10 each in year 0, acquires another firm by issuing 1 million shares at a price of \$100 each in year 10, and is acquired at a share price of \$80 in year 20. See the text for further details.

[†]TSR, cost of capital and excess return are annualized.

first decade, the annualized excess return⁴ is 16.6%, and in the second -9.5%. Using these figures leads to a LIVA of \$169M for the first decade and -\$272M for the second. The sum of these two is again -\$103M, the same as the figure from the discounted cash flow method.

Interestingly, total LIVA over the entire 20-year period is negative, even though both the TSR (of 11% per year) and the increase in market capitalization (of 15% per year) are well above the cost of capital (of 8% per year), and annualized excess returns are positive (2.7% per year). The reason is that though the negative excess returns in the second decade (-9.5%) are smaller in magnitude than the positive excess returns in the first (16.6%), the negative returns are on a much larger base than the positive ones, due to the acquisition. Hence, despite the fact that an investor holding a constant set of shares over the entire time period would have made a better return than the market, the company destroyed value for the investor base as a whole.

⁴This uses the continuous compounding equation for calculating excess return: $ER = (1+TSR)/(1+r) - 1$. The LIVA for each decade can then be calculated using the total, non-annualized excess returns. For instance, for the first period, the total excess return is $(1 + 16.6\%)^{10} - 1 = 363\%$, which multiplied by the initial discounted market cap of \$47M leads to a LIVA of \$169M in the first decade. Alternatively, one could employ equation (2.2) to the discounted market cap and excess returns in each individual year, leading exactly to the same results.

This example is not a mere theoretical possibility. For instance, in the merger of AOL and Time Warner, both companies had positive excess returns (5% per year for Time Warner from 1995 until 2000 when they were acquired, and 11% per year for AOL for the entire time period from 1995 to 2014). However, an analysis of LIVA shows that AOL destroyed more than \$150 billion of value, and still over \$100 billion after taking into account the value created for Time Warner shareholders. The reason for this result is similar to the above example: unlike excess return, the LIVA calculation takes into account that AOL's positive returns were on a much smaller base than its collapse after the merger.

2.3.3. Interpreting LIVA

The interpretation of LIVA is very similar to the interpretation of NPV. For instance, over the 20-year period 1995-2014, Apple generated a LIVA of \$658 billion. This means that Apple investors earned 658B more than if they had invested their money in a market-wide index over the same period. More precisely, if an investor had bought the outstanding shares of Apple in 1995 at the market price, borrowing the money at a cost equal to the average market return, sold all Apple shares again at the end of 2014 at the market price, and using that money as well as any intermediate cash returns (dividends and share buy backs) to repay her "debt", then she would have had 658B left in her bank account.

One important property to keep in mind when interpreting LIVA is that its average across all firms is zero. Though at first maybe surprising, this property follows immediately from the definition in equation (2.2): because excess return is a comparison to the market average, the (value-weighted) average excess return must be zero (hence the name excess return), and hence its discounted sum over time LIVA must be zero on average as well. Note that this property does not hinge on any assumptions about efficient markets, but rather results from the fact that returns are compared to the cost of capital, and the average cost of capital is by definition equal to the average market return. Because the average deviation from an average is always zero, the average return above the cost of capital is always zero, and thus the average LIVA is zero.

2.3.4. *LIVA and accounting returns*

Though we have defined LIVA in terms of shareholder returns, it is insightful to relate it to accounting returns. Intuitively, the two must be related, because ultimately the shareholder cash distributions (dividends and share buy backs) must be paid by the operating profits of the firm. In appendix A.1, we derive the relation between LIVA and a commonly used accounting-based operationalization of economic profit (EP):

$$LIVA = [EV_T - BV_T] - [EV_0 - BV_0](1 + r)^T + \sum_{t=1}^T \frac{EP_t}{(1 + r)^{t-T}} \quad (2.3)$$

In this equation, EV_t is the market value and BV_t the book value of the firm's operating assets. The economic profit EP is operationalized based on the net operating profit after tax (NOPAT)⁵ (Hawawini et al., 2003)⁶:

$$EP_t = NOPAT_t - r BV_{t-1} = (ROC_t - r)BV_{t-1} \quad (2.4)$$

The final part of this equation relates after tax return on capital (ROC) to economic profit. Substituting for EP in equation (2.3) yields:

$$LIVA = [EV_T - BV_T] - [EV_0 - BV_0](1 + r)^T + \sum_{t=1}^T \frac{(ROC_t - r)BV_{t-1}}{(1 + r)^{t-T}} \quad (2.5)$$

The summation in equation (2.5) is thus the accounting analogue of equation (2.2). The sum in equation (2.5) is over *accounting* returns above the cost of capital times accounting value of assets, while in equation (2.2) it is over *market* returns above the cost of equity times market value of equity. Finally, it is interesting to evaluate equations (2.1), (2.2), (2.3), and (2.5) over a firm's entire lifetime. This leads to particularly simple equations,

⁵NOPAT = EBIT (earnings before interest and taxes) minus income tax.

⁶Note that Hawawini et al. (2003) use the term capital employed (CE) instead of book value (BV); both reflect the balance sheet valuation of the net operating assets (i.e., operating assets minus operating liabilities, equal to balance sheet debt plus equity).

because before the inception of the firm and after its liquidation (e.g., through a take-over or bankruptcy) the firm's book and market values are zero:

$$LIVA = \sum_{t=1}^T \frac{FCF_t}{(1+r)^{t-T}} = \sum_{t=1}^T \frac{ER_t MC_{t-1}}{(1+r)^{t-T}} = \sum_{t=1}^T \frac{EP_t}{(1+r)^{t-T}} = \sum_{t=1}^T \frac{(ROC_t - r)BV_{t-1}}{(1+r)^{t-T}} \quad (2.6)$$

In other words, the LIVA over the entire lifetime of a firm is equal to the sum of discounted cash flows, equal to the sum of discounted (absolute) excess returns, equal to the sum of discounted economic profits, and equal to the sum over excess accounting returns times book value.

This equation sheds some further light on the relations between cash, market returns and accounting returns: when properly measured and discounted, the only difference between these measures is that they shift value backwards or forwards over time, without affecting the total. Moreover, this shows the origin of the leading two terms in the right hand side of equations (2.3) and (2.5): the differences between book and market value at time $t = 0$ and $t = T$ adjust for the fact that not all value as priced in the market (which is partly based on future expectations) might have been attributed in the accounting statements.

2.4. LIVA compared to other measures

In this section, we compare LIVA to common accounting and shareholder return measures. The purpose of this section is to show where and how LIVA can potentially lead to new strategic insights. The analyses in this section are based on a LIVA database that we have created for all US listed firms over the period 1995 to 2014, using the definition in equation (2.2) with monthly data from the CRSP database, for a total of over 15,000 company listings. Excess returns are calculated using the same data, and are annualized over the entire period, using a standard compounding procedure:

$$\text{Annualized Excess Return} = \left(\prod_{t=1}^T \frac{1 + TSR_t}{1 + r_t} \right)^{\frac{1}{n}} \quad (2.7)$$

In this equation, TSR_t is the monthly total shareholder return, r_t is the monthly average market return, and n is the number of years over which the return is calculated. Finally, return on capital is calculated with after tax profits, using Compustat data.

2.4.1. Finding top performing firms

Case studies provide an important method of inquiry for management research (Eisenhardt, 1989; Siggelkow, 2007). One way of selecting case studies is based on long-term performance, in order to find examples of particularly effective (or ineffective) strategic management. To assess the impact of different performance measures on finding top performing firms, we look for the top 10 firms in our database using several performance measures. Of course, a similar procedure can be employed for specific selections of firms, for instance within an industry or geographic region.

Table 3 shows the results for all companies in our database for which data were available for at least five years of the twenty-year period 1995-2014 and had at least \$10 million of balance sheet capital (in order to exclude companies with negative or extremely low capital, leading to unrealistically high ROA and ROC). TSR and Excess Return are annualized measures, using equation (2.7), and ROA and ROC are weighted annual averages over the period of data availability. Among the names on the LIVA list are many of “the usual suspects”: firms that have been in the news because of their successes, such as Apple, Exxon and Microsoft.

The shareholder-return-based top 10 lists in panels (b) and (c) are notably different from the LIVA list. The companies on these lists were certainly quite successful, but were only able to reach these very high returns because they started out so small: Optical Coating Lab had a market capitalization of \$55 million in 1995, and IDEC pharmaceuticals (the

Table 3: Top 10 companies 1995-2014 by LIVA and other common performance measures

a. LIVA (\$B)		d. Return on assets (ROA)	
Apple Inc	658	San Juan Basin Royalty Tr	167%
Exxon Mobil Corp	406	Bp Prudhoe Bay Royalty Trust	77%
Microsoft Corp	338	Cboe Holdings Inc	72%
Altria Group Inc	294	Texas Pacific Land Trust	68%
Genentech Inc	241	Great Northern Iron Ore Ppty	66%
International Business Machs Cor	210	Mesa Royalty Trust	64%
Wal Mart Stores Inc	194	Pzena Investment Management	62%
Johnson & Johnson	183	Terra Nitrogen Co -Lp	61%
Oracle Corp	181	Life Partners Holdings Inc	61%
Warner Lambert Co	151	Cross Timbers Royalty Trust	58%

b. Total shareholder return		e. Return on capital (ROC)	
Optical Coating Lab Inc	123%	San Juan Basin Royalty Tr	194%
Netoptix Corp	99%	Dun & Bradstreet Corp	99%
U T I Energy Corp	87%	Great Northern Iron Ore Ppty	87%
Burr Brown Corp	85%	Credit Suisse Usa Inc	83%
S D L Inc	84%	Terra Nitrogen Co -Lp	80%
Jones Pharma Inc	78%	Domino'S Pizza Inc	78%
Virtus Investment Partners Inc	76%	Bp Prudhoe Bay Royalty Trust	78%
Metal Management Inc	70%	Mesa Royalty Trust	75%
Medis E Ltd	67%	Lorillard Inc	75%
T V Guide Inc	63%	Pzena Investment Management	72%

c. Excess return	
Optical Coating Lab Inc	77%
Netoptix Corp	61%
U T I Energy Corp	59%
S D L Inc	57%
Metal Management Inc	50%
Advance P C S	49%
Burr Brown Corp	48%
Virtus Investment Partners Inc	48%
Jones Pharma Inc	45%
Medivation Inc	44%

Note. Top 10 lists based on a database of all listed US firms with at least \$10M balance sheet capital, and at least five years of data available. All relative measures are annual.

predecessor of Biogen IDEC) of just \$29 million. As a comparison, Apple was already a \$5 billion company in 1995, and it would have needed an impossibly high market capitalization of \$12 trillion in 2014 to reach the same growth rate as IDEC. Thus, the top positions on the shareholder return lists not only reflect a very strong performance but also a very small starting point.

Moreover, LIVIA takes size changes better into account than shareholder returns. For instance, Exxon Mobil is second on the LIVIA list, while having a shareholder return of only 2% above the market average. Over the last decade of the analysis, it even had a negative excess return, but still a positive LIVIA, because it had a significantly larger capital base when it realized positive returns than when it realized negative returns. This is the exact reverse of the example in Table 2.

Similar to the lists using market return measures, the top performers on the accounting return lists in panels (d) and (e) are typically relatively small companies. Particularly the extremely high ROA and ROC of most companies do not reflect very high profits, but rather very low balance sheet assets. For instance, the top performing company in terms of ROA and ROC is San Juan Basin Royalty Trust, a company that distributes profits from oil & gas royalty rights, which have not been capitalized on the balance sheet to their economic value. Thus, the high performance in terms of ROC and ROA appears to be as much a reflection of particular accounting conventions as of good underlying economic performance.

These examples show some of the potential advantages of using LIVIA when assessing long-term firm performance: it measures the absolute size of economic performance without favoring initially small companies (unlike relative shareholder return measures), it values both profits and growth (unlike accounting based return measures), and it does not hinge on accounting definitions.

2.4.2. Industry performance

Instead of applying LIVA to individual companies, it is also possible to assess industry performance. How correlated are “attractive” industries (industries usually defined as having a high ROC) with industries that have created a lot of value in absolute terms? Table 4 shows the industries with the most extreme values for LIVA: those with LIVA above \$100 billion or below minus \$100 billion over the period 1995 to 2014.⁷ The excess return has been calculated based on a value-weighted industry index, and the ROC is a weighted average over the entire period.

First comparing ROC and excess return for these industries, clearly the value appropriating industries on average had higher accounting returns than the value destroying industries. However, there are clear deviations too. For instance Biotechnology, which had one of the highest excess returns (7% per annum), had a relatively low return on capital (8%, which is 2% below the overall average of 10% over this period). These deviations can either be due to accounting regulations that do not properly capitalize assets on the balance sheet (such as brand value of firms in Household Products), or due to deviations between market and book value of the firms in an industry at the beginning or the end of the period.

The deviations between LIVA and excess return are primarily due to size. For instance, though the Oil & Gas industry realized “just” a 1% annual excess return, its economic impact was very large due to the sheer size of this industry. Particularly interesting in this respect is the enormous value destruction of telecommunication services and equipment firms: the companies in these two industries destroyed more than \$2 trillion of value over this period, not being able to obtain economic gains from the enormous investments that were needed in the transition to mobile.

⁷The Global Industry Classification Standard (GICS) by Standard & Poor’s is used for aggregation, because this system reflects current competitive arenas much better than the decades-old Standard Industry Classification (SIC) codes.

Table 4: Biggest LIVAs creating and destroying industries 1995-2014

GICS	Industry	LIVA (\$B)	Excess return	ROC
101020	Oil and Gas	715	1%	10%
352010	Biotechnology	704	7%	8%
352020	Pharmaceuticals	630	1%	17%
302030	Tobacco	421	5%	24%
452020	Technology Hardware	374	4%	13%
351010	Health Care Equipment	356	2%	14%
301010	Food Retailing	298	1%	11%
201010	Aerospace, Defense	297	2%	13%
451020	IT Services	207	2%	16%
303010	Household Products	182	0%	18%
402020	Consumer Finance	161	4%	8%
351020	Health Care Providers	154	0%	11%
402010	Diversified FS	148	1%	8%
451030	Software	135	2%	14%
302010	Beverages	131	0%	15%
201060	Machinery	128	2%	12%
255040	Specialty Retail	128	1%	13%
402030	Capital Markets	112	4%	14%
...				
401020	Thriffs, Mortgages	-103	-7%	7%
401010	Banks	-108	-1%	8%
151030	Containers, Packaging	-124	-3%	10%
252010	Household Durables	-136	-2%	6%
551050	Ind. Power Producers	-169	-8%	7%
251010	Auto Components	-182	-4%	11%
403010	Insurance	-188	-1%	8%
501020	Wireless Telco	-219	-7%	6%
452030	Electronic Equipment	-262	-2%	6%
251020	Automobiles	-263	-2%	6%
254010	Media	-274	0%	6%
202010	Commercial Services	-336	-4%	10%
451010	Internet Services	-466	0%	3%
151040	Metals, Mining	-537	-6%	11%
501010	Diversified Telco	-1163	-5%	9%
452010	Communications Equipment	-1252	-4%	7%

Note. GICS = Global Industry Classification Standard, a classification by Standard & Poor's.

2.4.3. Analyzing value destruction

As a final example, Table 5 shows an analysis of value destruction. Panel (a) lists the 20 companies with the most negative LIVA over the period 1995-2014. Consistent with the industry analysis in Table 4, many of the companies that destroyed value are in the telecom sector. Striking is the difference between the LIVA measures and excess returns. Some companies even have positive excess returns but negative LIVA (AOL TimeWarner and Pfizer) due to overly aggressive and costly expansion strategies. Moreover, LIVA provides a much more meaningful measure than shareholder returns can give in case of bankruptcy: in that case the TSR and excess return are always -100%, while LIVA provides the magnitude of the actual value destroyed.

Panel (b) of Table 5 shows an aggregation of all companies with negative LIVA by end of listing type. Interestingly, only 6% of value destruction was by companies that went bankrupt. In fact, more than 40% of value was destroyed by companies that are still active, and another 30% by companies that had been acquired. Note that this type of analysis would be difficult to perform with other measures: as shown earlier, both accounting measures as well as relative shareholder return measures typically do not reflect well the value destroyed in major corporate events such as acquisitions and bankruptcy. These results also suggest that LIVA might be a useful addition to bankruptcy and dissolution measures used regularly in industry life-cycle and population ecology studies: If we consider value destruction a form of “failure”, then most of the failure is actually experienced by firms that do not exit.

2.5. Drivers of M&A performance

In this section we show how LIVA can bring new strategic insights in a large-scale regression setting. We decided to re-analyze previously identified drivers of M&A performance, because LIVA potentially can account better for the long-term impact of the major cash transactions involved in mergers, as exemplified in Table 2. Moreover, LIVA is very suitable to assess long-term economic impact.

Table 5: Analysis of companies with negative LIVAs

a. Bottom 20 performers

Company	LIVA (\$B)	Excess return
A T & T Corp	-320	-13%
American International Group Inc	-290	-15%
Bank Of America Corp	-227	-4%
Citigroup Inc	-217	-7%
Motorola Solutions Inc	-206	-8%
Worldcom Inc	-204	Bankrupt
Lucent Technologies Inc	-175	-16%
Time Warner Inc	-152	11%
Wachovia Corp	-150	-11%
Nortel Networks Corp	-136	Bankrupt
J D S Uniphase Corp	-135	-3%
A T & T Inc	-126	-3%
General Motors Corp	-108	Bankrupt
Sprint Nextel Corp	-106	-9%
Daimler A G	-98	-4%
Pfizer Inc	-98	1%
C B S Corp	-93	-2%
Qwest Communications Intl Inc	-93	-4%
Coca Cola Co	-90	-1%
Electronic Data Sys Corp	-86	-12%

b. Analysis by end of listing type

Type	LIVA (\$B)	Of total
Bankrupt	-1,075	6%
M&A	-5,804	33%
Active	-7,385	42%
Other	-3,444	19%

The drivers of M&A performance have been widely studied in prior literature. In a pivotal paper, Haleblan and Finkelstein (1999) theorized and empirically confirmed that organizational learning processes would lead to a U-shaped relation between acquisition experience and acquirer performance. Later papers in this literature, however, have led to conflicting results regarding this finding (e.g., Laamanen and Keil, 2008; Zollo and Singh, 2004), while a meta-analysis did not find evidence of a significant relation between acquisition frequency and performance (King et al., 2004). Moreover, this meta-analysis “indicate[d] that [yet] unidentified variables may explain significant variance in post-acquisition performance” (p. 187). In this section, we re-analyze the relation between acquisition frequency and performance using LIVA, and assess potentially overlooked variables that could explain M&A performance. Moreover, we assess potential reasons for the conflicting findings across earlier studies, which usually used excess return, ROA, or related measures of performance.

2.5.1. Method

Following Laamanen and Keil (2008), we use a panel approach to analyze the performance of M&A activity in a given period (instead of an event study to analyze specific transactions). As antecedents of performance we use several of the most widely used variables in the M&A literature, such as deal frequency, deal size, and industry relatedness (Capron and Pistre, 2002; Haleblan and Finkelstein, 1999; King et al., 2004; Laamanen and Keil, 2008; Zollo and Singh, 2004).

Specifically, we use a panel by firm-year for all US listed firms that are in the CRSP database⁸ in the years 1990 to 2007—the latter cutoff is to ensure that we have at least 10 years of subsequent performance data available for each observation. This leads to 122,771 observations for 15,551 firms. Each firm is identified by its unique permco identifier in the CRSP database, which is then associated with one or more stock market listings (by permno) and all its CUSIP identifiers over the entire period. The CUSIP identifiers are

⁸We include only common stock with CRSP share code 10, 11 or 12, thus excluding any American depository receipts (ADRs), mutual funds, trusts, etc.

then matched to the ultimate acquiring parent⁹ for all mergers and majority acquisitions recorded in the Thomson SDC Platinum database that are marked as completed, for a total of 68,626 deals.

As dependent variable (DV) we use *LIVA*, and compare it with *Excess return*, a commonly used stock based measure of firm performance (e.g., Laamanen and Keil, 2008), similar to cumulative abnormal returns. Both measures are based on monthly total shareholder returns compared to the average return for all firms in the database, aggregated to 1-year, 3-year, and 10-year totals. *LIVA* is calculated using equation (2.2), while excess return is calculated as the compounded excess return (i.e., the dividend-adjusted “buy and hold” return in excess of the market average) over each period. To minimize any selection bias due to dropped observations, *LIVA* and excess return are set to zero after a firm is delisted, because, by definition, the average *LIVA* and the weighted average excess return across firms are zero. Thus, both of these performance measures capture the impact over the entire period of listing within the observation window.

We use several independent variables, all based on prior literature. Relating to the number of deals we use *Deal frequency* and *Deal experience*. Deal frequency is defined as the logarithm of one plus¹⁰ the number of deals in the focal year, while deal experience is the logarithm of one plus the total number of deals recorded in the prior five years. The *No deals* dummy captures if firms have no recorded deals in a certain year.

Absolute deal size and *Relative deal size* capture the size of all deals in a year in terms of the equity value in the SDC database. The absolute deal size is the logarithm of one plus the equity value deflated by the total market capitalization of all firms (at the beginning of that year), to account for changing monetary values of firms over time. The relative deal size is the equity value divided by the beginning of year market capitalization of the

⁹Matching to the acquiring entity or its immediate parent does not materially affect the main results.

¹⁰For variables that can be zero, we use the logarithm of one plus the variable to ensure that the logarithm is always defined. We then define a dummy for the variable being zero in order to separate out this potentially special case in the analysis.

acquiring firm (the latter extracted from the CRSP database). This value is maximized at one, to exclude the effect of reverse take-overs (for which the equity value of the acquisition is larger than the acquirer market capitalization) and also reduce the effect of some extreme outliers (acquiring firms with extremely small market capitalizations compared to the equity values of the acquisitions)¹¹. The *Deal size > market capitalization* dummy captures these reverse take-overs separately. Finally, The *No deal size* dummy captures missing equity values (which are set to zero).

Furthermore, *Industry change* captures the percentage of deals in a year for which the target had a different mid-level industry classification than the acquirer. *With advisor* captures the percentage of deals with an advisor. *Cash payment* captures the average cash payment, weighted by deal equity value. *Hostile deals* captures the percentage of deals listed as hostile.

Finally, we control for the logarithm of *Market capitalization*. Also, we include industry-year fixed effects (based on the CRSP 3-digit SIC codes). These control for all time-varying and time-constant effects at the industry level, such as differences in acquisition patterns, cost of capital, growth rates, and exposure to the economic cycle. We perform ordinary least squares (OLS) regressions for both dependent variables for the three different time horizons on all independent variables and controls, for a total of six regression models. We use cluster-robust standard errors to account for any potential serial correlation and heteroscedasticity; this is especially important for the models with three- and ten-year performance data, because the firm-year panel will in these cases have overlapping performance data for multiple within-firm observations, and thus almost certainly exhibit serial correlation. The cluster-robust standard errors ensure that this serial correlation does not incorrectly affect the reported significance levels.

¹¹Fewer than 1% of observations have a target equity value larger than the acquirer market capitalization, but the equity value can be almost 1000 times as large as the market capitalization for some observations.

2.5.2. Results

Table 6 shows the results of the regressions. Focusing on the LIVA regressions in models (1) to (3) first, the direction of the effects is the same for most variables across the three time horizons, with the effects becoming stronger and more significant over time. For the ten-year LIVA regression, in model (3), the main results are summarized below.

A one standard deviation increase in *Deal frequency*, which corresponds to 56% more deals in the focal year, is associated with a \$3B increase in LIVA ($p = 0.01$). The *No deals* dummy is weakly associated with an increase in LIVA, of \$2B ($p = 0.09$).

Deal experience is positive ($p = 0.01$) and its square negative ($p = 0.03$), indicating an inverted-U relation. The maximum of the inverted-U is at an average of 3.1 recorded deals over the past five years, which is associated with a \$1B increase in LIVA compared to no deals, while more than 9.7 recorded deals are associated with a decrease in LIVA.

Relative deal size is positive ($p = 0.009$) and its square negative ($p = 0.02$), again indicating an inverted-U shape. The maximum of the inverted-U is at a deal value that is 59% of the acquirer market capitalization, which is associated with a \$7B increase in LIVA compared to a relative deal size close to zero. *Absolute deal size* is significantly negative ($p = 0.002$). A separate regression (unreported), with the continuous deal size variables replaced by a single dummy, indicates the economic significance: deal values above \$10B are associated with a negative LIVA of \$26B ($p = 8 \cdot 10^{-4}$).

Industry change is associated with a \$1.5B decrease in LIVA ($p = 0.01$). *With advisors* is associated with a \$2.5B decrease in LIVA ($p = 0.002$). Full *Cash payment* is associated with a \$2.0B increase in LIVA ($p = 0.04$).

While the three LIVA regressions exhibit a relatively consistent pattern across the different time horizons, the excess return regressions in models (4) to (6) are more erratic. Unlike the LIVA regressions, there is no clear progression across the horizons, with some variables

Table 6: Regressions of LIVA and excess return on M&A characteristics

	LIVA			Excess return		
	1 year (1)	3 year (2)	10 year (3)	1 year (4)	3 year (5)	10 year (6)
Deal frequency [†]	-0.00 (0.39)	1.27 (0.85)	4.46** (1.72)	6.84*** (1.37)	4.71 (2.67)	2.76 (3.85)
No deals [‡]	0.11 (0.24)	0.64 (0.55)	1.88 (1.09)	-1.00 (1.63)	0.20 (2.39)	-2.17 (3.18)
Deal experience [†]	0.32* (0.15)	0.47 (0.30)	1.77* (0.71)	0.23 (0.74)	2.84 (1.69)	11.87*** (3.17)
Deal experience [†] square	-0.17* (0.08)	-0.19 (0.16)	-0.78* (0.36)	-0.43 (0.25)	-0.56 (0.58)	-4.10*** (1.16)
Relative deal size	5.93* (2.58)	10.26* (4.86)	23.43** (9.00)	38.45*** (10.44)	9.31 (17.82)	34.00 (26.39)
Relative deal size square	-6.52* (2.96)	-10.88 (5.56)	-20.00* (8.77)	-23.89 (14.12)	-5.96 (20.28)	-33.59 (31.03)
Absolute deal size [†]	-0.30 (0.56)	-3.43* (1.40)	-7.41** (2.42)	-2.47* (1.21)	1.99 (1.76)	-1.88 (2.47)
No deal size [‡]	0.75* (0.32)	1.03 (0.61)	2.12 (1.17)	4.34** (1.41)	9.18*** (2.63)	19.42*** (3.94)
Deal size > market capitalization [‡]	1.63 (0.92)	3.50 (2.02)	3.68 (2.49)	34.12*** (8.32)	11.65 (9.05)	12.83 (11.91)
Industry change	0.14 (0.15)	-0.46 (0.28)	-1.50* (0.59)	0.47 (0.33)	-1.80** (0.68)	-1.30 (1.04)
With advisor	0.13 (0.21)	-0.24 (0.37)	-2.45** (0.79)	-0.00 (0.46)	0.98 (0.89)	-1.05 (1.30)
Cash payment	0.40 (0.44)	1.31 (0.79)	1.99* (0.97)	-3.78* (1.47)	10.55** (3.48)	17.39*** (4.12)
Hostile deals	-6.08 (3.33)	3.69 (2.71)	5.83 (7.70)	-6.22 (5.09)	-8.35 (9.58)	-8.41 (15.92)
Market capitalization [†]	0.00 (0.02)	0.08 (0.06)	0.43** (0.15)	-2.52*** (0.19)	-3.62*** (0.41)	-0.40 (0.86)
Industry x year FE	X	X	X	X	X	X
Observations	122,771	122,771	122,771	122,771	122,771	122,771
R ² (full model)	0.05	0.06	0.08	0.14	0.13	0.10

Note. OLS regression of LIVA and excess return percentage on merger and acquisition characteristics by firm-year. Clustered standard errors robust to serial correlation are in parentheses.

[†]Indicates logged variable; [‡]Indicates dummy variable

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

highly significant in the one-year regression, and others in the three- or ten-year regressions. Also, effect sizes fluctuate strongly across time horizons, and sometimes their signs even change. Again, we summarize the main results.

Deal frequency is highly significant in the one-year excess return regression (positive, $p = 6 \cdot 10^{-7}$), but its significance strongly decreases for the longer time periods ($p = 0.08$ and $p = 0.47$, respectively). On the other hand, *Deal experience* is positive ($p = 1 \cdot 10^{-4}$) in the ten-year regression, and its square negative ($p = 4 \cdot 10^{-4}$), with strongly diminishing effect size and significance in the shorter time periods.

Relative deal size and its square are more significant in the shorter time horizon, becoming insignificant for the 10-year analysis ($p = 0.2$ and $p = 0.3$, respectively). *No deal size* becomes more pronounced over the longer time horizon, while *Deal size > market capitalization* over shorter time horizons again.

Industry change changes sign, and with strongest result a negative for the 3-year analysis ($p = 0.008$). *Cash payment* changes sign too, with the strongest result a positive over the 10-year time horizon ($p = 2 \cdot 10^{-5}$).

2.5.3. Discussion

One conclusion of the regressions is that excess return as a dependent variable can behave rather erratically, often producing inconsistent results both across time horizons and vis--vis the LIVA regressions. Earlier studies have found such conflicting results too. For instance, Laamanen and Keil (2008) note "... that although the acquisition rate with our three-year period of observation on average contributes negatively to performance, the longer term 10 to 13-year performance of the most active acquirers is significantly higher than the performance of the less frequent acquirers." (p. 670). Moreover, though individual studies often find significant effects of, for instance, deal experience, a meta-analysis across studies did not find significant effects on abnormal returns across studies for most of the commonly applied variables, including deal experience, payment method, and relatedness (King et al.,

Table 7: Characteristics of different size groups in regressions

Firm market capitalization	Number of observations	Percentage of total		
		Market capitalization	LIVA sum of squares	Ex. ret. sum of squares
<\$100M	35,648	0.4	0.0	50.0
\$100M-\$1B	52,235	4.6	1.1	37.9
\$1B-\$10B	27,430	20.1	8.0	10.9
\$10B-\$100B	6,764	42.8	34.0	1.2
> \$100B	694	32.2	56.9	0.0

Note. Sum of squares are shown for 10-year return measures, and provide an indication of the relative weight in a regression.

2004).

What explains the unstable results of excess returns across regressions? A potential answer can be found by assessing the weight in the regressions of different firm size groups. Because OLS minimizes squared residuals (i.e., $\sum(y_i - \hat{y}_i)^2$), the sum of squares $\sum(y_i - \bar{y}_i)^2$ of a group of observations provides an indication of its relative weight in a regression for a specific dependent variable y . Table 7 provides an overview of the relative percentage of the sum of squares across size groups for ten-year LIVA and ten-year excess returns, compared to the percentage market capitalization (which is a proxy for economic significance). The right column indicates that firms smaller than \$100M make up 50% of the sum of squares of ten-year excess returns, and firms smaller than \$1B almost 90%, even though this latter group covers just 5% of the market capitalization of all firms. Thus, regression results using excess returns as dependent variable are in a large part driven by effects in small firms. Because smaller firms have larger variance in relative performance, such as excess return, this provides a plausible cause for the unstable regression results: relatively few small firms with extreme performance can have a large influence on the regression coefficients.

The picture for LIVA is markedly different. Because LIVA is an absolute measure of performance, its variance is largest for large firms. According to Table 7, observations with a market capitalization larger than \$10B contribute 90% to the sum of squares in a regression, while their market capitalization constitutes about 75% of the total (in contrast, the

excess returns sum of squares for this group is just over 1%, so the results of these economically most significant firms hardly contribute to regression results using excess returns as a dependent variable).¹²

Another insight from this analysis is the importance of using long-term return measures. For one-year LIVA, most coefficients are relatively small and statistically only marginally significant, becoming larger and more significant for longer time horizons. This suggests that only over longer time periods the market correctly prices the full impact of M&A, and provides caution for the practice of using announcement returns, which are typically analyzed over time horizons of at most weeks or months, while our analyses suggest that many effects only become apparent after years or even a full decade.

Regarding the specific results of the M&A regression using ten-year LIVA, the deal frequency and industry change regression coefficients are mostly consistent with a capability-based perspective Zollo and Singh (2004). Firms performing more deals in the present year (*Deal frequency*), are associated with higher LIVA presumably because they gain more experience and thus become more skilled in doing acquisitions. Also, the negative performance associated with *Industry change* is consistent with a capability perspective: less similarity between target and acquirer would offer less potential for capability transfer and synergies.

Much more surprising vis--vis prior literature is the inverted-U relation between deal experience and performance, while, for instance, Haleblan and Finkelstein (1999) found the opposite U relationship. Though both the statistical and economic significance of these effects are much lower than the deal size effects discussed below, these findings clearly merit further research, given their theoretical and practical importance.

Potentially even more surprising is the large statistical and economic significant effect of deal size, with both small and large deals associated with lower LIVA. Very large deals (above

¹²Note that weighted least squares (WLS) could be used to adjust the weighting of different size groups when using, for instance, excess return as DV. However, this would not account for the long-term effects of significant changes in size (as exemplified in Table 2), and neither provide an estimate of the absolute deal size.

\$10B) are associated with particularly strong negative performance. A potential mechanism for the lower value of smaller deals could be that they take too much effort relative to their potential contribution to the firm. A potential mechanism for the lower value of (very) large deals could be that they are too difficult to execute successfully, particularly regarding the post-merger integration. Also, these large deals might be at least in part driven by managerial hubris and investment bankers' incentives for deal fees, which would be consistent with the negative LIVA associated with the use of advisors. Given the economic importance of large M&A deals, this appears to us a fruitful direction for further research.

A potential concern with the results from these analyses could be that there are many other effects that influence performance that are not included in the analysis, especially over longer time horizons. However, as long as these effects are not correlated with the covariates of interest, they only add noise to results, which would weaken rather than strengthen the results over longer time horizons. Still, as with any regression analysis, a key concern could be that such correlations might exist¹³—for instance, firms performing very large deals might underperform because they had inferior competitive positions to begin with rather than due to the execution of the deals. For that story to have an impact on the LIVA regressions though, the information about the *ex ante* inferior position should not have been known to the market at that time, and only have become known during the execution of the deal.

Summarizing, using LIVA as a dependent variable in a regression can offer three important benefits vis--vis excess returns or other relative performance measures such as ROA. First, results using LIVA are for the largest part driven by the economically most significant firms. Second, because results for the larger firms are less volatile, regression results are typically more stable across different specifications. Third, using LIVA produces easy-to-interpret

¹³Note that even event studies using short-term return measures are not immune to such omitted variable bias. The reason is that under the efficient market assumption on which event studies are based, events with certain characteristics could act as signals for other effects. Thus, a particular direction in which the share price moves upon an event (e.g., a merger) could be not due to the effects of that event, but rather due to effects that are signaled by the event but not directly related to the event itself.

effect sizes to assess economic significance, such as the result that a deal size above \$10B is associated with a negative LIVA of \$26B. This is a novel finding thanks to the use of LIVA, as most deal size coefficients are insignificant when using excess return as a dependent variable.

2.6. Case study: Circuit City vs. Best Buy

In this section, we want to illustrate how LIVA can be insightful for strategic analysis of a case study. We have selected Circuit City as a focal company, because it went bankrupt and thus shows how LIVA allows to provide a measure of performance over the entire lifetime of a company—unlike for instance excess return, which per the earlier discussion would just have been -100%. Moreover, Circuit City had a clear competitor, Best Buy, allowing for an interesting comparison.

2.6.1. LIVA decomposition

One attractive feature of LIVA is that it is an additive measure: different time periods and different parts of operations can be selected that jointly add up to LIVA. Decomposing LIVA over time is straight-forward—the summation can just be split into multiple time periods. For instance if one has LIVA computed over 1980 to 2010, then the sum can be split into two periods, from 1980 to 1995 (period I) and 1995 to 2010 (period II), respectively. Using equation (2.2):

$$\begin{aligned}
 LIVA &= \sum_{t=1981}^{2010} \frac{ER_t MC_{t-1}}{(1+r)^{t-2010}} \\
 &= \sum_{t=1981}^{1995} \frac{ER_t MC_{t-1}}{(1+r)^{t-2010}} + \sum_{t=1996}^{2010} \frac{ER_t MC_{t-1}}{(1+r)^{t-2010}} \\
 &= LIVA_I + LIVA_{II}
 \end{aligned}$$

Moreover, each of these can be decomposed into various elements using equation (2.1). For instance if one has a breakdown of free cash flow (FCF) into operating cash flow (OpCF)

minus investment cash flow (InvCF), and the Enterprise Value (EV) of a company (the market value of equity plus book value of debt) at the beginning and the end of a period, the LIVA for this time period can be decomposed as follows:

$$\begin{aligned}
LIVA_{II} &= \sum_{t=1996}^{2010} \frac{ER_t MC_{t-1}}{(1+r)^{t-2010}} \\
&= EV_{2010} - (1+r)^{15} EV_{1995} + \sum_{t=1996}^{2010} \frac{OpCF_t - InvCF_t}{(1+r)^{t-2010}} \\
&= [EV_{2010} - (1+r)^{15} EV_{1995}] + \sum_{t=1996}^{2010} \frac{OpCF_t}{(1+r)^{t-2010}} + \sum_{t=1996}^{2010} \frac{-InvCF_t}{(1+r)^{t-2010}}
\end{aligned}$$

The cash flow terms reflect direct NPV calculations from cash flows, while the term involving the two values for EV accounts for the market value of the assets at the beginning and the end of the period (similar to equation (2.1)). The EV term can be interpreted as the value created (or destroyed) for investors due to the growth (or decline) of firm value compared to the market return.

The above decomposition can be easily generalized to more time periods and more cash flow items. In the case study, we use six cash flow items for the LIVA decomposition (the italics in brackets refer to the Compustat item codes in the Wharton Research Data Services, WRDS, database):

1. Net operating profit after tax (NOPAT) [*sale - cogs - xsga - txt*]
2. Capital expenditure (CapEx) [*sppe - capx*]
3. Change in working capital [*wcap_t - wcap_{t-1}*]
4. Acquisitions [*aqc*]
5. Other items on 10-K [all cash flows recorded on the 10-K cash flow statement to share-

holders ($prstk + dv - sstk$) and debtholders ($xint - nopi - \text{change in net debt } (dltt + dlc - che)$), net of items 1 through 4 above]

6. Non-reported items [redistributions to shareholders based on CRSP stock market data that are not reported on the 10-K as recorded in Compustat]

2.6.2. Analysis

Using the definition in equation (2.2), the total LIVA of Circuit City amounts to negative \$3.4 billion over the period from its listing on the American Stock Exchange in 1968 as Wards, until its bankruptcy and delisting by the end of 2008. The LIVA of Best Buy amounts to positive 14.7 billion dollars over the period from its initial public offering on the NASDAQ in 1985, until the end of our data in 2014.¹⁴ Figure 3 displays the cumulative LIVA for both companies over time. For each point in time, the figure displays the total LIVA up to that time, discounted to 2015 values. Thus, the end points in the chart represent the total LIVA of -\$3.4B and \$14.7B, respectively.

Given that Circuit City went bankrupt, it might not be too surprising that its investors fared worse than the general market, leading to a negative LIVA. But, as mentioned earlier, bankruptcies can signify widely differing ranges of performance, depending on how much cash the company used from and redistributed to investors along the way. As a matter of fact, of the negative \$3.4 billion, already \$2.8 billion occurred before 1980, during the company's history as Wards. During this era, the company consisted mostly of smaller television and audio stores, operated under various store formats and brands (Wells and Danskin, 2012b). The decomposition in Table 8 shows that during this time from 1969 to 1980 the company did return significant cash to shareholders, but failed to grow its enterprise value at the rate of the market, leading to a negative LIVA.

¹⁴Both the LIVA of Circuit City and Best Buy are discounted to the year 2015, in order to make them directly comparable. Moreover, to calculate excess return we use a beta of 1.48, which is the average beta of Circuit City and Best Buy. The betas for the individual companies based on all monthly returns are respectively 1.54 and 1.42. These values are not statistically different from each other, and therefore we decide to average them given the similar operating and risk profiles of both companies.

Table 8: LIVA decomposition for Circuit City and Best Buy

a. Circuit City

	1969– 1980	1980– 1995	1995– 2010	2010– 2015	Total
1. Net operating profit after tax	6.3	26.0	15.0		47.3
2. Capital Expenditure	-2.6	-19.5	-10.1		-32.3
3. Change in Working capital	-2.2	-12.8	-5.7		-20.7
4. Acquisitions	0.0	-1.0	-0.7		-1.7
5. Other items in 10-K	2.8	6.0	1.1		9.9
6. Non-reported items	0.0	-1.6	4.5		2.9
Cash items	4.2	-2.9	4.1		5.5
Change in relative enterprise value*	-7.1	20.8	-22.6		-8.9
LIVA	-2.8	17.9	-18.5		-3.4

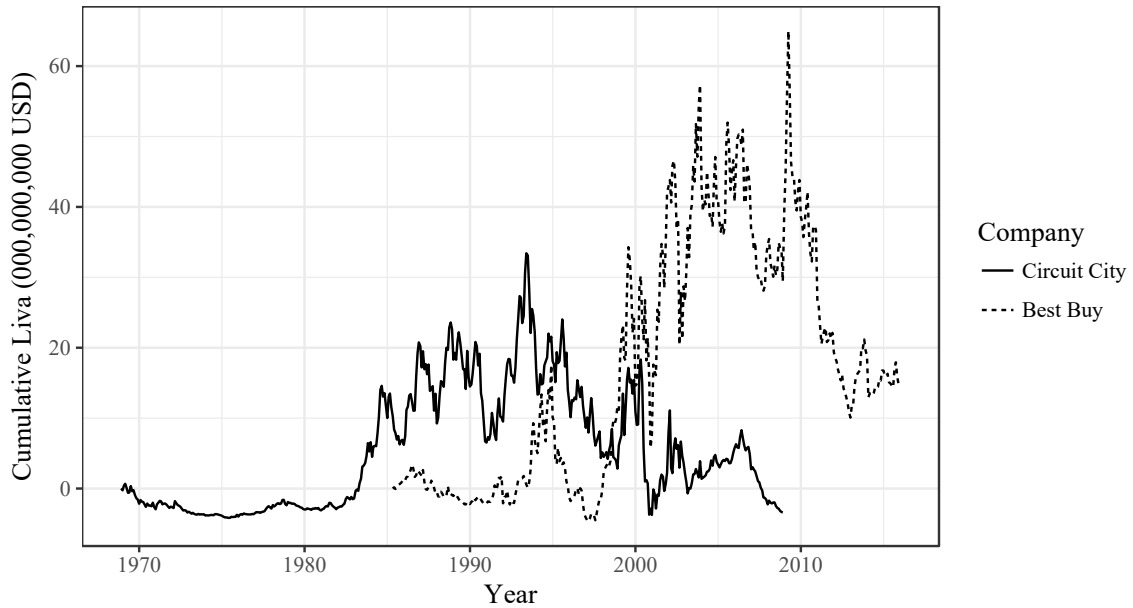
b. Best Buy

1. Net operating profit after tax		3.8	43.6	22.1	69.5
2. Capital Expenditure		-3.7	-26.1	-2.9	-32.7
3. Change in Working capital		-5.9	-0.7	-7.1	-13.7
4. Acquisitions		0.0	-14.6	-0.4	-15.0
5. Other items in 10-K		0.5	4.9	3.0	8.5
6. Non-reported items		-0.5	-4.5	-0.8	-5.8
Cash items		-5.7	2.6	13.9	10.7
Change in relative enterprise value*		9.7	29.2	-34.9	4.0
LIVA		4.0	31.7	-21.0	14.7

Note. Total LIVA is calculated using the definition in equation (2.2), and then decomposed over time and into various cash items using the derivations provided in the text. All figures in 2015 present value billion US dollars.

*Change in relative enterprise value for each period reflects the change in enterprise value (market value of equity plus book value of debt) compared to the market return. See equation (2.1) as well as the details in the LIVA decomposition section.

Figure 3: Cumulative LIVA over time for Circuit City and Best Buy



In the following period, from 1980 to 1995, Circuit City pioneered and rolled out the very successful electronics super-store format. This model started with the 40,000 square foot Wards Loading Dock store opened in Richmond in 1974. In 1981 it started to expand this successful store formula, rebranding the stores to Circuit City soon after. In 1984 the company itself was renamed, and listed on the New York stock exchange. The model with a wide selection of inventory kept in a store warehouse, a high-service commissioned sales force, and in-house offering of credit and repairs proved very successful, with Circuit City growing to 419 superstores in the mid-1990s. (Wells and Danskin, 2012b)

The LIVA decomposition in Table 8 corroborates this picture. The LIVA contribution of operating profits over this period was \$26 billion, clearly the highest of the three periods in Circuit City's history. Still, the company operated cash negative over the period, mainly due to the high capital expenditures and investments in working capital required for the company's aggressive expansion strategy, totaling \$32.3 billion in present value.

For much of this period Best Buy created relatively little value for its investors, as illustrated

by the mere \$3.8 billion LIVA contribution of its operating profits. That started to change in the early 1990s, when Best Buy modified its store format into a deep-discount retail model, with low service by hourly paid sales staff, and inventory stacked in the store display area instead of a separate warehouse (Wells and Danskin, 2012a). Within a few years this strategic change transformed Best Buy from a sub-scale follower into a market leader, overtaking Circuit City in revenue by 1995. Moreover, it was able to do so at a cumulative investment (capital expenditure and working capital) of \$9.6 billion in present value, less than a third of Circuit City's over the same period.

Over the following decade, Circuit City tried to catch up again through a series of expensive restructurings and store refurbishments. By the mid-2000s, Circuit City had laid off its commissioned sales force, and changed its store layout to a bright warehouse style, essentially becoming a copy-cat of the now entrenched discounter Best Buy. But it was too late, and Circuit City's competitive position continued to deteriorate. In the face of yet another failed turn-around attempt, the dawning financial crisis, and a drop of almost 50% in holiday sales, the company was forced into liquidation in January 2009. (Wells and Danskin, 2012a).

Table 8 illustrates the dramatic reversal of fortune in the period 1995-2010. While Circuit City's present value of operating profits dropped to \$15 billion, Best Buy's increased to \$43.6 billion, allowing the latter company to make higher investments in organic growth and acquisitions to further improve its competitive position. In the subsequent period 2010-2015, Best Buy was able to capitalize on those investments, generating cash worth close to \$14 billion. At the end of that period Best Buy's cumulative LIVA stood at \$14.7 billion, indicating a significant value creation for its investors.

Interestingly, Circuit City's total LIVA netted to negative \$0.6 billion over the superstore era, from 1980 until its liquidation—quite a small amount compared to the sizes of the cash flows and valuations involved. The LIVA decomposition indicates that this is largely thanks to significant cash redistributions in the period 1995 to 2010 (which were, in fact, larger than Best Buy's). A significant redistribution was the Car-Max spin-off in 2002, which

instantly contributed about \$5 billion to Circuit City's LIVAs, in a way saving the day for its investors. Note that this transaction happened in terms of a stock dividend and was not included in Circuit City's operating results, and thus would not have shown up in an analysis of classic performance measures such as return on assets or capital. Because in our LIVAs calculation we track actual returns to investors, it is included in our decomposition in Table 8 on line 6 ("non-reported items").¹⁵

Finally, the LIVAs decomposition points to the main culprit of Circuit City's demise: the very expensive but largely ineffective expansion and retaliation strategy in the late 1980s and early 1990s. This expansion cost the company well over \$30 billion worth of LIVAs, thus proving a very costly strategic mistake for its investors.

2.7. Discussion: LIVAs vs. other measures

In this section, we discuss the broader advantages and disadvantages of LIVAs vis--vis other potential measures of long-term firm performance. Using the two-by-two framework from the literature review in Figure 2, we provide a comparison across each of the four quadrants. LIVAs is an absolute measure of firm performance based on stock market data, and thus fits the top right quadrant—an empty quadrant in the 2016 SMJ volume. TSR (total shareholder return) and CAR (cumulative abnormal return) as relative stock market measures can both be extended to long-term measures by annualizing them using a compounded annual growth rate—note that in this paper we use the term excess return for the long-term equivalent of CAR. Equation (2.3) suggests that the (sum of) discounted economic profit (EP) would be the most appropriate absolute accounting measure of long-term firm performance. Finally, (average) ROC would be a logical relative accounting measure of long-term firm performance.

Table 9 provides a comparison for each of these four performance measures on various

¹⁵Best Buy too had shareholder transactions of more than \$5 billion not reported on the cash flow statements. According to the notes in the 10-K reports, these appear to be mainly share issues for management stock options and the repayment of convertible preferred securities.

Table 9: Comparison of LIVA with other common performance measures

	LIVA	TSR or excess return	Discounted economic profit	Average ROC
Increases with NPV-positive investments	✓	(✓)	(✓)	
Accounts well for mergers, bankruptcy, etc.	✓		(✓)	
Measures magnitude of economic impact	✓		✓	
Can be decomposed into different sources	✓		✓	
Is independent of stock market expectations			✓	✓
Measures returns relative to size		✓		✓

Note. Comparison of LIVA to other measures that can be used to measure long-term firm performance (over a decade or more) on various properties that might be desirable. TSR = total shareholder return (equivalent to buy and hold return: dividend yield plus share price appreciation with dividends reinvested), ROC = return on capital. Black tick marks indicate that a measure has a property, grey tick marks in brackets indicate that it has the property with some qualifications. Details are provided in the discussion section of the text.

potentially desirable properties (depending on the specific application). Four of them are potential advantages of LIVA and two are disadvantages vis-à-vis other metrics. We discuss each of them below.

2.7.1. Advantages of LIVA

LIVA increases with NPV-positive investments

We have derived LIVA as an empirical measure of long-term firm performance that measures *ex post* NPV. In other words, LIVA increases for NPV-positive investments or strategic actions, and decreases for NPV-negative investments or strategic actions. Because LIVA is closely related to excess return per equation (2.2), absent major cash redistributions, spin-offs, M&A, and bankruptcy events, the sign of the excess return also reflects whether the firm has made NPV positive or negative decisions over a period. However, in the case of major cash transactions, excess return can be positive for a company that has actually destroyed value and vice versa, as we have shown with the example in Table 2, as well as with the AOL-Time Warner merger.

Based on equation (2.3), the sum of discounted economic profit can be a good approximation

of LIVA (and thus of whether a firm has made NPV-positive investments), as long as the book value properly captures the market value of the firms investments at the beginning and the end of the period under consideration. This latter condition could be troublesome in industries in which intangibles play an important role, because investments in marketing and R&D can be very valuable but may not be captured on the balance sheet, and thus could artificially affect economic profit even over long time periods. Potentially this issue could be alleviated by using capitalization schemes for such intangible investments as suggested by Peters and Taylor (2017).

ROC and other relative accounting measures in general do not have the property that they increase if and only if firms make value-increasing investments: Firms can make positive-NPV investments that decrease ROC (Levinthal and Wu, 2010) and ROC can be below the cost of capital even if a firm creates economic value (Fisher and McGowan, 1983).

LIVA accounts well for mergers, bankruptcy, and other major corporate events

Related to the above discussion, LIVA can account well for the value created or destroyed in the presence of major corporate events such as major cash redistributions, spin-offs, M&A or bankruptcy events. This is not the case for TSR, excess return, and ROC. Economic profit could in principle be adjusted for the cash flows around major events, however this might require a case-by-case analysis of annual reports and merger announcements to correctly track the cash distributions. LIVA calculations are usually much simpler, because the widely used CRSP database has good information about the cash flows around these major corporate events.

LIVA measures the magnitude of economic impact

Because LIVA is an absolute measure (as is discounted economic profit), it captures the magnitude of a firms economic impact. As the M&A regressions show, using the absolute magnitude of investor value appropriation can be quite different from the relative (excess)

return generated for an investor, because very high (and very low) relative returns are typically only generated by small companies. When not properly accounted for in regressions, these companies can strongly influence the analysis, while having minimal economic impact.

LIVA can be decomposed into different sources

The case study of Circuit City and Best Buy shows how decomposition of LIVA into its constituent sources over time can yield valuable strategic insights. This could be similarly done using a sum of discounted economic profits as a measure of long-term firm performance. TSR, excess return and ROC cannot be decomposed into their constituent elements, because as relative measures they are not sums of different sources over time.

2.7.2. Disadvantages of LIVA

LIVA depends on stock market expectations

Equation (2.1) indicates that any measure of long-term value appropriation requires an assessment of V_0 and V_T , the firm value at the start and end point of the period under consideration. Because LIVA is defined based on the use of market valuations of V_0 and V_T , it requires careful attention to potential market distortions of these valuations. The value decomposition of Circuit City in Table 8 shows that this can have major consequences if investor expectations driving these valuations turn out to be incorrect: between 1980-1995 the enterprise value of Circuit City rose by more than \$20 billion over and beyond the return in the market (in present value), reflecting strong investor expectations about Circuit City's future; these turned out to be incorrect, and the company's bankruptcy in 2010 more than wiped out the growth in enterprise value from the previous period. On the one hand, this example shows that over the full investment cycle (in this case the 30 years between 1980 and 2010) LIVA provides an appropriate picture of the company's ability to generate value for its investors, because any deviations in between cancel out. On the other hand, it shows too that over shorter time periods the LIVA picture can be deceptive if investor expectations

strongly deviate from subsequent outcomes.

The appropriate time period over which to measure LIVA will depend on the specific context, and particularly on the time scales of the competitive dynamics. Because LIVA measures the extent to which investments have been successful in generating profits over time, a natural time period should span the time that strategic investments take to be reflected in competitive outcomes. For instance, in an industry life-cycle study, a natural time period would be a technology life-cycle, from initial investments to demise. When using stock market data to compute LIVA, care should be taken to avoid moments of great market turbulence as start or end points, such as bubbles and crashes. LIVA measures of technology firms ending in 2000, for instance, probably tell more about inflated market expectations than about strategic performance of the investments.

Because TSR and excess return are also based on stock market data, they suffer from the same issues if investor expectations strongly deviate from later outcomes. Accounting measures such as ROC and accounting-based operationalizations of economic profit are independent of stock market valuations, and thus do not face this issue (though they hinge on accounting rules for valuation instead).

LIVA does not measure returns relative to size

A final disadvantage of LIVA (and discounted economic profit) might be that it is an absolute measure, in dollar amounts, rather than a relative measure or ratio measuring a percentage return (such as TSR, excess return and ROC). However, this might mainly be a matter of custom. In fact, both our literature review and our M&A analysis suggest that relative measures can have serious drawbacks while potential issues with the use of absolute measures can be addressed in different ways than taking a ratio.

CHAPTER 3 : The dynamics of resource competition: Implications for firm and industry components of returns

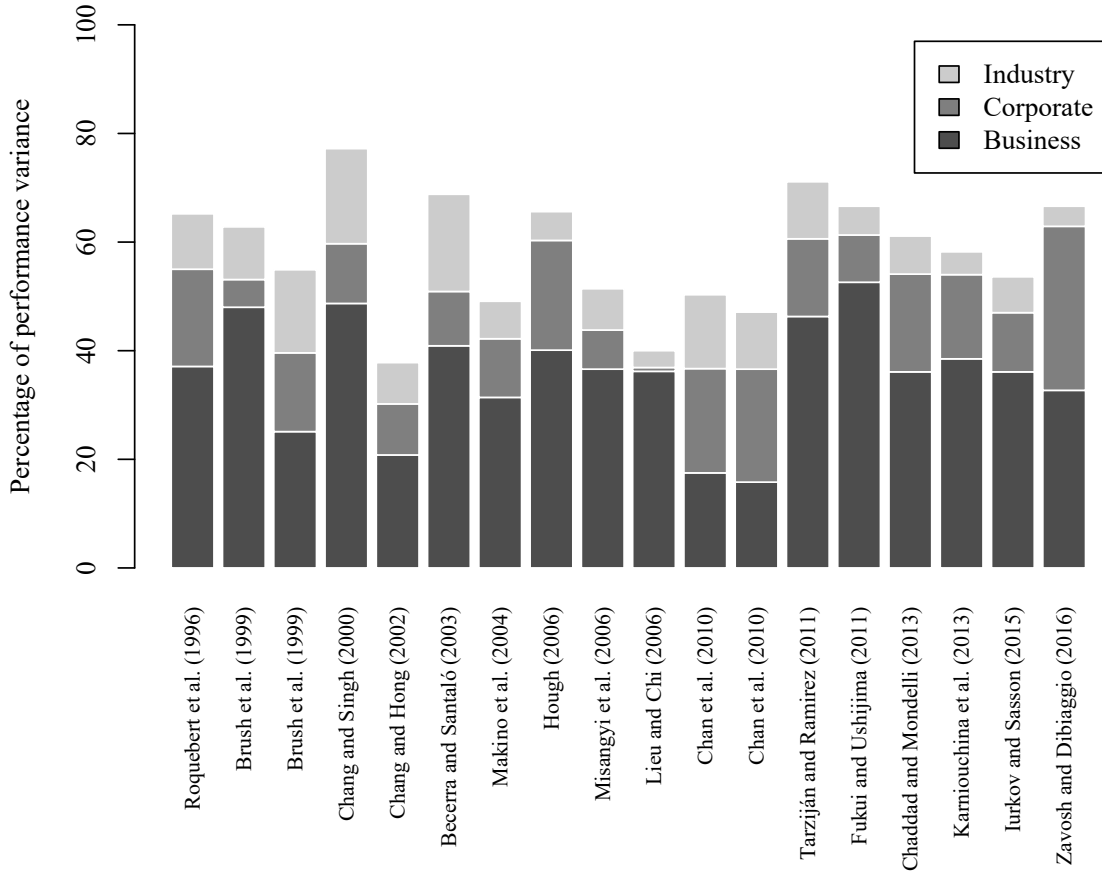
3.1. Introduction

What drives heterogeneity of firm performance has been a core theoretical and empirical question in the strategy literature. For instance, the variance component studies have consistently shown that firm-specific (i.e., business or corporate) effects explain a large part of performance variations across firms, significantly larger than industry effects (e.g., Rumelt, 1991; McGahan and Porter, 1997)—Figure 4 provides an overview based on Vanneste (2017). This empirical regularity has usually been interpreted as support for the resource-based view (RBV): in order to explain why some firms are persistently more profitable than their direct competitors, the better performing firms must have differential access to heterogeneously distributed resources that are not readily available on strategic factor markets (e.g., Rumelt, 1984; Wernerfelt, 1984; Barney, 1986, 1991; Peteraf, 1993).

However, though the RBV does explain why large performance differences among competitors *can* persist over long time periods, current theory does not explain whether or why these differences *should* be large. Moreover, we do not know theoretically or empirically whether the firm-specific component of performance differences is always (much) larger than the industry component, or that under certain circumstances it could be smaller. Finally, we do not know what effect size the RBV predicts; the variance decomposition studies in Figure 4 suggest that firm-specific components of performance variance are five to ten times as big as industry components, but current theory does not predict or explain this ratio, nor whether and how it should differ across industries. Thus, there is a gap in the literature to provide a quantitative bridge between the competitive dynamics of resource markets, and the empirical observations of firm and industry components of returns.

In this paper, I address this gap by introducing a formal model of resource competition and testing some of the model's predictions with empirically observed heterogeneity data. In

Figure 4: Overview of variance component studies



Note. Percentage of variance in accounting returns explained by industry, corporate and business components; excludes year and other components. The chart shows the data from all studies in Vanneste (2017, Table 4, pp. 130-132) that use VCA or HLM methods.

the model, based on the dynamic game framework (Ericson and Pakes, 1995), two *ex ante* symmetric firms make investment decisions during multiple periods to acquire resources. I find that large heterogeneity among direct competitors emerges under very general conditions. Specifically, four key conditions that jointly lead to large firm-specific heterogeneity are:

1. **Endogenous investment:** Firms choose the level of their sunk cost investments to optimize future expected value.
2. **Scarce resources:** There is a (long-term) supply curve for resources, such that resource cost increases when more resources have been acquired in the past.
3. **Time-compression diseconomies:** Acquiring the same amount of resources is more expensive over shorter time periods (Dierickx and Cool, 1989).
4. **Uncertainty:** For a given investment there is a probability distribution for gaining a better resource position.

The model indicates that under these four conditions, resource asymmetries are amplified: small differences in resource positions lead to large, persistent differences in performance—thus, these conditions are sufficient for the emergence of high firm-specific heterogeneity. Moreover, removing any of the four conditions removes this amplification effect—thus, they are not only sufficient, but also “minimal”.¹ Because these four conditions are plausibly satisfied in any competitive resource market, the model provides a general mechanism for the emergence of a high firm-specific component of resource variations. Additionally, the model predicts that the level of firm-specific heterogeneity will differ across industries, based on the industry characteristics of the resource dynamics, such as depreciation rates, level of scale economies, and growth stage.

¹They are not “necessary” in a strict sense of the word, as there are other conditions that can lead to high firm-specific heterogeneity; e.g., the mere assumption of large differences in resource endowments is clearly sufficient, too, to lead to large firm-specific differences in performance. The claim in this paper, however, is that removing any of these four conditions (without replacement by a different condition) will remove the amplification effect—hence the designation of “minimal”.

I test the model's predictions using a Bayesian analysis of stock price variations across different industries. I use a Bayesian method because it allows obtaining different estimates for different industries of the firm-specific and industry components of returns (Alcácer et al., 2013; Gelman et al., 2013, ch. 5), in contrast to classical variance decomposition techniques, which only provide estimates averaged over an entire sample. Consistent with the model's predictions, there are significant differences across industries in the heterogeneity of company returns and these differences are largely stable over time. Moreover, I find the highest firm-specific heterogeneity in industries that are typically associated with strong scale economies and resource depreciation rates, such as high-tech industries, while I find lower heterogeneity in industries with weaker scale economies and lower depreciation rates, such as utilities and financial services. A regression analysis confirms statistically significant effects that are consistent with the predictions of the model.

This study makes several contributions to the literature. First, this paper is one of the few papers to explicitly model the dynamics of resource markets and competition, instead of focusing on product markets and competition. Particularly, to my knowledge it is the first paper that incorporates all the above four elements of resource competition in a single model, as well as the first to explicitly link the dynamics of resource competition to observable variations in performance heterogeneity. Finally, this study empirically documents the variations of firm-specific vs. industry-wide heterogeneity at a high level of granularity across industries, and explain the observed empirical patterns using a dynamic model of resource competition.

The results of this paper solidify both the theoretical foundations of the RBV and its empirical support. This highlights the importance of resource-based analysis for managers, as relatively small resource differences vis-à-vis competitors could be amplified over time. At the same time, the paper also highlights the importance of industry analysis, because industry characteristics, such as sunk cost levels, depreciation rates, scale economies, and growth stage, determine how resource differences evolve over time and to what extent they

are amplified into performance differences. Thus, rather than viewing resource and industry analysis as opposing forces, this paper suggests that both are jointly needed to understand the evolution of firm performance.

3.2. Theory

3.2.1. Literature review

In this section I will develop a formal model of resource competition to analyze the development of performance variations over time, particularly to the extent they are firm- vs. industry-specific. I use the resource based view (RBV) as the basis for my model. Wernerfelt (1984) introduced the RBV to complement the product market perspective, arguing that “resources and products are two sides of the same coin” (p. 171). The RBV has been an important framework in the management literature to explain why some firms persistently outperform others, even if they have similar product market positions: the better performing firms must have differential access to heterogeneously distributed resources that are not readily available on strategic factor markets (e.g., Rumelt, 1984; Wernerfelt, 1984; Barney, 1986, 1991; Peteraf, 1993). Though there exist various definitions of “resources”, in this paper I will use a broad definition that incorporates, for instance, managerial capabilities, brands, technologies, trade secrets, or anything else “that the organization can draw upon to accomplish its aims” (Helfat et al., 2007, p. 4).

More specifically, I will focus on immobile resources, which firms cannot readily buy on strategic factor markets, but rather need to accumulate through strategic investments over time (Dierickx and Cool, 1989). In my model I assume that firms endogenously make these investment decisions, optimizing expected value. Second, I assume that these resources are scarce, because ultimately any resource must have a finite supply, which is a key element of understanding firm-specific profits (Gans and Ryall, 2017, p.22). Third, I assume time-compression diseconomies, which means that there are diminishing returns to resource investments in a given time period—this is equivalent to assuming strictly convex

adjustment costs of investments (Dierickx and Cool, 1989). Fourth and finally, I assume uncertainty, meaning that strategic investments increase the *probability* of gaining resources.

None of these four assumptions is new in itself; Table 10 provides a selection of resource investment models in the literature that incorporate some of these assumptions. For instance, Sutton (1991) uses an endogenous investment game to derive observable regularities about industry concentration. Sutton's model does not incorporate resource scarcity and uncertainty, and because it is a two-stage game it cannot incorporate time-compression diseconomies. Several other studies in the strategy literature do study models over multiple periods, incorporating time-compression diseconomies (e.g., Nelson and Winter, 1982; Ericson and Pakes, 1995; Knott et al., 2003; Denrell, 2004; Pacheco-de Almeida and Zemsky, 2007; Knudsen et al., 2014), but these models do not explicitly model resource scarcity either. Two papers that do explicitly model scarce resources through supply and demand curves either do not model longer horizons with time-compression diseconomies (Makadok and Barney, 2001) or do not incorporate uncertainty (Jacobides et al., 2012). To my knowledge, the model in this paper is the first to combine all these elements into a single model, while it turns out that it is the combination of the four above conditions that leads to the amplification of resource asymmetries in performance, and thus to an explanation of the empirical regularity of high firm-specific component in performance variations.

The model is based on a dynamic game (Ericson and Pakes, 1995), which is explicitly designed to study the long-term dynamics of endogenous investments. Though earlier dynamic game models primarily have been used to study product market dynamics, they are ideally suited to study the dynamics of resource competition, because dynamic games allow solving for the optimal investment policy based on the assumption of endogenous investments in an indefinite time-horizon dynamic model. Specifically, I will use the parsimonious model of Besanko and Doraszelski (2004) as the basis for my model.

I add three key elements to the dynamic game models in the current literature. First, this model explicitly incorporates competition for scarce resources, which manifests itself in

Table 10: Literature overview

	Endogenous investment	Scarce resources	Time-compression diseconomies	Uncertainty
Nelson and Winter (1982, Ch. 9)			✓	✓
Sutton (1991)	✓			
Ericson and Pakes (1995)	✓		✓	✓
Makadok and Barney (2001)	✓	✓		✓
Knott et al. (2003)			✓	
Denrell (2004)			✓	✓
Pacheco-de Almeida and Zemsky (2007)	✓		✓	
Jacobides et al. (2012)	✓	✓	✓	
Knudsen et al. (2014)			✓	✓

Note. Selection of formal models in the literature addressing resource investment (including, for instance, R&D). A check mark indicates that a model incorporates the specified attribute.

the state transition functions of the dynamic game. Second, I add an independent industry state, in order to be able to analyze the effects of firm-specific vis-à-vis industry-wide sources of firm heterogeneity. Third, I add an explicit link between the equilibrium solutions and observable patterns of performance heterogeneity.

3.2.2. Model

In the model, two firms compete for a set of scarce resources. I assume that the two firms compete on a single resource market, and that both firms are characterized by their respective resource positions q_i and q_j . Resource scarcity demands that the firms jointly cannot have their resource positions exceed some finite quantity Q ($q_i + q_j \leq Q$).

The gross profit (before investment) for each firm in the model is a quadratic function of the resource position, which can be viewed as the first two terms of a Taylor expansion of a more general profit function²—alternatively it can be derived directly from an economic product market model, see Appendix A.2:³

$$\pi(q_i, a_k) = a_k q_i + b q_i^2 \tag{3.1}$$

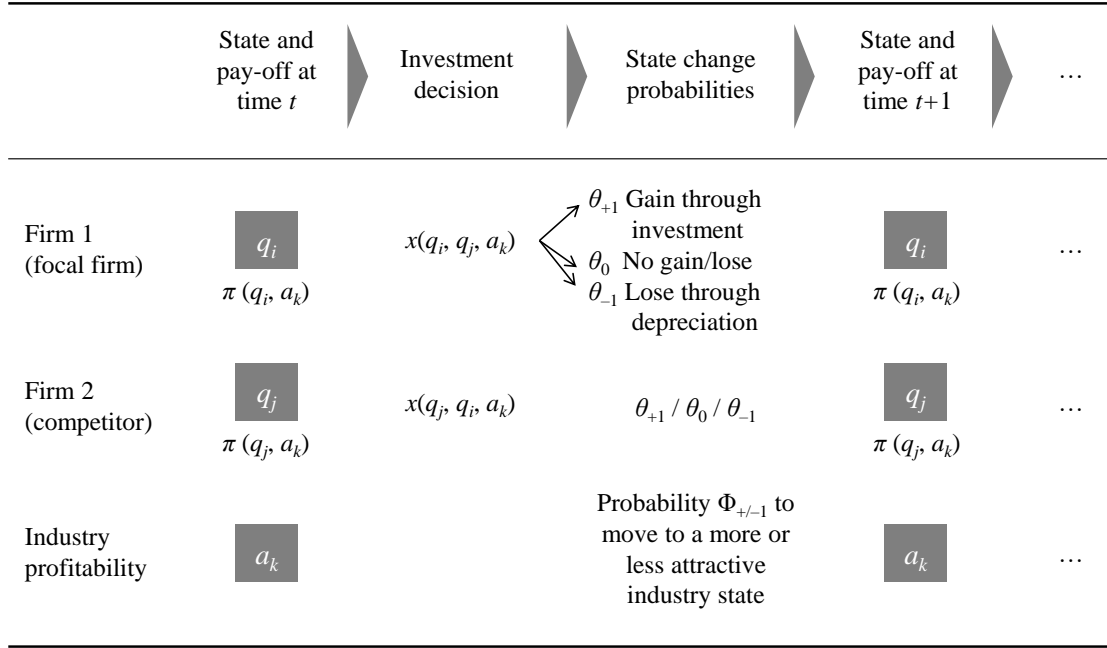
This equation posits that firm profit linearly increases with its resource position plus a quadratic term that allows for scale economies (which initially will be excluded, setting $b = 0$). Moreover, firm profits depend on exogenous factors that affect industry profitability, such as economic growth rates and commodity prices, which are captured in an industry profitability state a_k . This industry state is the same for both firms, and can change over time.

Thus, in the model there are three state variables: both firms' resource positions q_i, q_j , and the industry profitability a_k . These state variables can change over time, based on firms' investment decisions and the resulting resource gains and losses, which are probabilistic

²Any constant can be omitted without loss of generality, as it would not affect any of the dynamics.

³Here and in the following I specify the functions for firm 1, with index label i . As is standard in these models, the functions for firm 2 are symmetric, i.e. $\pi_2(q, a_k) = \pi_1(q, a_k) := \pi(q, a_k)$.

Figure 5: Time evolution of state variables and investment decisions in the model

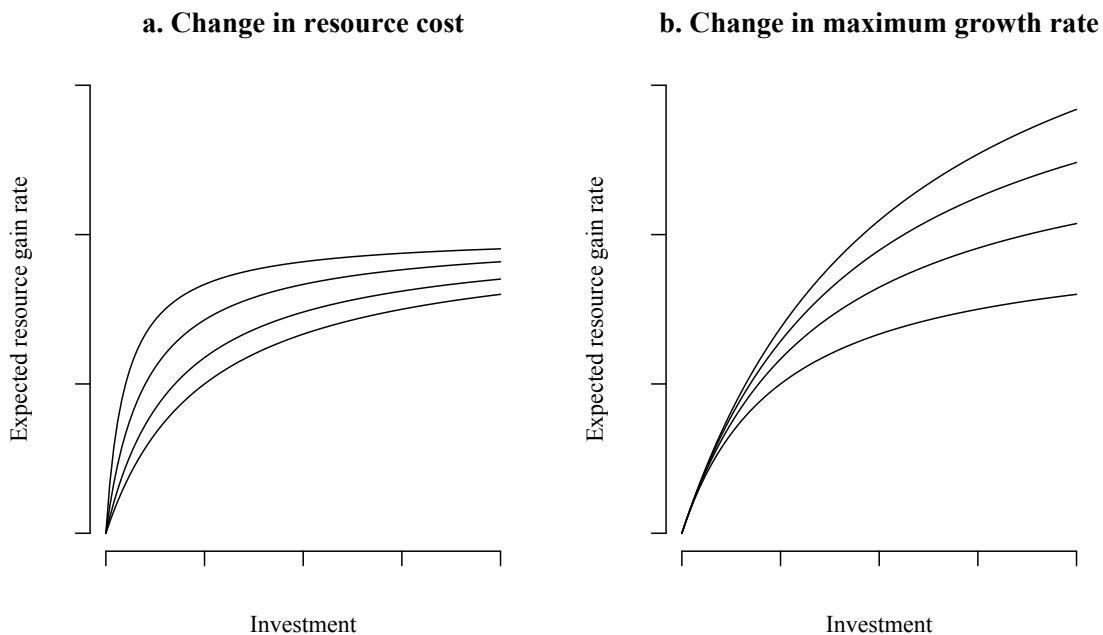


in nature. Any other parameters, such as the scale economies parameter b in the profit function, are assumed to be industry characteristics that stay fixed over time, and which ultimately determine the stochastic dynamics of the state variables q_i, q_j , and a_k .

In addition to specifying the state variables and the profit function, the input for a dynamic game model requires the probability function for the state transition as a function of a sunk cost investment x . To specify the state transition, it is common to assume that both the state space and the time are discrete. During each time interval Δt , the transition functions specify the probabilities θ and Φ of moving up or down a unit of resources q_i, q_j and industry profits a_k , given the current state $s = (q_i, q_j, a_k)$ and the investment x . Figure 5 conceptually shows the evolution of the state s for a single time period.

The change probabilities θ of the focal firm's resource state q_i depend on the overall state $s = (q_i, q_j, a_k)$ as well as the firm's investment x . These state change probabilities θ depend on the investment efficiency $f_s(x)$, which specifies the expected value of the rate of gaining resources given state s and investments x . Time-compression diseconomies demand this

Figure 6: Investment efficiency function



Note. Graphical depictions of the investment efficiency function $f_s(x)$ as defined in equation (3.2), for varying parameter values of resource cost c_s (panel a) and maximum growth rate g_s (panel b).

function be concave. For reasons of analytical tractability it is commonly chosen as some scaling of the function $\frac{x}{1+x}$ (e.g., Besanko and Doraszelski, 2004), which is strictly concave for $x \geq 0$. A convenient parametrization is:

$$f_s(x) = \frac{g_s x}{c_s g_s + x} \quad (3.2)$$

$c_s = \frac{1}{f'_s(0)}$ parametrizes the resource cost for small investments x , while $g_s = \lim_{x \rightarrow \infty} f_s(x)$ parametrizes the maximum growth rate, both depending on the state s . Figure 6 shows graphical depictions of function (3.2) for varying values of the parameters c_s and g_s .

The resource cost c_s codifies resource scarcity. A parsimonious specification is:

$$c_s = c_0 \frac{Q}{Q - q_i - q_j} \quad (3.3)$$

When firms have small resource positions ($q_i, q_j \approx 0$), resource cost for small investments x is close to some minimum resource cost c_0 . When the joint resource positions approach the total supply ($q_i + q_j \rightarrow Q$), resource cost becomes arbitrarily large.

The maximum growth rate g_s codifies the time-compression diseconomies. The growth potential of a firm should be a rate relative to the size of the firm: bigger firms, in general, will be able to grow faster in absolute terms (e.g., Knudsen et al., 2014):

$$g_s = g_0 + \gamma q_i \tag{3.4}$$

Note that the constant $g_0 > 0$ codifying the growth potential of small firms is necessary, otherwise firms without resources ($q_i = 0$) would never be able to grow. The parameter γ specifies the growth potential relative to size, which can be expressed as an annual percentage rate.

In addition to gaining resources through investment, firms can lose resources through depreciation. This is assumed to be a constant rate relative to size, i.e. the probability of depreciation is proportional to δq_i for some constant depreciation rate δ .

Thus, the probability of gaining a resource θ_{+1} in a given time period Δt is the probability of gaining a resource through investment ($f_s(x)\Delta t$) times the probability of not losing a resource through depreciation ($1 - \delta q_i \Delta t$), and *mutatis mutandis* for the probability of losing a resource θ_{-1} . Finally, the change rate in the industry profitability state a_k is assumed to be a constant ϕ , and thus the probability $\Phi_{\pm 1}$ in a given time period Δt equal to $\phi \Delta t$ —except that of course the probability of moving down Φ_{-1} in the lowest profitability state $a_k = a_1$ is zero, as is that of moving up Φ_{+1} in the highest profitability state $a_k = a_K$. Note that by design $\theta_{-1} = 0$ for $q_i = 0$ and $\theta_{+1} = 0$ for $q_i + q_j = Q$, because $\lim_{q_i + q_j \rightarrow Q} f_s(x) = 0$, as $c_s \rightarrow \infty$. Also note that Δt should be chosen sufficiently small such that all probabilities are between zero and one; this can always be done since the resource gain $f_s(x)$ is bounded by $g_{s,\max} = g_0 + \gamma Q$ and the depreciation by δQ .

Using the above specifications for the state change probabilities and the profit function (equation (3.1)), the Markov perfect equilibrium (MPE) is defined as a value function $V(s)$ and an optimal investment policy $x_o(s)$ that maximizes this value function, per the Bellman equation:

$$V(s) = \max_{x \geq 0} \left[\pi(s)\Delta t - x\Delta t + \beta \sum_{\ell=+1,0,-1} W_\ell(s)\theta_\ell(x; s) \right] \quad (3.5)$$

In this equation, the period discount rate $\beta = e^{-r\Delta t}$ follows from the cost of capital r . The function $W_\ell(s)$ is the expected value of changing the resource state of the focal firm to $q_{i+\ell}$, conditional on the policy function $x_2(s)$ of firm 2. Thus, the Bellman equation codifies that firms choose the investment level that optimizes net present value, which is the sum of profit net of investment and the discounted value of the future state.

The MPE for $V(s)$ and $x_o(s)$ can then be calculated using an iterative procedure. I use initial condition $V(s) = \pi(s)$, and verify that the resulting equilibrium is not affected by varying the initial conditions.⁴ Details for the derivation of the equilibrium conditions are in appendix A.3.

3.3. Model output

3.3.1. Base case

To calculate an equilibrium, I use a set of base case parameters (Table 11) that are in line with typically observed values. For instance, the base case depreciation rate $\delta = 10\%$ is close to the median observed accounting depreciation rate (8%, see Table 15); and the maximum growth rate for large firms $\gamma = 20\%$ is in the higher range of observed industry growth rates. Also the resulting firm and industry volatilities in Figure 7 are close to the empirically observed medians in Table 15, suggesting a reasonable correspondence between the model and the observed data. This set of parameters excludes any scale economies ($b = 0$). See appendix A.3 for computational details of the results, including the computational

⁴Though only existence is guaranteed, also Besanko and Doraszelski (2004, p. 30) find evidence that the resulting equilibria are unique.

Table 11: Base case parameter values

Parameter	Value	Description
δ	10%	Depreciation rate
γ	20%	Maximum average growth rate for large firms
r	10%	Cost of capital
ϕ	10%	Probability of industry profitability state change
a_5/a_1	4	Factor between highest industry profitability (a_5) and lowest (a_1)
c_0/a_3	1	Minimum cost of resources over median profitability (a_3)
b	0	Scale economies factor

Note. All parameter values are annual.

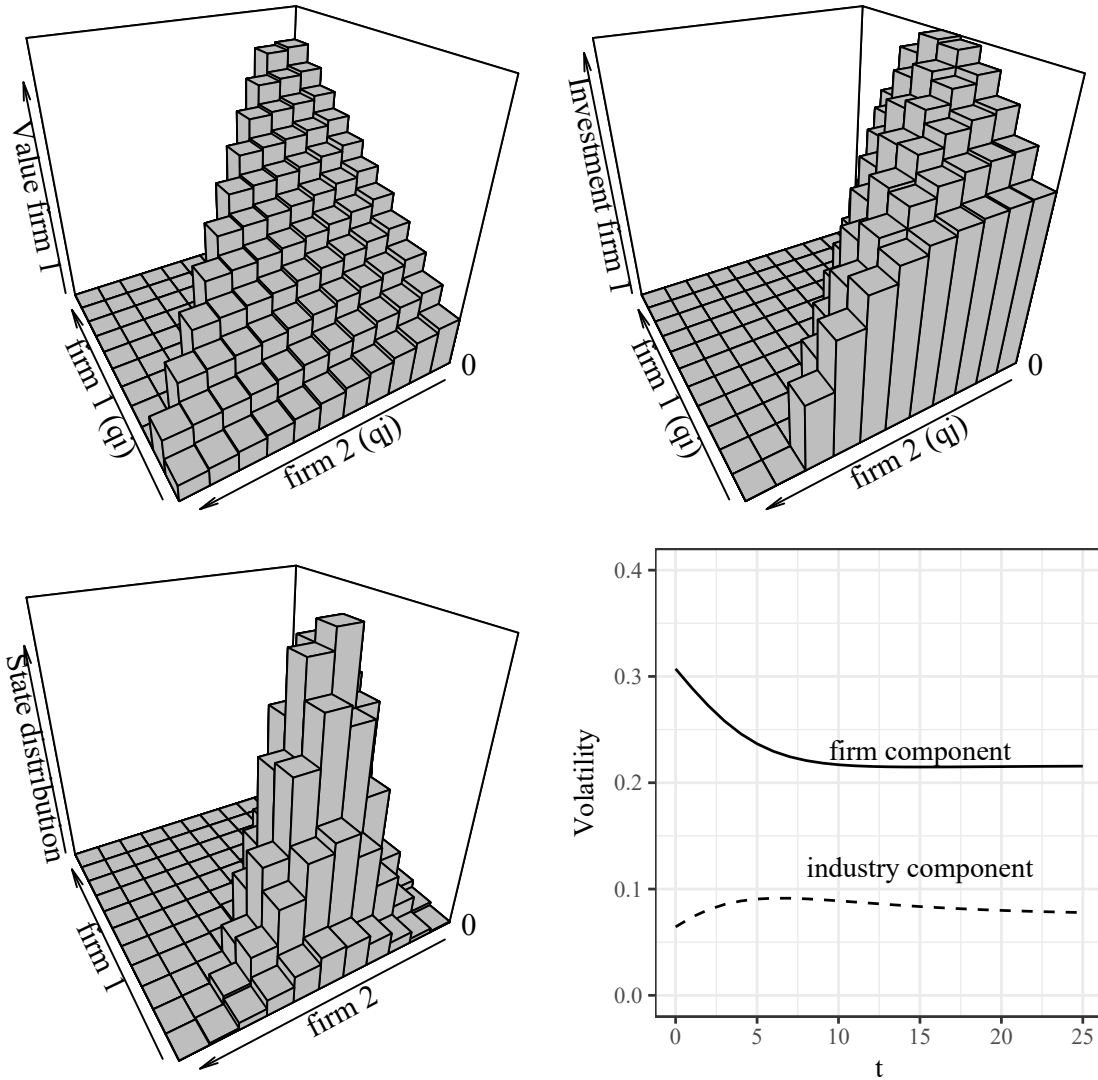
parameters used.

Figure 7 shows the Markov perfect equilibrium for these parameters. The top left panel depicts the value function $V(q_i, q_j)$, as defined in equation (3.5). The value of firm 1 is shown as a function of its own resource position as well as its competitor's (firm 2), and it is averaged over the industry states a_k . The bottom right corner of the chart in this panel represents the position $q_i = q_j = 0$. Note that the value function is only defined for attainable resource positions (i.e., $q_i + q_j \leq Q$), i.e. only for the bottom right half of the chart. States in which the focal firm has more resources have a higher value, and states in which its competitor has more resources have a lower value. The latter effect is solely due to resource competition, because firm 1's profit does not directly depend on firm 2's resource position, per equation (3.1). The reason for the effect on firm 1's value is that if firm 2 holds a better resource position, this precludes firm 1 from having these resources, thus making future resources more expensive and firm 1's growth position less favorable.

The top right panel of Figure 7 shows the optimal investment policy $x_o(s)$ in equilibrium from equation (A.2). Investments are the highest when there is significant opportunity for growth, and are lower when resources become depleted (when $q_i + q_j$ approaches Q and the resource cost becomes very high), or when the firm is still small ($q_i \ll Q$) and not able to grow so fast due to time-compression diseconomies.

The bottom left panel shows the equilibrium probability distribution of the resource states

Figure 7: Base case equilibrium output



Note. Value, investment and probability distribution at $t = 25$ as a function of the resource state variables q_i (firm 1) and q_j (firm 2), averaged over the industry state variables a_k , using parameter values in Table 11. The bottom right chart shows the firm and industry components of volatility over time. See the text for details.

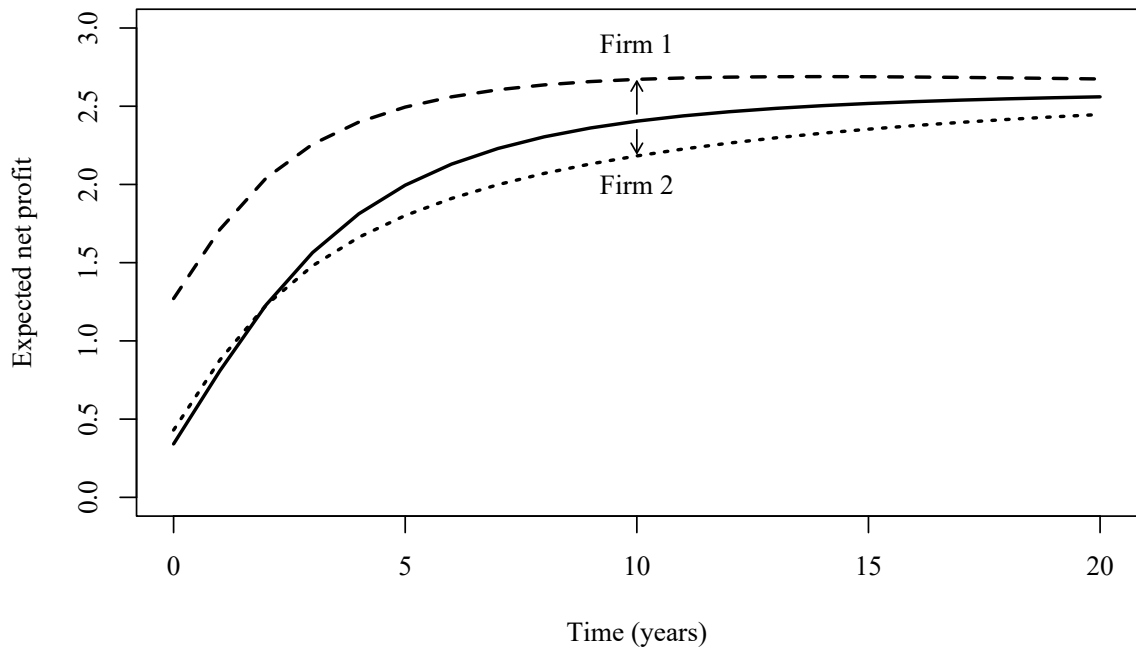
after 25 years starting from an initial state $q_i = 1, q_j = 1, a_k = \1 . After this time period the distribution is close to a *steady state*, meaning that the probability distribution does not change anymore (the resource position of individual firms still do change of course, it is only the stochastic distribution that remains constant). The state distribution is symmetric for firm 1 and firm 2, because the model is fully symmetric in both firms—any differences between the firms originate from random shocks. The modal states are (3, 3), (4, 3), and (3, 4), meaning that in most cases about 60% to 70% of the $Q = 10$ total available resources are depleted; after that they become too expensive to invest in.

The bottom right panel shows the evolution of firm and industry components of volatility (which is the square root of the variance), respectively. These volatilities are based on the state evolution as described above, and calculated from the expected variance and correlations of the next period total return above the cost of capital. Analogous to the calculation of total shareholder returns in stock markets, this excess return is calculated as any cash returns (the profits π minus investments x) plus appreciation in value V , discounted by the cost of capital r . The industry component reflects the volatility of returns that are common to both firms, while the firm component reflects the volatility of returns that are independent of the industry component. Details are in appendix A.4. Under the given parameters, across all time periods the firm-specific volatility is higher than the industry volatility, with the highest values in the earlier growth stages.

The example in Figure 8 helps to explain some of the mechanisms leading to the high firm-specific volatility. This figure shows the evolution of expected net profit $E(\pi - x)$ in the case that both firms start in the same resource position $q_i = q_j = 2$ (solid line) as well as the case that firm 1 has gained an advantage ($q_i = 3$) over firm 2 ($q_j = 2$) at time $t = 0$ (dashed lines). This indicates three reasons why a firm becomes much more valuable on gaining a more favorable resource position:

1. Firm 1 becomes much more profitable immediately at $t = 0$. Though it only gains 50% in its resource position (from $q_i = 2$ to $q_i = 3$), its net profit $\pi - x$ soars by a

Figure 8: Amplification of resource asymmetries in profit evolution



Note. Evolution of expected net profit $E(\pi - x)$ using the base case parameter values in Table 11. Solid line shows evolution for both firms with the same resource position ($q_i = q_j = 2$) at $t = 0$. Dashed lines show the profit evolution when firm 1 has gained an advantage ($q_i = 3$) over firm 2 ($q_j = 2$) at $t = 0$.

factor of four (from \$0.3 to \$1.2), because it needs to invest a lower percentage of its gross profit π to defend its resource position.

2. Firm 1 gains a better growth position. In the original situation it takes almost 15 years to reach an expected profit level of \$2.5, while in the advantaged position it takes firm 1 only 5 years, three times as short.
3. Firm 2 is precluded from this position, and has worse growth prospects. Interestingly, at first firm 2 earns slightly higher net profits as well when firm 1 gains an advantageous resource position; the reason is that it lowers its investment x due to its worsened growth prospects. Over the next 20 years these worsened growth prospects are reflected in the lower expected net profit of firm 2 vis-à-vis the original situation.

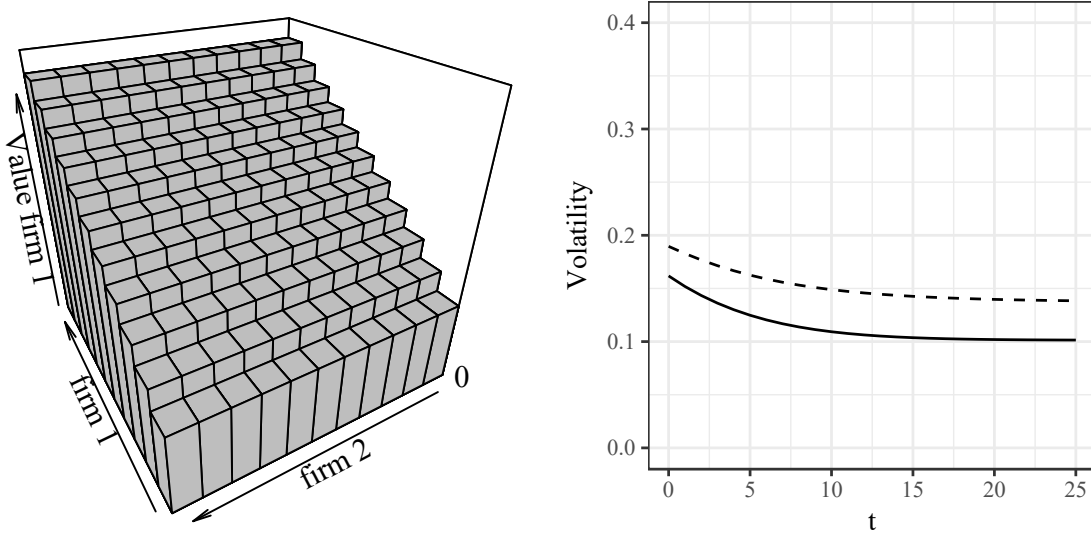
Thus, resource competition amplifies any changes in resource position into large and persistent changes in performance, leading to high heterogeneity in performance among direct competitors, reflected in the high firm-specific component of returns.

To verify that the high firm-specific component of returns is indeed due to the resource competition, and not due to specific choices of parameters or functional forms, I also run the model with equation (3.3) replaced by:

$$c_s = c_0 \frac{Q}{Q - q_i}$$

This shuts down the resource competition, because now each player has its own resource supply function, instead of both firms competing for a common set of resources. The resulting equilibrium without competition in Figure 9 looks dramatically different from the base case in Figure 7. In contrast to the base case, the value V is now fully determined by the focal firm's resource position, and not by firm 2's, because the firms do not compete on resource markets anymore. Moreover, the industry volatility in this case is now significantly higher than the firm-specific volatility. Hence, the high firm-specific volatility in the base case is attributable to the resource competition.

Figure 9: Equilibrium without resource competition ($d = 0$)

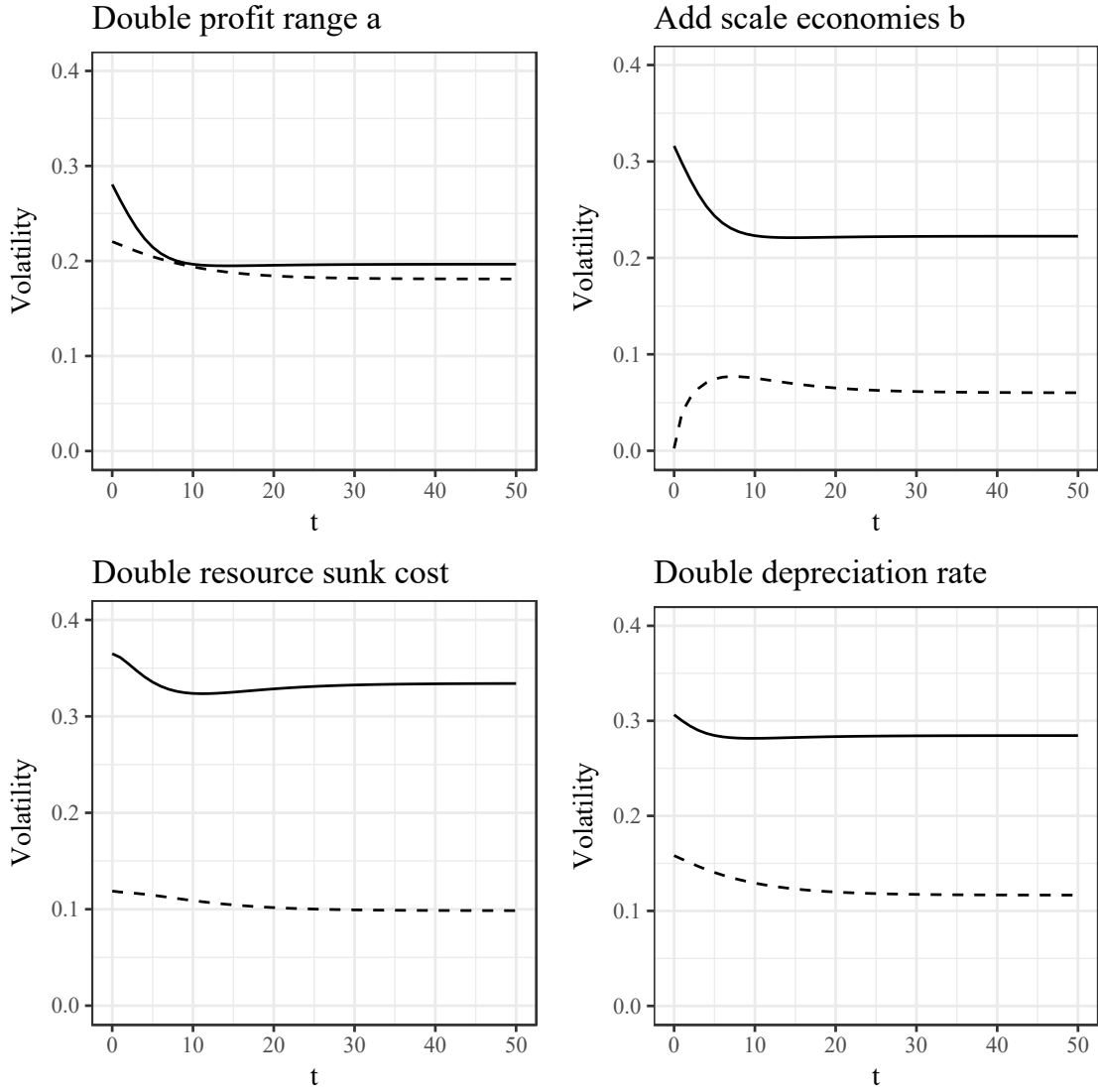


3.3.2. Variations from base case

Though resource competition in general will amplify any firm-specific, idiosyncratic shocks, the extent to which is affected by the dynamic characteristics of the resource. Figure 10 shows the evolution of firm and industry volatility over time for various changes of a single parameter compared to the base case in Figure 7. Shown in the top left panel, changing the range of industry states a_k markedly increases the industry volatility and to a lesser extent the firm-specific volatility, too. Thus, unsurprisingly, industries in which key resources have strong exposure to macro-economic shocks (such as natural resources and utilities) should be expected to have a larger share of performance variance explained by industry factors compared to firm-specific factors.

The three other panels each change one aspect of the competitive dynamics of the resource market. In general, scale economies and higher resource (sunk) cost lead to more intense resource competition and thus to higher firm-specific volatility (top right and bottom left panel). The effect of depreciation does change over time (bottom right panel). In mature stages a lower depreciation rate diminishes resource competition and firm-specific variance, because resource positions are more entrenched. In the nascent stages of the industry

Figure 10: Firm and industry components for various parameter changes



Note. Each figure changes one parameter compared to the base case from Table 11 and Figure 7. Solid lines represent the firm component and dashed lines the industry component of volatility.

though, high-depreciation resources lead to intense competition, exactly because they lead to later-entrenched and thus highly valuable positions.

Interestingly, for parameter changes within plausible ranges based on empirically observed data, the firm-specific component of volatility remains always higher than the industry component, even with, for instance, an extremely low depreciation rate of $\delta = 0.02$ per year (meaning a resource lifetime of more than $1/0.02 = 50$ years—not shown in the graphs). The reason is that even with these entrenched resource positions, whenever a set of resources is depreciated and becomes available, there will be intense competition for it, and the player that obtains the resource will get a strong increase in returns (and its competitor a decrease).

Finally, Figure 11 shows what happens in the presence of intense resource competition (strong scale economies and high resource cost). In this case, the steady state distribution (right panel) becomes “split”: the most likely state is that either one of the firms ends up with all the resources, deriving such strong benefits from its position that it will be unattractive for the other player to invest at all (left panel). In practice that would mean that in the presence of such strong resource competition, a monopoly would emerge. This has for instance been the case with Microsoft Windows. The critical resource here (installed base) has high sunk costs c and very strong network externalities, leading to strong scale economies b . For instance, in the race between Microsoft Windows and IBM OS/2 in the 1990s, indeed the latter at some point stopped investing, yielding a monopoly to Microsoft.

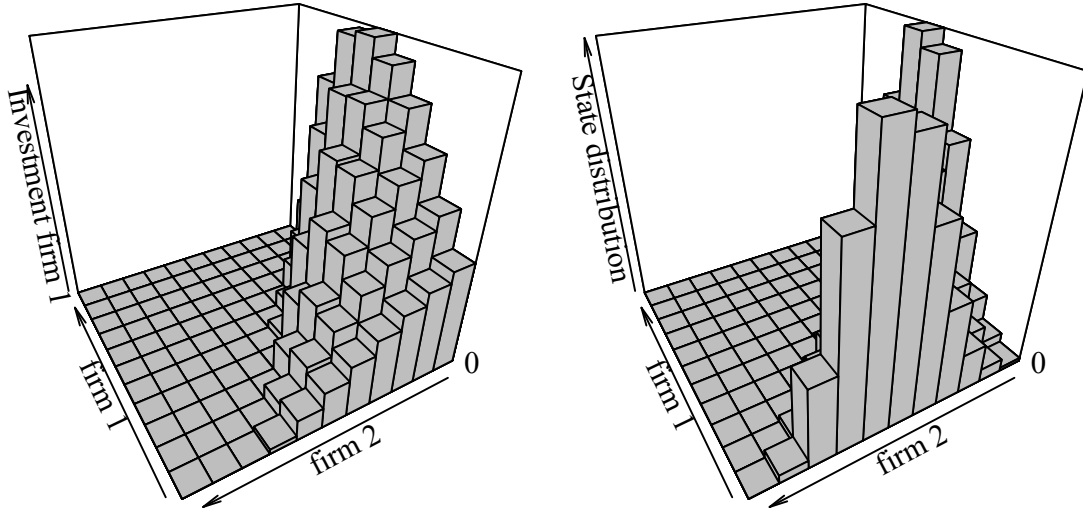
3.3.3. Propositions

The results from the model can be summarized in the following four propositions.

Proposition 1. *Resource competition will amplify idiosyncratic shocks, leading to high firm-specific volatility compared to industry-wide volatility, across a wide range of industry conditions.*

The above proposition is in line with the typical findings of the variance component studies. However, these studies assume a single number for the firm-specific component across

Figure 11: Equilibrium with scale economies and high resource cost



industries, which contrasts with the findings of the present model:

Proposition 2. *Though firm-specific variance will (almost) always be the largest component of performance variation, the specific percentage varies across industries.*

Moreover, the model makes predictions about the types of industries in which to expect high firm and industry volatility:

Proposition 3. *Several time-constant dynamic characteristics of resource competition drive persistent differences in firm-specific and industry-wide volatility, for instance:*

- a. *Industries that exhibit strong scale economies, high sunk costs and high depreciation will have a high firm-specific volatility (e.g., high-tech).*
- b. *Industries with a strong exposure to macro-economic variables such as commodity prices will have a high industry volatility (e.g., natural resources and utilities).*

Proposition 4. *Additionally, some time-varying dynamic characteristics of resource competition drive changing differences in firm-specific and industry-wide volatility, for instance:*

- a. *Rapidly growing industries will have a high firm-specific volatility (e.g., technology industries in the 1990s)*

b. Industries that face macro-economic uncertainty or turmoil will have a high industry volatility (e.g., financial services in the late 2000s and early 2010s).

3.4. Empirical evidence

The model in the previous section suggests that one should expect to find differences in firm-specific heterogeneity across industries. By contrast, most variance decomposition studies have implicitly assumed that the level of firm-specific vs. industry-level variance is constant. To test whether the model predictions hold, I perform a Bayesian analysis based on the classical variance decomposition studies, which allows for varying levels of firm-specific vs. industry-level variance.

3.4.1. Method and data

Most performance decomposition studies use RoA or some other accounting metric to measure performance. However, as usually acknowledged, accounting metrics are prone to specific conventions and only imperfectly measure actual economic returns (Fisher and McGowan, 1983; Lippman and Rumelt, 2003). These differences are likely to be systematic: because human capital is often not capitalized on the balance sheet, research and/or service intensive industries are likely to have systematically higher accounting returns than manufacturing industries. For instance, the high average RoA in the pharmaceuticals industry could be either due to truly higher economic returns, or due to the fact that a large part of research and marketing capital is not reflected on the balance sheet (or both). This systematic bias of accounting returns could lead to a spurious industry effect in the variance component studies.

Moreover, the complicated correlation structure of RoA panel data makes it hard to draw consistent inferences. For instance, RoA tends to exhibit a strong autoregressive component that is not explicitly accounted for in variance decomposition studies. Also the various nested effects (e.g., business unit and industry) have strong cross-correlations that make outcomes quite sensitive to the exact procedure and order used (McGahan and Porter,

2002). Indeed, many of the conflicting conclusions among various authors appear to stem from differences in procedures and interpretations of the decomposition results.

Finally, the commonly applied techniques in variance component studies, such as ANOVA, VCA and HLM (Vanneste, 2017), do provide single estimates across the entire sample, while based on the model in this paper it is plausible that in some industries the firm-specific effects are larger than in other industries. Indeed, McGahan and Porter (1997) find different effects across broad economic sectors, and Waring (1996) finds differences in profitability persistence rates across industries. However, with classical techniques these can only be explored by running separate analyses with different samples, which can greatly reduce statistical power.

In this study I use a Bayesian analysis of stock market data to address some of the concerns raised above. A major advantage of stock market data is that they are not prone to differences in accounting conventions. Of course they can be distorted by market sentiments and bubbles, but these aberrations tend to cancel out over longer time frames, while differences in RoA (or return on capital, RoC) due to accounting conventions can persist in the long run (Fisher and McGowan, 1983). Another major advantage of stock market data is that they tend to have very low serial correlation, especially when using monthly or annual (as opposed to daily or weekly) data (Tsay, 2005). This allows making a much cleaner decomposition into industry vs. firm effects, without the need to worry about the specific procedure used, or dealing with various cross-correlations.

Moreover, the use of Bayesian analysis overcomes some of the drawbacks of classical variance decomposition techniques. Though the essence of the analysis remains the same (a model in which variances in returns are decomposed into various components), Bayesian techniques naturally allow for variation in parameters within the same analysis; *in casu* the size of firm and industry effects can be allowed to vary across industries, by setting up a hierarchical model (Alcácer et al., 2013; Gelman et al., 2013, ch. 5). Moreover, because in Bayesian analysis all parameters are treated as probability distributions themselves, including the

variance parameters of the various effects, their statistical significance can be assessed.

Specifically, I use share total shareholder return data from the CRSP database, through WRDS (Wharton research data services), for all firms with all monthly data available from 1985 to 2015, split into three different decades to assess changes over time. I use 6-digit GICS (Global Industry Classification Standard) codes from the WRDS Compustat linked database to define industries—these codes are more balanced, consistent and up to date than the often-used SIC (Standard Industry Classification) codes.⁵ To focus the analysis on the economically most relevant companies, I exclude companies that have a market capitalization below 1 billion dollars (deflated to account for the overall growth in market cap over the past 30 years), as well as industries with fewer than three companies in them. Finally, I exclude non-primary listings such as ADRs (American Depository Receipts). Together, these exclusions amount to less than 20% of the total market capitalization of all US listed stocks. This leads to a database of 656,294 firm-month observations for 4,756 individual firm listings across 67 industries.

As is common in the finance literature, I use a four-factor model to account for aggregate risk (Fama and French, 1993; Carhart, 1997). The residuals from the stock prices regressed on the four factors yield the risk-adjusted excess total shareholder returns r_{ijt} , for each firm i in industry j and month t .

As is standard in variance component analyses, I am interested to decompose r_{ijt} in different factors; in this case into a firm-specific and an industry-wide component:

$$r_{ijt} = r_{jt}^{\text{ind}} + r_{ijt}^{\text{firm}}$$

Each of these components comes from a random distribution with mean zero (because I use risk-adjusted returns, which provide a measure of excess returns above or below the

⁵See <https://www.msci.com/gics> for more background behind the GICS codes, as well as full descriptions available for download.

average) and an unknown variance:

$$r_{jt}^{\text{ind}} \sim \mathcal{N}(0, \sigma_j^{\text{firm}}) \quad (3.6)$$

$$r_{ijt}^{\text{firm}} \sim t_\nu(0, \sigma_j^{\text{ind}}) \quad (3.7)$$

To identify the model, these components are all assumed to be independent. Note that I use a Student’s t distribution for the firm specific effects, because it is known that shareholder returns can have long tails and outliers that could strongly affect the analysis outcomes when using normal distributions (Tsay, 2005). In a Bayesian analysis, using a t distribution instead of a normal one makes the analysis more robust to such outliers—this is another advantage of Bayesian methods vis-à-vis classical inference (Gelman et al., 2013).

Note that in a classical analysis this decomposition would assume that both the firm and the industry components have a fixed variance (or equivalently a fixed standard deviation) across observations, which then can be estimated—i.e., σ_j^{firm} and σ_j^{ind} would not be able to depend on industry j . Because in a Bayesian analysis, parameters are treated as random variables themselves, it poses no problem to vary the parameters. The variance parameters are assumed to be draws from a random distribution again, which is commonly assumed to be an Inverse χ^2 distribution, as it is a so-called conjugate prior for a variance in a normal distribution (Gelman et al., 2013, p. 43):

$$(\sigma_j^{\text{ind}})^2 \sim \text{Inv-}\chi^2(\nu^{\text{ind}}, \sigma_0^{\text{ind}}) \quad (3.8)$$

$$(\sigma_j^{\text{firm}})^2 \sim \text{Inv-}\chi^2(\nu^{\text{firm}}, \sigma_0^{\text{firm}}) \quad (3.9)$$

All inference is performed using the Stan implementation in R (Carpenter et al., 2016; Gelman et al., 2013, Ap. C). Results are based on posterior inference of 4 Hamilton Monte Carlo (HMC) chains of 1,000 simulations each. The first 500 simulated draws have been discarded as warm-up⁶, and are thinned by a factor of 2, for a total of 1,000 draws from

⁶Because Bayesian simulation starts from some random initial values, the first simulation draws will

the posterior. Due to correlations between the different draws, the effective sample size is sometimes lower than the actual number of draws. For each parameter of interest, n_{eff} indicates the effective number of posterior draws, which should be at least 10 times the number of chains—40 in this case (Gelman et al., 2013, p. 287). Moreover, the \hat{R} statistic needs to be below 1.1 for all variables, to ensure that a sufficient number of draws have been discarded as warm-up, and that the sampling is indeed performed from the actual posterior (Gelman et al., 2013, p. 287).

3.4.2. Results

As an illustration of the data, Figure 12 shows the cumulative excess return evolution in four industries over a 5-year period.⁷ Each gray line represents a single firm and the thick black line represents the industry average (i.e., the average of all gray lines). Large variations in the black line indicate a high industry volatility σ_j^{ind} (expression (3.6)), while large variations of the gray lines vis-à-vis the industry average indicate a high firm-specific volatility σ_j^{firm} (expression (3.7)).

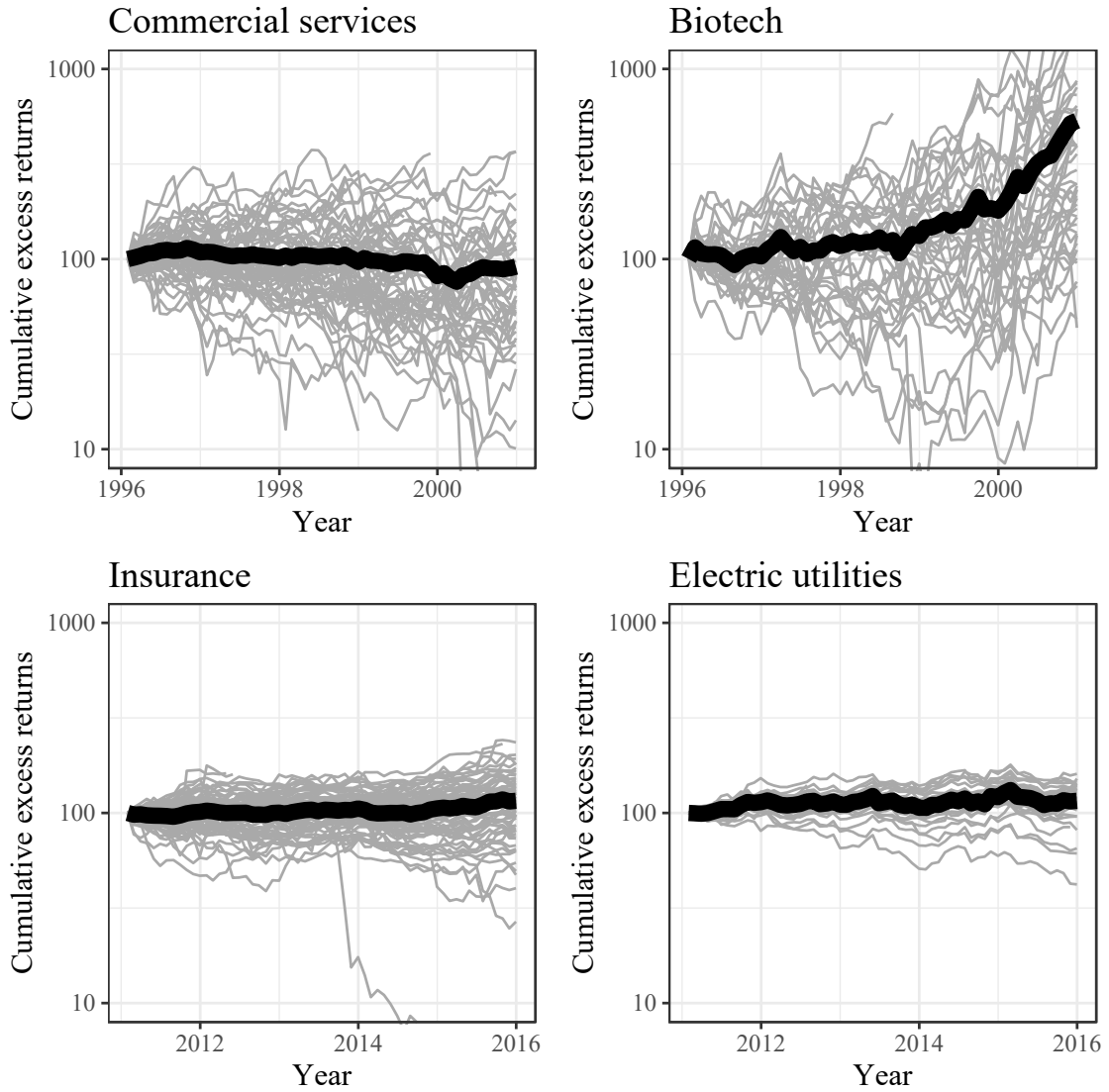
These data already suggest that there are marked differences in the heterogeneity patterns across industries, consistent with Proposition 2. For instance, the two industries on the left exhibit lower variations in the black line (i.e., lower σ_j^{ind}), than those on the right. Also, the bottom two industries exhibit much lower variations of the gray lines around the black line (i.e., lower σ_j^{firm}) than those on the top row.

The Bayesian analysis of all industry data over the full 30-year period corroborate these initial results. Table 12 shows the posterior inference of the parameters in the hierarchical priors in expressions (3.8) and (3.9). The posterior mean and standard deviation (in brackets) correspond to the classical maximum likelihood estimates and standard errors; the posterior interval corresponds to the classical confidence interval. As mentioned earlier,

reflect the initial starting point, rather than the actual posterior distribution. Therefore, the first draws (called the “warm-up” or “burn-in”) need to be discarded, until they can be expected to reflect the actual posterior (Gelman et al., 2013, p. 282).

⁷The cumulative excess return over a period from $t = 0$ to $t = T$ is defined by $100 \cdot \prod_{t=0}^T (1 + r_{ijt})$.

Figure 12: Excess return patterns in four industries



Note. Each figure shows the evolution of cumulative excess returns (logarithmic scale) for a given industry over a five year period. Each grey line represents a single firm, and the thick black line represents the industry average of all firms in that industry (i.e., the average of the gray lines).

Table 12: Bayesian posterior inference for main parameters

Parameter	Posterior mean (standard deviation)	Posterior 95% interval	n_{eff}	\hat{R}
Decade 1985-1995:				
ν	3.50 (0.03)	[3.45, 3.56]	836	1.00
ν^{firm}	6.59 (1.07)	[4.67, 8.74]	926	1.00
σ_0^{firm}	19.89 (0.69)	[18.51, 21.20]	1000	1.00
ν^{ind}	4.02 (0.75)	[2.72, 5.62]	296	1.02
σ_0^{ind}	6.44 (0.35)	[5.72, 7.12]	256	1.01
Decade 1995-2005:				
ν	3.13 (0.02)	[3.08, 3.18]	896	1.00
ν^{firm}	5.19 (0.84)	[3.61, 6.88]	624	1.00
σ_0^{firm}	22.13 (0.85)	[20.43, 23.71]	740	1.00
ν^{ind}	5.28 (0.97)	[3.59, 7.39]	538	1.01
σ_0^{ind}	10.07 (0.43)	[9.25, 10.96]	458	1.00
Decade 2005-2015:				
ν	2.82 (0.02)	[2.78, 2.86]	949	1.00
ν^{firm}	5.00 (0.80)	[3.47, 6.71]	806	1.00
σ_0^{firm}	16.27 (0.65)	[15.03, 17.60]	895	1.00
ν^{ind}	4.06 (0.72)	[2.75, 5.58]	790	1.00
σ_0^{ind}	7.90 (0.37)	[7.18, 8.66]	773	1.00

Note. Bayesian posterior inference for the hierarchical prior parameters in expressions (3.8) and (3.9). Posterior mean and standard deviation correspond with the classical maximum likelihood estimate (MLE) and standard error; the posterior interval corresponds with the classical confidence interval. n_{eff} and \hat{R} are measures of convergence, and should be above 40 and below 1.1, respectively (Gelman et al., 2013, p. 287). See text for details.

n_{eff} and \hat{R} are measures of convergence, and are well within their desirable ranges of above 40 and below 1.1, respectively (Gelman et al., 2013, p. 287).

Across the different decades of analysis, the typical firm-specific volatility σ_0^{firm} is two to three times as large as the industry volatility σ_0^{ind} . This corresponds well with earlier results using RoA data in Figure 4. Because the latter contains variances, the ratio of volatilities from Table 12 needs to be squared in order to facilitate a direct comparison; thus, the firm-specific variance from the present analysis is a factor $2^2 = 4$ to $3^2 = 9$ higher than the industry variance. In his meta-analysis, Vanneste (2017, Table 6) finds an industry component of variance between 0.06 and 0.10 and a firm component (business plus

corporate) between 0.41 and 0.58. The resulting ratio of firm to industry variance between $0.41/0.10 \approx 4$ and $0.58/0.06 \approx 10$ corresponds very well with the ratios found in the present analysis.

Moreover, the relatively small values and standard errors of ν^{ind} and ν^{firm} indicate significant differences across industries. Table 13 makes this even more clear. This table shows the Bayesian estimates for σ_j^{firm} and σ_j^{ind} for each industry j , averaged over all three decades. Clearly, the differences between industries are highly significant, both statistically and economically—the difference between the industry with the lowest firm-specific volatility (multi-utilities) and the highest (internet software & services) is more than a factor of four.

Finally, Table 14 shows the industry pattern over time by decade. The industries are averaged into eleven 2-digit GICS sectors to facilitate the comparison. Both the firm-specific (σ_j^{firm}) and industry component of volatility (σ_j^{ind}) are shown. Consistent with Proposition 3 there is significant persistence in these measures over time, with the energy and utilities sectors always having a much lower firm-specific component than the other sectors. Moreover, as expected from the model, industries with strong scale economies and high depreciation rates (such as information technology) exhibit relatively high firm volatility, while industries with strong exposure to macro-economic variables (in particular energy) exhibit high industry volatility. Consistent with Proposition 4, the firm volatility in information technology is particularly high in the boom period 95-05. A finer grained analysis (unreported) shows that, similarly, the banking industry shows a clear rise in industry volatility around the financial crises between 2008-2010, while after that returning to stability.

3.4.3. Regression

The previous results provide suggestive empirical evidence that is consistent with the model. To ensure that these effects are statistically robust, I additionally perform formal hypothesis

Table 13: Firm-specific and industry-wide volatility by industry

Industry	σ_j^{firm}	σ_j^{ind}
Multi-Utilities	9.86 (0.20)	10.13 (0.70)
Electric Utilities	10.10 (0.13)	10.04 (0.68)
Gas Utilities	11.47 (0.23)	8.58 (0.64)
Water Utilities	14.18 (0.81)	8.51 (1.25)
Household Products	14.40 (0.47)	6.89 (0.83)
Industrial Conglomerates	14.46 (0.55)	6.12 (0.83)
Banks	15.46 (0.13)	9.32 (0.61)
Beverages	16.21 (0.44)	7.23 (0.74)
Insurance	16.21 (0.16)	7.35 (0.52)
Tobacco	16.65 (0.67)	12.56 (1.24)
Containers & Packaging	17.75 (0.39)	8.18 (0.74)
Thrifts & Mortgage Finance	18.16 (0.42)	9.17 (0.86)
Food Products	18.28 (0.28)	6.89 (0.57)
Aerospace & Defense	18.28 (0.33)	7.93 (0.69)
Capital Markets	18.75 (0.31)	7.05 (0.64)
Machinery	18.78 (0.23)	8.18 (0.61)
Equity Real Estate Investment Trusts (REITs)	18.78 (0.60)	5.90 (0.88)
Construction Materials	18.82 (0.69)	11.24 (1.26)
Distributors	19.18 (0.92)	6.35 (1.16)
Chemicals	19.48 (0.25)	7.50 (0.57)
Transportation Infrastructure	19.71 (1.40)	7.80 (2.05)
Trading Companies & Distributors	19.82 (0.65)	7.37 (1.04)
Road & Rail	19.98 (0.41)	10.19 (0.85)
Food & Staples Retailing	20.24 (0.37)	6.42 (0.64)
Paper & Forest Products	20.69 (0.71)	10.82 (1.31)
Marine	21.00 (0.97)	9.09 (1.69)
Independent Power and Renewable Electricity Producers	21.00 (0.78)	13.60 (1.45)
Building Products	21.34 (0.60)	8.20 (0.99)
Electrical Equipment	21.43 (0.49)	6.81 (0.77)
Professional Services	21.43 (0.66)	8.67 (1.13)
Diversified Financial Services	21.49 (0.56)	6.24 (0.80)
Multiline Retail	21.52 (0.51)	11.78 (1.00)
Media	21.76 (0.25)	6.49 (0.52)
Personal Products	21.84 (0.66)	7.72 (1.03)
Automobiles	21.88 (0.87)	9.93 (1.53)
Commercial Services & Supplies	21.91 (0.29)	5.06 (0.49)
Auto Components	21.97 (0.46)	10.93 (0.98)
Oil, Gas & Consumable Fuels	22.23 (0.23)	16.83 (1.15)

Table 13 continued

Industry	σ_j^{firm}	σ_j^{ind}
Life Sciences Tools & Services	22.32 (0.56)	7.35 (0.91)
Real Estate Management & Development	22.40 (0.91)	6.78 (1.41)
Air Freight & Logistics	22.58 (0.90)	10.62 (1.61)
Health Care Equipment & Supplies	23.11 (0.32)	6.71 (0.59)
Leisure Products	23.16 (0.80)	7.63 (1.28)
Household Durables	23.16 (0.37)	11.04 (0.88)
Diversified Telecommunication Services	23.60 (0.49)	9.53 (0.88)
Energy Equipment & Services	23.70 (0.38)	25.79 (1.72)
Consumer Finance	23.71 (0.72)	9.34 (1.17)
Hotels, Restaurants & Leisure	23.90 (0.31)	8.91 (0.72)
IT Services	24.11 (0.38)	6.44 (0.66)
Diversified Consumer Services	24.21 (0.62)	8.63 (0.99)
Textiles, Apparel & Luxury Goods	24.59 (0.45)	10.24 (0.89)
Construction & Engineering	25.10 (0.71)	10.57 (1.25)
Electronic Equipment, Instruments & Components	25.12 (0.37)	8.38 (0.75)
Airlines	25.30 (0.69)	20.15 (1.56)
Health Care Providers & Services	25.39 (0.32)	10.81 (0.84)
Pharmaceuticals	25.85 (0.43)	8.97 (0.83)
Metals & Mining	26.59 (0.35)	18.59 (1.30)
Specialty Retail	27.41 (0.32)	11.70 (0.87)
Wireless Telecommunication Services	28.47 (0.70)	13.79 (1.43)
Semiconductors & Semiconductor Equipment	29.90 (0.41)	19.53 (1.40)
Technology Hardware, Storage & Peripherals	32.57 (0.55)	11.95 (1.11)
Health Care Technology	32.73 (1.33)	11.82 (2.31)
Software	33.32 (0.38)	9.93 (0.84)
Communications Equipment	35.07 (0.53)	10.33 (0.99)
Biotechnology	35.36 (0.58)	17.27 (1.35)
Internet & Direct Marketing Retail	37.57 (1.37)	9.61 (2.20)
Internet Software & Services	43.17 (1.45)	11.98 (2.16)

Note. Annualized percentage of firm-specific and industry-wide volatility by six-digit GICS industry, sorted by increasing firm-specific volatility. Shows averages over three decades for the Bayesian posterior mean, with standard deviation in brackets; these correspond with the classical maximum likelihood estimate (MLE) and standard error.

Table 14: Analysis by decade

GICS	Industry name	Firm component			Industry component		
		85-95	95-05	05-15	85-95	95-05	05-15
10	Energy	19.9	22.9	24.2	14.0	22.8	18.7
15	Materials	20.4	22.3	22.2	10.0	12.2	15.0
20	Industrials	19.7	21.2	16.2	6.5	9.5	9.0
25	Consumer Discretionary	24.1	26.0	21.0	7.4	11.0	9.9
30	Consumer Staples	18.5	20.1	15.2	5.9	9.9	7.4
35	Health Care	25.4	30.1	23.3	7.7	13.0	9.1
40	Financials	19.0	18.0	15.4	7.3	9.1	8.5
45	Information Technology	30.5	40.7	22.3	11.4	16.6	7.5
50	Telecommunication Services	17.9	36.1	18.9	7.1	13.2	10.1
55	Utilities	10.6	15.6	9.7	7.9	12.0	11.6
60	Real Estate	19.6	20.1	18.5	5.6	6.2	6.3

Note. Firm-specific and industry components of volatility by decade, aggregated by two-digit GICS industry. Volatilities are annualized.

tests using a regression analysis. As dependent variables I use logarithms of the posterior median estimates from the Bayesian analysis of the *firm component* and *industry component* of volatility (annualized percentages), by industry-decade at the 6-digit GICS level. As is common in this type of analysis, I exclude financial services (GICS 40xxxx), because these industries often use different accounting standards.

All independent variables are calculated using the Compustat database. I use *R&D intensity* and *advertising intensity* as proxies for sunk resource costs, both defined as the relevant item from the income statement as a percentage of revenues. I use *capital intensity*, defined as capital expenditure over sales, as a proxy for asset exposure to macro-economic shocks. *Depreciation rate* is defined as the income statement depreciation and amortization as a percentage of net property, plant and equipment plus intangibles on the balance sheet. *Demand shocks* are defined as the logged standard deviation of EBITDA (earnings before interest, taxes, depreciation and amortization) margins. *Industry growth* is defined as the compounded annual growth rate of the total revenues of all companies combined in the industry. I do not use a proxy for scale economies, because prior literature has found it to be highly endogenous, and thus hard to estimate (e.g., Basu and Fernald, 1997).

Table 15 shows distributional summary statistics (median and end points of the 95% interval) as well as correlations for all variables. Table 16 shows the regression analyses for the hypothesis tests.⁸ All models use ordinary least squares (OLS) with decade fixed-effects (to remove the effect of any changes in volatility common to all industries) and cluster-robust standard errors at the industry level (to account for any serial correlation).

Models (1) and (4) are pure fixed effect models to assess the extent to which firm and industry components of volatility reflect time-constant characteristics of the industry. Both R-squares are high and very significant ($p < 0.001$ for the F-statistics); the firm component is for 92% explained by time-constant or economy-wide effects, and the industry component for 78%. This provides strong support for the first part of Proposition 3: especially firm-

⁸Table made with the help of the `texreg` package (Leifeld, 2013).

Table 15: Summary statistics and correlations

	Distribution				Correlations							
	2.5%	50%	97.5%	1	2	3	4	5	6	7		
1. Firm component (%) [†]	9.70	25.39	49.04									
2. Industry component (%) [†]	4.25	7.97	21.70	0.25								
3. R&D Intensity	0.00	0.01	0.15	0.39	0.20							
4. Advertising intensity	0.00	0.00	0.07	0.19	-0.05	0.05						
5. Depreciation rate	0.03	0.08	0.21	0.59	0.08	0.32	-0.01					
6. Industry growth [†]	-0.03	0.05	0.19	0.40	0.01	0.28	0.05	0.36				
7. Capital intensity	0.02	0.06	0.23	-0.18	0.34	0.00	-0.21	-0.15	0.06			
8. Demand shocks [†]	0.00	0.01	0.07	0.19	0.46	0.30	-0.00	0.08	0.18	0.43		

Note. Distributional statistics and correlations for variables by industry-decade. Distribution shows 95% percentile endpoints and median.

[†]Logged variable used in correlations and regressions (but not in distribution); one plus log used for industry growth.

Table 16: Regression results

Dependent variable:	Firm component [†]			Industry component [†]		
	(1)	(2)	(3)	(4)	(5)	(6)
	FE	Full	Winsor	FE	Full	Winsor
R&D Intensity		0.91*** (0.24)	1.05 (0.77)		0.35 (0.69)	-1.00 (1.28)
Advertising intensity		2.83* (1.29)	2.97* (1.38)		-0.05 (1.39)	0.22 (1.54)
Depreciation rate		3.34*** (0.60)	3.32*** (0.76)		1.33 (0.69)	1.82 (0.96)
Industry growth [†]		1.11*** (0.33)	1.49** (0.45)		0.10 (0.49)	0.69 (0.64)
Capital intensity		-0.72 (0.69)	-1.10 (0.68)		1.78*** (0.53)	1.81** (0.62)
Demand shocks [†]		0.05 (0.03)	0.06 (0.03)		0.16** (0.05)	0.18** (0.06)
Industry fixed-effects	x			x		
Decade fixed-effects	x	x	x	x	x	x
Num. obs.	182	182	182	182	182	182
R ² (full model)	0.92	0.54	0.54	0.78	0.39	0.40

Note. OLS regression of firm component and industry component of volatility by GICS industry-decade. Models (1) and (4) are pure fixed effects models; models (2) and (5) include all untransformed variables; and models (3) and (6) use winsorized regressors to reduce the effect of any outliers. Clustered standard errors robust to serial correlation and heteroscedasticity in parentheses. R² for full model including fixed effects.

[†]Logged variable.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

specific volatility is for a large part determined by time-constant industry characteristics.

Models (2) and (5) regress respectively the firm and industry component of volatility onto the covariates. The results are remarkably consistent with propositions 3 and 4. Industries with high firm components of volatility are associated with high sunk costs (R&D and advertising intensity), high depreciation rates, and high growth. Industries with high industry components of volatility are associated with a high capital intensity and exposure to demand shocks. Moreover, most of these associations are highly significant (at the 0.01 or 0.001 level), even though the sample size is relatively small (fewer than 200 data points, with standard errors clustered into fewer than 70 industries).

Models (3) and (6) use winsorized regressors to reduce the effect of any outliers. The results are very similar to the unwinsorized models, except for the strong reduction in statistical significance of R&D Intensity on the firm component of volatility. The significance of this effect in model (2) is largely driven by Biotechnology (GICS 352010), which indeed has high firm-specific volatility and very high R&D intensity for each of the three decades.

3.5. Discussion

The model in this paper suggests that the dynamics of resource competition provide a very general mechanism for the emergence of high a firm-specific heterogeneity. Of course, many such mechanisms have been suggested in the past, such as mobility barriers between industry groups (Caves and Porter, 1977), evolutionary path dependence (Nelson and Winter, 1982), resource endowments (Barney, 1991), dynamic capabilities (Teece et al., 1997), and unique asymmetries (Miller, 2003). However, none of these mechanisms provides a reason why firm-specific heterogeneity should be high under all circumstances. This study does provide an answer to that question: high firm-specific heterogeneity should be expected to emerge under very general conditions, because resource competition amplifies small differences in resource positions into large differences in performance.

The assumptions of the model that prove minimally sufficient for the emergence of high

firm-specific heterogeneity are endogenous investment, scarce resources, time-compression diseconomies, and uncertainty. They are “minimal” in the sense that each is needed in conjunction with the others to lead to the amplification resulting in high firm-specific heterogeneity; removing any of the four conditions (without replacing it by a different one) will remove the amplification effect.

First, *endogenous investment* is a critical part of a model for resource markets, because firms principally make sunk cost investments in resources because of their long term effects, and not myopically to improve current profits. For instance, in order to build a resource position in a market, firms almost always have to suffer initial losses before becoming profitable, which can’t be squared with pure myopia. The endogenous investment model in this paper does assume forward looking managers and is consistent with this phenomenon: in the initial resource state $q_i = q_j = 1$ firms make investments of $x = \$1.5$ on profits of just $\pi = \$1$ (using the base case parameters in Table 11).

Second, *scarce resources* are modeled through increasing cost in equation (3.3) when the resources used by both firms ($q_i + q_j$) approaches some finite stock Q . As Figure 9 shows, removing the competition for scarce resources makes the value function trivial, and the firm-specific heterogeneity low.

Third, *time-compression diseconomies* are needed on a much more fundamental level. If there were no time-compression diseconomies, each firm could obtain any resource position overnight, thus competing any profits immediately away. The time-compression diseconomies in the present model are implemented through the concavity of the expected value of the resource efficiency function in equation (3.2). This means that if a firm invests twice the amount of money x in a given time period, it increases the probability (and thus expected value) of resource gain θ_{+1} by less than a factor of two. In other words, spreading an amount of money x over two periods increases the expected resource gain vis-à-vis investing the same amount in a single period.

Fourth and finally, *uncertainty* is required to provide the seeds of heterogeneity. In absence of all uncertainty, the symmetry of the model in firm 1 and firm 2 would lead to two firms that are exactly the same, without any heterogeneity. Of course, in actuality there will always be differences across firms, if only because no two people are exactly the same. The example in Figure 8 shows that any such small differences will be amplified in performance due to resource competition when the uncertainty is combined with the other three conditions.

As an interesting aside, the first three conditions can be seen as principles covering the functioning of resource markets, particularly demand, long-term supply and short-term supply, respectively. The long-term supply curve driven by the scarcity of resources incorporates increasing costs as resources become more depleted—quite analogous to ‘ordinary’ product markets. However, due to time-compression diseconomies, there will be a short-term supply curve that is much steeper, limiting the amount of resources that a firm can obtain in a short period of time depending on its size—in general, larger firms will be able to grow faster in absolute terms (Knudsen et al., 2014). In resource markets, there is no direct analogue of the demand curve though. While in product markets the demand curve is independent of the supply curve—product demand depends solely on utility functions, and not on production functions—in resource markets firms take future supply and potential growth rates into account in making investment decisions. This provides a fundamental reason why resource markets cannot be analyzed through ordinary supply and demand functions, but require some version of the machinery of dynamic games used in this paper.

Apart from the above four key assumptions underlying the model, there are of course several other specific assumptions needed to develop the model. For instance, the model in this paper assumes two firms without entry or exit, in order to keep the model parsimonious while capturing the essence of competition—one could think of the two firms as the focal firm and its strongest competitor. This is a common assumption in the literature for models investigating specific competitive mechanisms (e.g., Makadok and Barney, 2001; Besanko and Doraszelski, 2004; Adner and Zemsky, 2006). For similar reasons, the model focuses on a

single resource with a fixed amount of inimitable resources (following, for instance, Makadok and Barney, 2001; Jacobides et al., 2012), and assumes a “winner takes all” competition in each segment (see appendix A.2), similar to the strong product market competition assumed in for instance value based strategy (Brandenburger and Stuart, 1996).

Also, the dynamic game model used in this paper comes with rather strong assumptions about the rationality of managers and the information that is available to them: managers in actual companies are not known to optimize a Bellman equation (3.5), nor would they have the exact information to do so. However, because the key industry characteristics of resource competition turn out to be relatively stable over time, managers and investors can be expected to develop investment heuristics that are consistent with rational optimization—if only due to evolutionary processes: firms and investors that deviate from the equilibrium strategy will in the long term be less successful. Indeed, the equilibrium investment strategies appear to be consistent with actually observed investment behavior; for instance, in markets with strong scale economies and large sunk cost (such as technology industries), firms tend to make major investments early onward to rapidly gain favorable competitive positions at the expense of competitors, consistent with the model predictions in Figure 11.

As a direction for further research it would be fruitful to study models of resource competition under less strict assumptions. For instance, evolutionary or learning models could be used to alleviate the strong rationality assumptions. This would also reduce the computational burden when allowing more firms and more resource markets, as the Markov perfect equilibrium calculation would rapidly become unfeasible due to the exponential growth of the state space with the addition of state variables. Therefore, evolutionary models with alleviated rationality assumptions could be ideally suitable to study resource competition, guided by the key four conditions outlined in this paper.

CHAPTER 4 : Performance persistence in the presence of higher order resources

4.1. Introduction

The persistence of profitability differences even among direct competitors is a well-documented empirical regularity in the strategy literature (e.g., Mueller, 1977; Jacobsen, 1988; McGahan and Porter, 1999). The resource based view (RBV) provides an explanation for why these profitability differences are not rapidly competed away, invoking firms' access to heterogeneous and immobile resources with limits to competition. (e.g., Wernerfelt, 1984; Barney, 1991; Peteraf, 1993). The RBV can be augmented with the notion of *higher order resources*, providing an explanation for how certain firms are able to systematically develop and maintain favorable resource positions in their product markets, without having to pay their full economic price on strategic factor markets. Unlike "ordinary" or "operating" resources, higher order resources do not allow a firm to make higher profits directly, but instead allow it to systematically obtain other superior resources that over time allow it to increase profits. This concept of higher order resources (or higher order capabilities) provides a way to characterize dynamic capabilities within the RBV (Winter, 2003; Helfat et al., 2007).

However, despite the extensive literature on the RBV and dynamic capabilities, little is known about how higher order resources theoretically and empirically affect profit persistence patterns, which is a particularly puzzling gap given that the relation between resources and profitability has been a central theme in the strategy literature. In this paper, I address this gap with a formal stochastic model of the dynamics between higher order resources, operating resources and profits. I find that operating resources, which affect profits directly, lead to persistence in the *level* of profit differences; the addition of higher order resources, which affect other resources but not directly profits, leads to persistence in the *growth* of profit differences.

The model also provides an empirical test of the presence of higher order resources in a

30-year panel of profit data. The model shows that the presence of higher order resources versus merely operating resources, translates into different classes of autoregressive moving average (ARMA) models. I test for these different models using both classical statistical time series and Bayesian hierarchical methods, which allow for estimation of differences across industries (Alcácer et al., 2013; Mackey et al., 2017; Gelman et al., 2013). I do find significant evidence for the persistence in both level and growth of profits, indicating the presence of higher order resources shaping profit persistence patterns.

This paper makes several contributions to the literature. First, the formal framework offers precise relations between resources and performance, without equating the two, thus addressing the often-cited tautology critique on the RBV (Priem and Butler, 2001). Resources must *persistently* affect *expected* profit over time—providing two critical distinctions between resources and performance as such. Moreover, resources can be distinguished into those affecting expected profit directly (operating resources) and those affecting expected profit through other resources over time (higher order resources). The mathematical definitions and derivations in this paper allow deriving how the dynamics of operating resources and higher order resources individually as well as in conjunction affect profit persistence patterns in terms of stochastic time series models, as summarized in Table 18.

Second, this study offers a large scale empirical test of whether and to what extent higher order resources affect profit persistence patterns. This addresses an often heard call “... to make greater use of empirical methodologies beyond qualitative case analyses and analysis of survey data, such as ... econometric analysis of ‘big’ archival data, to further broaden the toolkit used in dynamic capabilities research.” (Schilke et al., 2018, p. 392). I find that higher order resources significantly affect profit persistence patterns over time, for a wide range of industries. A notable exception is that the effect of higher order resources on profit persistence patterns is much smaller, or even absent in some industries in the decade 1995-2004, providing an explanation for the findings of “hypercompetition” literature in the 2000s (e.g., D’Aveni et al., 2010), as well as recent findings of a reversion of this trend (e.g.,

Bennett and Gartenberg, 2016).

Third and finally, this paper offers new findings and methods for the persistence of profitability literature (e.g., Mueller, 1977, 1986; Jacobsen, 1988; Geroski and Jacquemin, 1988; Waring, 1996; Goddard and Wilson, 1999; McGahan and Porter, 1999; Bou and Satorra, 2007; Bottazzi et al., 2008; Madsen and Walker, 2017). For instance, the formal model of resource dynamics suggests that an autoregressive moving average (ARMA) model should be used to estimate the time series behavior of profit persistence, instead of the earlier used pure autoregressive (AR) models, which likely have led to an attenuation bias in prior studies. Indeed, I find significantly higher values for the first order persistence of profit, around 0.95. In some industries and time periods the first order persistence is even above one, consistent with recent findings of unit-root behavior in profit data (Canarella et al., 2013). Unlike traditionally used maximum likelihood estimators (MLE), the Bayesian inference used in this paper allows estimation of such non-stationary behavior without problems, providing direct estimates of the first-order persistence over time. I find that first order profit persistence above one (i.e., diverging profit differences) never last for more than a single decade, thus suggesting the existence of 'non-stationary episodes', rather than long-term non-stationary profit behavior.

4.2. Theory

The primary subject of this paper is how the dynamics of operating and higher order resources affect the persistence of profitability differences among firms. In this section, I will precisely define the notions of operating resources and higher-order resources, relate them to previous literature, and mathematically derive how they affect empirically observable persistence of profits over time.

4.2.1. Operating resources

Winter (2003) defines ordinary or “zero-level” capabilities as those “that permit a firm to ‘make a living’ in the short term” (p. 991). The idea to specifically distinguish capabilities

that contribute to short-term profits can be logically extended to other types of resources, defined as “something that the organization can draw upon to accomplish its aims” (Helfat et al., 2007, p. 4). Examples of such resources that can directly affect profits could be brands, patents, captive customers, or specialized plants. I will call them *operating resources*—the word “operating” is to be interpreted broadly, as a generalization of “operating routines” (King and Tucci, 2002; Zollo and Winter, 2002), thus not merely pertaining to production operations, but also to any other element of operating in a certain product market. Presumably, such operating resources are the ones that the original RBV authors mainly referred to in order to explain why certain firms are able to make above-normal profits for prolonged periods of time (Wernerfelt, 1984; Rumelt, 1984; Barney, 1986; Dierickx and Cool, 1989). Specifically, I define:

Definition 1. *Operating resources are resources that directly and persistently affect a firm’s expected profit.*

Two words are critical in this definition: “persistently” and “expected”. The latter means that the effect of a firm’s operating resources x_t in period t on its profit y_t can be described by $x_t = E y_t$. Conversely, this means that firm profit can be written in terms of the effect of its operating resources x_t plus a random noise term u_t that is uncorrelated with the firm’s resource base:¹

$$y_t = x_t + u_t \tag{4.1}$$

The meaning of “persistent” is that operating resources in subsequent periods must be correlated: $\text{cor}(x_{t-1}, x_t) > 0$. To first order approximation² this is equivalent to $x_t =$

¹More formally: $\text{cor}(u_t, x_t) = \text{cor}(u_t, u_{s \neq t}) = E u_t = 0$.

²The persistence of profit literature usually assumes an autoregressive model for profits. All these models are linear time series models: excess profits y_t in a given year are described in terms of a linear combination of its lags (y_{t-1}, y_{t-2}, \dots) as well as an error term ϵ_t and potentially its lags ($\epsilon_{t-1}, \epsilon_{t-2}, \dots$). In order to keep the model mathematically tractable as well as to make it comparable with the earlier literature, I want to stay within this framework of linear time series models, and thus require linear relations between profit and resources.

$\mu + \lambda x_{t-1} + v_t$, for some parameters $\mu, \lambda > 0$ and some random noise term v_t .³ This equation can be simplified by redefining x_t as the difference compared to the competitive average.⁴

$$x_t = \lambda x_{t-1} + v_t \tag{4.2}$$

Because in this paper I am primarily interested in the dynamics vis-à-vis the competitive average, rather than the (constant) level of the average itself, from now on I will work with demeaned values: the resource base x_t is to be understood as the resource advantage (or disadvantage) vis-à-vis competitors, and the profit y_t is to be understood as the profit in excess to the competitive average.

Equation (4.2) shows the importance of the persistence requirement in the definition. If resources were not persistent, this would mean $x_t = v_t$, and thus $y_t = v_t + u_t$: the profit y_t would be the sum of two uncorrelated noise terms, which is equivalent to a random noise process $y_t = u_t$. Intuitively this makes sense: a short spike in profit due to some very short-lived “resource” is indistinguishable from a spike in profit due to some other random event. Therefore, operating resources are defined as resources that persistently affect (expected) profit, i.e. for a prolonged period of time, captured by the requirement $\lambda > 0$. The value of λ captures to what extent operating resources are persistent.

As will be shown later, equations 4.1 and 4.2 are equivalent to a so called first-order autoregressive and first-order moving average (ARMA(1,1)) model for profits y_t . Previous literature has usually just considered a first-order autoregressive (AR(1)) model. However, it turns out that the moving average (MA) term is not inconsequential, and has likely led to a downward bias in the estimation of profit persistence in earlier studies.

³Similarly: $\text{cor}(v_t, x_t) = \text{cor}(v_t, u_t) = \text{cor}(v_t, v_{s \neq t}) = E v_t = 0$.

⁴Formally, this amounts to shifting x_t and y_t by a constant, redefining $x_t \mapsto x_t + \frac{\mu}{1-\lambda}$ and $y_t \mapsto y_t + \frac{\mu}{1-\lambda}$. In the case $\lambda = 1$ one can add or subtract an arbitrary constant without affecting the dynamics.

4.2.2. Higher order resources

Analogous to extending the concept of “zero-level” capabilities to other resources that directly affect profit, dynamic or “higher-order” capabilities (Winter, 2003) can be extended to other resources that govern the “rate of change” of operating resources (Collis, 1994; Winter, 2003):

Definition 2. *Higher order resources are resources that do not directly affect current expected profit, but that do persistently affect the expected rate of change of other resources.*

Similarly to the definition of operating resources, the notions of persistence and expected values are critical. As Winter (2003, p. 992) notes, the requirement of such patterns beyond random noise is critical to determine whether higher order resources exist in some non-trivial sense. If higher order resources were just defined as the rate of change of operating resources, they would trivially exist, because evidently firms’ resource positions change all the time. The interesting question is if such change is patterned, and hence the rate of change of operating resources persists over time. The primary purpose of this paper thus is deriving the effect of such persistent higher order resources on the time series behavior of profit, as well as whether and how broadly this effect can be empirically observed.

To mathematically formalize the effect of higher order resources on profit persistence, I will only consider second order resources, i.e., resources that directly affect operating resources.⁵ Conceptually the arguments can be easily extended to third order resources (resources that affect the rate of change of second order resources), and so forth.

By definition, and in exact analogy to operating resources, higher order resources z_t can be described by their effect on the rate of change of the operating resources, beyond the baseline from equation (4.2):

$$x_t = \lambda^{(1)}x_{t-1} + z_t + v_t \tag{4.3}$$

⁵The terminology for numbering of the higher orders in the literature is inconsistent; for instance, what I call “second order resources” corresponds to “first order dynamic capabilities” in Winter (2003) and to “second order competences” in Danneels (2016).

The evolution of z_t itself is described by:

$$z_t = \lambda^{(2)} z_{t-1} + w_t \tag{4.4}$$

Note that $\lambda^{(1)}$ and $\lambda^{(2)}$ respectively capture the persistence of operating and higher order resources. In the next subsection I will show with a couple of numerical examples how these parameters affect profit persistence, and afterward more formally derive the implied time series patterns.

4.2.3. Numerical examples

To provide some intuition for the meaning of the model in equations (4.1), (4.3), and (4.4), Table 17 shows several numerical examples for various values of $\lambda^{(1)}$ and $\lambda^{(2)}$. For all examples, the initial stock of both operating and higher order resources $x_0 = z_0 = 100$, and I suppress the random shocks $u_t = v_t = w_t = 0$ in order to focus exclusively on the effect of resource persistence on profits—in the more formal treatment in the next subsection I will include the random shocks to precisely derive their impact on the time-series behavior of profit.

Panel a shows the evolution of profit y , operating resources x and higher order resources z in the case that operating resources fully persist ($\lambda^{(1)} = 1$), while higher resources decay immediately ($\lambda^{(2)} = 0$); thus, only operating resources affect the evolution of profit in this example. In this simple case, operating resources and thus profit levels persist over time.

Next, panel b shows what happens if the roles of operating and higher order resources are interchanged: higher order resources fully persist ($\lambda^{(2)} = 1$), while operating resources decay immediately ($\lambda^{(1)} = 0$). Interestingly, this leads to the exact same pattern with constant profit. Though maybe surprising at first, this can be understood intuitively: the profit pattern that derives directly from a stable operating resource is equivalent to the profit pattern that derives via a short-lived operating resource from a stable higher order

Table 17: Numerical examples.

a. Operating resources ($\lambda^{(1)} = 1, \lambda^{(2)} = 0$)						
t	0	1	2	3	4	5
y	100	100	100	100	100	100
x	100	100	100	100	100	100
z	0	0	0	0	0	0

b. Higher order resources ($\lambda^{(1)} = 0, \lambda^{(2)} = 1$)						
t	0	1	2	3	4	5
y	100	100	100	100	100	100
x	100	100	100	100	100	100
z	100	100	100	100	100	100

c. Both resources ($\lambda^{(1)} = \lambda^{(2)} = 1$)						
t	0	1	2	3	4	5
y	100	200	300	400	500	600
x	100	200	300	400	500	600
z	100	100	100	100	100	100

d. Eroding resources ($\lambda^{(1)} = 0.5, \lambda^{(2)} = 0.9$)						
t	0	1	2	3	4	5
y	100	140	151	148	140	129
x	100	140	151	148	140	129
z	100	90	81	73	66	59

e. Eroding resources ($\lambda^{(1)} = 0.9, \lambda^{(2)} = 0.5$)						
t	0	1	2	3	4	5
y	100	140	151	148	140	129
x	100	140	151	148	140	129
z	100	50	25	13	6	3

Note. Numerical examples for the model in equations (4.1), (4.3), and (4.4), showing the evolution of profit y_t , operating resources x_t , and higher order resources z_t over time for different values of $\lambda^{(1)}$ and $\lambda^{(2)}$. For all examples $x_0 = z_0 = 100$ and $u_t = v_t = w_t = 0$.

resource.

Panel c shows the pattern in case both operating and higher order resources fully persist ($\lambda^{(1)} = \lambda^{(2)} = 1$). This leads to a markedly different pattern of profit evolution: now not merely the profit *level* persists, but profit *growth* persists. Thus, the interaction of operating and higher order resources leads to a unique pattern that is distinct from what would be possible if only operating or higher order resources were present.

Panel d shows the evolution when resource advantages erode over time. Specifically, $\lambda^{(1)} = 0.5$, meaning that operating resources on average decline with 50% per year, and $\lambda^{(2)} = 0.9$, meaning that higher order resources on average decline with 10% per year. This leads to a decline in z_t from a value of 100 in year 0 to 59 in year 5. The evolution of the operating resource advantage x_t (which still equals excess profits y_t in the absence of random shocks) is shaped by two opposing forces, the positive higher-order resource advantage z_t drives a growth in operating resources, while the direct effect due to $\lambda^{(1)} < 1$ drives a decline. In the first two years, the first force is stronger, while later the second become stronger, as the advantage in higher order resources diminishes.

Panel e shows an example similar to panel d, but with the values of $\lambda^{(1)}$ and $\lambda^{(2)}$ interchanged. Interestingly, this does not affect the evolution of the operating resource advantage x_t and excess profit y_t (though the evolution of higher order resources z_t is different). This is another instance of the equivalence observed between panel a and b. In general, for the time series behavior of profits it does not matter whether a firm possesses a rapidly deteriorating operating resource advantage and a slowly deteriorating higher order resource advantage, as compared to the other way around.

These examples suggest first that profit behavior in the presence of both operating and higher order resources is markedly different than in the presence of either one alone, and second that the roles and persistence rate of operating and higher order resources can be interchanged without affecting profit persistence patterns. Below I will derive these

statements more formally.

4.2.4. Transformation into ARMA time series

The persistence of profitability literature often uses autoregressive (AR) models to describe the time series behavior of profits (e.g., Mueller, 1977, 1986; Jacobsen, 1988; Geroski and Jacquemin, 1988; Waring, 1996; Goddard and Wilson, 1999; McGahan and Porter, 1999; Bou and Satorra, 2007; Bottazzi et al., 2008; Madsen and Walker, 2017). These time series models are part of a larger class of *autoregressive moving average* (ARMA) models. ARMA models have been extensively studied and used in various scientific disciplines, because of their parsimony, wide ranging applications and well-known properties (e.g., Cowpertwait and Metcalfe, 2009; Hamilton, 1994).

Autoregressive (AR) processes describe y_t in terms of the sum of lagged variables up to a certain lag p , plus a random shock: $y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \epsilon_t$. Because the random variable y_t at a certain time t feeds into the equation for the next year, autoregressive processes have a “long-term memory”: even for an autoregressive process with just one time lag (AR(1)), a random shock ϵ_t in a certain year can affect many future years, through the various time lags. These properties, along with the fact that AR(1) models can easily be estimated with OLS (by regressing a variable on its first lag) make them popular for time series analyses, for instance in the earlier mentioned persistence of profitability literature.

Moving average (MA) processes⁶ describe y_t in terms of the sum of lagged error terms up to a certain lag q : $y_t = \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$. In contrast to autoregressive models, moving average models have a “short-term memory”: any random shock can only affect the outcome variable up to the maximum lag of q time periods later, after which it literally drops out of the equation. Thus moving average processes provide a natural model for noise with a short lived impact.

Because the linear model in equations (4.1), (4.3), and (4.4) consists of a combination of

⁶General autoregressive and moving average processes can also include a constant term (e.g., $y_t = c + \phi y_{t-1} + \epsilon_t$ for an AR(1) process). Because in this paper I work with demeaned variables, I omit the constant.

autoregressive processes (the decay of resource (dis-)advantages x_t and z_t) and short lived noise terms (e.g., the annual variation u_t), it might be expected that the implied time series behavior of y_t consists of a combination of autoregressive and moving average terms. This is indeed the case:

Proposition 5. *Assume the stochastic model as defined in equations (4.1), (4.3), and (4.4). Then the time series y_t is second-order equivalent to an ARMA(2,2) process with zero mean. In other words, there exist identically and independently distributed error terms ϵ_t and parameters θ_k, ϕ_k such that the time series y_t defined by*

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \quad (4.5)$$

has the same expected values and variance-covariance structure as the time series y_t defined by equations (4.1) to (4.4). Moreover, the values of $\lambda^{(1)}$ and $\lambda^{(2)}$ determine the autoregressive parameters ϕ_1 and ϕ_2 as follows:

$$\phi_1 = \lambda^{(1)} + \lambda^{(2)} \quad (4.6)$$

$$\phi_2 = -\lambda^{(1)}\lambda^{(2)} \quad (4.7)$$

Proof. See appendix A.5. ⁷ \square

Thus, equations (4.1), (4.3), and (4.4) describe an ARMA(2,2) model for y_t . Moreover, from lemma 1 in appendix A.5 it follows that a model with only operating resources and no higher order resources (i.e, $\lambda^{(2)} = 0$) is equivalent to an ARMA(1,1) model. A great advantage of the ARMA formulation over the original one is that the ARMA estimation properties are well-known: the model is identified as long as there are no jointly redundant AR and MA components present in the model (e.g., Hamilton, 1994).

Intuitively the relation between the earlier model and the ARMA formulation can be understood as follows. In a model with no higher order resource advantages ($z_t = w_t = \lambda^{(2)} = 0$)

⁷The appendices are provided in a separate document.

and no annual noise term ($u_t = 0$), the excess profit y_t follows an AR(1) model, because in that case $y_t = x_t$ and thus $y_t = \lambda^{(1)}y_{t-1} + v_t$. The addition of the annual noise term u_t essentially means that the excess profits y_t become a noisy measure of the operating resource advantage x_t , which leads to the addition of a moving average term to the time series model: ARMA(1,1). The addition of higher order resource advantages leads to both an additional autoregressive term and an additional error term, which translates into a second moving average term, thus defining an ARMA(2,2) model.

Note that $\lambda^{(1)}$ and $\lambda^{(2)}$ are fully interchangeable, in the sense that exchanging them leads to the exact same autoregressive parameters, consistent with the examples provided in Table 17 panels d and e. Therefore, in the empirical inference one needs to decide which value to assign to which λ -parameter in order to uniquely determine them. For the remainder of the paper, I will assign the lower value to $\lambda^{(2)}$, as this provides a lower bound of the persistence rate of higher order resource advantages, and thus a more conservative estimate of this value.

Table 18 summarizes the results from this section. If no operating and higher order resources are present ($\lambda^{(1)} = \lambda^{(2)} = 0$), then profit evolution is equivalent to $y_t = u_t$, a random walk, without any autoregressive components. If only operating resources are present ($\lambda^{(1)} > 0, \lambda^{(2)} = 0$), then profit levels persist over time, leading to a first order autoregressive component. Because of the interchangeability of $\lambda^{(1)}$ and $\lambda^{(2)}$ the case with only higher order resources present ($\lambda^{(2)} > 0, \lambda^{(1)} = 0$) is exactly equivalent to the case with only operating resources. Finally, when both operating and higher order resources are present this leads to the addition of higher order autoregressive and moving average terms, described by a model with two autoregressive components; intuitively this can be understood as the persistence of not only profit levels but also profit growth over time, per the example in Table 17 panel c.

These results provide an empirical test for the presence of higher order resources in shaping profit persistence. In a world with only operating resources one would expect excess profits

Table 18: Summary of the effects of operating and higher order resources on profit persistence.

Presence of resources:		Persistence of profit:		Time-series
Operating	Higher order	Level	Growth	# AR components
-	-	-	-	0
✓	-	✓	-	1
-	✓	✓	-	1
✓	✓	✓	✓	2

to follow an ARMA(1,1) process, while in a world in which also higher order resources play a significant role, one would expect higher order autoregressive components to be present in profit time series. Note that, while the AR terms determined by the $\lambda^{(k)}$ parametrize the long term profit persistence patterns, the MA parameters θ_k only parametrize short-term noise. Therefore the MA parameters are typically treated as “nuisance parameters” in this type of model (Harris, 1999, p. 158), analogous to for instance the standard deviation of the error term in OLS: such parameters are needed in the model for correct inference, but their values are not of direct interest by themselves.

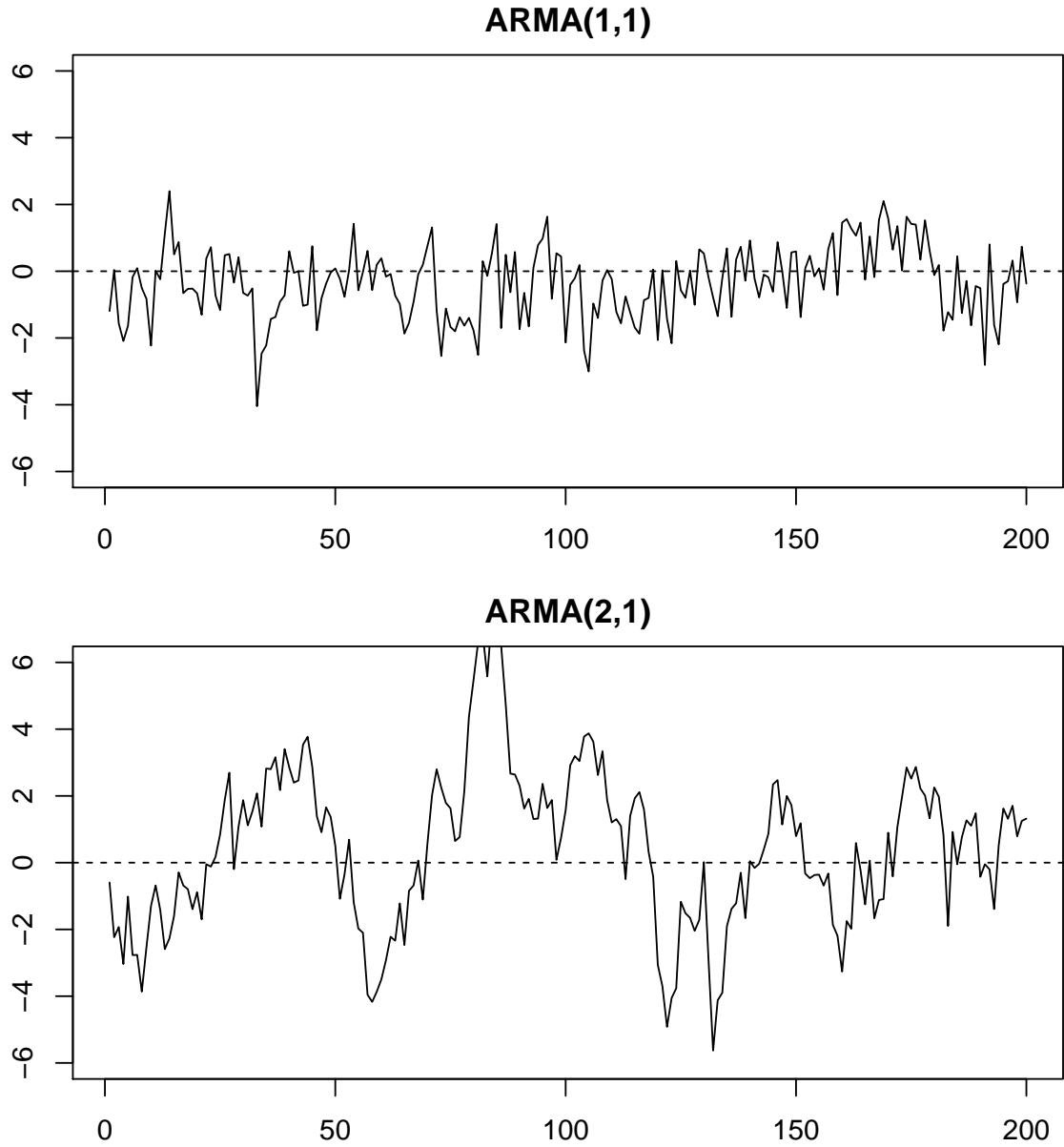
Figure 13 shows two simulations, first of an ARMA(1,1) time series, and second of the same model with the addition of an AR(2) term. Similar to the highly stylized examples in Table 17, the addition of the AR(2) term leads to markedly different behavior. With the addition of the AR(2) term, profit changes become persistent over time, leading to larger and longer deviations from the competitive average. The remainder of this paper will be devoted to measuring the presence of absence of the AR(2) term in observed profit data, and thus of the importance of higher order resources in shaping profit persistence patterns.

4.3. Methods

4.3.1. Data and measures

To determine whether higher order ARMA components are present in actual profit time series, I use the Compustat North-America database—the *de facto* standard for this type

Figure 13: ARMA time series simulations



Note. Simulations of ARMA time series models with $\lambda^{(1)} = 0.9$, $\theta_1 = -0.7$, and a standard normal distribution for the error term ϵ_t ; in the ARMA(2,1) model a second order component with $\lambda^{(2)} = 0.7$ has been added.

of analysis—over the 30-year period from 1985 to 2014 (inclusive). Consistent with many other studies using performance data (e.g., Rumelt, 1991; McGahan and Porter, 1997; Ruefli and Wiggins, 2003), for each firm i in industry j and year t , I measure assets K_{ijt} through total balance sheet assets (Compustat field code `at` in Wharton Research Data Services (WRDS)), and profits π_{ijt} through operating income (Compustat field code `oiadp`; also known as earnings before interest and taxes (EBIT)). I operationalize industry definitions through 6-digit GICS codes (as a robustness check, I also use SIC, see appendix A.7; I use GICS codes for my main specification, because they provide more contemporary industry definitions than the decades-old SIC codes, and lead to more balanced distributions of firms across industries). All monetary figures are deflated to 2010 real US dollars using annual consumer price index figures. The excess profits $y_{ijt} = \pi_{ijt} - r_{jt}K_{ijt}$ are calculated for each company-year using the industry average returns for each industry-year r_{jt} .

Note that the measure of excess profits differs by a size factor from the commonly used return on assets (RoA) vis-à-vis competitors. Specifically, denoting a firm’s RoA with $\tilde{r}_{ijt} = \pi_{ijt}/K_{ijt}$, my measure of excess profits is equal to $y_{ijt} = (\tilde{r}_{ijt} - r_{jt})K_{ijt}$: the firm’s RoA above or below the industry average, times the firm’s assets. This measure is closely related to the notion of economic profit, replacing the cost of capital (which is on average equal to the average returns of all firms), by an industry-year average return. The reason to use this measure is that firms ought to optimize a dollar amount of excess profits, instead of a ratio such as RoA (Levinthal and Wu, 2010). For related reasons, Hawawini et al. (2003) use economic profit instead of RoA to analyze profit patterns.

In the sample, I exclude all financial services companies, because of their different accounting standards, as well as all observations with assets below 100 million real 2010 US dollars, to limit the sample size (and thus computing time) to the economically most relevant companies. Also, I exclude any observations after there had been a missing observation for a specific company, so I obtain an unbalanced panel with contiguous financial data per company. Finally, I exclude companies with fewer than ten observations and all industries

that have fewer than five companies in some years. This leads to a sample of 79,203 observations by company-year for 4,321 companies in 57 industries. In appendix A.7, I check for the robustness of the results to the specific cut-off points used.

4.3.2. Box-Jenkins method

Using the Compustat data, I want to estimate the parameters of the model defined in equations (4.1) to (4.4), or equivalently the parameters of the ARMA transformed model in equation (4.5). However, an ARMA model in general can't be estimated using regression techniques such as OLS (Harris, 1999, p. 149). Instead, the Box-Jenkins method is the canonical approach for time series analysis (Box and Jenkins, 1970; Hamilton, 1994). This method relies on an analysis of (partial) autocorrelation functions and maximum likelihood estimation (MLE) of the ARMA parameters θ_k and ϕ_k in equation (4.5).

The autocorrelation function $\rho_s = \text{Cov}(y_t, y_{t-s}) / (\sigma(y_t) \sigma(y_{t-s}))$ is simply the correlation between time lags s of the time series y_t . The autocorrelation function can be estimated by just taking the sample autocorrelations. The partial autocorrelation function α_s is the correlation between the y_t and y_{t-s} after projecting out linear combinations time series y_τ in between times t and $t - s$. The partial autocorrelation function can be estimated by taking a regression of y_t on the time series y_{t-1}, \dots, y_{t-s} ; the partial slope of y_{t-s} provides an estimate of the partial autocorrelation α_s (Hamilton, 1994, p. 111). These functions have standard implementations in the major statistical packages. I will use the implementation in R, using the functions `acf` and `pacf`, respectively (R Core Team, 2016).

The (partial) autocorrelation functions can be used as a first model-free assessment of the time-series structure. For instance for a pure MA(q) process, the autocorrelation function ρ_s is zero for $s > q$, while for a pure AR(p) process, the *partial* autocorrelation α_s is zero for $s > q$. If both ρ_s and α_s have several significant terms, this usually indicates a mixed ARMA(p, q) process.

The initial assessment using (partial) autocorrelation functions can then be used to estimate

specific ARMA models using maximum likelihood estimation (MLE), providing confidence intervals for the ARMA parameters θ_k and ϕ_k . In R this estimation procedure is implemented in the `arima` function. The significance of the terms and the quality of model fit (assessed through e.g. the Akaike Information Criterion, AIC) also provide an opportunity for selecting the best ARMA model. Moreover, the error terms ϵ_t that remain after fitting the ARMA model can again be tested for (partial) autocorrelations—if these are still significant, this indicates the need for a higher order ARMA model.

One problem with the Box-Jenkins approach in this setting is that it is designed for a single time series y_t , while in this analysis there is a separate time series y_{ijt} for each firm i . In order to still employ the Box-Jenkins approach, I analyze all time series together as if they were generated from a single ARMA model. I implement this by dividing each individual time series $t \mapsto y_{ijt}$ by its sample standard deviation, so each time series is at the same scale, and then inserting a sufficient number of “missing” observations between the different time series of each individual firm. For instance, when making assessments of up to lag five in either the (partial) autocorrelation functions or the MLE, at least five “missing data” points are needed in order to make sure that estimates are not contaminated between time series from different companies. However, this approach still makes the assumption that all ARMA parameters are the same across industries, which is not plausible given earlier studies (e.g., Waring, 1996).

4.3.3. Bayesian inference

Hierarchical modeling using Bayesian estimation provides a way to overcome the potentially troublesome assumption that all parameters are equal across industries, while still obtaining full statistical power from all data (Alcácer et al., 2013; Gelman et al., 2013, ch. 5). In the Bayesian framework, parameters are modeled as a vector of random variables ϑ . The statistical model consists of two parts: the prior distribution $p(\vartheta)$ and the likelihood $p(y|\vartheta)$ of the observations y given certain parameters ϑ . Bayesian inference then proceeds by assessing the posterior $p(\vartheta|y) \propto p(y|\vartheta)p(\vartheta)$, which describes the probability distribution

of the parameters given the actually observed data. This contrasts with the classical maximum likelihood estimator (MLE), which merely describes the mode (i.e., maximum) of the likelihood function $\vartheta \mapsto p(y|\vartheta)$, while the standard errors provide an approximation of the width of the peak of the likelihood function. Thus, when taking a flat prior ($p(\vartheta) = 1$) in Bayesian inference, the posterior mode and second derivative are equal to the MLE. Thus, MLE (and thus regression) can be seen as a specific case of Bayesian inference.

Because parameters in Bayesian inference are modeled stochastically, parameter values for different industries can be assumed to be draws from some common probability distribution—the *hierarchical prior*. Usually, such hierarchical models provide more reasonable parameter estimates than non-hierarchical models (Gelman et al., 2013, p. 101). For instance, the persistence parameters $\lambda_j^{(k)}$ for each industry j can be drawn from a normal distribution:

$$\lambda_j^{(k)} \sim \mathcal{N}(\mu_{\lambda,k}, \sigma_{\lambda,k}) \quad (4.8)$$

The AR parameters ϕ_k for each draw can be calculated from the persistence parameters $\lambda^{(k)}$ using equations (4.6) and (4.7).

The standard deviations $\sigma_{\lambda,k}$ in expression (4.8) provide estimates of the spread in parameters across industries, thus giving an estimate of the extent to which resource dynamics differ across industries. Such measures of variance can be interesting measures for strategic research that are hard to obtain using classical techniques such as OLS and MLE (Alcácer et al., 2013).

The advantage of Bayesian analysis to do hierarchical modeling comes at the disadvantage of a much bigger computational burden to calculate the posterior distribution than to calculate the classical OLS, ML or GMM estimates (which essentially are approximations of a full Bayesian posterior). Because the posterior is analytically intractable except for the simplest models, it usually needs to be simulated: hundreds or thousands of randomized instances of the model are calculated, which represent draws from the posterior distribution.

Inference then proceeds with calculation of statistics from these draws, such as the mean, the median, the standard deviation and the 95% intervals—this is the *posterior inference*. The posterior mean or median correspond with the classical point estimates, such as the MLE. The posterior standard deviation and 95% interval correspond with the classical standard error and 95% confidence interval, respectively. The central limit theorem guarantees that for large N all distributions are normal, with the same parameters using either Bayesian or classical inference (Gelman et al., 2013, ch. 4).

With the advent of modern IT, the computational costs of such simulations have greatly diminished. For instance, estimating the hierarchical model used in this paper on close to 100,000 observations takes up to a few days, using a parallelized algorithm on designated hardware. Thus, in this case, the computational burden is manageable while the benefit of being able to model parameters by industry are significant.

Appendix A.6 describes the full specification of the Bayesian model used in this paper. It combines the ARMA transformation of the original model as outlined in proposition 5 with the hierarchical model as outlined in this section. The prior $p(\vartheta)$ is designed to be minimally informative: all posterior information comes from the data. Note that because of the large data set used to estimate the model, the specific choice of prior should have minimal influence on the posterior distribution in any case.

4.4. Results

4.4.1. Descriptive statistics, autocorrelations, and ML estimates

Because in this paper I perform a time series analysis, there is only one variable of interest: the excess profits y_{ijt} for each company i in industry j and year t , measured in million 2010 US dollars. By definition, the average of y_{ijt} is zero (in fact, the average for each industry-year equals zero, because each company's profits are measured compared to the industry average in each year). Table 19 shows the percentile distribution. Note that though the mean of y is exactly zero by definition, the median has a small negative value

Table 19: Percentile distribution of excess profits y .

0%	2.5%	25%	50%	75%	97.5%	100%
-25,153.6	-1,003.3	-46.4	-1.0	43.5	973.7	32,298.0

Note. Data for 79,203 company-years for 4,321 companies in 57 industries. Figures in million constant 2010 US dollars.

indicating a slight skew. Moreover, due to the highly skewed size distribution of companies, most companies make a small profit or loss (half of the companies between -46 and $+43$ million), while a few companies make a large profit or loss (about five percent of companies more extreme than plus or minus one billion).

In order to bring the distribution across different companies at the same scale when they are analyzed as a single joint time series, each time series is divided by its standard deviation ($y_{ijt}/\text{s.d.}(y_{ij.}) \mapsto y_{ijt}$). Moreover, this also greatly reduces the influence of any outliers: if a particular time series contains an extreme value, the standard deviation for this company will be large, and the absolute scale of this time series thus strongly diminished. Note that this transformation does not change the time series pattern (i.e. the λ and θ parameters), because it merely rescales each time series by a constant. All analyses are performed on this rescaled variable y .

Table 20 lists the autocorrelation and partial autocorrelation function of y_{ijt} for various time lags. Given that $N = 79,203$, values more extreme than $\pm 2/\sqrt{79,203} \approx \pm 0.007$ are significantly different from zero at the 95% level (Hamilton, 1994, p. 111). Clearly, all autocorrelations and partial autocorrelations up to lag 5 are significant, suggesting the presence of both autoregressive and moving average components in the data.

The ML estimates in table 21 corroborate this picture. A comparison of both the Akaike information criterion (AIC) and the parameter estimates indicate that at least two AR components and one MA component are highly significant. The AIC is lowest for the ARMA(2,2) model—indicating the best model fit for that specification. Moreover, the $\lambda^{(2)}$, θ_1 and θ_2 estimates are significantly different from zero in each of the models, further

Table 20: (Partial) autocorrelation functions.

Lag (years)	ACF	PACF
1	0.798	0.798
2	0.677	0.112
3	0.601	0.087
4	0.547	0.066
5	0.510	0.064

Note. Autocorrelation function (ACF) and partial autocorrelation function (PACF) of excess profits y .

suggesting that the ARMA(2,2) model provides the best ML fit for this time series.

The $\lambda^{(1)}$ estimate for the AR(1) model is in line with the values commonly reported in the literature, which are usually in the range 0.6–0.9 for models without firm fixed effects (e.g., Bottazzi et al., 2008; Waring, 1996, figure 3). This value is significantly lower though than the $\lambda^{(1)}$ value for the other models, particularly ARMA(2,1) and ARMA(2,2). This suggests that the reported values in the literature are too low due to misspecification of the model—i.e. leaving out the moving average and second order autoregressive components. Indeed, leaving out a moving average term should lead to attenuation of the parameter estimates. Essentially, when using profitability in a regression to determine the AR(1) coefficient, this introduces a measurement error due to annual variations that are independent of the “true” underlying resource position. The resulting measurement error in the independent variable is a source of endogeneity and creates an attenuation bias in the parameter estimates (e.g, Wooldridge, 2010). Hence it should be expected that the reported values in the literature are lower than the actual persistence, consistent with the findings in table 21.

I have also analyzed the autocorrelations in the resulting error terms ϵ_t from the various ML inferences. The error terms from the ARMA(1,1) model still have significant (partial) autocorrelations for first and second order lags of 0.025 and -0.056 respectively, well outside the 95% confidence interval of ± 0.007 expected from a null model, providing further evidence for higher order components present in the data. The (partial) autocorrelations of the

Table 21: Model comparison for MLE.

	AR(1)	ARMA(1,1)	ARMA(2,1)	ARMA(2,2)
$\lambda^{(1)}$	0.809 (0.002)	0.867 (0.002)	0.928 (0.002)	0.925 (0.002)
$\lambda^{(2)}$			0.477 (0.011)	0.402 (0.024)
θ_1		-0.172 (0.005)	-0.708 (0.011)	-0.626 (0.026)
θ_2				-0.031 (0.008)
AIC	206,707.8	205,637.7	204,878.2	204,866.0

Note. Maximum likelihood estimates for various autoregressive moving average (ARMA) models of company excess profits. Standard errors in brackets. The Akaike information criterion (AIC) provides an estimate of model fit: lower values indicate a better fit to the data, accounting for the number of parameters in the model.

ARMA(2,2) model are much smaller, suggesting that those models account for most or all of the correlation structure in the data.

Finally, the normal probability plots for the error terms exhibit roughly a straight line, suggesting that the MLE does not suffer from outliers or other issues due to deviations from the normal distribution.

4.4.2. Bayesian inference

To better account for variations across industries, I have also performed a hierarchical Bayesian estimation as described in the methods section. All results are based on posterior inference of 4 Hamilton Monte Carlo (HMC) chains of 5,000 simulations each. The first 2,500 simulated draws have been discarded as warm-up⁸, and the remainder is thinned by a factor 20 (to conserve memory and disk space), for a total of 500 posterior draws for each model. Due to correlations between the different draws, the effective sample size is sometimes lower.

⁸Because Bayesian simulation starts from some random initial values, the first simulation draws will reflect the initial starting point, rather than the actual posterior distribution. Therefore, the first draws (called the “warm-up” or “burn-in”) need to be discarded, until they can be expected to reflect the actual posterior (Gelman et al., 2013, p. 282).

For each parameter of interest, \hat{n}_{eff} indicates the effective number of posterior draws, which should be at least 10 times the number of chains—40 in this case (Gelman et al., 2013, p. 287). Moreover, the \hat{R} statistic needs to be below 1.1 for all variables, to ensure that a sufficient number of draws have been discarded as warm-up, and that the sampling is indeed performed from the actual posterior (Gelman et al., 2013, p. 287). For each model and variable of interest I check that these values are in their desired ranges, and they are reported in table 23 for the main specification. All inference is performed using the Stan implementation in R (Carpenter et al., 2016; Gelman et al., 2013, Ap. C).

Table 22 shows the Bayesian hierarchical equivalent of table 21 to compare the different model specifications. The parameters shown are the means of the hierarchical priors—e.g., the values of the μ_λ 's in equation (4.8). The values in the table are the posterior means, which asymptotically correspond with the ML or OLS point estimates in classical inference. The figures in brackets are the posterior standard deviations, which asymptotically correspond with the standard errors in classical inference. Finally, the AIC is replaced with the WAIC (Watanabe-Akaike or widely applicable information criterion), because this accounts better for the effective number of parameters used in hierarchical models, and is in general preferred in Bayesian analyses (Gelman et al., 2014, 2013, p. 173).

As was the case for the ML estimates, also in the Bayesian hierarchical model, the moving average and second order autoregressive components are highly significant, and the WAIC markedly decreases when these components are added. Both suggest that the ARMA(2,2) specification best fits the data.

Table 23 shows more details for the ARMA(2,2) specifications for the main variables of interest: the mean of the resource advantage persistence rates $\lambda^{(k)}$, as well as the standard deviation across industries. Note that the posterior standard deviation in brackets is quite different from the standard deviation of the $\lambda^{(k)}$ parameters in the bottom two rows: the standard deviations in brackets refer to the the variation in posterior samples and are analogous to the classical standard errors, while the bottom rows with hierarchical standard

Table 22: Model comparison for Bayesian inference.

	AR(1)	ARMA(1,1)	ARMA(2,1)	ARMA(2,2)
Mean $\lambda^{(1)}$	0.866 (0.012)	0.886 (0.010)	0.949 (0.008)	0.954 (0.009)
Mean $\lambda^{(2)}$			0.560 (0.031)	0.528 (0.031)
Mean θ_1		-0.052 (0.018)	-0.696 (0.036)	-0.648 (0.033)
Mean θ_2				-0.053 (0.011)
WAIC	182,279.8	181,772.2	181,490.2	181,296.5

Note. Bayesian posterior means for various hierarchical autoregressive moving average (ARMA) models of company excess profits. Posterior standard deviations in brackets (asymptotically, these correspond to standard errors in classical inference). The Watanabe-Akaike or widely applicable information criterion (WAIC) provides a Bayesian estimate of model fit: lower values indicate a better fit to the data, accounting for the effective number of parameters in the model.

deviations have no direct classical analogue, and refer to the variation of parameters across industries, thus providing an estimate to what extent industries are similar or different in terms of persistence patterns.

All variables are significantly larger than zero⁹, and the posterior estimate for the mean of $\lambda^{(1)}$ is significantly higher than the 0.7 to 0.9 range common in the literature. The variation across industries can be seen from the relatively large standard deviation priors, especially for $\lambda^{(2)}$. Assessing the full distribution of all posterior draws of $\lambda_j^{(k)}$ across industries indicates a positive $\lambda_j^{(2)}$ for the far majority (99.3%) of posterior draws across industries— and thus the importance of the AR(2) component of the model, signifying that higher order resources significantly affect profit persistence patterns in (almost) all industries.

Appendix A.7 includes several robustness checks to the analyses in this section.

⁹Strictly speaking, significance testing is not possible in Bayesian analysis, and instead one should speak of ‘posterior probability’. Specifically, I use the term ‘significant variable’, to signify a variable that has a high posterior probability to be strictly greater (or strictly smaller) than zero. Asymptotically, this is consistent with the use of the term ‘statistical significance’ in classical inference.

Table 23: Posterior inference for ARMA(2,2) model.

	Mean (s.d.)		2.5%	97.5%	n_{eff}	\hat{R}
Mean $\lambda^{(1)}$	0.954	0.009	0.937	0.972	427	1.00
Mean $\lambda^{(2)}$	0.528	0.031	0.466	0.587	431	1.00
S.d. $\lambda^{(1)}$	0.048	0.007	0.036	0.064	500	1.00
S.d. $\lambda^{(2)}$	0.122	0.016	0.091	0.152	500	1.00

Note. Posterior mean, standard deviation, and 95% interval bounds for the eigenvalues of the autoregressive components in the hierarchical ARMA(2,2) model of excess profits. The effective sample size n_{eff} and the scale reduction factor \hat{R} are used to assess convergence of the posterior draws; Gelman et al. (2013) recommend these to be above 40 and below 1.1, respectively.

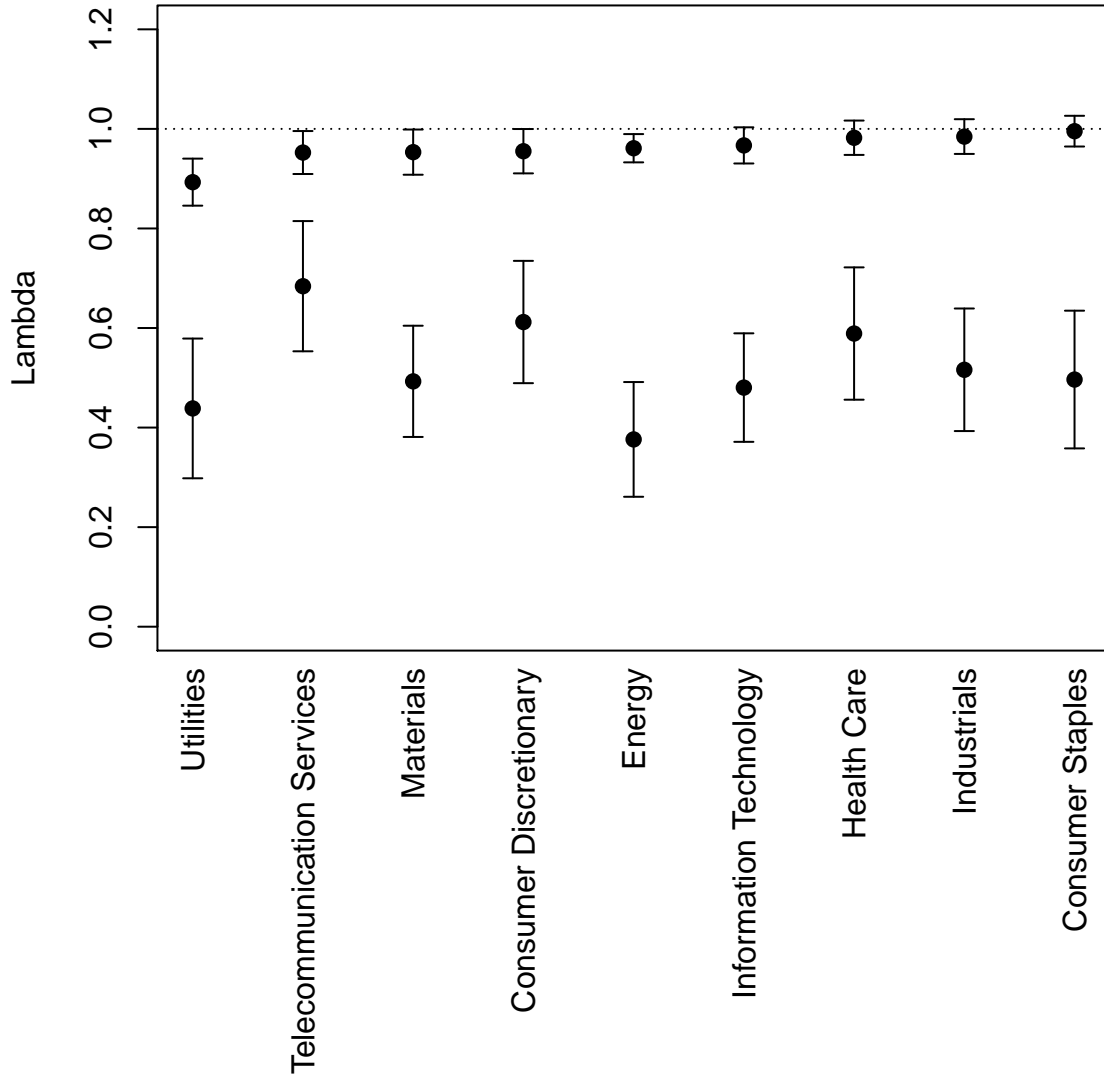
4.4.3. Industry analysis

An attractive feature of hierarchical Bayesian modeling is that it allows inference at the industry-level.¹⁰ Figure 14 shows the mean and 95% interval of the λ -parameters, averaged by industry sector.¹¹ Consistent with the results of the previous section, all 95% intervals are well above 0, signifying the importance of both autoregressive components. Moreover, the results indicate some clear differences across industries. Materials, Energy, and Utilities have relatively low values for both λ -parameters, suggesting lower persistence of profit differences in these industries—especially the low $\lambda^{(1)}$ value for utilities stands out. In these industries industry-level factors, such as energy prices, play an important role in profitability patterns, presumably diminishing the importance of both operating and higher-order resources, and thus profit persistence. On the other hand, Health Care, Industrials and Consumer Staples have relatively high values for both λ -parameters, with the value of $\lambda^{(1)}$ close to 1, suggesting a strong persistence of any profitability differences. Indeed, these industries are often associated with strong firm-specific resources (e.g., brands for consumer goods companies), which according to the RBV should lead to higher persistence of profits, consistent with the findings in, for instance Villalonga (2004).

¹⁰Thanks to the editor Brian Wu and two anonymous reviewers for suggesting this addition to the paper.

¹¹The figure shows the average by 2-digit GISC sector of the original 6-digit GICS industry estimates, weighted by the inverse of the posterior variance for each estimate (which is the optimal weighting).

Figure 14: Inference of λ -parameters by industry



Note. Weighted average of posterior industry means (dots) and 95% intervals (error bars) of $\lambda^{(1)}$ and $\lambda^{(2)}$ by 2-digit GICS industry sector, sorted by increasing $\lambda^{(1)}$.

Additionally, it is interesting to analyze how industry-level patterns evolve over time. Therefore, I have performed a Bayesian inference of the ARMA(2,2) model for each of the three decades between 1985 and 2014. Figure 15 shows the posterior mean and 95%-intervals for $\lambda^{(1)}$ and $\lambda^{(2)}$, both for the average across industries (panel a), as well as for selected individual industries with particularly low or high values in some of the decades for either of the λ -parameters (panels b-f).

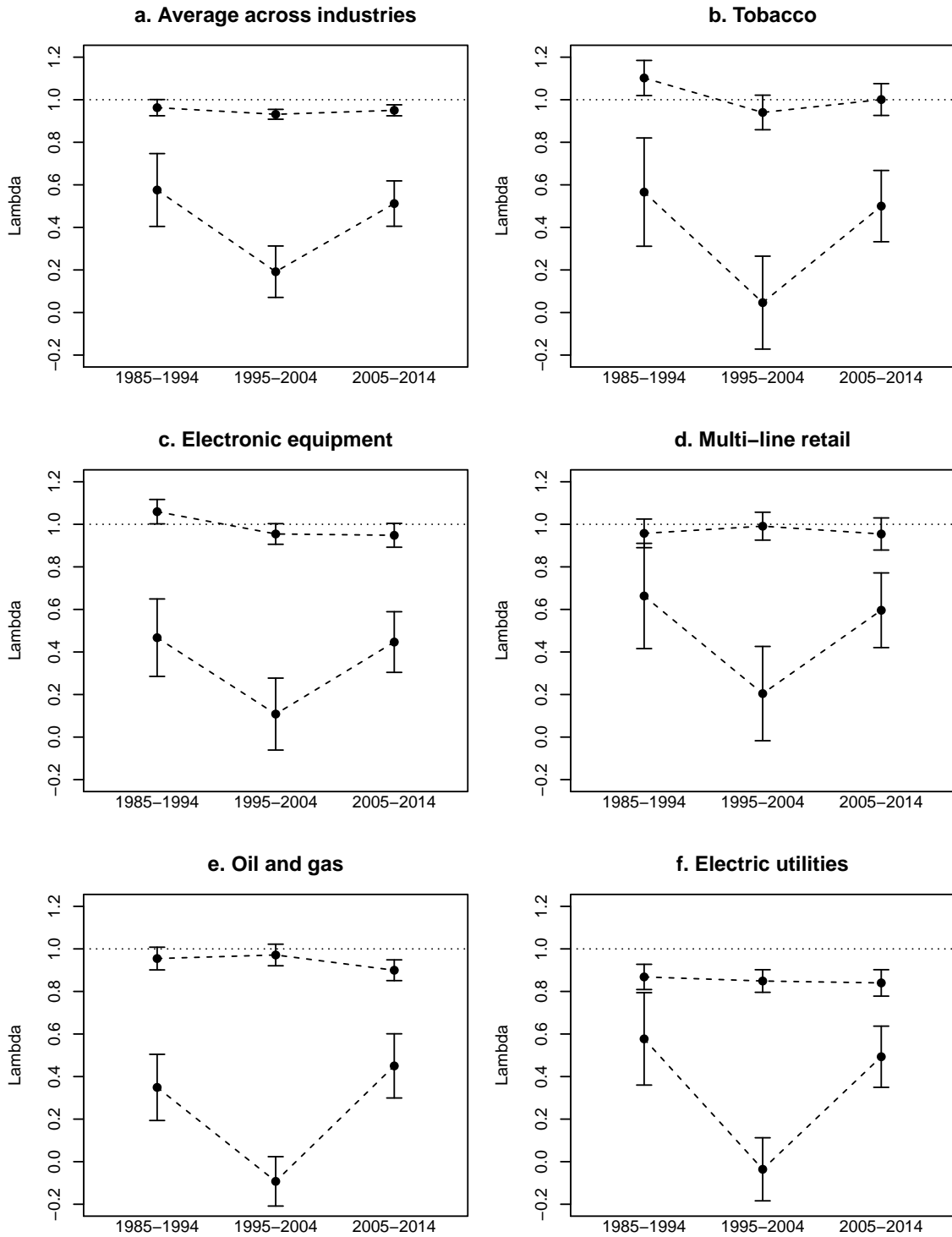
A clear time pattern stands out from this figure: $\lambda^{(2)}$ is lower in the decade 1995-2004 than in the decades before and after. This is consistent with both the findings of “hypercompetition” literature in the 2000s (e.g., D’Aveni et al., 2010), as well as recent with findings of a reversion of this trend (Bennett and Gartenberg, 2016). Interestingly, the decline in persistence of profits in the decade 1995-2004 is almost exclusively in $\lambda^{(2)}$ (the persistence of profit *growth*), rather than in $\lambda^{(1)}$ (the persistence of profit *levels*). For some industries, $\lambda^{(2)}$ in the decade 1995-2004 is actually not significantly different from zero.

Another interesting finding is that some industries had periods with a $\lambda^{(1)} > 1$, such as Tobacco in 1985-1994 ($p = 0.02$).¹² This means that during this decade any differences in profits were, on average, amplified over time. Note that this would have been troublesome in case I had used a traditional maximum likelihood estimator (MLE), because the model would have been non-stationary (e.g., Hamilton, 1994). The issue is that the time series models used in the MLE are based on the assumption that the time series can be extended *ad infinitum*. Indeed, this would be troublesome, as any differences in profits would eventually grow arbitrarily large, which is impossible. Conceptually there is no problem with a $\lambda > 1$ though for a finite period of time; indeed it appears quite plausible that such “non-stationary episodes” could exist. In a Bayesian model it is rather easy to alleviate the assumptions that cause the trouble in the MLE, and thus provide correct inference for data with such non-stationary episodes.¹³

¹²Thanks to an anonymous reviewer for investigating potentially non-stationary behavior.

¹³Using simulated data, I confirm that the Bayesian model that I use correctly provide correct inference for non-stationary ARMA models, while the classical MLE fails. Details are available from the author upon request.

Figure 15: Inference of λ -parameters by decade



Note. Posterior mean (dots) and 95% interval (error bars) of $\lambda^{(1)}$ and $\lambda^{(2)}$ over time. Panel a shows average across industries, and panels b-f show inference for selected 6-digit GICS industries.

4.5. Discussion

The evidence from the autocorrelation analysis, the ML estimates, the Bayesian analysis and the robustness tests all provide support for (1) the existence of at least one moving average term and a resulting higher AR(1) component than found in preceding studies of profit persistence, sometimes even above unity, leading to 'non-stationary episodes', (2) the existence of a higher autoregressive component in profitability persistence patterns, and (3) a significant variation across industries in the persistence of operating and higher order resources. Based on the model in this paper, all these findings have a direct interpretation in the RBV framework. First, the higher AR(1) parameter suggests that profits are more persistent than found hitherto, further strengthening the importance of the RBV for strategy. Second, the existence of a higher order autoregressive component suggests the importance of higher order resources in shaping profitability patterns. Third, the industry variation suggests that differences across industries not only manifest themselves in *levels* of profits, but also in the temporal dynamics of profits, and that resource-based theories can inform the differences in these dynamics.

In addition to these empirical findings, the paper shows how Bayesian hierarchical analysis—a method that is increasingly common in other disciplines, but still little used in strategy—can provide new insights to core strategic questions. The hierarchical model allows obtaining full statistical power across industries, while not having to make the assumption that competitive dynamics are the same across industries. Moreover, the Bayesian framework allows making inferences at different levels of the hierarchy: both at the aggregate level and at the level of the industry or even of the individual company. The Bayesian inference also provides more flexibility in, for instance, the estimation of non-stationary models, with autoregressive components above unity.

Moreover, the theoretical model provides a different perspective on the relation between managerial agency and random variation. Commonly, these two are viewed as mutually exclusive: the random variation is the null hypothesis, and managerial agency is the al-

ternative explanation if the null hypothesis is rejected (e.g., Denrell et al., 2014). Instead, in this paper, managerial agency can be endogenized in the stochastic model, through the variation in higher order resources z_t . In this view, the quality of managers is itself a distribution that yields shocks to the stock of a firm’s higher order resources (w_t in the model). In other words, from the standpoint of the researcher, a particularly skilled managerial team can be seen as a favorable draw from a distribution, providing the firm higher order resources that allow it to obtain favorable operating resources generating profits at a fraction of their economic costs. This view is perfectly consistent with the view from the standpoint of the manager that she or he can influence the company (and the broader world) beyond mere random variation.

The findings of this paper are also directly relevant for managers. For instance, the significance of resource dynamics in shaping long-term profits, stipulates the importance for managers to analyze their firms’ resource positions vis-à-vis competitors, and improving them as possible. Moreover, the analysis suggests that these resource positions are hard to infer from the short-term profit data that managers and investors often rely on. Also for managers, the Bayesian approach to make inferences about probabilistic inferences about individual company positions could be a useful tool to better separate short-term noise from longer-term underlying resource positions.

Naturally, the study has several limitations, which could be addressed in future research. First, this study is based on a single data source of accounting data for US firms; this could be extended to other geographies as well as other types of data—in fact a Bayesian modeling structure would be uniquely suitable to incorporate data from multiple data sources (e.g., accounting and stock market data) in a single model. Another improvement would be to consider a broader class of models, for instance including interactions between operating and higher order resources, the addition of non-linear terms, and models that would allow for changing error term distributions over time—e.g., when a firm gets bigger it is very plausible that its profitability shocks in absolute terms also get larger, and it would be

helpful to incorporate that in the model. Finally, it would be interesting to further explore the differences in resource dynamics resulting from business unit vs. corporate vs. industry effects, and investigate their origins—for instance to explain the drivers of the industry variation.

CONCLUSIONS

In this dissertation I examine, using formal modeling, how resource dynamics in competitive markets shape long term profit patterns. Chapter 1 provides an introduction and summarizes a key empirical regularity of long-term profit: it closely follows a distribution generated by a geometric random walk, which also shapes the size distribution across firms. This provides a motivation for studying long-term profit patterns using dynamic stochastic processes.

Chapter 2 introduces a new measure of long-term firm performance: LIVA (long-term investor value appropriation). This measure helps to address a disconnect between the common theoretical assumption that managers optimize firm value, and the widespread empirical practice of measuring performance using short-term ratios such as ROA (return on assets). LIVA has four primary advantages over commonly used performance measures such as TSR (total shareholder return) and ROA. First, LIVA increases if and only if firms make NPV-positive investments, unlike for instance ROA and ROC (return on capital). Second, LIVA accounts well for mergers, bankruptcies, and other major corporate events, while even a measure such as total shareholder return above the cost of equity can be misleading because it does not account well for the change in size of the firm. Third, LIVA measures the size of economic impact, because it is an absolute measure (in monetary amounts). Fourth and finally, LIVA can be decomposed into different sources, as illustrated by the Best Buy versus Circuit City case study.

In chapter 3, I study how *ex ante* similar firms develop heterogeneous resource positions that lead to persistent performance differences, even among direct competitors. The model in this chapter shows that resource competition drives amplification of small differences in resource positions into large performance differences. This finding provides a general theoretical mechanism for the empirical regularity that most performance variations are within industries, which is consistent with the fact that top LIVA performers in Table 3

are distributed across a variety of industries.¹⁴ Moreover, the model predicts that several industry-level resource characteristics, such as depreciation rate, scale economies, and growth stage affect the extent to which variations in returns are driven by firm versus industry effects. Empirical results from a Bayesian hierarchical analysis of stock market returns indeed indicate clear differences in the firm-specific component of performance across industries, which are persistent over time and broadly consistent with the predictions of the model.

In chapter 4, I introduce a formal model of how higher order resources affect profit persistence. Higher order resources are resources that do not affect profits directly, but can affect other resources that in turn affect profits over time. The model shows that higher order resources lead to persistence not only in the level of profits, but also in their growth. The derived stochastic time series structure of profits provides an empirical test for the presence of higher order resources. Empirical estimation of the model using both classical and Bayesian hierarchical methods on a 30-year panel of more than 2,400 US firms provides support for the importance of higher order resources in shaping profit persistence.

A broader question that this dissertation addresses is what we gain from the formalization of resource dynamics. One thing it does not address, is the tautology of the theory. In fact, any good formal theory has to be a tautology (otherwise it would contain a contradiction). This is, for instance, also true for very successful theories in other fields such as quantum mechanics in physics and utility theory in economics. Note that both these theories—despite their widespread use in their respective disciplines—have very little empirical content by themselves, because these theories are very general and the underlying fundamental constructs (wave functions and utility functions, respectively) are not observable. Only specific theories within these frameworks relate to observable constructs and can be empirically tested, such as quantum electrodynamics or Cournot competition. This is

¹⁴Note that though tech firms are overrepresented in the list of top performers, they are also overrepresented in the list of bottom performers, illustrating Proposition 3 on page 68 that technology industries should be expected to exhibit high firm-specific variance due to their high scale economies and large sunk cost investments.

the power of these formal frameworks: they allow developing theories to study more specific phenomena, and specify the rules these more specific theories need to obey.

This is also the case for the formal framework presented in this dissertation: the equations (1.2), (1.3), and (1.4) have little empirical content in themselves, but the specific models developed in chapters 3 and 4 make clear empirical predictions, which are tested using observed performance data. Moreover, the formal theories in these two chapters make predictions that would have been hard if not impossible to derive with verbal theories. For instance, the effect size predictions in chapter 3 of firm versus industry level heterogeneity based on simulations using the characteristics of resource markets are virtually impossible to derive without mathematical modeling. Similarly, the proposition in chapter 4 that higher order resources lead to a second order autoregressive term in profitability data hinges on a formal proof.

Thus, the formal framework allows specifying RBV-based theories with very precise definitions, and clear empirical implications. The framework is quite general, and mainly specifies that RBV-based theories need to be stochastic, and that its dynamics follow the fundamental laws of causality (this is essentially the content of equation (1.3)). At the same time, the framework offers a specific stepwise program of how specific RBV-based formal theories can be developed:

1. **Specify the state space**, i.e. define the elements of \mathbf{X}_t in terms of firms' potential resource bundles, as well as potential exogenous factors of relevance (such as the industry profitability a_k in chapter 3).
2. **Define the relation to observables**, such as the relations between resources on the one hand, and profit and value on the other in equations (3.1) and (3.5).
3. **Specify how the state influences managerial action**, which in turn determines the stochastic evolution of the state space. For instance, in chapter 3 managerial action is defined by the investment level that optimizes firm value, while in chapter 4

the level of higher order resources can be seen as the quality of managerial capabilities, which in turn affects the change rate of operating resources.

4. **Derive the stochastic dynamics of the observables**, such as the firm and industry level profits in chapter 3 and the time series structure of profits in 4.
5. **Make inferences about the underlying resource dynamics** based on empirical data of the observables, such as the volatility data implying the importance of resource competition in chapter 3 and the profit time series patterns implying the importance of higher order resources in chapter 4.

Based on the above program, the models in this dissertation help shed light on the multi-trillion dollar question why just a few firms are able to appropriate a large share of value, as exemplified in Figure 1 and Table 9 Panel a. The following picture emerges. Any small differences across firms in the resources they have access to, such as the quality of a new technology or plant design, are amplified over time, because the firm with the better resources has more incentives to invest in more rapid growth, leading to persistent differences in performance and large differences in value. The presence of higher order resource further increases these persistent differences. The resulting evolutionary process, in which more profitable firms invest more and grow faster, leads to the log-normal size distribution specified by Gibrat's law, and ultimately to the observed heavily skewed LIVA distribution as documented in Figure 1.

Of course, the models in this paper only provide a few small extra pieces to the puzzle of performance heterogeneity. There are several potential extensions for future research, for instance:

- The relationship between growth, profitability, and size could be further investigated both theoretically and empirically, to better understand how exactly it shapes the skewed distribution of long-term returns (i.e., LIVA) as documented in Figure 1.

- The dynamic models in this dissertation can serve as a basis for developing stochastic performance measures that, for instance, indicate the 95% interval of the “true” underlying performance.
- The industry dynamics in the model in chapter 3 could be extended with more firms, entry & exit, competition in multiple resource markets, and mergers & acquisitions.
- The ideas in chapters 3 and 4 could be integrated to study how investment competition shapes the emergence of higher order resources.
- The model in chapter 3 currently relies on very strong rationality assumptions; it would be interesting to alleviate these and have the investment policy instead be determined through learning processes.

Hopefully, the models in this dissertation serve as a helpful starting point for future work to investigate the origins of performance variations across firms, and exemplify the value of mathematical modeling in strategy.

APPENDIX

A.1. Algebraic derivations LIVA

In this appendix, we include the derivations of equations (2.2) and (2.3) from equation (2.1).

A.1.1. Equivalence of NPV and excess return definitions

Here we show the equivalence between the LIVA definition in equations (2.1) and (2.2) when using enterprise value for the firm value V_t , using the following identities:

$$FCF_t = div_t + nsbb_t + iND_{t-1} - \Delta ND_t$$

$$V_t = MC_t + ND_t$$

$$\Delta MC_t = \Delta sp_t MC_{t-1} - nsbb_t$$

$$TSR_t = \Delta sp_t + \frac{div_t}{MC_{t-1}}$$

$$r V_t = r_l MC_t + i ND_t$$

The first equation states that all cash flows from operations must be paid out to investors in some way, through dividends (div), share buy-backs net of share repurchases ($nsbb$), interest (the rate i times net debt ND), or change in net debt (i.e. re-paying debt or putting cash in the bank net of issuing debt). The second equation defines enterprise value: the market capitalization MC plus the net debt ND . The third describes that changes in market capitalization are either due to changes in share price (relative change Δsp) or changes in shares outstanding. The fourth is the defining equation of total shareholder return TSR : changes in share price plus dividend yield. The final equation is the defining equation of cost of capital: it is the weighted average of the cost of capital for shareholders (r_l) and debt holders (i). Because of this identity, the unleveraged cost of capital r is also sometimes called the weighted average cost of capital (WACC). Using these identities, we

have:

$$\begin{aligned}
LIVA &= V_T - V_0(1+r)^T + \sum_{t=1}^T \frac{FCF_t}{(1+r)^{t-T}} \\
&= \sum_{t=1}^T \frac{V_t - (1+r)V_{t-1}}{(1+r)^{t-T}} + \sum_{t=1}^T \frac{FCF_t}{(1+r)^{t-T}} \\
&= \sum_{t=1}^T \frac{FCF_t + \Delta V_t - r \Delta V_{t-1}}{(1+r)^{t-T}} \\
&= \sum_{t=1}^T \frac{ER_t MC_{t-1}}{(1+r)^{t-T}}
\end{aligned}$$

In the above, the second equation follows from the general identity for a geometric series, and the final equation follows from substitution of the earlier identities:

$$\begin{aligned}
FCF_t + \Delta V_t - r V_{t-1} &= (div_t + nsbb_t + i ND_{t-1} - \Delta ND_t) \\
&\quad + (\Delta sp_t MC_{t-1} - nsbb_t + \Delta ND_t) - (r_l MC_t + i ND_t) \\
&= div_t + \Delta sp_t MC_{t-1} - r_l MC_{t-1} \\
&= \left(\frac{div_t}{MC_{t-1}} + \Delta sp_t - r_l \right) MC_{t-1} \\
&= (TSR_t - r_l) MC_{t-1} \\
&= ER_t MC_{t-1}
\end{aligned}$$

A.1.2. Relation with economic profit (EP)

For the derivation of equation (2.3), we use the following accounting identities:

$$\begin{aligned}
FCF_t &= Rev_t - OpEx_t - CapEx_t \\
\Delta BV_t &= CapEx_t - D_t \\
NOPAT_t &= Rev_t - OpEx_t - D_t
\end{aligned}$$

The first identity states that the free cash flow to investors is the same as the free cash flow from operations: revenues, minus operating expenditure (including taxes) and capital expenditure. The second states that the change in book value is equal to the capital expenditure minus depreciation (including amortization). The final identity states that the net operating profit after tax is revenue minus operating expenditure (including taxes) and depreciation (including amortization). Then using enterprise value from the definition in equation (2.1), we have:

$$\begin{aligned}
LIVA &= EV_T - EV_0(1+r)^T + \sum_{t=1}^T \frac{FCF_t}{(1+r)^{t-T}} \\
&= [EV_T - BV_T] - [EV_0 - BV_0](1+r)^T + \sum_{t=1}^T \frac{FCF_t + \Delta BV_t - r BV_{t-1}}{(1+r)^{t-T}} \\
&= [EV_T - BV_T] - [EV_0 - BV_0](1+r)^T + \sum_{t=1}^T \frac{EP_t}{(1+r)^{t-T}}
\end{aligned}$$

The second equation follows again from the geometric series, and the final from substitutions of the above accounting identities and the definition of economic profit in equation (2.3).

A.2. Economic product market model

In this appendix I derive the profit function in equation (3.1) directly from a set of consumer preferences and production functions.

Assume a market with n consumers divided over N segments of equal size. Each segment consists of $\Delta q = \frac{Q}{N}$ homogeneous consumers that have a willingness to pay of wp_k for a single unit of product from either firm (and 0 for any further units), which depends on the industry state k (changes in the industry state can thus be interpreted as demand shocks). In order to be able to produce for a certain segment, a firm needs to have acquired the Δq resources that are specific for that segment (e.g., established customer relations). For the segments for which the firm has acquired the required resources it can produce $A(L + \beta L^2)$

units of product for a labor¹ input of L ; the parameter β indicates any potential increased returns to scale (due to asset mass efficiencies) and is assumed to be small ($0 \leq \beta \ll \frac{1}{L}$). Thus the overall production function for a firm which has acquired q_i resources is:

$$q^{(p)} = \max(q_i, A(L + \beta L^2)) \quad (\text{A.1})$$

Now, given two firms engaging in Bertrand competition in each segment, a firm will only invest in obtaining the resources required for a segment in which the competitor does not yet have the resources, because in Bertrand competition any marginal profits would be competed away, while there are strictly positive investments needed to acquire the required resources, per equation (3.2).

Finally, assume a cost of labor c_L such that the firm is always able to make a profit ($wp_k > \frac{c_L}{A}$ for all k). Then in order to maximize its profit, each firm will set a price $p_k = wp_k$ and produce Δq units in each segment for which it has the required resources. Thus, in total it will produce q_i goods. The labor input required for this production are such that $q_i = A(L + \beta L^2)$. Using the quadratic formula to solve this equation leads to total cost of:

$$c(q_i) = c_L L = c_L \frac{-1 + \sqrt{1 + \frac{4\beta q_i}{A}}}{2\beta} \approx \frac{q_i}{A} - \frac{\beta q_i^2}{A^2}$$

The approximation uses the fact that β is small, applied to a second order Taylor expansion of the square root ($\sqrt{1+x} \approx 1 + \frac{1}{2}x - \frac{1}{8}x^2$). Then the profit function becomes:

$$\pi(q_i, wp_k) = wp_k q_i - c_L \left(\frac{q_i}{A} - \frac{\beta q_i^2}{A^2} \right)$$

This can be rewritten to obtain equation (3.1) by substituting $a_k := wp_k - \frac{c_L}{A}$ (which is positive by assumption) and $b := \frac{\beta c_L}{A^2}$.

¹Of course, labor can be substituted for any other type of input in this model.

A.3. Markov perfect equilibrium derivation

In this appendix I will derive the equilibrium conditions for the model derived in section 3.2.2. The derivation closely follows Besanko and Doraszelski (2004, pp. 29-30). See Doraszelski and Pakes (2007) for a more extensive and general treatment for computing equilibria of dynamic games. Technicalities about boundary solutions, second order conditions, and existence proofs are treated in these references and are omitted here.

Specifically, assume that the profit function is given by equation (3.1), and the state transition probabilities for q_i as a function of the investment x for firm 1 and the state $s = (q_i, q_j, a_k)$:

$$\begin{aligned}\theta_{+1}(x; s) &= f_s(x)\Delta t \cdot (1 - \delta q_i \Delta t) \\ \theta_0(x; s) &= f_s(x)\Delta t \cdot \delta q_i \Delta t + (1 - f_s(x)\Delta t) \cdot (1 - \delta q_i \Delta t) \\ \theta_{-1}(x; s) &= (1 - f_s(x)\Delta t) \cdot \delta q_i \Delta t\end{aligned}$$

The functional form of $f_s(x)$ is specified in equations (3.2), (3.3), and (3.4).

The function $W_\ell(s = q_i, q_j, a_k)$ in the Bellman equation (3.5) is the expected value of the resource state $s = (q_{i+\ell}, q_j, a_k)$, conditional on a given policy function $x_2(s)$ for firm 2:

$$W_\ell(s = q_i, q_j, a_k) = \sum_{m,r=+1,0,-1} V(q_{i+\ell}, q_{j+m}, a_{k+r}) \theta_m(x_2(s); s) \Phi_r(s)$$

The first order condition for equation (3.5) is that the partial derivative of the value function with respect to x equals 0:

$$\begin{aligned}\frac{\partial V(x; s)}{\partial x} &= -\Delta t + \beta \sum_{\ell=+1,0,-1} W_\ell(s) \frac{\partial \theta_\ell(x; s)}{\partial x} \\ &= -\Delta t + \beta f'_s(x)\Delta t [(1 - \delta q_i \Delta t)(W_{+1}(s) - W_0(s)) + \delta q_i \Delta t(W_0(s) - W_{-1}(s))]\end{aligned}$$

Then using the derivative of equation (3.2), $f'_s(x) = \frac{c_s g_s^2}{(c_s g_s + x)^2}$, this leads to solving the following equation (abbreviating the constant term in square brackets above with [...]):

$$(c_s g_s + x)^2 = \beta c_s g_s^2 [\dots]$$

This is a quadratic equation in x , with solution:

$$x_o(s) = -c_s g_s + \sqrt{\beta c_s g_s^2 [(1 - \delta q_i \Delta t)(W_{+1}(s) - W_0(s)) + \delta q_i \Delta t(W_0(s) - W_{-1}(s))]} \quad (\text{A.2})$$

All models are calculated on a discrete state space ($q_i, q_j = 0, 1, \dots, Q$ and $a_k = a_1, \dots, a_K$), with grid spacing parameters $Q = 10$, $K = 5$, $\Delta t = 0.2$. Moreover $g_0 = a_3 = 1$, and the values a_1, \dots, a_K are determined as a geometric series based on the ratio a_K/a_1 for the given simulation. E.g., for $a_5/a_1 = 4$, the values are $a_k = (0.5, 0.7, 1, 1.4, 2.0)$. I check robustness against varying any of these parameters, which for none of them materially affects the results.

The equilibrium is then calculated using an iterative procedure, in each step providing better estimates for $x(s)$ and $V(s)$, both of which can be represented as vectors, because they are functions on the discrete state space. First, the updated estimate $\tilde{x}(s)$ is calculated using equation (A.2), assuming the initial estimates $V(s)$ and $x(s)$, which is used to calculate $W_\ell(s)$, assuming the policy function for player 2 is $x_2(q_i, q_j, a_k) = x(q_j, q_i, a_k)$. Then, assuming the new estimate for the policy function $\tilde{x}(s)$, the updated value function $\tilde{V}(s)$ is calculated using the expression in square brackets in the Bellman equation (3.5). This procedure is repeated until the changes in updating $V(s)$ and $x(s)$ fall below a specified value, indicating conversion to an equilibrium. Computations have been performed in R.

A.4. Firm and industry volatilities in model

Let the excess returns of firm $n = 1, 2$ be:

$$Y_n = X_{\text{ind}} + X_n \quad (\text{A.3})$$

With X_{ind} , X_1 and X_2 independently and normally distributed, with mean 0 and standard deviations σ_{ind} and σ_{firm} respectively. Then the covariance between the firms' returns is:

$$\begin{aligned} \text{Cov}(Y_1, Y_2) &= \text{Cov}(X_{\text{ind}} + X_1, X_{\text{ind}} + X_2) \\ &= \text{Cov}(X_{\text{ind}}, X_{\text{ind}}) \\ &= (\sigma_{\text{ind}})^2 \end{aligned}$$

So the total volatility and the correlation between the firms' returns are:

$$\begin{aligned} \sigma^2 &= \text{Var}(Y_1) = (\sigma_{\text{ind}})^2 + (\sigma_{\text{firm}})^2 \\ \rho &= \text{Cov}(Y_1, Y_2) / \text{Var}(Y_1) = \frac{(\sigma_{\text{ind}})^2}{(\sigma_{\text{ind}})^2 + (\sigma_{\text{firm}})^2} \end{aligned}$$

Vice versa:

$$\begin{aligned} \sigma_{\text{ind}} &= \sigma \sqrt{\rho} \\ \sigma_{\text{firm}} &= \sigma \sqrt{1 - \rho} \end{aligned}$$

Note that in the case of $\rho < 0$ the model specified in equation (A.3) is not meaningful: if there is a common industry effect, the correlations between the returns must be positive.

A.5. Proof of ARMA transformation

I first derive two lemmas. Using those, I then prove proposition 5.

Lemma 1. *Let x_t, y_t, u_t and v_t be stochastic time series, such that:*

$$y_t = x_t + u_t \tag{A.4}$$

$$x_t = \lambda x_{t-1} + v_t, \tag{A.5}$$

with u_t and v_t each i.i.d. error terms that are also independent of each other. Moreover, assume initial conditions such that y_t is stationary. Then y_t is second-order equivalent to an ARMA(1,1) process, with AR(1) coefficient λ .

Proof. The autocovariance structure of y_t is:

$$\begin{aligned} \gamma_t &:= \text{Cov}(y_{t+s}, y_s) \\ &= \text{Cov}(y_t, y_0) \\ &= \text{Cov}(x_t + u_t, x_0 + u_0) \\ &= \text{Cov}(x_t, x_0) \\ &= \text{Cov}(\lambda x_{t-1} + v_t, x_0) \\ &= \lambda \text{Cov}(x_{t-1}, x_0) \\ &= \lambda \gamma_{t-1} \end{aligned}$$

The first step follows from the stationarity of y_t , the second substitutes equation (A.4), the third follows from the independence assumptions, the fourth substitutes equation (A.5), the fifth follows again from the independence assumptions, and the final substitutes the definition of γ_t . Moreover, $\gamma_0 = \text{Cov}(x_0, x_0) = \sigma_x^2$, with σ_x^2 denoting the variance of x_t . Combining these two results yields for $t \in \{0, 1, 2, \dots\}$:

$$\gamma_t = \lambda^t \sigma_x^2$$

This autocovariance structure is the same as that of an ARMA(1,1) time series with AR(1) parameter λ (Cowpertwait and Metcalfe, 2009, p. 129). The MA(1) parameter and variance

of the error term of the ARMA(1,1) model can be chosen such that they match the variances of u_t and v_t (Harris, 1999, p. 158). \square

Lemma 2. *Let q_t be an ARMA(1,1) time series with AR(1) parameter $\lambda^{(1)}$, i.e.*

$$q_t = \lambda^{(1)}q_{t-1} + \epsilon_t - \kappa_1\epsilon_{t-1}, \quad (\text{A.6})$$

for some i.i.d. error term ϵ_t and some MA(1) parameter κ_1 . Moreover, let y_t have the same ARMA(1,1) structure, with AR(1) parameter $\lambda^{(2)}$, but with the error term itself now being an ARMA(1,1) time series:

$$y_t = \lambda^{(2)}y_{t-1} + q_t - \kappa_2q_{t-1} \quad (\text{A.7})$$

Then y_t is an ARMA(2,2) process, with the reciprocal eigenvalues of the AR(2) characteristic polynomial equal to $\lambda^{(1)}$ and $\lambda^{(2)}$.

Proof. Let L be the lag operator, which shifts a time series back by one unit (e.g., $Ly_t = y_{t-1}$). After rearranging terms, equation (A.7) then becomes:

$$(1 - \lambda^{(2)}L)y_t = (1 - \kappa_2L)q_t.$$

Similarly, equation (A.6) can be rewritten as:

$$q_t = \frac{1 - \kappa_1L}{1 - \lambda^{(1)}L}\epsilon_t.$$

Substituting this expression for q_t into the one before yields, after rearranging terms:

$$(1 - \lambda^{(1)}L)(1 - \lambda^{(2)}L)y_t = (1 - \kappa_1L)(1 - \kappa_2L)\epsilon_t.$$

This manifestly defines y_t as an ARMA(2,2) time series (Hamilton, 1994, p. 59), with AR(2) eigenvalues $(\lambda^{(1)})^{-1}$ and $(\lambda^{(2)})^{-1}$. \square

Proof of proposition 5. Define the following time series:

$$q_t := z_t + v_t.$$

This equation and (4.4) together are in the form specified in lemma 1, and thus q_t is equivalent to an ARMA(1,1) time series with AR(1) parameter $\lambda^{(2)}$.

Moreover, equations (4.1) and (4.3) yield:

$$\begin{aligned} y_t &= x_t + u_t \\ x_t &= \lambda^{(1)}x_{t-1} + q_t. \end{aligned}$$

These equations again are in the form specified in lemma 1, but with q_t itself now an ARMA(1,1) time series. Thus, by lemma 2, y_t is second-order equivalent to an ARMA(2,2) process with reciprocal eigenvalues $\lambda^{(1)}$ and $\lambda^{(2)}$. Specifically, the characteristic polynomial of the autoregressive part can be written as:

$$\forall s : 1 - \phi_1 s - \phi_2 s^2 = (1 - \lambda^{(1)}s)(1 - \lambda^{(2)}s). \quad (\text{A.8})$$

Equations (4.6) and (4.7) follow immediately from the requirement that the characteristic polynomial holds for all s . \square

A.6. Bayesian ARMA model specification

Likelihood (ARMA(2,2), cf. Stan Development Team (2016, sec. 7.4)):

$$\begin{aligned}\phi_{1,j} &:= \lambda_j^{(1)} + \lambda_j^{(2)} \\ \phi_{2,j} &:= -\lambda_j^{(1)}\lambda_j^{(2)} \\ \epsilon_{ij,t=1} &:= y_{ij,t=1} \\ \epsilon_{ij,t=2} &:= y_{ij,t=2} - \phi_{1,j}y_{ij,t=1} - \theta_{1,j}\epsilon_{ij,t=1} \\ \epsilon_{ij,t\geq 3} &:= y_{ij,t} - \phi_{1,j}y_{ij,t-1} - \phi_{2,j}y_{ij,t-2} - \theta_{1,j}\epsilon_{ij,t-1} - \theta_{2,j}\epsilon_{ij,t-2} \\ \epsilon_{ij,t\geq 3} &\sim \mathcal{N}(0, \sigma_i)\end{aligned}$$

Hierarchical priors:

$$\begin{aligned}\lambda_j^{(k)} &\sim \mathcal{N}(\mu_{\lambda,k}, \sigma_{\lambda,k}) \\ \theta_{k,j} &\sim \mathcal{N}(\mu_{\theta,k}, \sigma_{\theta,k})\end{aligned}$$

Hyperprior (uninformative):

$$p(\mu_{\lambda}, \mu_{\theta}, \sigma_{\lambda}, \sigma_{\theta}, \sigma_i^2) \propto \prod_i \frac{1}{\sigma_i^2}$$

Using Bayes' formula, the posterior distribution $p(\vartheta|y_{ijt})$ is the product of the likelihood and the prior distributions, where ϑ denotes the vector of all stochastic parameters. Inference in this paper is based on random draws from this posterior distribution.

A.7. Further robustness tests ARMA estimation

In this section I present two further robustness tests: alternative model specifications and testing the models on synthetic data. Table 24 shows five alternative specifications, each changing one aspect of the main specification in table 23: using the 4-digit SIC industry

Table 24: Alternative specifications for ARMA(2,2) model.

	Main	SIC	EBITDA	\$150M+	≥ 8 yrs	Vs. total	RoA
Mean $\lambda^{(1)}$	0.954 (0.009)	0.923 (0.009)	0.962 (0.007)	0.952 (0.008)	0.952 (0.008)	0.980 (0.008)	0.943 (0.009)
Mean $\lambda^{(2)}$	0.528 (0.031)	0.365 (0.046)	0.545 (0.030)	0.525 (0.034)	0.525 (0.032)	0.611 (0.027)	0.568 (0.031)
S.d. $\lambda^{(1)}$	0.048 (0.007)	0.049 (0.007)	0.042 (0.006)	0.047 (0.007)	0.049 (0.007)	0.045 (0.007)	0.034 (0.005)
S.d. $\lambda^{(2)}$	0.122 (0.016)	0.143 (0.017)	0.100 (0.013)	0.121 (0.016)	0.120 (0.017)	0.106 (0.015)	0.090 (0.012)

Note. Posterior mean with standard deviation in brackets for alternative specifications, each changing one aspect of the model presented in table 23. See the text for details.

classification system instead of GICS; using EBITDA instead of EBIT to operationalize profits; looking only at firms of at least 150M assets instead of 100M; including firms with at least 8 consecutive years of data instead of 10; calculating excess profit vs. the total average instead of an industry average; and using return on assets (RoA) compared to the industry average as profit measure instead of excess profit y . The results across these different specifications are broadly in line with the main specification in table 23; in particular the mean of $\lambda^{(2)}$ remains (highly) significant across specifications.

Table 25 shows the results of running the model on two synthetic data sets: one data set that is simulated from ARMA(2,2) parameters that are close to the inferred parameters from the main specification in table 23, and one that is simulated from an ARMA(1,1) parameter set, in which I have set the hierarchical means and standard deviations for the eigenvalues of the AR(2) and MA(2) components to zero. The values in the table represent the 95% posterior intervals to facilitate an assessment of whether the actual values used in the simulation are indeed within that interval.

For the ARMA(2,2) generated data, the ARMA(2,2) model indeed finds values for the λ -parameters that are well within the 95% interval. In this case the mis-specified ARMA(1,1) model exhibits the same attenuation bias as was present in both the MLE and the Bayesian

Table 25: Bayesian inference on synthetic data.

Data:	ARMA(2,2)			ARMA(1,1)			ARMA(1,1)			
	Actual	2.5%	97.5%	2.5%	97.5%	Actual	2.5%	97.5%	2.5%	97.5%
Mean $\lambda^{(1)}$	0.95	0.924	0.957	0.858	0.909	0.95	0.925	0.960	0.924	0.957
Mean $\lambda^{(2)}$	0.52	0.364	0.530			0	-0.097	0.059		
Mean θ_1	-0.60	-0.590	-0.416	-0.046	0.049	-0.70	-0.747	-0.579	-0.723	-0.649
Mean θ_2	-0.07	-0.099	-0.054			0	-0.081	0.028		
WAIC		90,568.2		90,777.7			114,254.2		114,287.1	

Note. 95% posterior interval limits for Bayesian inference on synthetic data. The left and right half of the table represent generated data from an ARMA(2,2) and ARMA(1,1) simulation respectively, with the values used shown in the column "Actual". Inference using the two different models has been performed on both generated data sets. A lower WAIC for a certain inference model indicates a better model fit for that specific data set (WAICs cannot be compared for different data sets).

inference on the Compustat dataset.

For the ARMA(1,1) generated data, the correct ARMA(1,1) model again infers 95% intervals that include the actual simulation values. The ARMA(2,2) model does too, but with very wide intervals for the second order components, because ARMA models which have redundant components in both the AR and MA parts are ill-identified.

The WAIC also predicts the right model for the ARMA(2,2) data, it is significantly lower for the ARMA(2,2) model. It is also lowest when using the AMRA(1,1) data, but then the difference between the WAIC estimates for both models is insignificant—16.8 with a standard error of 19.8. These results should put further confidence in the finding that the observed persistence data truly include higher order autoregressive components, and that the findings are not artifacts from the inference techniques.

BIBLIOGRAPHY

- Adner, R., L. Polos, M. Ryall, and O. Sorenson (2009). The case for formal theory. *Academy of Management Review* 34(2), 201–208.
- Adner, R. and P. Zemsky (2006). A demand-based perspective on sustainable competitive advantage. *Strategic Management Journal* 27(3), 215–239.
- Alcácer, J., W. Chung, A. Hawk, and G. Pacheco-de Almeida (2013). Applying random coefficient models to strategy research: Testing for firm heterogeneity, predicting firm-specific coefficients, and estimating strategy trade-offs. Harvard Business School Strategy Unit Working Paper 14-022.
- Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management* 17(1), 99–120.
- Barney, J. B. (1986). Strategic factor markets: Expectations, luck, and business strategy. *Management Science* 32(10), 1231–1241.
- Basu, S. and J. G. Fernald (1997). Returns to scale in us production: Estimates and implications. *Journal of Political Economy* 105(2), 249–283.
- Bennett, V. M. and C. M. Gartenberg (2016). Changes in persistence of performance over time. Duke I&E Research Paper No. 2016-41.
- Besanko, D. and U. Doraszelski (2004). Capacity dynamics and endogenous asymmetries in firm size. *RAND Journal of Economics* 35(1), 23–49.
- Bottazzi, G., A. Secchi, and F. Tamagni (2008). Productivity, profitability and financial performance. *Industrial and Corporate Change* 17(4), 711–751.
- Bou, J. C. and A. Satorra (2007). The persistence of abnormal returns at industry and firm levels: Evidence from Spain. *Strategic Management Journal* 28(7), 707–722.
- Box, G. E. and G. M. Jenkins (1970). *Time series analysis: forecasting and control*. Holden-Day.
- Brandenburger, A. M. and H. W. Stuart (1996). Value-based business strategy. *Journal of economics & management strategy* 5(1), 5–24.
- Brealey, R., S. Myers, and F. Allen (2006). *Corporate finance* (8th ed.). McGraw-Hill/Irwin.
- Canarella, G., S. M. Miller, and M. M. Nourayi (2013). Firm profitability: Mean-reverting or random-walk behavior? *Journal of Economics and Business* 66, 76–97.
- Capron, L. and N. Pistre (2002). When do acquirers earn abnormal returns? *Strategic Management Journal* 23(9), 781–794.

- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance* 52(1), 57–82.
- Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell (2016). Stan: A probabilistic programming language. *Journal of Statistical Software* 76(1), 1–32.
- Caves, R. E. and M. E. Porter (1977). From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition. *Quarterly Journal of Economics* 91(2), 241–261.
- Coff, R. W. (1999). When competitive advantage doesn't lead to performance: The resource-based view and stakeholder bargaining power. *Organization Science* 10(2), 119–133.
- Collis, D. J. (1994). Research note: how valuable are organizational capabilities? *Strategic Management Journal* 15(S1), 143–152.
- Cowpertwait, P. S. and A. V. Metcalfe (2009). *Introductory time series with R*. Springer Science & Business Media.
- Danneels, E. (2016). Survey measures of first-and second-order competences. *Strategic Management Journal* 37(10), 2174–2188.
- D'Aveni, R. A., G. B. Dagnino, and K. G. Smith (2010). The age of temporary advantage. *Strategic Management Journal* 31(13), 1371–1385.
- Denrell, J. (2004). Random walks and sustained competitive advantage. *Management Science* 50(7), 922–934.
- Denrell, J., C. Fang, and C. Liu (2014). Perspectivechance explanations in the management sciences. *Organization Science* 26(3), 923–940.
- Dierickx, I. and K. Cool (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science* 35(12), 1504–1511.
- Doraszelski, U. and A. Pakes (2007). *A framework for applied dynamic analysis in IO*, Volume 3, Chapter 30, pp. 1887–1966. Elsevier.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review* 14(4), 532–550.
- Ericson, R. and A. Pakes (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies* 62(1), 53–82.
- Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3–56.

- Fama, E. F. and K. R. French (1997). Industry costs of equity. *Journal of Financial Economics* 43(2), 153–193.
- Fisher, F. M. and J. J. McGowan (1983). On the misuse of accounting rates of return to infer monopoly profits. *American Economic Review* 73(1), 82–97.
- Gans, J. and M. D. Ryall (2017). Value capture theory: A strategic management review. *Strategic Management Journal* 38(1), 17–41.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.
- Gelman, A., J. Hwang, and A. Vehtari (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing* 24(6), 997–1016.
- Geroski, P. A. and A. Jacquemin (1988). The persistence of profits: a european comparison. *Economic Journal* 98(391), 375–389.
- Gibrat, R. (1931). *Les inégalités économiques*. Recueil Sirey.
- Goddard, J. A. and J. O. Wilson (1999). The persistence of profit: a new empirical interpretation. *International Journal of Industrial Organization* 17(5), 663–687.
- Goetzmann, W. N. and S. J. Garstka (1999). The development of corporate performance measures: benchmarks before EVA. Yale ICF Working Paper.
- Haleblian, J. and S. Finkelstein (1999). The influence of organizational acquisition experience on acquisition performance: A behavioral learning perspective. *Administrative Science Quarterly* 44(1), 29–56.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton university press.
- Harris, D. (1999). GMM estimation of time series models. In L. Matyas (Ed.), *Generalized method of moments estimation*, Chapter 6, pp. 149–170. Cambridge University Press.
- Hawawini, G., V. Subramanian, and P. Verdin (2003). Is performance driven by industry-or firm-specific factors? A new look at the evidence. *Strategic Management Journal* 24(1), 1–16.
- Helfat, C. E., S. Finkelstein, W. Mitchell, M. A. Peteraf, H. Singh, D. J. Teece, and S. G. Winter (2007). *Dynamic capabilities: Understanding strategic change in organizations*. Blackwell Publishing.
- Ittner, C. D. and D. F. Larcker (1998). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research* 36, 1–35.

- Jacobides, M. G., S. G. Winter, and S. M. Kassberger (2012). The dynamics of wealth, profit, and sustainable advantage. *Strategic Management Journal* 33(12), 1384–1410.
- Jacobsen, R. (1988). The persistence of abnormal returns. *Strategic Management Journal* 9(5), 415–430.
- King, A. A. and C. L. Tucci (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.
- King, D. R., D. R. Dalton, C. M. Daily, and J. G. Covin (2004). Meta-analyses of post-acquisition performance: Indications of unidentified moderators. *Strategic Management Journal* 25(2), 187–200.
- Knott, A. M., D. J. Bryce, and H. E. Posen (2003). On the strategic accumulation of intangible assets. *Organization Science* 14(2), 192–207.
- Knudsen, T., D. A. Levinthal, and S. G. Winter (2014). Hidden but in plain sight: The role of scale adjustment in industry dynamics. *Strategic Management Journal* 35(11), 1569–1584.
- Knudsen, T., D. A. Levinthal, and S. G. Winter (2017). Systematic differences and random rates: Reconciling Gibrat’s law with firm differences. *Strategy Science* 2(2), 111–120.
- Koller, T., M. Goedhart, D. Wessels, and McKinsey & Company (2010). *Valuation: Measuring and managing the value of companies* (5th ed.). John Wiley & Sons.
- Kraaijenbrink, J., J.-C. Spender, and A. J. Groen (2010). The resource-based view: a review and assessment of its critiques. *Journal of Management* 36(1), 349–372.
- Laamanen, T. and T. Keil (2008). Performance of serial acquirers: Toward an acquisition program perspective. *Strategic Management Journal* 29(6), 663–672.
- Leifeld, P. (2013). texreg: Conversion of statistical model output in R to L^AT_EX and HTML tables. *Journal of Statistical Software* 55(8), 1–24.
- Levinthal, D. A. (1991). Random walks and organizational mortality. *Administrative Science Quarterly* 36(3), 397–420.
- Levinthal, D. A. and B. Wu (2010). Opportunity costs and non-scale free capabilities: profit maximization, corporate scope, and profit margins. *Strategic Management Journal* 31(7), 780–801.
- Lieberman, M. B., R. Garcia-Castro, and N. Balasubramanian (2017). Measuring value creation and appropriation in firms: The VCA model. *Strategic Management Journal* 38(6), 1193–1211.

- Lippman, S. A. and R. P. Rumelt (2003). A bargaining perspective on resource advantage. *Strategic Management Journal* 24(11), 1069–1086.
- Mackey, T. B., J. B. Barney, and J. P. Dotson (2017). Corporate diversification and the value of individual firms: A Bayesian approach. *Strategic Management Journal* 38(2), 322–341.
- Madsen, T. L. and G. Walker (2017). Competitive heterogeneity, cohorts, and persistent advantage. *Strategic Management Journal* 38(2), 184–202.
- Makadok, R. and J. B. Barney (2001). Strategic factor market intelligence: An application of information economics to strategy formulation and competitor intelligence. *Management Science* 47(12), 1621–1638.
- McGahan, A. M. and M. E. Porter (1997). How much does industry matter, really? *Strategic Management Journal* 18(S1), 15–30.
- McGahan, A. M. and M. E. Porter (1999). The persistence of shocks to profitability. *Review of Economics and Statistics* 81(1), 143–153.
- McGahan, A. M. and M. E. Porter (2002). What do we know about variance in accounting profitability? *Management Science* 48(7), 834–851.
- Miller, D. (2003). An asymmetry-based view of advantage: Towards an attainable sustainability. *Strategic Management Journal* 24(10), 961–976.
- Mueller, D. C. (1977). The persistence of profits above the norm. *Economica* 44(176), 369–380.
- Mueller, D. C. (1986). *Profits in the long run*. Cambridge University Press.
- Nelson, R. R. and S. G. Winter (1982). *An evolutionary theory of economic change*. Belknap Press of Harvard University Press.
- Oxley, J. E., J. W. Rivkin, M. D. Ryall, and Strategy Research Initiative (2010). The Strategy Research Initiative: Recognizing and encouraging high-quality research in strategy. *Strategic Organization* 8(4), 377–386.
- Pacheco-de Almeida, G. and P. Zemsky (2007). The timing of resource development and sustainable competitive advantage. *Management Science* 53(4), 651–666.
- Pearson, K. (1897). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proceedings of the Royal Society of London* 60, 489–498.
- Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal* 14(3), 179–191.

- Peters, R. H. and L. A. Taylor (2017). Intangible capital and the investment-q relation. *Journal of Financial Economics* 123(2), 251–272.
- Porter, M. E. (1980). *Competitive strategy: techniques for analyzing industries and competitors*. Free Press.
- Priem, R. L. and J. E. Butler (2001). Is the resource-based view a useful perspective for strategic management research? *Academy of Management Review* 26(1), 22–40.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Richard, P. J., T. M. Devinney, G. S. Yip, and G. Johnson (2009). Measuring organizational performance: Towards methodological best practice. *Journal of Management* 35(3), 718–804.
- Ruefli, T. W. and R. R. Wiggins (2003). Industry, corporate, and segment effects and business performance: A non-parametric approach. *Strategic Management Journal* 24(9), 861–879.
- Rumelt, R. P. (1984). Toward a strategic theory of the firm. In R. B. Lamb (Ed.), *Competitive strategic management*, pp. 556–570. Prentice-Hall.
- Rumelt, R. P. (1991). How much does industry matter? *Strategic Management Journal* 12(3), 167–185.
- Schilke, O., S. Hu, and C. E. Helfat (2018). Quo vadis, dynamic capabilities? A content-analytic review of the current state of knowledge and recommendations for future research. *Academy of Management Annals* 12(1), 390–439.
- Siggelkow, N. (2007). Persuasion with case studies. *Academy of Management Journal* 50(1), 20–24.
- Stan Development Team (2016). *Stan modeling language: User’s guide and reference manual, version 2.10.0*.
- Sutton, J. (1991). *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. MIT press.
- Teece, D. J., G. Pisano, and A. Shuen (1997). Dynamic capabilities and strategic management. *Strategic Management Journal* 18(7), 509–533.
- Tsay, R. S. (2005). *Analysis of financial time series*. John Wiley & Sons.
- Vanneste, B. S. (2017). How much do industry, corporation, and business matter, really? A meta-analysis. *Strategy Science* 2(2), 121–139.

- Villalonga, B. (2004). Intangible resources, Tobin's q, and sustainability of performance differences. *Journal of Economic Behavior & Organization* 54(2), 205–230.
- Waring, G. F. (1996). Industry differences in the persistence of firm-specific returns. *American Economic Review* 86(5), 1253–1265.
- Wells, J. R. and G. Danskin (2012a). The fall of Circuit City Stores, Inc. Harvard Business School case study 9-713-402.
- Wells, J. R. and G. Danskin (2012b). The rise of Circuit City Stores, Inc. Harvard Business School case study 9-713-401.
- Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal* 5(2), 171–180.
- Williamson, O. E. (1979). Transaction-cost economics: the governance of contractual relations. *Journal of Law and Economics* 22(2), 233–261.
- Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal* 24(10), 991–995.
- Wiseman, R. M. (2009). On the use and misuse of ratios in strategic management research. Volume 5 of *Research methodology in strategy and management*, pp. 75–110. Emerald Group Publishing Limited.
- Wooldridge, J. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.
- Zollo, M. and H. Singh (2004). Deliberate learning in corporate acquisitions: Post-acquisition strategies and integration capability in us bank mergers. *Strategic Management Journal* 25(13), 1233–1256.
- Zollo, M. and S. G. Winter (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science* 13(3), 339–351.